THE MARKET FOR TEACHER QUALITY

Eric A. Hanushek
John F. Kain
Daniel M. O'Brien
Steven G. Rivkin

The Market for Teacher Quality
Eric A. Hanushek, John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin
NBER Working Paper No. 11154
February 2005
JEL No. I2, J4, H4

ABSTRACT

Much of education policy focuses on improving teacher quality, but most policies lack strong research support. We use student achievement gains to estimate teacher value-added, our measure of teacher quality. The analysis reveals substantial variation in the quality of instruction, most of which occurs within rather than between schools. Although teacher quality appears to be unrelated to advanced degrees or certification, experience does matter — but only in the first year of teaching. We also find that good teachers tend to be effective with all student ability levels but that there is a positive value of matching students and teachers by race. In the second part of the analysis, we show that teachers staying in our sample of urban schools tend to be as good as or better than those who exit. Thus, the main cost of large turnover is the introduction of more first year teachers. Finally, there is little or no evidence that districts that offer higher salaries and have better working conditions attract the higher quality teachers among those who depart the central city district. The overall results have a variety of direct policy implications for the design of school accountability and the compensation of teachers.

Eric A Hanushek
Hoover Institution
Stanford University
Stanford, CA 94305-6010
and NBER
hanushek@stanford.edu

John F. Kain
School of Social Sciences
University of Texas at Dallas
Richardson, TX 75083

Daniel M. O'Brien
School of Social Sciences
University of Texas at Dallas
Richardson, TX 75083
obri@utdallas.edu

Steven G. Rivkin
Amherst College
Department of Economics
P.O. Box 5000
Amherst, MA 01002-5000
and NBER
sgrivkin@amherst.edu

The Market for Teacher Quality

by Eric A. Hanushek, John F. Kain, Daniel M. O'Brien and Steven G. Rivkin

Given the emphasis parents, students, and educators place on teachers, evidence that observable measures of teacher quality explain little of the variation in student performance provides a conundrum for researchers. It may be that parents and students overstate the importance of teachers, but an alternative explanation is simply that measurable characteristics such as experience, certification, advanced degrees, and even scores on standardized tests explain little of the true variation in teacher effectiveness.

In this paper we use matched panel data on teachers and students for a large district in Texas to estimate variations in teacher quality through a semi-parametric approach based on value added to student achievement. These estimates confirm the existence of substantial variation in teacher effectiveness, much of it within as opposed to between schools. This within school heterogeneity has direct implications for the design of accountability and teacher incentive programs. Our findings on teacher effectiveness are also consistent with prior evidence showing that certification and experience explain little of the quality variation, with the exception of sizeable improvement following the initial year of teaching. The pattern of results supports the view that good teachers tend to be superior for students across the achievement distribution, but there is strong evidence that students benefit from being matched with teachers of the same race.

The development of teacher-specific quality measures permits investigation of important aspects of the market for teacher quality. Administrators in large urban districts often bemoan the loss of teachers to the suburbs, private schools, and other occupations, and available evidence suggests that exit probabilities are higher for teachers with better alternative earning opportunities or more education. However, there is no prior evidence on instructional effectiveness on school leavers, a void we can fill by comparing movers and stayers by their performance in the classroom.

Our analysis brings some commonly held beliefs into question. First, teachers who remain in our urban schools are as good as or better on average than those leaving. Second, there is little evidence that higher salaries and more desirable student characteristics raise the quality of new hires among teachers who switch districts in this labor market. This finding lends itself to two alternative though not mutually exclusive interpretations that carry different policy ramifications: 1) districts are unable to rank applicants on the basis of actual classroom effectiveness; or 2) quality plays a secondary role in the hiring process.

Empirical Model

A student's performance at any point in time reflects not only current educational inputs but also the history of family, neighborhood, and school factors. Even as databases begin to follow students over time, it is generally not possible to parse the separate current and historical influences that combine to determine the level of student performance. This leads many to focus on achievement gains rather than levels, an approach we adopt in this paper.[1]

The basic model relates achievement growth to the flow of current inputs such as:

$$(1) \qquad \Delta A_{isg} = A_{is_g g} - A_{is_{g-1} g-1} = f(X_{ig}, S_{ig}, \gamma_i, \varepsilon_{isg})$$

where A_{isg} is achievement of student i, in school s and grade g, and ΔA is the gain in achievement across grades; X is a vector of nonschool factors including family, peers, and neighborhoods; S is a vector of school and teacher factors; γ is individual differences in achievement growth; and ε is a random error.[2]

---

[1] Random assignment or instrumental variables techniques might be used to purge the estimates of confounding influences, although serial correlation in the variables of interest often complicates the interpretation of the results. Another alternative is estimation models of test score levels with student fixed effects. While this removes fixed unobserved factors that affect the performance level, it does not control for time varying influences in the past including variation in the quality of recent teachers.

[2] Alternative growth formulations including placing the earlier achievement, $A_{is_{g-1}g-1}$, on the right hand side have been employed; see Hanushek (1979). We discuss these alternatives explicitly below when we consider measurement issues.

This formulation controls for past individual, family, and school factors and permits concentration on the contemporaneous circumstances that are generally measured along with student achievement. Nonetheless, focusing on annual gains does not eliminate the difficulties in separating the various inputs from unmeasured confounding factors. A series of specification and measurement issues must be addressed before it is possible to obtain credible estimates of the influence of teachers on student achievement.

General Specification Issues

In past analyses, parametric models based on observed school and teacher characteristics have not yielded a reliable characterization of important aspects of schools and teachers, even when very detailed data on students and teachers are available (Hanushek (2003), Hanushek and Rivkin (2004)). The alternative, which we pursue here, is the semi-parametric estimation of teacher and school effects. Consider:

$$(2) \qquad \Delta A_{isg} = f'(X_{ig}, \widetilde{S}_{ig}) + \sum t_j T_{ijg} + (\gamma_i + \varepsilon_{isg})$$

where T_{ijg}=1 if student i has teacher j in grade g and is 0 otherwise. $\widetilde{S}_{ig}$ represents school factors other than teachers, and we combine the unmeasured individual and idiosyncratic terms (γ, ε) into a single error term. In this formulation teacher fixed effects, t_j, provide a natural measure of teacher quality based on value added to student achievement.[3]

Although this approach circumvents problems of identifying the separate components of teacher effectiveness, it must address a variety of selection issues related to the matching of

---

[3] For previous analyses of this sort, see among others Hanushek (1971, 1992), Murnane (1975), Armor et al. (1976), Murnane and Phillips (1981), Aaronson, Barrow, and Sander (2003), and Rockoff (2004). Rivkin, Hanushek, and Kain (2005) address the various selection factors along with providing a lower bound on the variations in teacher quality specified in this way. A different but complementary strand of research comes out of the Tennessee value-added work of William Sanders and his co-authors (Sanders and Horn (1994); Sanders, Saxton, and Horn (1997)); see also the methodological discussions in Ballou, Sanders, and Wright (2004).

teachers and students. Because of the endogeneity of community and school choice by families and of administrator decisions about classroom placement, the unmeasured influences on achievement are likely not orthogonal to teacher quality. In particular, students with family background and other factors conducive to higher achievement tend to seek out better schools with higher quality teachers. Administrative decisions regarding teacher and student classroom assignments may amplify or dampen the correlations introduced by such family choices. The matching of better students with higher quality teachers would tend to increase the positive correlations produced by family decisions, while conscious efforts to place more effective teachers with struggling students would tend to reduce them. Importantly, conditioning on prior score in the gains formulation does eliminate the first order selection problems because any placement by observed achievement will be accounted for by the early test score. Nonetheless, more subtle placement by unobserved characteristics will not be captured by prior achievement.

Another source of correlation between teacher quality and student performance results from the matching of teachers with schools. Teacher preferences for schools with higher achieving, nonpoor students in addition to higher salaries potentially introduce a positive correlation between teacher quality and family contribution to learning (Hanushek, Kain, and Rivkin (2004)). A desire to work in better managed schools likely amplifies the correlation between teacher and school quality. On the other hand, the failure of schools to hire the best available candidates would dampen this relationship (Ballou (1996)). Importantly, district assignment practices tend to give the newest teachers the lowest priority in terms of deciding where to teach, frequently leading to the pairing of the most inexperienced teachers with the most educationally needy children (see, for example, Clotfelter, Ladd, and Vigdor (forthcoming)).

In each of these cases, the central issue is whether it is reasonable to presume:

(3) $$E(\gamma_i + \varepsilon_{isg} \mid f'(X_{ig}, \tilde{S}_{ig}), T_{ijg}) = 0$$

4

The requirement that teacher fixed effects are orthogonal to the error highlights the importance of accounting for systematic elements of families and schools that explicitly or implicitly affect teacher-student matching.

The potential failure of observable characteristics such as family income, parental education, principal education, certification or experience, peer demographic composition, and other readily available variables to capture all school, peer, neighborhood, and family influences related to teacher quality leads us to include school fixed effects in many specifications. In these models, the teacher coefficients reflect only within school variation in the quality of instruction.[4]

This approach goes too far to the extent that the typical teacher in some schools is better than the typical teacher in others, perhaps in part because an important dimension of administrator skill is the ability to identify and develop good teachers. By eliminating all between school variation in teacher quality, the estimator implicitly attributes all aggregate school variation to factors other than the teachers.

Although the concentration on within school variation does not guard against all potential selection problems, we interpret its as a lower bound estimate of the variation in teacher quality. Only the purposeful matching of students and teachers in a very specific manner could conflate teacher and student influences in a way that biases upward estimates of the within school variation in teacher quality. Our empirical model focuses on student gains, includes a series of student covariates, employs a test metric (described below) that compares teachers on the basis of their performance with comparable students, and considers just teacher effects that persist over time as reflecting quality differences. Thus, only persistent sorting of students on the expected rate of learning conditional on initial score and the included covariates would introduce any upward bias. We believe strongly that the downward biases introduced by ignoring all between

---

[4] This approach of concentrating on within school variation is related to that in Rivkin, Hanushek, and Kain (2005), which develops a lower bound on the variation in teacher quality. That analysis, however, uses a different methodology that systematically guards against a variety of possible dimensions of selection, although it does not permit quality estimation for individual teachers.

school variation in teachers and year to year differences in a teacher's performance (see below) dwarfs any remaining contamination induced by within school sorting. Nonetheless, we also present estimates based on the variations across the entire district (and including between school variation) in order to indicate the range of plausible estimates.

An alternative approach that we do not pursue is the simultaneous estimation of student and teacher fixed effects in adjusted gains.[5] In addition to the computational demands and introduction of substantial noise in the estimation of the teacher coefficients, extensive school sorting by ethnicity and income raises a serious concern about the interpretation of the teacher fixed effect estimates generated by such a model. In the extreme case of complete school segregation by race, for example, the model imposes the assumption that average teacher quality for black students equals average teacher quality for white students. Of course the focus on within school differences ignores all average differences among schools, but in that case the identifying assumption is transparent and a much more tractable model is estimated.

Test Measurement Issues

Although psychometricians have long been concerned about the properties of cognitive tests and the implications for research on teachers and schools, economists have not paid much attention to these issues until recently.[6] Our analysis relies on the achievement tests used in Texas (the TAAS tests) to assess basic student proficiency in a range of subjects.[7] These tests, designed for the Texas accountability system, focus on minimum competency as defined by

---

[5] In sorting out the effects of teacher-student matching by race, we do introduce student fixed effects. See the discussions below about the specific interpretation of those estimates.
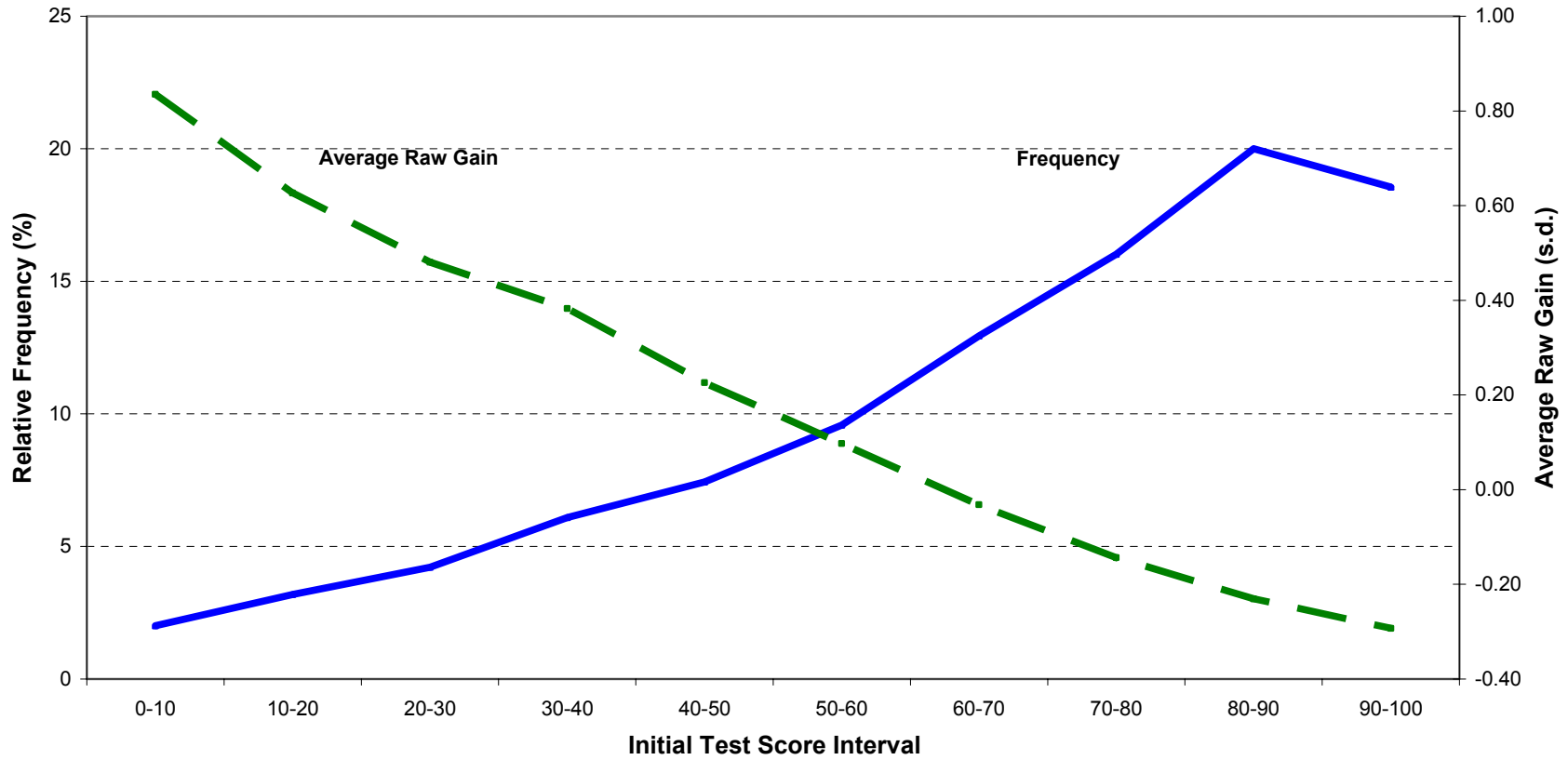
[6] One exception is Kane and Staiger (2002, 2002), who have discussed measurement error and test reliability in the context of school accountability systems. Bishop et al. (2001) considers the structure of tests from the perspective on student incentives. Bacolod and Tobias (2003) and Tobias (2004), as discussed below, do put testing into the context of achievement modeling.

[7] The TAAS tests are generally referred to as criterion referenced tests, because they are devised to link directly to pre-established curriculum or learning standards. The common alternative is a norm referenced test that covers general material appropriate for a subject and grade but that is not as closely linked to the specific state teaching standards. In principle, all students could achieve the maximum score on a criterion referenced test with no variation across students, while norm referenced tests focus on obtaining information about the distribution of skills across the tested population. In practice, scores on the two types of tests tend to be highly correlated.

subject matter standards for each grade. This focus dictates that the variation in test score gains generated by differences in instructional quality differs across the initial achievement distribution. For example, the additional gain in test scores resulting from a substantial improvement in the quality of instruction may be quite sizeable for a student who begins at the lower end of the skill distribution and for whom the test covers much of the knowledge gained by virtue of any higher teacher quality. On the other hand, a student higher up the initial skill distribution may answer most of the questions correctly even if taught by a quite low quality teacher. Better instructional quality may translate into only a few additional correct answers if the test does not concentrate on or cover the additional knowledge generated for this student by the superior instruction.

Figure 1 provides some insight into the potential magnitude of this problem. This graph plots two elements of the test score distribution. The test scores for students on the TAAS mathematics test are divided into ten equal score intervals. The solid line (corresponding to frequencies on the left axis) shows the distribution of scores across students. This distribution is highly skewed with a significant proportion in the highest two score intervals of correct test answers, where improvements would be quite difficult because of the test score ceiling. The dashed line (corresponding to the axis of raw gains measured in standard deviations of the test) portrays the distribution of average gains during a school year of students whose initial test scores fall into each score interval. This graph makes it very clear that typical gains at the bottom are much higher than in the upper ranges. Part of this could reflect regression to the mean induced by measurement error, but the problem is not simply one of the bounds on the tests as seen by the consistent pattern of gains in the middle of the distribution. The distribution of test questions in terms of degree of difficulty also suggests that those who begin at a lower level likely gain more knowledge about items examined on these tests. A rank order statistic such as percentile mitigates problems introduced by any correlation between initial knowledge and expected gain, but this transformation does not eliminate the problem that identical differences in teacher quality do not produce identical variations in average student improvement (in this case percentile

7

Figure 1. Relative Frequencies and Achievement Gains by Score Interval of Initial Test Scores

changes) across the initial skill distribution. Permitting the pretest score to have a coefficient less

than one by moving it to the right hand side provides an alternative solution. However, the

assumption of linearity between pretest and post-test scores is likely to be inappropriate (e.g., the

test may focus on the center of the skill distribution), and the inclusion of a noisily measured,

endogenous variable on the right hand side introduces additional problems.[8]

We adopt a direct approach for standardizing gains that permits comparisons of teachers

in classrooms with very different initial test score distributions. First, we divide the initial test

score distribution into ten equal score intervals (c_m for m=1,…,10) and for each year compute the

mean and the standard deviation of the *gains* for all district students starting in that interval.

Specifically, suppressing the notation for year and school, for all students with A_{ig-1} in the

interval c_m defined by [$\underline{A}_{g-1}^{c_m}, \widehat{A}_{g-1}^{c_m}$] for the given year,

$$(4) \qquad\qquad \mu_g^{c_m} = \overline{(A_{ig} - A_{ig-1})}, \text{ and}$$

$$(5) \qquad\qquad \sigma_g^{c_m} = \sqrt{\sum\left((A_{ig} - A_{ig-1}) - \mu_g^{c_m}\right)^2 \Big/ n_{c_m}}$$

The standardized gain score for each student in interval c_m is then calculated as:

$$(6) \qquad\qquad G_{isg} = \left[(A_{isg} - A_{isg-1}) - \mu_g^{c_m}\right]\Big/\sigma_g^{c_m}$$

Consequently, gains are distributed with mean zero and standard deviation one *for each*

score interval in each year, and teachers are judged based on their students' gains relative to other

---

[8] If the pretest score is measured with error, the usual problems of errors-in-variables occurs, and this will
be exacerbated by persistent errors for individuals. Bacolod and Tobias (2003) actually estimate a semi-
parametric nonlinear relationship between pre- and post-test scores and subsequently rely on residuals from
predicted scores for individuals to assess schools; see also Tobias (2004). Their approach is quite similar to
our own in spirit, although problems related to the inclusion of the pretest as a regressor remain. Hanushek
(1992) does not, however, find that test measurement errors are important in biasing the estimates of
educational production functions that include pretest scores on the right hand side.

students in the same place in the initial test score distribution.[9] This normalization also permits us to explore the possibility that effectiveness for teachers varies across the distribution, such as might occur if teachers tend to specialize in certain types of students.

This transformation of the dependent variable does not eliminate problems caused by errors in measurement and test reliability. Importantly, in contrast to parametric models for which classical measurement error in the dependent variable does not introduce bias, such error does contaminate estimates of the variance in teacher quality derived from teacher fixed effects. As previously discussed in Aaronson, Barrow, and Sander (2003) and Rockoff (2004), any such error inflates estimates of the variance in teacher quality.[10]

Our estimates of teacher quality, $\hat{t}_j$, are conditional means of student performance and necessarily include any aggregate test measurement error for classrooms. We now rewrite $\hat{t}_j$ as the sum of the true quality index (t_j) and an error component (v_j).

$$(7) \qquad\qquad \hat{t}_j = t_j + v_j$$

Assuming classical measurement error, the variance of $\hat{t}_j$ is the sum of the true variance of quality (t_j) and the variance of the error.

We use of multiple years of information for teachers to obtain a direct estimate of the variation in teacher quality. Specifically, we estimate a separate fixed effect for each teacher/year combination and use the year-to-year correlation among the estimated fixed effects for each teacher to eliminate the contribution of random error. If we have estimates for teacher effects for two years ($\hat{t}_j^{(1)}, \hat{t}_j^{(2)}$) and if the measurement error in teacher quality is independent across years, then the expected value of the correlation of teacher estimates, $E(r_{12})$, is:

---

[9] Note that simple gains models that include student fixed effects implicitly compare students at similar places in the skill distribution.
[10] Kane and Staiger (2001, 2002) also highlight problems introduced by measurement error in the related context of school aggregate scores.

$$(8) \qquad E(r_{12}) = var(t)/var(\hat{t})$$

Multiplication of the estimated variance of *t* by the year-to-year correlation thus provides an estimate of the overall variance in teacher quality corrected for measurement error. Importantly, this approach addresses problems related to both the noisiness of tests as measures of learning and any single year shocks (either purposeful or random) in classroom average student quality.[11]

The removal of all non-persistent variation also eliminates some portion of the true variation in quality resulting from changes in actual effectiveness. To address one portion of this problem, we remove first year teachers from the correlation calculations, because of the sizeable gains to experience in the initial year (see Rivkin, Hanushek, and Kain (2005)and below).

Yet the estimate of a teacher's quality based on within school variation may also change over time even in the absence of any true difference in performance, because the within school estimates of quality are defined relative to other teachers. Any turnover can alter a teacher's place in the quality distribution in her school. Therefore, by considering just the persistent quality differences (equation 8), some true systematic differences in teachers are masked by a varying comparison group and are treated as random noise – amplifying the downward bias in the estimation of the variation in teacher quality.

Texas Schools Project Data

The cornerstone of the analysis of teacher quality is a subset of the unique stacked panel data set constructed by the Texas Schools Project of the University of Texas at Dallas. The data on students, teachers, schools and other personnel in the Texas Schools Microdata Panel (TSMP)

---

[11] This differs significantly from corrections based on the sampling error of estimated fixed effects. Those approaches are directly related to the equation error variance in the sample, so that anything that reduces that error – including sample selection, the classroom sorting of students on the basis of unobservables, and simply adding more regressors to the model – tends to reduce the estimated error and thus the measurement error correction. See Aaronson, Barrow, and Sander (2003) and Rockoff (2004). Simply splitting the sample and using the correlations among the samples to correct the variance would also fail to remove the influences of annual shocks, implying that the split samples would have correlated errors.

have been augmented to include the details of classroom assignment of students and teachers for one large urban district – the "Lone Star" district.

TSMP contains administrative records on students and teachers collected by the Texas Education Agency (TEA) from the 1989-90 school year through 2001-2002. Student and teacher records are linked over time using identifiers that have been encrypted to preserve data confidentiality. These identifiers enable us to follow both teachers and students who switch schools within the Lone Star district and who move to another public school in Texas.

The student background data contain a number of student, family, and program characteristics including race, ethnicity, gender, and eligibility for a free or reduced price lunch (the measure of economic disadvantage). Students are annually tested in a number of subjects using the Texas Assessment of Academic Skills (TAAS), which was administered each spring to eligible students enrolled in grades three through eight. These criterion referenced tests evaluate student mastery of grade-specific subject matter, and this paper presents results for mathematics. The results are qualitatively quite similar for reading, although, consistent with the findings of our previous work on Texas, schools appear to exert a larger impact on math than reading in grades 4 through 8 (see Hanushek, Kain, and Rivkin (2002) and Rivkin, Hanushek, and Kain (2005)).[12]

Teacher and administrative personnel information in the TSMP include characteristics such as race/ethnicity, degrees earned, years of experience, certification test results, tenure with the current district, specific job assignment, and campus.

Importantly, for this paper we employ a significant extension to the basic TSMP database that permits the linkage of students and teachers for a subset of classrooms. TEA does not collect information linking individual students and teachers, but for this study we use additional information obtained from the Lone Star district to match a student with her mathematics teacher in each year. This is typically a general classroom teacher for elementary school and a specialist

---

[12] Part of the difference between math and reading might relate specifically to the TAAS instruments, which appear (as seen in Figure 1) to involve some truncation at the top end. For math, the outcomes are less bunched around the highest passing scores than they are for reading.

mathematics teacher for junior high.

In this paper we study students and teachers in grades 4 through 8 for the school years 1995/1996 to 2000/2001. We eliminate any student without a valid test score and teachers with fewer than ten students having valid test score gains.

The Distribution of Teacher Quality

A fundamental issue is how much variation in teacher quality exists. If small, policies to improve student performance should concentrate on issues other than the hiring and retention of teachers. This section describes the distribution of teacher quality in Lone Star district and examines the sensitivity of the estimates to controls for student, peer, and other differences across schools and years. By using the persistence of teacher effects over time, we account for the contribution of measurement error to estimated differences among teachers. A complementary analysis considers the roles experience, certification examination scores, and educational attainment play in explaining differences in teacher quality. Finally, we extend the basic modeling to consider the role of match quality that potentially leads teachers to be differentially effective with students of varying ability or backgrounds.

Performance Variations Across Teachers

We begin with a basic description of the contributions of teachers, principals, and other institutional features of district schools in shaping student achievement. Table 1 reports the between classroom (teacher by year) variance, the adjacent year correlation of estimated teacher value added, and the measurement error adjusted estimate of the variance in teacher quality for different specifications. The first and second columns use both within and between school variation, while the third and fourth use only within school variation. In addition, the second and fourth specifications regression-adjust for differences in observable characteristics. Differences

Table 1. Classroom and Teacher Differences in Student Achievement Gains

	Within district		Within school and year	
	Unadjusted	Demographic controls[b]	Unadjusted	Demographic controls[b]
Teacher-year variation[a]	0.210	0.179	0.109	0.104
Adjacent year correlation	0.500	0.419	0.458	0.442
Teacher quality variance / (s.d.)	0.105 **(0.32)**	0.075 **(0.27)**	0.050 **(0.22)**	0.047 **(0.22)**

Notes: a. The columns provide the variance in student achievement gains explained by fixed effects for teachers by year.
b. Student characteristics are eligibility for free or reduced lunch, gender, race/ethnicity, grade, limited English proficiency, special education, student mobility status, and year dummy variables.

among the specifications provide information on the extent of student sorting and on the magnitude of within relative to between school and year variation in classroom average gains.

A comparison of the estimated variance across columns indicates the potential importance of factors correlated with classroom differences in achievement. Controlling for observable student characteristics and using only within school and year variation noticeably reduces the between teacher variance in standardized gain. As expected given that most sorting occurs among schools, controls for measured student heterogeneity have a much larger effect in specifications not restricted to within school and year variation.[13]

The second row reports the adjacent year correlations in estimated teacher value added. The magnitudes range from 0.50 to 0.42, suggesting that roughly half of the variance is persistent. These correlations show considerable stability in the impact of individual teachers, particularly when just compared to other teachers in the same school. Again the controls for student heterogeneity reduce the correlations much less in the within school and year specifications. (Since some of the year to year variation in estimated teacher value added relative to others in the school comes from random differences in students, these controls could potentially increase the estimated persistence).

The most conservative estimate of the variance in teacher quality in Column 4 is based entirely on within-school variations in student achievement gains controlling for measured student heterogeneity and measurement error. Although this specification eliminates any between school variation in teacher quality and changes over time in the quality of instruction for a given teacher, it protects against variations in the effectiveness of principals, school average student characteristics, random measurement error, year to year differences in student ability and the like.

Despite the fact that it almost certainly understates the true variations in the quality of instruction, the variance estimate of 0.047 indicates the presence of substantial differences in

---

[13] Student characteristics are eligibility for free or reduced lunch, gender, race/ethnicity, grade, limited English proficiency, special education, student mobility status, and year dummy variables.

teacher quality when put in the context of student achievement growth. This implies that a one standard deviation increase in teacher quality raises standardized gain by 0.22 standard deviations. In other words, a student who has a teacher at the 85th percentile can expect annual achievement gains of at least 0.22 s.d. above those of a student who has the median teacher. Since these quality variations relate to single years of achievement gains for students, they underscore the fact that the particular draw of teachers for an individual student can accumulate to huge impacts on ultimate achievement.

The range of ambiguity about the causal impact of teacher quality differences is also relatively small. The best bounds on the standard deviation of teacher quality are 0.22 to 0.27. The higher estimate controls for measured differences among students (race, subsidized lunch status, LEP, and special education status), but allows between school teacher quality differences.

These estimates can be compared to the bound on teacher quality differences reported in Rivkin, Hanushek, and Kain (2005). That lower bound of 0.11 standard deviations estimate is based entirely on within school differences over time for the same students, subject to measurement error that almost certainly attenuates the estimate, and based on specifications that control comprehensively for possible sources of upward bias. Importantly, however, that estimate is not directly comparable to the estimates here, because it is based on the distribution of raw gains, which have a standard deviation of approximately two-thirds of the standard deviation of the standardized gain used here. Putting our current estimates on the same scale, two-thirds of our within-school estimate of 0.22 equals slightly less than 0.15.

The finding of significant quality variation within school and years coupled with the large annual turnover of teachers (below) enters directly in discussions of teacher performance incentives and teacher personnel practices more generally. First, the regression adjusted teacher average test score gains in a given year capture important variation among teachers within and almost certainly between schools, even if they are noisy measures of teacher quality. In contrast,

an incentive program having just the test score level provides far less information on teacher

value added.

Second, any incentive program (such as incorporated in many state accountability

systems) that focuses just on between school performance ignores the primary source of quality

variation that occurs within schools. The importance of within school variations also highlights

problems with the suggestion in Kane and Staiger (2002) that accountability systems should

aggregate test scores over time in a way that produces the least noisy estimate of school average

performance.[14] The changing cadre of teachers in a school certainly contributes to the year-to-

year variation in school average performance, and turnover is likely to lead to larger year to year

differences in small schools in which the school annual average variation in teacher quality will

generally be larger. Not only does such intertemporal averaging miss the majority of real

variation in teacher quality, but, more importantly, it also confuses the performance of current

teachers with that of their predecessors and thus provides a weaker incentive for improvement.

Measurable Teacher Characteristics

Prior studies suggest that most observable characteristics other than experience explain

little of the variation in teacher quality (Hanushek (2003)). This section adopts the common

education production function framework but extends it to circumvent analytic problems that

have plagued prior work. The basic formulation is:

$$(9) \qquad\qquad G_{isg} = f(X_{ig}) + \alpha_{TC} TC_{ijg} + (\gamma_i + \varepsilon_{isg})$$

where *TC* is a vector of measurable teacher characteristics with associated impact parameters,

α_{TC}. Although the adjusted gain specification deals directly with fixed factors that enter into

achievement differences, the inclusion of time varying family effects and, at times, individual

---

[14] Kane and Staiger (2002) consider a variety of dimensions of measurement error and appropriately
highlight the error introduced by separately evaluating small subgroups of the students. The text discussion
refers to aggregate school measures.

student fixed effects addresses issues of nonschool factors that systematically affect the growth in student performance.

We concentrate on the effects of a master's degree, a passing score on certification examinations, and teacher experience on standardized math score gains. These are particularly important characteristics because they are directly linked to teacher compensation. The initial estimates concentrate on the sample of teachers with information on certification tests. While we do not know the precise scores on the relevant certification tests, we do know whether they passed the first time they took the test and whether they have ever passed the test.[15]

The hypotheses of no significant differences in teacher value added on the basis of teacher education or certification examination performance cannot be rejected at conventional levels regardless of whether student fixed effects are included. These findings shown in Table 2 reinforce prior studies and raise serious questions both about the desirability of requiring or rewarding with higher pay those with a post-graduate degree and about the efficacy of the existing certification procedures in Texas and similar systems in other states.

To describe the impact of teacher experience, we return to the more complete sample of teachers (which does not exclude teachers without certification data).[16] Existing evidence suggests that most improvement occurs very early in the career (see Hanushek and Rivkin (2004), Rivkin, Hanushek, and Kain (2005)), but experience may affect student achievement gains through a number of channels. The first is learning by doing. The second is nonrandom selection: If less talented teachers are either more or less likely to quit than more talented peers on

---

[15] Note that there are many different certificates, and thus many different tests, for which teachers can qualify. We make no attempt to distinguish among alternative certificates. The testing information is not available for all teachers, thus reducing the sample to those with valid test information. In general, this information is more commonly available for younger teachers, reflecting the history of the certification program. The sample used in the certification regressions is roughly 60 percent as large as the sample used in the experience regressions.

[16] Note that the estimation in Table 2 also includes teacher experience, and the results for experience are qualitatively similar to those reported below. Our measure of experience is total years of teaching in any Texas public school, not just in the Lone Star district. Experience in private schools or in schools outside of Texas is not observed. Note also that the multiple observations on experience are used to correct inconsistencies in the raw data; see Rivkin, Hanushek, and Kain (2005).

Table 2. Estimated Impacts of Advanced Degree and Passing score on the Certification Examination (absolute value of t-statistics in parentheses)

	No fixed effects	With student fixed effects
Master's degree	0.015	0.004
	(1.17)	(0.27)
Pass certification exam	0.002	-0.002
	(0.08)	(0.07)
Pass certification exam on first attempt	0.015	-0.039
	(0.91)	(1.85)

Note: Each coefficient comes from a separate regression. All specifications include full sets of experience, year, and grade dummy variables and indictors for special education classification, eligibility for a subsidized lunch, limited English proficient classification, a structural move from elementary to junior high school prior to the academic year, and a non-structural school change prior to or during the academic year. Specifications without student fixed effects also include gender and race/ethnicity dummy variables. Sample sizes are 230,000 for the Master's degree regressions and 141,744 for the certification examination regressions.

average, estimates of the return to experience capture the change in the average quality of the teaching pool. Finally, teachers may vary effort systematically with experience in response to tenure decisions or other institutional and contractual issues. Each of these causal links raises the possibility of a highly nonlinear relationship between the quality of instruction and experience. Therefore we include a series of dummy variables indicating first, second, third, fourth and fifth year teachers. (Teachers with more than five years experience are the omitted category. Preliminary analysis, not shown, found no experience effects beyond five years of experience).

The results in Table 3 highlight the much lower average performance of first year teachers. Notice that the inclusion of student fixed effects does not alter significantly the return to experience, but the addition of teacher fixed effects reduces the penalty for first year teachers by roughly 25 percent and also eliminates any quality deficit for second year teachers.[17] This pattern indicates some selection effects in that inexperienced teachers who exit following the school year are systematically less effective than other teachers. We will return to this issue in the transition section below. Finally, it appears that fourth year teachers perform systematically better than others, suggesting the possibility that average incentives are quite strong in the fourth year. While Texas schools do not have formal collective bargaining, the character of contracts and teacher reviews may influence the incentives at this career point.

These experience effects indicate that the high turnover among U.S. teachers, and particularly urban teachers, has detrimental effects on student achievement. For Texas, some ten percent of teachers with 0-2 years of experience and 7 percent of all teachers leave teaching each year, requiring replacements who generally arrive with no experience.[18] For the Lone Star

---

[17] A different pattern appears if raw gains are used instead of the standardized gains. The first year effect is smaller in raw gains with no fixed effects than in Table 3, but increases by almost 40 percent with the introduction of student fixed effects. We interpret this pattern as confirming the standardization of gains for the TAAS test. The fixed effects serve the same function as the test score transformation by restricting the estimation to comparisons of students with others (in this case oneself) at the same part of the test score distribution.

[18] These figures refer to 1994-96 in Texas (Hanushek, Kain, and Rivkin (2004)). The rate of new hires

Table 3. Estimated Impact of Experience Level
(comparisons to teachers with 6 or more years experience)

Year teaching	No fixed effects	With student fixed effects	With student and teacher fixed effects
1st year	-0.16	-0.16	-0.12
	(8.36)	(9.60)	(8.01)
2nd year	-0.03	-0.03	0.00
	(1.55)	(1.82)	(0.29)
3rd year	0.03	0.02	0.02
	(1.43)	(1.28)	(1.26)
4th year	0.05	0.08	0.06
	(2.38)	(3.42)	(3.14)
5th year	0.04	0.03	0.01
	(1.76)	(1.51)	(0.55)

Note: All specifications include full sets of year and grade dummy variables and indictors for special education classification, eligibility for a subsidized lunch, limited English proficient classification, a structural move from elementary to junior high school prior to the academic year, a non-structural school change prior to or during the academic year, and that the teacher has an M.A. Specifications without student fixed effects also include gender and race/ethnicity dummy variables. The sample size is 230,000.

district, similar to other large urban districts, the annual exit rates from the district for teachers with 0-3 years experience are close to 20 percent (see Appendix A, below). The first year effects estimated here show that having a first year teacher on average is roughly equivalent to having a teacher a half standard deviation down in the quality distribution.

Teacher-Student Matching

The analysis presumes that each teacher can be ranked according to a single underlying dimension of quality and that we can infer this from information about student progress. This may, however, not be a satisfactory characterization. Teachers may specialize with particular skill groups or merely decide to target a particular skill range. Student and teacher demographic characteristics may also influence the quality of student/teacher interactions. Each of these cases alters the concept of teacher quality by raising the possibility that it is not constant for all students. Importantly, the distribution of quality that we trace out may also be influenced by differences in the nature of classroom matching across schools and principals.

To investigate specialization by student achievement, we divide students into three academic preparation classifications (based on initial scores) and compute the correlation between the teacher average gain for students in one category with the teacher average standardized gain for students in the other categories.[19] The positive correlations of 0.45 between the low and middle categories, 0.57 between the high and middle categories, and 0.31 between the low and high categories refute the notion that the effects of any curricular targeting or matching are large relative to the impact of overall teacher quality. These correlations are in the same general range as those across years (Table 1). The strong positive correlation between the average standardized gains in the top and bottom categories is particularly striking given the

---

varies some over time, depending on student demographics, the extent of teacher retirement, and the numbers of returning teachers who have prior experience.

[19]The ten categories used to produce the standardized gain measure are aggregated into three using the district average distribution of students over all years as the fixed weights for all teachers.

relatively small number of students in the bottom category in schools with large numbers of students in the top category and the large error variance described above.

The possibility of differential effects by the matching of teacher and student race also exists, if for example students respond better to teachers in the same ethnic group. The re-analysis of the Coleman Report data by Ehrenberg and Brewer (1995) suggested a positive race matching effect but raised questions about how quality differentials of black teachers might work against race matching with students. Dee (2004), using the random assignment data from the Tennessee STAR, finds strong evidence for beneficial effects from matching the race of teacher and students.

Table 4 provides descriptive information about the matching of teachers and students in the Lone Star district. (The table identifies teachers and students who are black, Hispanic, and white and omits separate tabulations of the small numbers of both teachers and students of other ethnic backgrounds). Few teachers are Hispanic, making the black and white comparisons the interesting part.

Black teachers are disproportionately matched with black or Hispanic students. In contrast, white teachers, who make up approximately 40 percent of the teaching force, end up teaching two-thirds of the white students. These patterns are consistent with the race differences in teacher mobility patterns found by Hanushek, Kain, and Rivkin (2004) . In an analysis of the entire state of Texas, they found that white teachers typically moved to schools with fewer minority students and with higher achieving students. In contrast, black teachers moved to schools with more minority students, although they still sought higher achieving students.

Table 5 turns to the achievement implications of this sorting. The table reports results for regressions that include those covariates used in the experience regressions as well as indicators for black and Hispanic teachers and interactions between teacher and student race and ethnicity. (Asian and Native American students and teachers are excluded because of their small numbers).

19

Table 4. Joint Distribution of Black, Hispanic, and non-Hispanic White Students and Teachers by Race/Ethnicity

Teachers	Students			
	Black	Hispanic	White	**All**
Black	27.7	21.0	4.2	**54.1**
Hispanic	0.5	3.2	0.3	**4.0**
White	10.0	18.2	9.4	**39.5**
All	**38.6**	**43.9**	**14.1**	

Note: Neither the joint nor marginal probabilities sum to 100 because of the omission of Asian and Native American Teachers and Students.

Experience dummy variables are excluded to capture the full effect of race differences, but the results from specifications with experience controls are almost identical.

The clearest finding is that, regardless of whether the specification includes student fixed effects, black teachers tend to be more effective with minority students. Estimates range in size from 0.05 to 0.10 standard deviations, or somewhere between 30 percent and 70 percent of the cost of having a first year teacher. The final column of Table 5 suggests that the benefit of a same race teacher is higher for girls than boys, but the difference is small and not statistically significant.

On the other hand, the table reveals contradictory findings on the question of ethnic differences in average teacher quality. Without student fixed effects black teachers appear to be significantly less effective than white teachers (the omitted category), but with student fixed effects the average effectiveness of black teachers appears to exceed that of white teachers. Moreover, the first specification (without fixed effects) suggests that despite the benefits derived from being matched with a same race teacher, average quality of instruction received by black students with white teachers exceeds the average quality received by black students with black teachers. In contrast, the fixed effect specifications show just the opposite.

Which specification is more informative? Two factors enter the interpretation of the models with student fixed effects. First, and most straightforward, the change in the apparent impact of black teachers partially reflects assignment of black teachers to schools with more difficult to educate students. Table 4 shows that in the Lone Star district white teachers tend to teach fewer minority students but the matching is even more skewed than that suggests. Only 56 percent of black students taught by white teachers are disadvantaged (as measured by free and reduced price lunch status), while 71 percent of black students taught by black teachers are disadvantaged; further, black teachers teach twice the proportion of disadvantaged white students as the white teachers do. In short, the measured, and likely the unmeasured, characteristics of the

20

Table 5. Effects of teacher and student race matching on standardized achievement gains

	No fixed effects		With student fixed effects		With student and teacher fixed effects	
Teacher black	-0.057	-0.082	0.025	-0.015		
	(5.12)	(5.13)	(2.34)	(1.02)		
Teacher Hispanic	-0.034	-0.087	-0.011	-0.068		
	(1.34)	(2.11)	(0.48)	(1.65)		
Teacher black*student black		0.047		0.102	0.105	.119
		(2.40)		(5.31)	(7.89)	(7.74)
Teacher black*student male						0.001
						(0.06)
Teacher black*student black*student male						-0.031
						(1.54)
Teacher black*student Hispanic		0.040		0.036	0.030	0.030
		(2.30)		(2.10)	(2.72)	(2.61)
Teacher Hispanic*student black		0.049		0.076	0.019	0.020
		(0.83)		(1.30)	(0.47)	(0.48)
Teacher Hispanic*student Hispanic		0.079		0.084	0.020	0.020
		(1.85)		(1.94)	(0.82)	(0.82)

Note: All specifications include full sets of year, and grade dummy variables and indictors for special education classification, eligibility for a subsidized lunch, limited English proficient classification, a structural move from elementary to junior high school prior to the academic year, and a non-structural school change prior to or during the academic year. Specifications without student fixed effects also include gender and race/ethnicity dummy variables. Asian and Native American teachers and students are excluded. The sample size is 216,958.

student bodies tend to be correlated with teacher race. The introduction of student fixed effects is designed to control for systematic but unobserved student heterogeneity.

The inclusion of student fixed effects, however, does more than simply control for unobserved student heterogeneity. With student fixed effects, the estimates of the average impact of a black teacher come entirely from variations in average achievement patterns of students who have both a black and a white teacher over the observation period. Consequently, if the extensive sorting of students and teachers among schools described in Table 4 and the distribution of subsidized lunch students is accompanied by systematic differences in teacher effectiveness, the estimated racial gaps in teacher effectiveness could differ across specifications even in the absence of unobserved student differences. (This discussion is an extension of the earlier discussion about identification of teacher fixed effects when school fixed effects are also included).

To see the complications, consider the following case that is consistent with the aggregate move patterns of teachers: 1) The average effectiveness of white teachers weighted by the share of students who also have a black teacher is below the overall average effectiveness of white teachers; and 2) the average effectiveness of black teachers weighted by the share of students who also have a white teacher is roughly equal to the overall average for black teachers. Assuming no unobserved student or school differences, these patterns imply that the addition of student fixed effects would raise the estimate of the average effectiveness of black teachers relative to white teachers. Thus, the estimates in Column 3 indicate that the typical black teacher is better than the typical white teacher in the same school setting, but it is not possible to provide a precise comparison to the average white teacher found across the district.

Notice that these same ambiguities from the inclusion of student fixed effects potentially complicate the interpretation of the differences between the estimated benefits of having a teacher with the same ethnicity. Nevertheless, the fixed effect specification shows clearly that black

students who have had both a black and white teacher perform better relative to classmates during the year in which they had the teacher of their own race.

The underlying mechanism behind the matching gains (to both white and black students) from having a same race teacher cannot be identified from this analysis. It is not possible, for example, to distinguish between positive mentoring effects that raise the performance of a black student with a black teacher and negative learning effects that lower the performance of a white student with black teachers.

Teacher Transitions

The finding of significant variations in teacher quality that are not captured by simple observed characteristics is interesting from a research perspective but does not necessarily have any direct policy implications. Even if researchers do not fully understand teacher quality differences, policy makers, educators, and parents may understand these differences and may act effectively on them. To investigate implications for policy, we turn to the operation of the teacher labor market.

The high rate of teacher turnover in large urban districts engenders considerable concern among educators, but the absence of evidence on the link between actual performance in the classroom and transitions makes it difficult to judge the importance of this concern. One aspect of this – the effect of initial teacher experience – has already been noted. Here we complement that with a comparison of the overall effectiveness of teachers who exit the Lone Star District with those who remain.

We divide teachers into four mobility categories for each year that they teach: remaining in the same school, moving to a new school in the Lone Star District, moving to a new school

outside of Lone Star, or exiting the Texas public schools entirely.[20] These categories correspond to those common to policy discussions about teacher turnover.

Three features of teacher mobility rates are important. First, consistent with a number of prior studies, teacher turnover is large (see, for example, Boyd et al. (2002), Hanushek, Kain, and Rivkin (2004), and Podgursky, Monroe, and Watson (2004)). As described in detail in Appendix A, the turnover of inexperienced teachers is especially high. Only 70 percent of teachers with less than three years of experience remain in the same school from year to year. Second, teacher turnover is systematically related to characteristics of the student body, most importantly the achievement level of students in a school. Third, and relevant for the subsequent estimation of mobility patterns, teachers who change districts on average see lower salary increases in the year of transition than those who remain in the Lone Star district.

Our direct estimates of quality differences permit tracing the achievement implications of the observed transitions. Table 6 reports estimates of differences in teacher quality by transition type for a series of regressions that differ according to whether or not they control for student fixed effects, school-by-year fixed effects, and the status of women teachers who return following a one year hiatus (which may have been a maternity leave). The first three columns ignore any subsequent return, while for the final specification women returnees are reclassified based on where they teach in the year following their return. Note that the school-by-year fixed effect specifications generate coefficients based on achievement differences within schools. All estimates compare those who leave a school with those who remain.

The estimates in Table 6 provide little or no evidence that more effective teachers have higher exit probabilities. On the contrary, those who exit are significantly less effective on average than stayers regardless of whether they are compared to all stayers in the district or only those in the same school and year. Moreover, those who switch campuses within the same

---

[20] There is no distinction between involuntary and voluntary separations, because such information is not available, but past analysis suggests that virtually all transitions out of teaching are teacher initiated.

Table 6. Estimates of Mean Differences in Teacher Quality by Transition Status

(Standardized Gains compared to teachers remaining in same school; absolute value of t statistics in parentheses)

		With student fixed effects		
			With school by year fixed effects	With reclassification of women returnees
Change campus	-0.089	-0.061	-0.054	-0.060
	(3.96)	(2.69)	(2.59)	(2.65)
Change district	-0.011	-0.031	-0.023	-0.028
	(0.36)	(1.05)	(0.78)	(1.02)
exit public schools	-0.044	-0.089	-0.072	-0.095
	(1.90)	(3.83)	(3.53)	(3.77)

Note: All specifications include full sets of experience, year, and grade dummy variables. The sample size is 230,000.

district are also significantly less effective, while teachers who switch districts do not appear to differ significantly from the stayers.

These mean differences are certainly informative, but they do not paint a comprehensive picture of the distributions of stayers and movers. It is important to know if movers come disproportionately from the tails of the distribution. Are inner city schools actually losing a large number of the most promising teachers to other districts? Do those who really struggle in the classroom have a very high rate of attrition?

In order to learn more about quality differences by transition status, Figures 2 plots kernel density estimates of the distributions of teacher fixed effects by move status based on regressions of adjusted student gain on a full set of teacher-by-year fixed effects, teacher experience dummies included above, and student characteristics. Because of the aforementioned sorting of teachers among schools, we also plot estimated teacher fixed effects produced by specifications that include school fixed effects (Figure 3). Regardless of the specification, however, the distributions of those who either change campuses or exit public schools fall distinctly below those who stay, while quality distributions for those who change districts are quite similar to those of the stayers.

Although the specifications control for experience effects, the differences across transition categories may differ systematically by experience. Therefore, Table 7 reports separate estimates of differences by transition type for teachers with one, two, and three years of experience. Unfortunately most of these coefficients are not precisely estimated, but two distinct experience patterns do emerge. In the case of within district campus changes and exits out of the Texas public schools, the largest quality gaps arise for teachers who transition out following their second and third years. In the case of district switchers on the other hand, the younger movers tend to be slightly above average in performance, although this difference is not statistically significant and any quality premium appears to decline (and even reverse) with experience.

Interestingly, plots of the full distribution of teachers in the lower experience categories (not shown) give some idea of the source of the mean differences that were identified. The

Figure 2. Kernal Density Estimates of Teacher Quality Distribution: Standardized Average Gains by Teacher Move Status
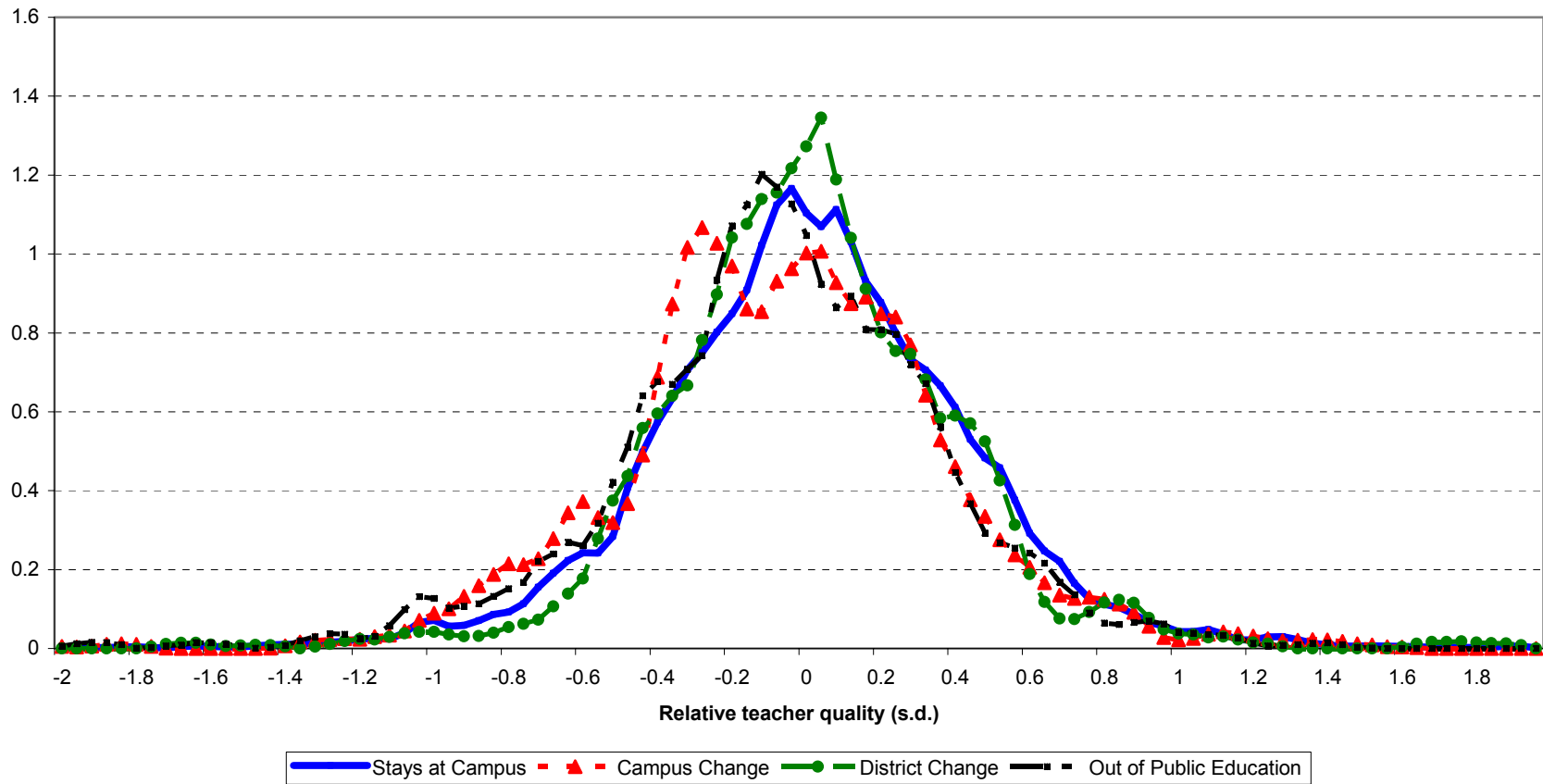
Figure 3. Kernal Density Estimates of Teacher Quality Distribution: Standardized Average Gains Compared to Other Teachers at the Same Campus by Teacher Move Status
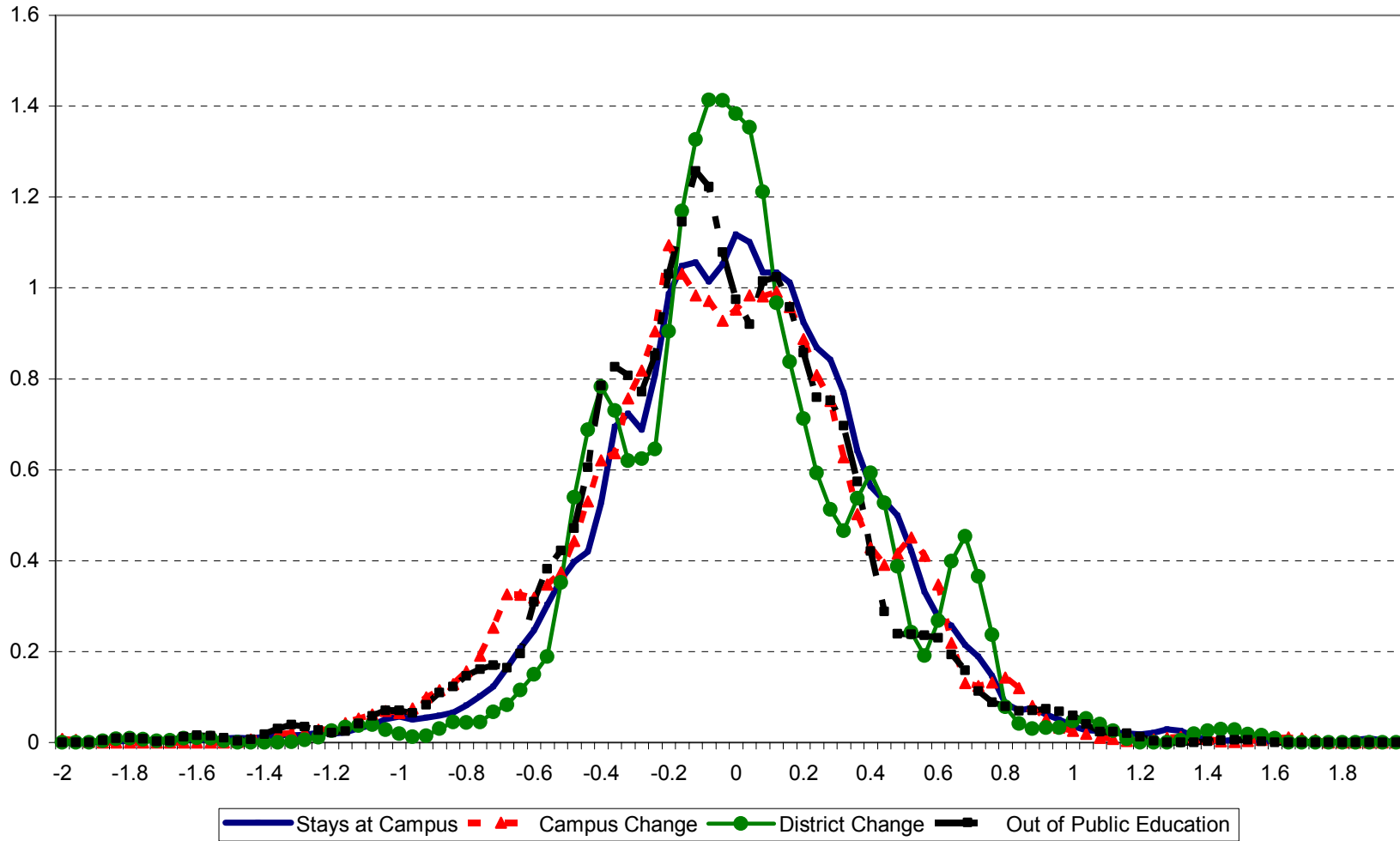
Table 7. Estimates of Differences in Teacher Quality by Transition and Experience
(Standardized Gains compared to teachers remaining in same school including student fixed effects; absolute value of t statistics in parentheses)

Teacher experience	Change campuses	Change districts	Exit Texas public schools
First year experience	-0.031	0.107	-0.071
	(0.45)	(1.51)	(1.40)
Second year experience	-0.130	0.062	-0.159
	(1.27)	(0.07)	(2.31)
Third year experience	-0.089	0.021	-0.173
	(1.46)	(0.28)	(2.73)
More than three years experience	-0.057	-0.082	-0.059
	(2.14)	(2.21)	(1.91)

Note: All specifications include full sets of experience, year, and grade dummy variables. The sample size is 230,000.

numbers of teachers in the transition groups by experience get rather small, but the positive mean for the inexperienced district changers appears to be driven by a small number of very good teachers who leave, and the distribution for the bulk of district switchers falls slightly to the left of those who do not move. For those who exit teaching, the right hand tail of quality is very similar to that for the stayers, but there is a noticeably thicker left hand half of the quality distribution for exiters.

A final issue is the interpretation of the finding that teachers who exit the Texas public schools are systematically less effective than those who remain. While these teachers may have been less effective in the classroom throughout their careers, it is also possible that the exit year was anomalous and not indicative of typical performance. For example, the exiting teacher might have had a particularly unruly class or might have reacted to some other bad situation in the school such as conflict with a new principal. An alternative possibility is that effort is reduced once the decision is made not to return and that at least a portion of the transition quality gap arises from the feedback effect of the decision to exit.

To investigate these possibilities, we measure teacher quality by student gains in the year prior to each transition. For example, we describe the distribution of quality for transitions following the 1999 school year with average student gains during the 1998 school year, meaning that any change in circumstances or effort following the decision not to return for the subsequent year does not affect the quality calculations. Note that this reduces sample size by eliminating student performance information on the final year taught for each teacher and all who teach only a single year in Lone Star district.

Table 8 reports two sets of coefficients, one based on lagged achievement gains and the second based on current achievement gains for the same sample of transitions. The table also compares movers both to all teachers and to just those in the same school through the inclusion of school fixed effects. Note that, although the point estimates for the current scores without school

Table 8. Differences in Teacher Quality by Transition and Year Quality Measured
(Standardized Gains compared to teachers remaining in same school; absolute value of t statistics in parentheses)

	Current year quality estimates		Lagged year quality estimates	
	Within district comparisons	Within school comparisons	Within district comparisons	Within school comparisons
change campus	-0.067	-0.033	-0.032	0.002
	(2.29)	(1.32)	(1.12)	(0.08)
change district	-0.021	-0.038	-0.024	-0.023
	(0.54)	(1.01)	(0.54)	(0.60)
exit public schools	-0.060	-0.067	0.004	0.001
	(1.90)	(2.46)	(0.12)	(0.05)

Note: All specifications include full sets of experience, year, and grade dummy variables. The sample size is 149,420.

fixed effects differ some from the comparable estimates in Table 6 that use the entire sample, the patterns are qualitatively the same.

Two findings stand out in the examination of exit year effects. First, those who leave a school within the Lone Star District tend to be below average in both the district and the specific school they are leaving during their final year. Second, and more important, the performance during the exit year is noticeably worse than in the previous year. This strongly suggests that those who exit are not systematically worse in a longer term sense but only in the year in question. Whether this reflects a reduction in effort or particular difficulties in that year (that might contribute to an exit decision) cannot be fully ascertained at this time. Nonetheless, the fact that the differences holds for the within school comparisons suggest that it is not simply a new principal or any school wide problem that is driving the results.

Who Hires More Effective Teachers?

The large changes in student characteristics observed for teachers who leave Lone Star district for another Texas public school are similar to those described in other work and strongly suggest that teachers consider these factors in making decisions about where to work.[21] Moreover, even though salary declines on average following a transfer out of Lone Star District, research indicates that current salary and alternative opportunities each affect transition probabilities once compensating differentials have been adequately accounted for, although the majority of teachers who exit the profession, at least in Georgia, do not procure jobs with higher salaries.[22] The crucial, unanswered question is whether schools take advantage of the attractiveness of their student body (or, more likely, amenities correlated with student characteristics) or higher pay to procure a higher quality teacher.

---

[21] See Boyd et al. (2002) for evidence on New York State teachers, Hanushek, Kain, and Rivkin (2004) for evidence on Texas teachers, and Scafidi, Sjoquist, and Stinebrickner (2002) for evidence on Georgia teachers.

[22] See Boyd et al (2002) and Hanushek, Kain, and Rivkin (2004) for discussions of compensating differentials. Scafidi, Sjoquist, and Stinebrickner (2002) follow teachers when they leave teaching in Georgia to ascertain the change in income for those exiting.

Ballou (1996) presents evidence that districts do not hire the best candidates based on measurable characteristics and that administrators may trade off quality for lower expected turnover. Since experience appears to have substantial impacts early in the career, such a tradeoff could be justified if the quality differential is not too large – but prior analyses are mute on this..

To learn more about the value of quality in the teacher labor market, we examine the linkage between salary and school demographic characteristics on the one hand and teacher quality on the other for the 245 teachers who move to a new district. Previous work with flows of teachers suggests that salary and the composition of students are each, on net, an attraction to teachers. On the other hand, such districts might squander these advantages either because they pursue other objectives or because they are not able to measure quality.

Table 9 reports salary and student demographic coefficients for a series of teacher characteristic regressions that differ according to the "quality" measure that is the dependent variable. The first is our outcome based measure (the teacher value added coefficients from earlier regressions on standardized gain with an adjustment for experience), while the latter two are common observable teacher characteristics – advanced degrees and certification test results.

It is important to recognize that we do not offer direct evidence on district choices. That is, we do not observe which teachers applied to which schools or even which teachers looked to change schools or districts. Nonetheless, the apparent preferences of teachers for certain types of school environments and higher salaries strongly suggests that schools able to offer better compensation packages (including non-pecuniary amenities) can choose from a deeper applicant pool and thus can have a better chance of hiring high quality teachers. Consequently, if schools are able to determine teacher effectiveness *and* they choose to hire the most effective teachers, we should observe a systematic relationship between teacher quality and all aspects of compensation.

The results in Table 9 provide evidence that higher salaries and fewer black students may raise the probability of hiring a teacher with an advanced degree. There is, however, little systematic evidence that districts use higher salaries to procure better quality teachers defined in

27

Table 9. Estimated Effects of Salary and Student Demographic Characteristics on the Quality of Newly Arrived Teachers (absolute value of t statistics in parentheses)

Destination campus characteristics	Direct quality measure	Teacher quality proxies	
	teacher standardized gains	passed certification examination	advanced degree
log salary	0.12	-0.15	0.22
	(0.63)	(1.72)	(1.88)
% Black	0.0000	-0.0009	-0.0024
	(0.01)	(1.18)	(2.67)
% Limited English Proficient	0.0016	-0.0017	-0.0017
	(0.89)	(1.86)	(1.49)

Note: All specifications include experience, year, teacher gender, and teacher race/ethnicity dummy variables. The sample size is 245. Teacher fixed effects are obtained from the regression of adjusted gain on a full set of teacher dummy variables, student characteristics, and a full set of teacher experience dummy variables.

terms of student outcomes. The salary coefficient in the specification with our measure of quality
as the dependent variable does not approach statistical significance. Although the small sample
size certainly contributes to the imprecise estimates in the quality specifications, the consistency
with the evidence that shows that an advanced degree has little systematic relationship with the
quality of instruction reinforces the result.

Conclusions

Much policy debate revolves around the importance of teacher quality, but little
consistent evidence has been available about the importance or character of quality variations.
This paper investigates a number of dimensions of the distribution of teacher quality: the
magnitude of the variation in quality; the contributions of experience and teacher education to
that variation; and the importance of student teacher matching. This description expands
considerably what we know about the importance of teacher quality.

This analysis of teacher quality is, however, largely an input into understandings how
market forces play out in the teacher labor market. We employ our direct quality measures to
ascertain differences between those who remain in the large, urban district and those who leave
for other districts or professions. Further, we consider the extent to which suburban districts use
salary and student characteristics to attract better teachers.

Three methodological issues are important. First, because of the psychometric
characteristics of the Texas tests, it is crucial to standardize the test for the initial level of
achievement in order to compare teachers across the entire distribution. Second, we control for
potential nonteacher factors through both direct measurement of student body characteristics and,
at times, by restricting attention just to within-school variations in quality. Third, we use repeated
measures of teacher performance to obtain estimates of measurement error versus systematic
quality differences.

The analysis of the determinants of quality reveals some very important insights. First, there is significant learning about the craft of teaching in the first few years of teaching. The largest impact is the first year of experience, and experience effects disappear quickly after the first year. Second, teachers who do well with students in one part of the achievement distribution also do well with students in other parts of the distribution, suggesting that teacher-student matching by initial preparation is not a particularly important policy issue. Third, students benefit from having a same race teacher, quality held constant, and that benefit is sizeable.

The impact of these factors differs systematically by student ethnicity. Because of the pattern of changing schools by teachers, 12 percent of the black and Hispanic students are taught by teachers in their first year as compared to 7 percent of white students.[23] A new teacher lowers achievement growth by 0.12-0.16 standard deviations, holding overall teacher quality constant. If we cumulate these effects across six years of primary schooling, the experience differential could account for 0.04 s.d. average difference between whites and either blacks or Hispanics.

A commonly voiced policy concern is that large urban districts lose their better teachers to other occupations or to suburban schools. We find little if any support for the notion that the better teachers are the most likely to exit the public schools entirely. To the contrary, teachers exiting Texas public schools are significantly less effective on average in the year prior to leaving than those who remain, and those changing districts are quite similar in terms of effectiveness.

Similarly, there is little systematic evidence in support of the view that the urban district loses its better teachers to the suburbs. Much has been made of the fact that salary differentials in metropolitan areas exist and that these may frequently lead to a drain of high quality teachers from urban centers. This view is reinforced by analyses that show urban areas to be net suppliers of teachers to other districts and that show urban districts to lose teachers disproportionately from schools with low achievement and high minority populations. Although high turnover hurts

---

[23] The difference for teachers in their second year of experience is much smaller: 7 percent for black students, 8 percent for Hispanic students, and 6 percent for white students.

students because of the lower performance of inexperienced (replacement) teachers, the evidence does not support the related concern that the best teachers are those most likely to leave.

At the same time, looking at hiring patterns for districts employing teachers from the Lone Star district fails to find a strong relationship between quality and salary, suggesting that it may be quite difficult to discern quality from a teacher's past performance in another district. (The small sample size does, however, limit the precision of the estimates).

The identification of large variation in the quality of instruction within schools has at least two additional implications related to school accountability. Current state accountability systems typically aggregate the level of student scores for an entire school, in part on arguments about averaging out any measurement error. First, even if the average level of test scores has a larger signal to noise ratio, the appropriately constructed achievement gain is a much more appropriate measure of value added and school effectiveness. It is more closely related to current teacher performance and controls for important family and community differences that tend to confound estimates of teacher value added based on achievement levels. And second, any formal or informal school evaluation program that aggregates performance to the school level or across years misses the majority of the variation in the quality of instruction. This weakens the incentives for good teachers to enter and remain in teaching, for ineffective teachers to leave, and for all teachers to put forth greater effort. These costs should be compared with the superiority of measuring performance at the school level in terms of fostering greater teacher cooperation within the school, a concept frequently advanced in policy discussions.

Separating the elements of teacher quality from other possible influences on achievement is difficult, and some ambiguity will necessarily remain. Moreover, this analysis is further limited in important ways: It is clearly conditional upon both the test instruments and the institutional structure of Lone Star schools and their hiring patterns. Nonetheless, it provides both a methodology and a set of initial investigations of the market of teacher quality – something that has defied past analysis.

Appendix A. Description of Teacher Mobility in the Lone Star District

Appendix Table A1 shows the substantial teacher mobility in the Lone Star District, particularly among inexperienced teachers. Only 70 percent of teachers with fewer than three years of experience remain in the same school from one year to the next. New teachers have the highest transition rates: 12 percent switch schools within Lone Star District, 4 percent switch districts, and 14 percent exit the Texas public schools entirely following their first year as teachers. As teachers acquire experience, the probability of exiting declines steadily prior to rising again near retirement. On the other hand, experience does not substantially reduce movement within Lone Star District, and a substantial fraction of teachers with fewer than 10 years of experience move to a different Texas public school outside of Lone Star District.

Consistent with prior studies, the probability of moving differs by a number of classroom characteristics. Appendix Table A2 provides school average student demographic characteristics by annual teacher move status. This shows vividly that average math score is much lower for teachers who switch schools within Lone Star District (-0.38 standard deviations) and lower for those who exit the public schools entirely (-0.33 s.d.) or switch districts (-0.29 s.d.) than for teachers who remain in the same school.[24] Somewhat surprisingly based on our prior work for the state as a whole (Hanushek, Kain, and Rivkin (2004)), the differences in racial and ethnic composition and income are smaller and less systematic.

Because transition rates are higher for less experienced teachers, the two panels report these classroom characteristics for teachers in their first year and in their second or third years of experience, i.e., for the groups with the highest exit rates from their current school. While anecdotal stories of teacher placement suggest that beginning teachers work in classrooms that uniformly have the lowest achievement levels and the most disadvantaged students in terms of income or the race and ethnicity of the students, within the Lone Star District these patterns do not completely hold. As with all teachers, those switching schools within the district have the lowest average achievement in their classes (compared to other categories of move status). The racial composition and income composition, however, do not show much in the way of systematic patterns, quite possibly reflecting the very high minority enrollment and the generally low income levels of the district.

While Appendix Table A2 presents the characteristics of Lone Star District classrooms from which teachers exit, an alternative perspective is how campus characteristics change with a move. Appendix Table A3 reports the year-to-year changes in salary, average math achievement, and student demographic characteristics by experience and transition type for those who switch schools within the district and those who leave the district for another Texas public school. Not surprisingly, those who exit Lone Star District for another district see the most dramatic change in student characteristics. Average math achievement increases by 0.13 standard deviations and percentages Black, Hispanic, and students eligible for a subsidized lunch fall by 8, 15 and 28 percentage points, respectively. This compares to much smaller changes for those who switch schools within Lone Star District (or remain in the same school).

In contrast to the dramatic improvements in student achievement and increases in student SES, district switchers tend to experience smaller salary increases than those who remain in Lone Star District. For all teachers who do not move, the average salary increase is $2,137; first year

---

[24] All test scores, as described previously, are normalized to mean zero and standard deviation one for the state as a whole in each grade-test year. The overall scores for all teachers simply indicate that Lone Star District, like most large urban districts, has overall performance that is 0.25 standard deviations below the Texas state mean.

teachers who stay in the same school have salary increases averaging $3,204. But, the average salary increase for all district movers is $2,085 and for new teachers it is only $1,807. This pattern is consistent with research on the State of Texas as a whole (Hanushek, Kain, and Rivkin (2004)). Given that teachers initiate the vast majority of transitions and undoubtedly prefer higher salaries, this pattern indicates the existence of large compensating differentials which complicate the identification of the relationship between salary and the supply of teachers.

Overall, the changes by experience group are most interesting in their similarities to the "all teacher" results. While there are a few differences such as the decline in campus math scores for new teachers that change districts, most of the overall patterns are quite consistent.

Appendix Table A1. Teacher Transitions by Teacher Experience
(annual rates in percent)

Teacher Experience	No Move	Change Campus	Change District	Exit Public Schools
1 year	70.4	11.5	4.0	14.0
2-3 years	70.8	11.2	5.0	13.0
4-6 years	77.0	10.4	5.4	7.2
7-11 years	79.7	10.6	4.3	5.4
12-21 years	86.2	8.3	2.0	3.5
> 21 years	86.5	5.7	0.7	7.2

Appendix Table A2. Average Classroom Characteristics by Transition Status and Teacher Experience[b]

	No Move	Change Campus	Change District	Exit Public Schools
All teachers				
Average Math z Score[a]	-0.23	-0.38	-0.29	-0.33
Percent Black	33.4	35.3	33.7	33.5
Percent Hispanic	52.3	55.5	53.0	52.1
Percent Lunch Program	73.6	79.3	75.9	73.2
1 year experience				
Average Math z Score[a]	-0.47	-0.57	-0.40	-0.56
Percent Black	31.4	37.8	43.1	40.8
Percent Hispanic	59.7	57.1	45.7	49.2
Percent Lunch Program	80.5	85.4	78.8	75.6
2-3 years experience				
Average Math z Score[a]	-0.29	-0.39	-0.25	-0.34
Percent Black	31.1	34.7	32.0	31.3
Percent Hispanic	57.6	57.9	53.6	57.1
Percent Lunch Program	77.0	80.6	75.4	76.4

Notes: a. Scores are normalized in each grade and year such that the entire state has mean zero and standard deviation one.

   b. Characteristics pertain to classroom characteristics for each year teachers are observed.

Appendix Table A3. Change in Average Campus Characteristics and Pay By Transition Status and Experience

	No Move	Change Campus	Change District
All teachers			
Average Math z Score	0.00	0.03	0.13
Percent Black	-0.57	-2.79	-8.17
Percent Hispanic	1.04	-0.65	-14.85
Percent Lunch Program.	1.17	-4.65	-27.63
Average Salary	2,137	2,435	2,085
1 year experience			
Average Math z Score	0.03	0.07	-0.03
Percent Black	-0.65	-6.72	-9.36
Percent Hispanic	1.07	5.22	-8.57
Percent Lunch Program..	1.25	-1.56	-27.40
Average Salary	3,204	3,580	1,807
2-3 years experience			
Average Math z Score	0.02	0.06	0.17
Percent Black	-0.56	-2.10	-6.73
Percent Hispanic	1.00	-5.04	-17.26
Percent Lunch Program.	1.08	-9.24	-27.84
Average Salary	1,631	1,604	1,340

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2003. "Teachers and Student Achievement in the Chicago Public High Schools." WP 2002-28, Federal Reserve Bank of Chicago (June).

Armor, David J., Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman. 1976. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corp.

Bacolod, Marigee P., and Justin L. Tobias. 2003. "Schools, school quality and academic achievement: Evidence from the Phillipines." (mimeo), University of California, Irvine (March).

Ballou, Dale. 1996. "Do public schools hire the best applicants?" *Quarterly Journal of Economics* 111,no.1 (February):97-133.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for student background in value-added assessment of teachers." *Journal of Educational and Behavioral Statistics* 29,no.1 (Spring):37-65.

Bishop, John H., Ferran Mane, Michael Bishop, and Joan Moriarty. 2001. "The role of end-of-course exams and minimal competency exams in standards-based reforms." In *Brookings papers on education policy 2001*, edited by Diane Ravitch. Washington, DC: Brookings:267-345.

Boyd, Don, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2002. "Do high-stakes tests affect teachers' exit and transfer decisions? The case of the 4th grade test in New York State." Stanford Graduate School of Education (mimeo)

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. forthcoming. "Who teaches whom? Race and the distribution of novice teachers." *Economics of Education Review*.

Dee, Thomas S. 2004. "Teachers, race, and student achievemen tin a randomized experiment." *Review of Economics and Statistics* 86,no.1 (February):195-210.

Ehrenberg, Ronald G., and Dominic J. Brewer. 1995. "Did teachers' verbal ability and race matter in the 1960s? Coleman revisited." *Economics of Education Review* 14,no.1 (March):1-21.

Hanushek, Eric A. 1971. "Teacher characteristics and gains in student achievement: Estimation using micro data." *American Economic Review* 60,no.2 (May):280-288.

———. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14,no.3 (Summer):351-388.

———. 1992. "The trade-off between child quantity and quality." *Journal of Political Economy* 100,no.1 (February):84-117.

———. 2003. "The failure of input-based schooling policies." *Economic Journal* 113,no.485 (February):F64-F98.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2002. "Inferring program effects for specialized populations: Does special education raise achievement for students with disabilities?" *Review of Economics and Statistics* 84,no.4 (November):584-599.

———. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39,no.2:326-354.

Hanushek, Eric A., and Steven G. Rivkin. 2004. "How to improve the supply of high quality teachers." In *Brookings Papers on Education Policy 2004*, edited by Diane Ravitch. Washington, DC: Brookings Institution Press:7-25.

Kane, Thomas J., and Douglas O. Staiger. 2001. "Improving school accountability measures." WP 8156, National Bureau of Economic Research (March).

———. 2002. "The promise and pitfalls of using imprecise school accountabilty measures." *Journal of Economic Perspectives* 16,no.4 (Fall):91-114.

———. 2002. "Volatility in school test scores: Implications for test-based accountability systems." In *Brookings Papers on Education Policy 2002*, edited by Diane Ravitch. Washington, DC: Brookings:235-269.

Murnane, Richard J. 1975. *Impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

Murnane, Richard J., and Barbara Phillips. 1981. "What do effective teachers of inner-city children have in common?" *Social Science Research* 10,no.1 (March):83-100.

Podgursky, Michael, Ryan Monroe, and Daniel Watson. 2004. "The academic quality of public school teachers: An analysis of entry and exit behavior." *Economics of Education Review* 23,no.5 (October):507-518.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73,no.2 (March).

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94,no.2 (May):247-252.

Sanders, William L., and Sandra P. Horn. 1994. "The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment." *Journal of Personnel Evaluation in Education* 8:299-311.

Sanders, William L., Arnold M. Saxton, and Sandra P. Horn. 1997. "The Tennessee Value-Added Assessment System: A Quanititive, Outcomes-Based Approach to Educational Assessment." In *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* Thousand Oaks, CA: Corwin Press, Inc.

Scafidi, Benjamin, David Sjoquist, and Todd R Stinebrickner. 2002. "Where do teachers go? Georgia State University (mimeo), October.

Tobias, Justin L. 2004. "Assessing Assessments of School Performance: The Case of California." *American Statistician* 58,no.1:55-63.