Preliminary version

The Market for Teacher Quality

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

July 2004

abstract

While many policy discussions focus on teacher quality, most research has failed to identify systematic components of quality.. This paper builds on the non-parametric approaches of recent work that measure quality on the basis of teacher value added to student achievement. The focus is learning more about the distribution of teacher quality in a large urban district, about the quality of those who choose to leave the district vis-à-vis stayers, and about the extent to which districts make use of higher salary or more desirable working conditions to attract more effective teachers. The empirical analysis provides additional support for the existence of large variation in teacher effectiveness that is not explained by information on certification or degree type. And, while public rhetoric suggests that urban districts lose many of their best teachers to suburban and private schools and other occupations, there is little evidence that better teachers are more likely to exit the large district. Finally, there is not strong evidence that higher paying districts systematically attract the more effective teachers.

[*] Stanford University, University of Texas at Dallas, and NBER; University of Texas at Dallas; University of Texas at Dallas, and Amherst College, University of Texas at Dallas, and NBER, respectively. John Kain participated in the development of this project but sadly died before its completion. This paper benefited from early discussions with Larry Katz and from comments on the draft by Chris Robinson. This research has been supported by grants from the Smith Richardson Foundation and the Packard Humanities Institute.

The Market for Teacher Quality

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

Given the emphasis of parents, students, and educators on the importance of 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 supported by recent research is that measurable characteristics of teachers such as experience, certification, advanced educational background, and even scores on standardized tests explain little of the true variation in teacher effectiveness. This paper builds on the non-parametric approaches of recent work that measure quality on the basis of teacher value added to student achievement in an effort to learn more about the market for teacher quality. Specifically, we use matched panel data on teachers and students for a large district in Texas to describe the distribution of teacher quality; to identify systematic differences between teachers who remain in the same school, switch schools within the district, switch districts and exit the public schools entirely; and to investigate the extent to which districts make use of higher salary or more desirable working conditions to attract more effective teachers.

A growing body of research attempts to identify the variance of teacher quality through the estimation of teacher fixed effects, and we make use of the matched panels of teachers and students to overcome the major impediments to the consistent estimation of the variance in teacher quality: non-random sorting of students and measurement error.[1] Any non-random student sorting would lead to different rates of learning across classrooms and schools even in the absence of variation in the quality of instruction, and tests are a noisy measure of student knowledge and academic progress. In addition, the standardized test used in Texas focuses on

[1] While the development of this methodology is discussed below, recent methodological work on the approach is found in Rivkin, Hanushek, and Kain (2001) and Aaronson, Barrow, and Sander (2003). The interaction with measurement error issues can be traced to discussions in Kane and Staiger (2002)

basic concepts, meaning that differences in teacher quality likely translate into larger test score differences for those whose students are concentrated at the lower end of the initial achievement distribution. Therefore it is imperative to control for differences in the distribution of initial student achievement and non-school determinants of academic growth and to consider the influence of measurement error in order to construct a valid index of teacher quality.

The availability of repeated observations for students enables us to control for unobserved as well as observed determinants of learning, and repeated observations for teachers provide a framework for the estimation of an upper bound on the contribution of random measurement error. In addition, we develop an adjusted test score gain measure that accounts for differences in test score gains systematically related to the location in the test score distribution and nature of the tests. We compare these estimates, which can be interpreted as upper bounds on the variance in teacher quality, to the lower bound variance estimates reported in Rivkin, Hanushek, and Kain (2001) and obtain a reasonably tight bound on the variance on teacher quality.

Because we can characterize the entire distribution of teacher quality, we can now analyze important but previously unobservable characteristics of the distribution. For example, while there it has been asserted that students from disadvantaged backgrounds are further disadvantaged by receiving poorer school resources – particularly poorer teachers – these discussions have always been based on relatively unreliable proxies of quality. Put in terms of classroom effectiveness, we find that there are significant differences in teacher quality by background, and that these are most pronounced in terms of the race and ethnicity of the student. The average black student in our large urban district has a teacher who is significantly less able than the average white student.

The variance in quality of the stock of teachers is an important determinant of school effectiveness, but for policy purposes a comprehensive description of teacher flows into and out of districts would also be quite valuable. This is particularly true for large urban districts, where high teacher turnover is a chronic problem, and policy makers often bemoan the loss of many of the best teachers to suburban districts, private schools and other occupations. While there is some evidence to suggest that the probability of leaving is higher for teachers with better alternative earning opportunities or more education, there is no evidence on the question of whether better teachers as measured by their effectiveness in educating students are more likely to switch schools or districts or to exit public schools entirely.[2] The average quality of those who exit urban schools clearly affects the impact of turnover on education quality, and the paucity of useful information leaves a large void in the policy discussion. This paper attempts to fill at least part of the information vacuum by comparing the value added to student achievement of teachers who remain in their same school with those who switch schools within a large urban district, those who switch districts, and those who exit public schools entirely. Importantly, we allow for the possibility that teacher quality is not a fixed attribute but may vary in response to incentives or as teachers gain experience.

A related and more often studied question is whether districts effectively use salaries and non-pecuniary amenities to attract and to retain the best teachers. There can be little doubt that teachers prefer higher pay, shorter travel time to work, and better working conditions as proxied by student achievement, income and other factors.[3] Whether schools use these factors to their advantage is far less clear. Ballou (1996) presents strong 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 the return to experience appears to be substantial early in the career, such a tradeoff may be justified if the quality differential is not large. Note, however, that

---

[2] In a series of papers, Dolton and van der Klaauw (1995, (1999) investigate the impact of alternative opportunities on teacher transitions. They find evidence that opportunity wages affect the probabilities of both entry and exit. These results are consistent with earlier work by Murnane and Olsen (1989, (1990), which found that opportunity wages affected duration in teaching in both Michigan and North Carolina.
[3] Boyd et al. (2002) and Hanushek, Kain, and Rivkin (2004b) each show that teachers respond to both pay and student demographic characteristics in their choice of schools.

information on actual performance in the classroom is not available to Ballou, and the observable characteristics may not be good predictors of instructional quality.

Studies that investigate the relationship between student outcomes and school, district or state average salaries provide perhaps the most compelling evidence on the degree to which higher salaries improve the quality of education. A survey of such research produces a very mixed set of findings, with several authors (Aaronson, Barrow, and Sander (2003; Hanushek (1997; Murnane (1975; Murnane and Phillips (1981b)) emphasizing the difficulty identifying the causal link between teacher salaries and student outcomes. Not only must researchers address problems related to the nonrandom allocation of students to schools and salary setting policies of districts and states, they must also consider the rigidity of the teacher labor market.[4] Specifically, mobility across districts appears to be limited as careers unfold by both formal policies and informal preferences for new teachers, meaning that teachers may be earning substantial rents at any point in time. Consequently, salaries paid at the time each cohort of teachers was hired may be as or more predictive of quality than current salary, complicating any attempts to estimate the link between quality and pay using student data and the stock of existing teachers.

By focusing on the salaries of job changers, we can use past performance to estimate quality and circumvent entirely the stock/flow problem that complicates the specification of salary. In addition, the availability of student demographic information on the receiving school provides proxy measures for working conditions and other aspects of schools that have been shown to be related to teacher transition probabilities. The relatively small number of district switchers does limit sample size and statistical power; nevertheless the analysis can expand our understanding of the extent to which districts make use of desirable student demographics or

[4] Recent work by Loeb and Page (2000) adopts an appealing instrumental variables strategy to overcome some of the key methodological difficulties and finds that salaries significantly improve student outcomes, but concerns related to the aggregation and timing of the data and the validity of the instruments raise questions about the findings.

salary in order to procure more effective teachers or those with particular characteristics such as post-graduate degrees or passing scores on certification examinations.

Empirical Model

The modern study of student achievement focuses on growth in student performance in an effort to eliminate the influences of past family, neighborhood, and school factors. A student's performance at any point in time reflects not only current educational inputs but also the past history of those inputs. Thus, even as databases begin to follow students over time, it is generally not possible to parse the separate influences that combine to determine the level of student performance.

The simplest formulation of models designed to understand school factors generally places achievement growth in terms of the flow of contemporaneous inputs such as:

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

where A_{isg} is achievement of student i in school s and grade, 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.

This formulation eliminates the influences of past individual, family, and school factors and permits concentration on the contemporaneous circumstances that are generally measured along with student achievement.[5] Nonetheless, just focusing on current inputs does not eliminate the difficulties in separating the various inputs. 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.

Specification Issues

[5] 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.

Past analyses, even those with extensive detailed data about schools and teachers, have been unable to characterize reliably the important aspects of schools (Hanushek (2003), Hanushek and Rivkin (2004)). The general approach of fitting simple parametric models based on commonly observed school characteristics have been quite unsuccessful and have been unable to identify much in the way of systematic production relationships.

The alternative, which we pursue here, is the semi-parametric estimation of teacher and school effects. In a simple formulation, which leads into our estimation of teacher effects, consider:

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

where T_{ijg}=1 if student i has teacher j in grade g and =0 otherwise. $\tilde{S}_{ig}$ represents school factors other than individual teachers, and we combine the unmeasured individual and idiosyncratic terms (γ,ε) into a common error term. In this, we separate teacher fixed effects, and t_j is a natural measure of teacher quality that is based on effectiveness of individual teachers in raising student achievement.[6]

This formulation circumvents problems of identifying the separate components of teacher but does not necessarily provide unbiased estimates of teacher quality. First, a variety of selection issues related to the matching of teachers and students are important. Because of the endogeneity of community and school choice for families and of administrator decisions, the unmeasured influences on achievement are not orthogonal to teacher quality. In particular, students with family background and other factors conducive to higher achievement will 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

[6] For previous analyses of this sort, see among others Hanushek (1971, (1992), Murnane (1975), Armor et al. (1976), Murnane and Phillips (1981a), and Aaronson, Barrow, and Sander (2003). Rivkin, Hanushek, and Kain (2001) address the various selection factors along with providing a lower bound on the variations in teacher quality specified in this way.

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.

Second, another source of correlation between teacher quality and student circumstances results from the matching of teachers with schools. Teacher preferences for better working conditions and higher achieving, non-poor, non-racial/ethnic minority students in addition to higher salaries potentially introduces a positive correlation between teacher quality and family contribution to learning (Hanushek, Kain, and Rivkin (2004b)). The previously mentioned failure to hire the best available candidates would, however, reduce the magnitude of this relationship. Within districts, the assignment practices tend to give the newest teachers the lowest priority in terms of deciding where to teach.

In each of these cases, the central issue is whether or not it is reasonable to presume the orthogonality of the combined error:

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

This requirement highlights the importance of accounting for systematic elements of families and schools that explicitly or implicitly affect teacher-student matching. While portions of the measurement and sorting concerns are ameliorated by analyzing growth formulations – thus eliminating a variety of factors that enter into both A_{ig} and A_{ig-1} – the specification of components affecting differential growth becomes important. The empirical work involves estimation of a series of increasingly complex models in an effort to learn more about the extent to which the various sorting processes introduce correlations between these components.

Measurement Issues

An important set of concerns revolving around measurement of student performance has received less attention than the specification issues. First, because of the inability to describe achievement growth fully, estimates of teacher quality, $\hat{t}_j$, necessarily include error:

$$\hat{t}_j = t_j + \nu_j \tag{4}$$

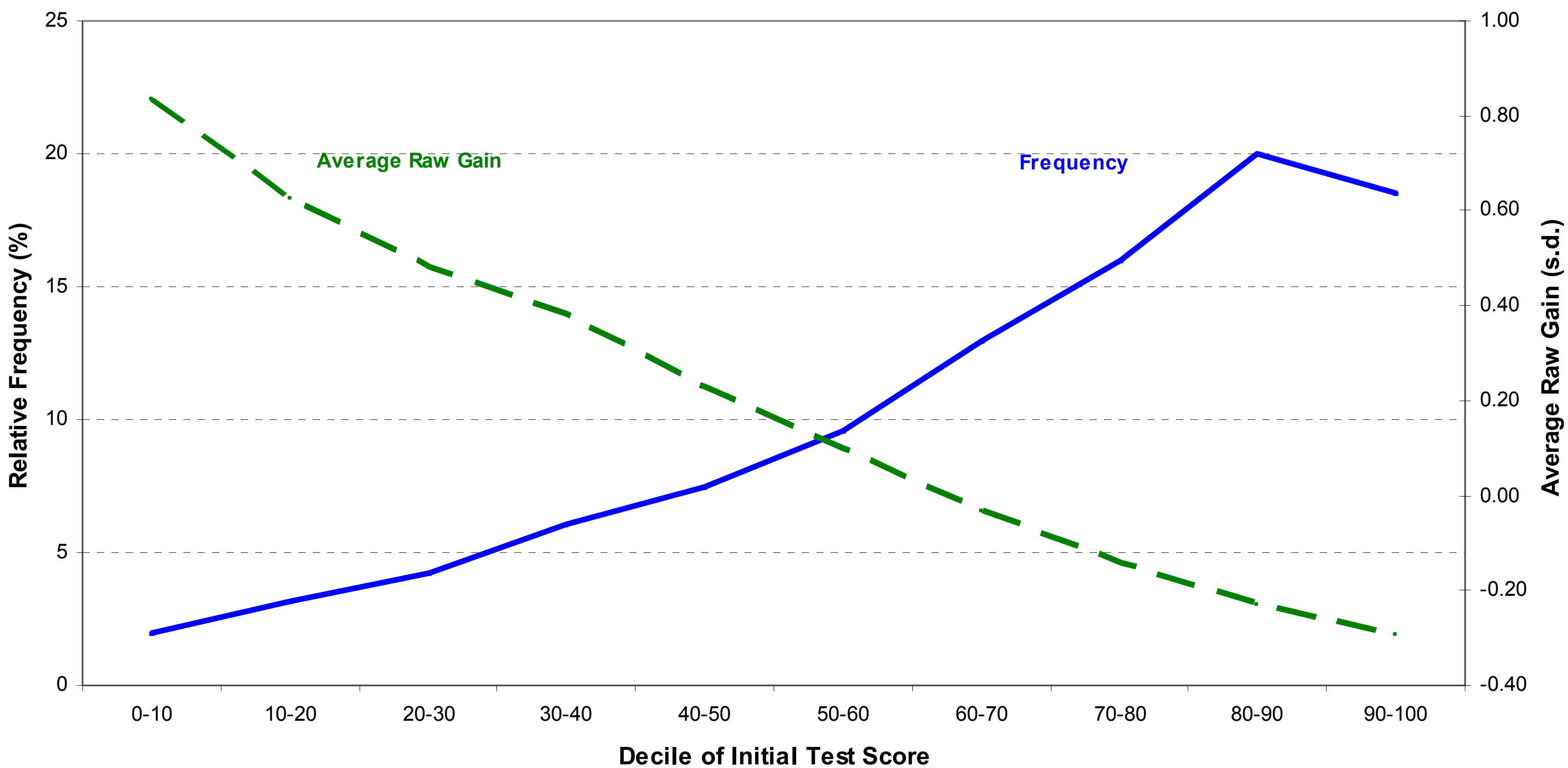
The error, ν, includes pure measurement error arising from imperfect testing along with other idiosyncratic features.[7] This problem, previously discussed in Aaronson, Barrow, and Sander (2003), implies that the estimated distribution of teacher quality will be inflated by estimation errors. Because we have multiple measures of teacher quality for each teacher generated by different years of teaching, it is possible to estimate the contribution of measurement error to the estimated distribution of teacher quality.

Second, some consideration must be given to the metric of achievement and the precise way in which the TAAS test used by the state of Texas indexes student knowledge. Because the test focuses on basic skills, differences in instruction quality likely translate into very different rates of improvement in measured test score gains across the initial achievement distribution. For example, the difference in test scores generated by 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 higher teacher quality. On the other hand, consider a student higher up the initial skill distribution who would 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 better instruction.

Some insight into the potential magnitude of this problem can be seen from Figure 1.

[7] In a somewhat different context, Kane and Staiger (2002) consider test measurement error in the development of school accountability measures. There analysis is directly related because it involves separating measurement error from systematic school components, of which teacher quality is an important dimension.

Figure 1. Relative Frequencies and Achievement Gains by Decile of Initial Score



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 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 score decile, where improvements are 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 average gain of students whose initial test scores fall in each decile. This graph makes it very clear that gains at the bottom – where several test items cover small gradations of math skills – are much higher than in the upper ranges where scores are much more concentrated and the skill gradation is not as large.

In the context of equation 2, this problem suggests that teacher quality is not constant for each teacher across the student achievement distribution but instead is a function of initial scores. Specifically, the magnitude of the difference in test score gains produced by two different teachers may depend critically on the distribution of initial student achievement. To mitigate problems introduced by differences in student academic preparation across teachers, we generate a standardized gain for each student based on comparisons between a student's gain and the average gain in achievement for all students in the district at the same initial achievement level. First, we divide the initial test score distribution into ten equal 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 $[\underset{\smile}{A}_{g-1}^{c_m}, \overset{\frown}{A}_{g-1}^{c_m}]$ for the given year,

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

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

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

$$(7) \qquad G_{isg} = \left[(A_{isg} - A_{isg-1}) - \mu_g^{c_m} \right] \Big/ \sigma_g^{c_m}$$

Consequently, gains in each interval are distributed with mean zero and standard deviation one in each year.

The analysis concentrates on estimating models such as equation 2 using the standardized gains defined by equation 7. The objective is to estimate teacher effects in a metric that provides quality estimates consistent across teachers. In addition, we also investigate whether teachers specialize, being better with students in one part of the distribution than those in other portions.

Although standardized gains serve as our primary measure of teacher quality, we also present a series of other measures for comparison. First, in parts we difference out the campus mean standardized gains to eliminate any influences of systematic differences in students or schools that might contaminate estimates of teacher contributions to value added. (Note that such differencing will also remove actual differences in teacher effectiveness if there is any systematic sorting by teacher quality across schools). Second, we present estimates based on the raw, unadjusted gains in achievement in order to provide a direct comparison to existing research.[8]

Texas Schools Project Data

The cornerstone of the analysis of teacher quality is a unique stacked panel data set of school operations constructed by the Texas Schools Project of the University of Texas at Dallas. The data on students, teachers, schools and other personnel come from the Texas Schools Microdata Panel (TSMP) which has 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 individual 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 individual identifiers that have been encrypted to preserve data confidentiality.

[8] In analyzing raw gains, we first standardize scores to mean zero and variance one for each grade and year.

The student 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) and Title I services. Students are also observed when they switch schools and can be followed across schools in the Lone Star District.

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, role, and campus. Importantly, teachers switching schools or districts can be followed as long as they remain in the Texas public schools.

Student performance is assessed by 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. This paper presents results for mathematics, although the results are qualitatively quite similar for reading. Consistent with the findings of our previous work on Texas, schools appear to exert a much larger impact on math than reading in grades 4 through 7 (see Hanushek, Kain, and Rivkin (2002) and Rivkin, Hanushek, and Kain (2001)).[9]

Importantly, for this paper we employ a significant extension to the basic TSMP database. The student database can be linked to information on teachers and schools for a subset of classrooms. TEA does not collect information linking individual students and teachers. Using just TSMP data, we can place students and teachers in the same campus, grade and year, but not in the same classroom. However, for this study we use additional information for a large urban district ("Lone Star District") to match a student with her mathematics teacher in each year she appears in the sample. This is typically the whole classroom teacher for elementary school and a mathematics teacher for junior high.

In this paper we study students and teachers in grades three through eight for the school

[9] Part of the difference between math and reading might relate specifically to the TAAS instruments, which appear 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.

years 1995/1996 to 2000/2001. We eliminate any student without a valid test score and teachers with fewer than ten students with valid test score gains.

Though the classroom matches are used in the estimation of teacher quality, teachers are assigned average campus characteristics as measures of the student demographic composition of a campus. In combination with salary, these characteristics provide information on changes in both pecuniary and non-pecuniary aspects of jobs following a transition to a new campus or district. Note that, because we investigate aspects of moving, we employ these data for all Texas public schools and not just schools in the Lone Star District.

Distribution of Teacher Quality

This section describes the distribution of teacher quality in Lone Star district using both standardized gains and raw gains to measure quality. It examines the sensitivity of the estimates to controls for student and peer differences and attempts to account for the contribution of measurement error to differences among teachers. An initial extension is the extent to which experience explains differences in teacher quality. Next it describes quality differences by student income and race/ethnicity and considers whether teachers who are very effective with less well-prepared students also tends to be very effective with high achieving students or whether teachers appear to target students at a given skill level at the expense of others.

Variance of Student Gains

It is natural to begin with an analysis of the source of variation in student achievement gains. This analysis is clearly conditional upon the test instruments on the one hand and the institutional structure of Lone Star schools and their hiring patterns on the other. Nonetheless, it places rough bounds on the potential contributions of teachers, principals, and other institutional features of district schools.

Table 1 reports the components of the variances in standardized gains and raw gains along with their sensitivity to controls for student and school characteristics. The top row begins

Table 1. School and Teacher Differences in Student Achievement Gains[a]

	Total	Between teacher and year	Between teacher	Between school
Standardized gains				
no student controls	0.97	0.19	0.12	0.03
controls for student characteristics[b]		0.18	0.12	0.04
Raw gains				
no student controls	0.59	0.13	0.07	0.01
controls for student characteristics[b]		0.13	0.07	0.02

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

with an uncontrolled analysis of variance, i.e., equation 2 with only teacher or school fixed effects. Interestingly, while 19 percent of the variance in standardized gain occurs between teacher-years and 12 percent occurs between teachers, only 3 percent occurs between schools. (The calculation for teacher-years treats each year a teacher teaches as an entirely independent experience, while the calculation for teachers aggregates across all classroom experiences of each teacher). Clearly, this simple decomposition can mask the influence of student sorting by school and teacher. Nonetheless, the teacher-by-year variance component is virtually unchanged in the second row that eliminates measured student characteristics by regressing the standardized gains on income, gender, race/ethnicity, grade, limited English proficiency, special education, mobility,[10] and year dummy variables along with the vector of teacher-by-year fixed effects as in equation 2, despite the strong relationship between standardized gain and a number of demographic characteristics (see Appendix Table A1). Controls for these same student variables plus peer income, percent black, percent Hispanic, percent Asian, percent special education, percent limited English proficient, and percent of students who are new entrants to the school variables also leaves the teacher variance component almost unchanged.[11] This suggests that selection involving student heterogeneity and peer composition account for little of the observed variance in gains across teachers.[12] A similar pattern emerges for the raw gains measure, including the fact that measured student and classroom characteristics have little or no effect on variations in student gains (see Appendix Table A2).

The variance decomposition suggests the existence of large differences in teacher quality within schools and of much smaller differences between schools. This implies, for example, that

[10] Following the findings in Hanushek, Kain, and Rivkin (2004a), we incorporate indicator variables for student moves by type of move.
[11] Note that classroom composition measures cannot be separated in the first estimation but can in the second where teachers face different classroom composition across years.
[12] An alternative method for controlling for student heterogeneity is to include student fixed effects along with the teacher fixed effects. Unfortunately the majority of a teacher's students had only a small number of teachers in the other grades, making it difficult to identify separately the teacher and student fixed effects and dramatically reducing the signal to noise ratio.

analyses concentrating on just between school variations in performance ignore a substantial amount of the systematic impacts of schools on achievement.

Nonetheless, as highlighted in equation 4, measurement error certainly accounts for a portion of the estimated between teacher variance. Importantly, the structure of our data allow us to generate an upper bound on the between teacher error variance under the very strong assumption that teacher quality is fixed for each teacher over time and that measurement error is proportional to the number of students per teacher.[13] In this case the difference between the between teacher and year variance and the between teacher variance in standardized gains is attributed to the decline in the error variance resulting from aggregation. The student weighted average of the ratio of the number of students per teacher/year and the total number of students per teacher across all years equals 3.8. Therefore the between teacher and year variance of the measurement error component would be expected to be 3.8 times as large as that for the between teacher error variance.

Given that aggregation across years decreases the between teacher variance from 0.19 to 0.12, we estimate an error variance of less than 0.025, meaning that measurement error accounts for less than 20 percent of the between teacher variance in standardized gains.[14] Thus, concerns that the estimated teacher effects are driven predominantly by measurement error appear unfounded.[15]

These estimates show dramatic differences in school quality within schools. A one standard deviation improvement in teacher quality leads to a 0.31 standard deviation increase in

---

[13] Note that this rules out gains from experience plus all other changes in effectiveness over time. We return to this below and demonstrate, consistent with Rivkin, Hanushek, and Kain (2001), that there are significant early career experience effects and that instructional effectiveness varies over time.

[14] From equation 4, $\sigma_{\hat{t}}^2 = \sigma_t^2 + \sigma_\nu^2$ where the measurement error, σ_ν^2 is inversely related to student sample size in the estimation of teacher effects.

[15] Part of the discussion of teacher quality estimates relate to the analysis of Kane and Staiger (2002) that emphasizes the importance of small numbers of observations in judging school performance under many accountability systems. Their concern about small cells in school accountability is correct, but it also ignores the importance of variations in school performance that relate to systematic within-school variation that does not follow a systematic pattern.

standardized test score gains. 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.

Despite the fact that the observed student differences account for virtually none of the between teacher variance and instructional effectiveness definitely varies across time for teachers, there may be unobserved differences among students, teachers and schools that inflate (or deflate) the estimated variance of teacher quality.[16] Thus, it is natural to view this estimate as an upper bound on the true variance in teacher quality and compare it with the lower bound estimate of 0.11 standard deviations reported in Rivkin, Hanushek, and Kain (2001). That lower bound estimate is based entirely on within school differences, subject to measurement error that almost certainly attenuates the estimate, and based on specifications that control comprehensively 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 0.67 or two thirds of the standard deviation of the standardized gain.[17] Putting our current estimates on the same scale, two-thirds of 0.31 equals slightly more than 0.20. In other words, a reasonable range for the standard deviation of the distribution of teacher quality based on test score gain is 0.11 to 0.20.

Importantly, common proxies of teacher quality appear to explain little of the actual variation in the quality of instruction. The correlation of our estimated teacher quality with having an advanced degree is -0.0007, insignificant by any standard. While we do not have the actual certification test scores, we do know whether each teacher passed the test on her first attempt.

[16] An alternative approach is to add student fixed effects to the aforementioned regressions. The hypothesis that there is no systematic variation in teacher quality is strongly rejected even in the presence of student fixed effects. Yet because of the structure of grade transitions in which the majority of students continue along the same path, it is quite difficult to obtain precise estimates of teacher quality in the presence of student fixed effects.

[17] That analysis implicitly corrects for the concerns about the test construction by including individual fixed effects in raw gains. Since the individual fixed effect will adjust expected gains for the individual's test level, they represent an alternative approach to the adjusted gains used here.

The correlation of this with our quality measure is 0.03, which is statistically different from zero but explains less than one tenth of one percent of the variation in quality. Thus, we reaffirm prior findings that measured characteristics of teachers do a poor job of identifying teacher quality defined in terms of student learning.

Experience Effects on Teacher Quality

Recent work shows that the returns to experience are concentrated very early in the career and also suggests that the lower effectiveness of inexperienced teachers results from their inexperience rather than the weeding out of the worst teachers following the first and second years on the job (Rivkin, Hanushek, and Kain (2001)). The results in Table 2 based on regressions that compare performance in the first five years of teaching to those teaching 6 or more years generally support both notions. Three sets of experience coefficients are produced for each of the quality measures. Columns 1 employs specifications with no fixed effects, Column 2 adds student fixed effects, and finally Column 3 adds teacher fixed effects.

The pattern of estimates highlights clearly the concentration of learning in the first year on the job. None of the other experience coefficients are negative and significant. There is evidence of additional small improvements until the fourth year, after which there is a small decline. Notice that the estimation for raw gains in the right hand panel yield qualitatively similar effects of experience.[18]

The pattern of results across specifications also suggests that at least a portion of the first year effect results from the disproportionate departure of less effective teachers following the first year on the job: Controlling for teacher fixed effects in addition to student fixed effects reduces

[18] Table 2 also provides some additional evidence in support of the belief that the structure of the test tends to produce higher gains for those with lower starting scores. Notice that with the sample gains measure but not with the standardized gain, the inclusion of student fixed effects increases the magnitude of the initial experience effect by roughly 33 percent. Because less well-prepared students are more likely to have teachers with no prior teaching experience, any test related higher gains would tend to offset the experience effect.

Table 2. Estimated Returns in Standardized Gains and Raw Gains to Experience
(comparisons to teachers with 6 or more years experience)

	Standardized Gains			Raw Gains		
		With student fixed effects	With student fixed effects and teacher fixed effects		With student fixed effects	With student fixed effects and teacher fixed effects
1st year	-0.18	-0.16	-0.12	-0.09	-0.12	-0.08
	(9.06)	(9.76)	(8.09)	(6.54)	(9.01)	(6.82)
2nd year	-0.04	-0.04	-0.01	-.01	-0.02	0.01
	(1.90)	(1.93)	(0.34)	(0.31)	(0.99)	(0.47)
3rd year	0.03	0.02	0.02	0.02	0.02	0.02
	(1.13)	(1.26)	(1.21)	(1.01)	(1.19)	(1.12)
4th year	0.06	0.08	0.06	0.05	0.06	0.05
	(2.24)	(3.59)	(3.14)	(2.66)	(3.27)	(2.91)
5th year	0.04	0.03	0.01	0.04	0.04	0.
	(1.46)	(1.53)	(0.55)	(2.54)	(2.36)	

the cost of having a new teacher by roughly 25 percent. Thus although teachers appear to learn a great deal in the first year on the job, composition also contributes to the steep experience profile early in the career.

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 enter with no experience.[19] For the Lone Star 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 Table 3, 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.

Racial, Ethnic, and Income Differences in Teacher Quality

Many policy discussions explicitly worry that disadvantaged students have poorer teachers. Much of this concern is generated by the disparate housing locations that has poor and minority students concentrated in urban areas with a perceived more difficult time in hiring teachers. Part also results from the fact that teacher turnover is higher in schools with high concentrations of poor and minority students (Hanushek, Kain, and Rivkin (2004b)) and the well founded concern that having teachers in their initial years is problematic.

Our comparisons here of teacher quality look just within the Lone Star District, so they do not include any possible urban-suburban variations in quality.[20] Nonetheless, the distributions sorted by student characteristics show pronounced differences by race and income. The average quality for teachers of African American students and Hispanics as measured by the teacher fixed

[19] These figures refer to 1994-96 in Texas (Hanushek, Kain, and Rivkin (2004b)). The rate of new hires varies some over time, depending on student demographics, the extent of teacher retirement, and the numbers of returning teachers who have prior experience.
[20] Prior work on achievement differences related to urban-suburban moves of students indicate systematically higher school quality in suburban schools, although no attempt is made to disentangle the contributions of teacher quality to these differences (Hanushek, Kain, and Rivkin (2004a)).

effects controlling for student characteristics including race/ethnicity and income (-0.02) is roughly one eighth of a standard deviation below the average for whites (0.10).[21] In terms of income, the average quality for students eligible for a subsidized lunch is 0.06 of a standard deviation below the average for higher income students not so eligible.

Figures 2 and 3 provide kernel density plots of the teacher quality distributions by race/ethnicity and income, respectively, using the teacher by year fixed effects produced for Table 1 to measure teacher quality. Interestingly, the mean differences do not come from outliers in the distributions but instead from a noticeable shift in the entire distributions. Moreover, the same pattern appears even if the quality distribution is adjusted for experience differences of teachers.[22] The plots reveal pronounced differences by both race and income.

Importantly, the racial differences in teacher quality are not nearly as apparent in estimates relying just on raw gains. The average black student in the Lone Star district scores XX standard deviations below the average white (roughly equivalent to the differences observed throughout the state). The difference in average quality of teachers between white and black or Hispanic students in terms of raw gains is just XX, significantly different from the 0.12 s.d. in terms of standardized gains. The larger average gains in the lower parts of the achievement distribution offset the fact that blacks and Hispanics have lower quality teachers. While the plot of quality in terms of raw gains that corresponds to Figure 2 (not shown) shows a slight advantage of white students, it does not stand out in the way that it does with the standardized gain measures.

The exact mechanism behind these race and ethnic differences in teacher quality is unclear. White students are a distinct minority in the Lone Star district (less than 20 percent) and are unevenly distributed across the schools. This distribution interacts with teacher choices (see

[21] Note that these estimates come from models that incorporate main effects of race differences, so they do not reflect any systematic average differences in student gains by race or ethnicity over time.
[22] To adjust for experience, the teacher fixed effects are regressed on a series of single year experience dummies, and the residuals are plotted as the long term quality differences.

Figure 2. Kernal Density Estimates of Teacher Quality Distribution: Standardized Average Gains by Student Race and Ethnicity

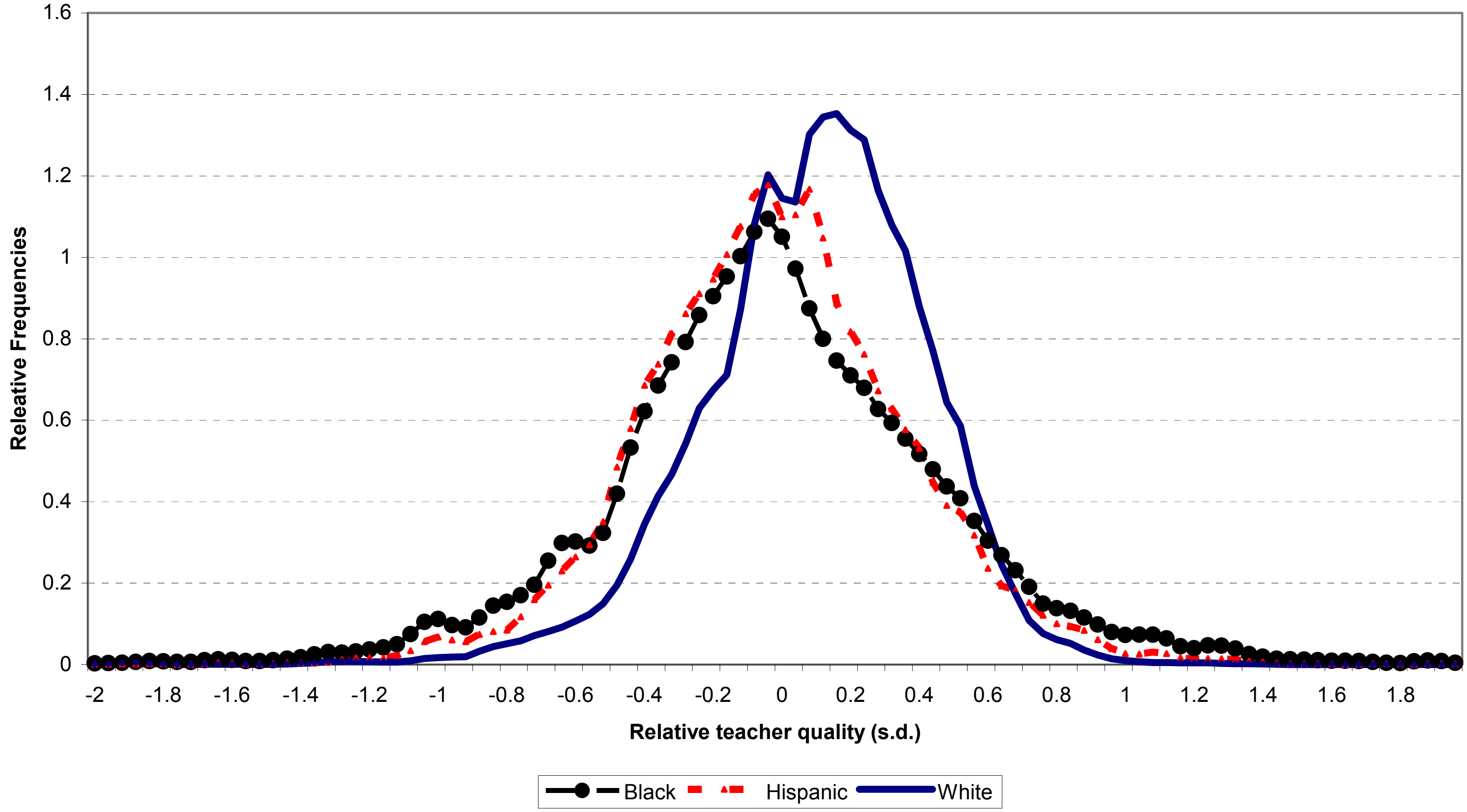
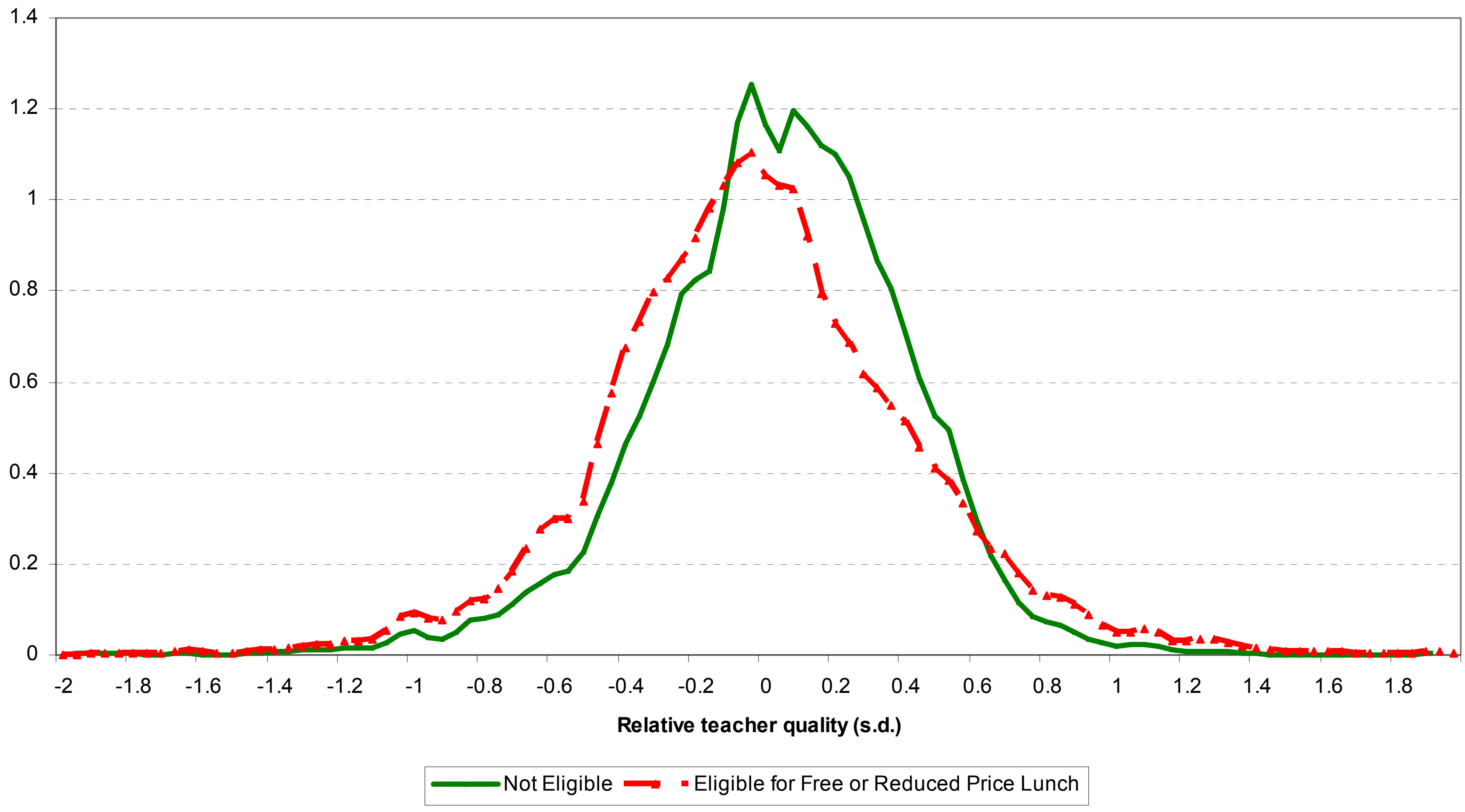


Figure 3. Kernal Density Estimates of Teacher Quality Distribution: Standardized Average Gains by Eligibility for Free or Reduced Price Lunch



Hanushek, Kain, and Rivkin (2004b) and below) and with the within-school sorting of students. But understanding exactly how these distributions occur is beyond our capacity here.

Teacher-Student Matching

An important question is whether there exists substantial variation in the quality of instruction across students of different academic preparation levels within the classroom relative to between teacher quality differences. Consider two otherwise identical teachers who aim their lesson plans at different level of difficulty – one at the bottom of the class and one at the top of the class. Then, depending on the classroom academic preparation, they could well appear to be different in the average quality distribution identified above. Or, perhaps more importantly, consider two equally competent teachers who are specialists with either bottom or top students such that the matching of teachers with students is particularly important. The distribution of quality that we trace out may be influenced by differences in the quality of classroom matching across schools and principals.

We investigate this possibility by dividing students into three academic preparation classifications (based on initial scores) and computing the correlation between the teacher average gain for students in one category with the teacher average standardized gain for students in the other categories.[23] 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. The strong positive correlation between the average standardized gains in the top and bottom categories is particularly striking given the relatively small number of students in the bottom category in schools with large numbers of students in the top category.

---

[23] Because students are initially divided into ten categories to produce the standardized gain measure, we aggregate these ten categories into three using the district average distribution of students over all years as the fixed weights for all teachers.

Teacher Transitions

The high rate of teacher turnover in large urban districts engenders considerable concern among educators, in large part because of the belief that such districts tend to lose their most skilled teachers. Although existing work does consider the impact of salaries, alternative opportunities, working conditions, and other observable characteristics on transition probabilities, there exists no evidence on the link between actual performance in the classroom and transitions. One aspect of this – the effect of initial teacher experience – has already been noted. Here we use the previous estimates to compare the long term quality of teachers who exit the Lone Star District between 1996 and 2000 with those who remain.

Each year teachers fall into four distinct categories: 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.[24] Prior to describing quality differences by transition and by experience, we describe mobility rates and differences in school characteristics for movers and stayers.

Table 3 describes 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.

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

Table 3. 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

Consistent with prior studies, the probability of moving differs by a number of school characteristics. Table 4 provides school average demographic characteristics for each teacher by annual move status. This shows vividly that school 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.[25] Somewhat surprisingly based on our prior work for the state as a whole (Hanushek, Kain, and Rivkin (2004b)), the differences in racial and ethnic composition and income are smaller and less systematic.

Because transition rates are higher for less experienced teachers, the next panel reports these same school characteristics for teachers in their first year and for those with 2-3 years experience, the groups with the highest exit rates from their current school. While anecdotal stories of teacher placement suggest that beginning teachers work in schools 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 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 Table 4 presents the characteristics of Lone Star District schools from which teachers exit, an alternative perspective is how campus characteristics change with a move. Table 5 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

[25] 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 XX standard deviations below the Texas state mean.

Table 4. Average Student Characteristics by Transition Status and 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 state has mean zero and standard deviation one.

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

Table 5. 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

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).

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.

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 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 (2004b)). Given that teachers initiate the vast majority of transitions and undoubtably 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.

Tables 4 and 5 provide information on differences and changes in school characteristics by transition status, but they provide no information on quality differences of those moving. Table 6 reports 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. In the first three columns the transition classification ignores the subsequent return, while for the final specification returnees are reclassified on the basis of where they teach following their return. Note that the school by year fixed effect specifications generate coefficients based on differences within schools. All estimates compare those who exit with those who remain.

Table 6. Differences in Teacher Quality by Transition Status
(Standardized Gains compared to teachers remaining in same school)

		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)

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 or only those in the same school and year. Moreover, those who switch campuses within the same 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 have a very high rate of attrition? Figure 4 plots kernel density estimates of the distribution of teacher quality by move status. The solid dark plot identifies the quality distribution of the nonmovers. The distributions of those who either change campuses or exit public schools fall distinctly below those who stay, while those who change districts are insignificantly different in both mean and distribution of quality.

Table 7 reports differences in quality across transitions by experience, focusing on young teachers. Unfortunately most of these estimates 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 above average in performance, although any quality premium appears to decline with experience.

Overall, while there is ambiguity introduced by the imprecision of the estimates, the pattern of exits does not show strong average quality differences. The estimates provide some support for the belief that young teachers who leave for other districts are drawn more from the

Figure 4. Kernal Density Estimates of Teacher Quality Distribution: Standardized Average Gains by Teacher Move Status

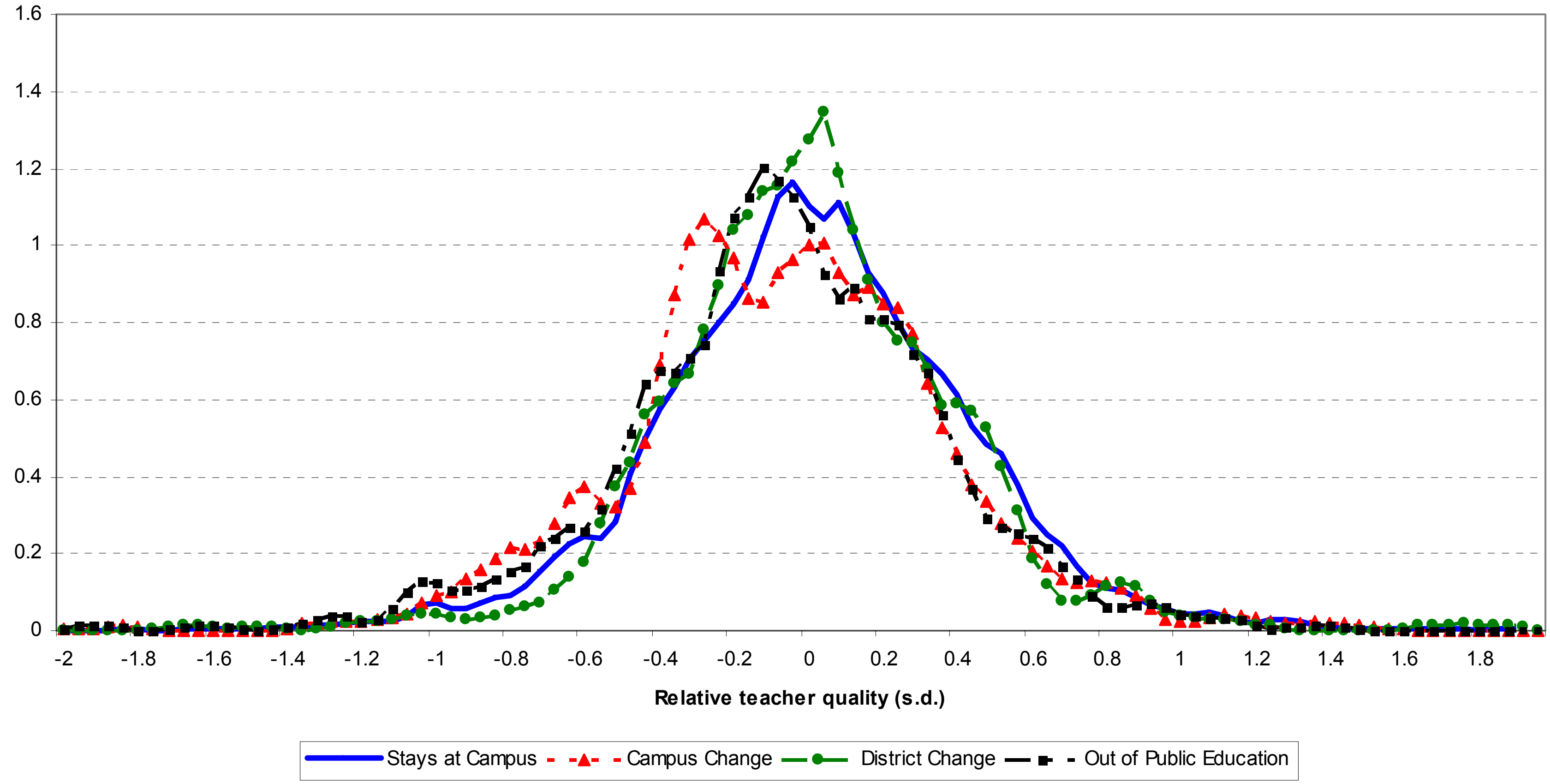


Table 7. Differences in Teacher Quality by Transition and Experience

(Standardized Gains compared to teachers remaining in same school including student fixed effects)

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)

upper part of the quality distribution and stronger evidence that less effective teachers are more likely to exit public school teaching entirely.

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 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.

Although Tables 6 and 7 show that teachers who exit the public schools are less effective in the classroom than those who remain in their original school, it is possible that the exit year was anomalous and not indicative of their 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 on the basis of student gains in the year prior to each transition. In other words, we describe the distribution of quality for transitions following the 1999 school year with teacher 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

Table 8. Differences in Teacher Quality by Transition and Year Quality Measured

(Standardized Gains compared to teachers remaining in same school)

	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)

compares movers 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 fixed effects differ some from the comparable estimates in Table 6, the patterns are the same.

Two findings stand out in the estimates. First, those that change campuses within the Lone Star District tend to be above average in the school they are leaving. Second, and more important, the performance in the year immediately prior to exiting is noticeably worse than that 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 the year (that might contribute to an exit decision) cannot be ascertained at this time.

Who Hires the Most 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.[26] 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.[27] The crucial, unanswered question is whether or not 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.

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

[27] See Boyd et al (2002) and Hanushek, Kain, and Rivkin (2004b) 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.

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. Thus, we might expect that districts who are more attractive in these dimensions would have a better chance to hire high quality teachers. On the other hand, they might squander these attributes in order to attract other characteristics of teachers.

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 two are the teacher value added coefficients from earlier regressions with an adjustment for experience. The final two rely on common proxies, albeit ones not closely connected to teacher quality: the third dependent variable is an indicator for whether teachers pass the certification examination, and the final dependent variable is an indicator variable for the possession of a post-graduate degree. The two demographic variables added with salary in the regressions are percent black and percent classified as limited English proficient.

Prior to considering the empirical 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) choose from a deeper, higher average quality applicant pool. Consequently, if schools are able to determine teacher effectiveness *and* 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 some suggestive 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

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	Teacher quality measures		Teacher quality proxies	
	standardized gains[a]	raw gains[a]	passed certification examination	advanced degree
log salary	0.12	0.21	-0.15	0.22
	(0.63)	(1.24)	(1.72)	(1.88)
% Black	0.0000	0.0020	-0.0009	-0.0024
	(0.01)	(1.17)	(1.18)	(2.67)
% Limited English	0.0016	0.0007	-0.0017	-0.0017
Proficient	(0.89)	(0.55)	(1.86)	(1.49)

a. Quality measures are adjusted for single years of experience.

teachers defined in terms of student outcomes. Neither of the salary coefficients in the specifications with a measure of quality as the dependent variable approach statistical significance. Although the small sample size certainly contributes to the imprecise estimates in the quality specifications, the contrast with the findings for advanced degree is consistent with evidence that shows such a degree has little systematic relationship with the quality of instruction.

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. Our prior analysis made an effort to establish a lower bound on quality variations – a lower bound that guarded against possible selection of schools and teachers by students, against administrative actions to match students and teachers, and against a variety of measurement erros on omitted other factors. That work established that the lower bound was quite substantial – a standard deviation of teacher quality equaled 0.11 standard deviations of student gains.

Here we consider the performance of individual teachers, based upon classroom rather than school average gains, and attempt to estimate the full distribution of teacher quality. With adjustment for student backgrounds and for measurement errors, we find that a comparable upper bound of the standard deviation of quality is approximately 0.2 standard deviations of student gains.

The information on the quality distribution of teachers in our large urban district reveals some very important insights. First, we confirm that there is significant learning about the craft of teaching that goes on 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, even within a single large urban district, there are significant differences in the quality of teachers by race and ethnicity and by income. The average white student has a teacher who is about one eighth of a standard deviation higher quality than the average black or Hispanic student. Similar, but smaller, differences exist between students eligible for free or reduced price lunch and those not eligible. These differences come from a shift in the entire quality distribution and not just from a small number of teachers at the tails of the distribution.

A related concern is the possibility that large urban districts lose their better teachers to other occupations or to suburban schools. Here 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 the Texas public schools are significantly less effective on average than those who remain, and those moving to other 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 students because of the lower performance of inexperienced teachers, the evidence does not support the related concern that the best teachers are those most likely to leave.

Finally, the last empirical section examines whether districts make use of salary and demographic characteristics to procure more effective teachers. The small sample size limits the precision of the estimates. Nonetheless, the failure to find a strong relationship between quality and salary suggest that it is quite difficult to discern quality from past performance in another district.

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.

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

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)

Dolton, Peter J., and Wilbert van der Klaauw. 1995. "Leaving teaching in the UK: A duration analysis." *The Economic Journal* 105(March):431-444.

———. 1999. "The turnover of teachers: A competing risks explanation." *Review of Economics and Statistics* 81,no.3 (August):543-552.

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.

———. 1997. "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19,no.2 (Summer):141-164.

———. 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.

———. 2004a. "Disruption versus Tiebout improvement: The costs and benefits of switching schools." *Journal of Public Economics* Vol 88/9-10:1721-1746.

———. 2004b. "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. 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.

Loeb, Susanna, and Marianne E Page. 2000. "Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation." *Review of Economics and Statistics* 82,no.3 (August):393-408.

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

Murnane, Richard J., and Randall Olsen. 1989. "The effects of salaries and opportunity costs on length of stay in teaching: Evidence from Michigan." *Review of Economics and Statistics* 71,no.2 (May):347-352.

———. 1990. "The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina." *Journal of Human Resources* 25,no.1 (Winter):106-124.

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

Murnane, Richard J., and Barbara R. Phillips. 1981b. "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance." *Economics of Education Review* 1,no.4 (Autumn):453-465.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2001. "Teachers, schools, and academic achievement." Working Paper No. 6691, National Bureau of Economic Research (revised)

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