Do Disadvantaged Urban Schools Lose Their Best Teachers?

Eric A. Hanushek and Steven G. Rivkin*
May 2009

Abstract

Recent research documents substantial variation in teacher quality as measured by value added to student achievement, much of which is not captured by characteristics typically used in the hiring and salary determination processes including post-graduate schooling and experience. Consequently an understanding of the impact of teacher transitions on school quality cannot focus simply on such observed characteristics. This paper begins by estimating the variance in teacher value added for a large urban district in Texas using methods that mitigate bias potentially introduced by test measurement error and by the nonrandom allocation of students among classrooms. It then describes average differences in value added between teachers who switch schools within or across districts, exit the Texas public schools altogether, or remain in the same school. The results show significant variation in teacher value added after allowing for the various measurement issues. Perhaps the more important result, however, is that teachers who switch schools within a district, switch districts, or exit the Texas public schools entirely do not appear more effective than those who remain in their school and quite possibly are less effective. This finding is clearest for the typical teacher who exits the Texas Public Schools. Moreover, teachers leaving the most disadvantaged schools in terms of student populations are consistently less effective than those who stay.

* Hoover Institution/Stanford University, University of Texas at Dallas, and National Bureau of Economic Research; Amherst College, University of Texas at Dallas, and National Bureau of Economic Research, respectively. We thank Dan O'Brien for help with the data development and early analytical work. This research has received support from the Spencer Foundation, the Hewlett Foundation, and the Packard Humanities Institute.

Do Disadvantaged Urban Schools Lose Their Best Teachers?

By Eric A. Hanushek and Steven G. Rivkin

Current views about the distribution of teachers across schools are largely based on observations about experience and education differences rather than differences in actual classroom effectiveness. Recent research documents substantial variation in teacher quality as measured by value added to student achievement, much of which is not captured by characteristics typically used in the hiring and salary determination processes including post-graduate schooling and experience. Consequently an understanding of the impact of teacher transitions on school quality cannot focus simply on such observed characteristics. In this paper we study differences in teacher value added among teachers who remain in the same school, school switchers, and those who exit Texas public schools in order to understand better the impact of such transitions on large urban districts.

One strand of literature emphasizes the importance of community type, wealth, crime, and the other factors on the choice of jobs and presents a *prima facie* case that teacher quality is not distributed equitably. Boyd, Lankford, Loeb, and Wyckoff (2005) show that teacher labor markets tend to be highly localized, which complicates recruitment efforts in both urban centers and rural areas. Teachers also appear to prefer schools with higher achieving, higher income students, in addition to higher salaries (Lankford, Loeb, and Wyckoff (2002), Hanushek, Kain, and Rivkin (2004), Scafidi, Sjoquist, and Stinebrickner (2007)). Moreover, there is also evidence that teacher exit probabilities are higher for those with better alternative earning opportunities or more education (Dolton and van der Klaauw (1995, (1999)). These findings support the notion that high poverty and geographically isolated schools face myriad impediments to teacher hiring and retention, a view reinforced by administrators in rural areas and large urban districts who often bemoan both the difficulty of attracting teachers and the loss of teachers to the suburbs, private schools, and other occupations.

Importantly, none of these studies provides direct evidence on classroom effectiveness, thus constituting a large void in our ability to understand the dynamics of teacher labor market. Even if teachers with better alternative earnings opportunities are more likely to quit teaching and inner city and rural schools experience higher turnover than suburban schools, the implications for policy remain unclear. Any impact on the wellbeing of students depends crucially on the actual effectiveness of leavers and of their replacements. To begin with, teacher effectiveness may not be strongly correlated with outside opportunities. Moreover, if teaching performance is a primary determinant of a teacher's job satisfaction and desire to stay in a school, leavers come disproportionately from the lower end of the teacher quality distribution even when exiting is positively related to alternative earnings opportunities. As an empirical matter, Scafidi, Sjoquist, and Stinebrickner (2006) show that majority of exiting teachers from public schools do not move to higher paying jobs outside of teaching but instead are more likely either to exit the labor market entirely or switch to a lower paying job in a private school – facts that are consistent with the possibility that job satisfaction is an important determinant of teacher retention.

We do not investigate the underlying motivation for teacher exits from given schools but instead concentrate simply on the relative effectiveness of those leaving. The central focus is understanding differences in average teacher value-added by school transition status and school characteristics for a large urban district in Texas. First, we consider how much teachers differ in effectiveness by estimating the variance of teacher value-added using a set of teacher-by-year fixed effect specifications that progressively take more and more steps to protect against bias introduced by non-random sorting.[1] We estimate both the overall variance and the within school, grade, and year variances; we also divide schools by the degree to which students are sorted systematically by classrooms based on two different empirical criteria. Finally, we conduct some nullification tests suggested by {Rothstein, 2008 #5302} in order to provide additional evidence on the likelihood that the non-random sorting of students among schools and classrooms introduces substantial bias into the estimates.[2] Systematic sorting of students among communities and schools does not introduce bias into the within school-grade-year estimates, and these are the preferred models. However, because these specifications provide value added estimates relative to others in the same school, grade, and year and do not permit cross school comparisons, we also provide estimates that compare teachers with others in the district as a whole. Note that we use the correlation of adjacent year fixed effect estimates for the same teacher to separate the true variance from the variance produced by test error.

Following this analysis we turn to a study of quality variation by transition status and school characteristics. We estimate the mean quality of stayers, of teachers who move within the district, of those who move to a different district, and of teachers who exit teaching using a value-added model similar to that used in the study of the variance of teacher quality except that it substitutes transition indicators for the teacher-by-year fixed effects. Again we estimate a series of specifications that take increasing steps mitigate bias from the non-random sorting of students. We also compare specifications that use teacher performance immediately prior to the transition to specifications that use performance in earlier years to examine the extent to which temporary shocks in productivity rather than more permanent differences determine any quality differences by transition. We conclude with a discussion of policy implications and possible extensions for future work.

I. Empirical Model

The primary analytical task is the separation of teacher contributions to achievement from other student, family, school, and community factors. This analysis builds on a cumulative model of learning, and highlights the specific issues relevant to the estimation of teacher fixed effects.

A. Cumulative Model of Learning

We focus on the growth in learning that occurs during a specific grade and relate this to the flows of educational inputs from schools and elsewhere. Equation (1) models achievement of student *i* in grade G and year y (suppressed in the equation since year is unique to grade G for student i) as a function of initial student skill at entry to grade G (α_{iG}), of family background and other influences outside of schools (X), of peer composition (P), of school factors – including resources, principal quality, and school or district determined curriculum – (S), of teacher quality (τ), and of a random error (*e*).

---

[1] For other estimates of teacher value-added, see Hanushek (1971, (1992), Armor et al. (1976),Murnane (1975), Murnane and Phillips (1981), Aaronson, Barrow, and Sander (2003), Rockoff (2004), Rivkin, Hanushek, and Kain (2005), Boyd et al. (2006), and Kane, Rockoff, and Staiger (forthcoming).

[2] The interpretation and use of estimated differences in individual teacher effectiveness, often under the heading using "value-added measures" for teacher evaluation and possibly compensation, has been the subject of considerable recent discussion (e.g., Sanders and Horn (1994), Wainer (2004), McCaffrey, Lockwood, Louis, and Hamilton (2004), Ballou, Sanders, and Wright (2004), Rothstein (2008), Rivkin (2008)). While this discussion introduces additional issues, a portion is relevant for the estimation of the aggregate distribution of teacher effectiveness and is discussed below.

(1) $$A_{iG} = \alpha_{iG} + \beta X_{iG} + \lambda P_{iG} + \delta S_{iG} + \tau_{jy} + e_{iG}$$

Rather than using observed teacher characteristics as proxies for quality, we adopt a less parametric approach and include a full set of indicators (fixed effects) for each teacher-year combination (i.e., τ_{jy} for each teacher j in year y), permitting teacher effectiveness to vary with experience and other factors that change over time but constraining it to be constant for all students having a given teacher in a given year. In this formulation teacher quality is estimated by netting out the contributions of initial student skill (α_{iG}), family background (X), peer composition (P), and school factors (S) from the average achievement of students taught by a teacher in a given year.

Without loss of generality, think of each of the terms as a scalar index of the respective characteristics that increase in value as the characteristic becomes more conducive to achievement. For example, a higher value of P indicates a better peer composition (perhaps fewer disruptive students). Therefore all of the parameters are non-negative, as higher skill, family characteristics that support achievement, better peer compositions, better schools, and higher teacher quality all raise test scores.

In the absence of random assignment, unobserved peer and school factors for a given class could confound estimates of the quality of the teacher assigned to that class. The problems can come from a variety of behavioral outcomes: principal assignment of better teachers to classrooms with better students (or worse students, if seeking to equalize achievement across classes); better teachers gravitating toward higher resource schools; families with the most educational concerns and most resources to support children moving to the school districts with the best teachers. All complicate the estimation of teacher value-added to achievement, as teacher quality becomes intertwined with characteristics of students or schools.

The desirability of any particular approach to isolating the value-added of teachers depends upon the extent to which it accounts for the potential confounding factors. Teacher quality is identified only if all potentially confounding factors are included and properly specified as explanatory variables in the regression. Either omission or misspecification of factors that determine α or in e corrupts the estimates of teacher quality.

We model α in a one dimensional framework in which differences in cognitive skills are assumed to evolve over time with educational experiences at home, in school, and in the community in a manner consistent with Equation (1). Equation (2) describes this cumulative process,

(2) $$\alpha_{iG} = \beta\sum_{g=1}^{G-1}\theta^{G-g}X_{ig} + \delta\sum_{g=1}^{G-1}\theta^{G-g}P_{ig} + \lambda\sum_{g=1}^{G-1}\theta^{G-g}S_{ig} + \sum_{g=1}^{G-1}\theta^{G-g}\tau_{jy} + (\gamma_i + \sum_{g=1}^{G-1}\theta^{G-g}\gamma_i)$$

where γ_i is "innate ability" of i, which is assumed to be constant and to affect achievement over time.[3]

A good teacher likely raises achievement in the current year and subsequent years by increasing the stock of knowledge, and a very supportive parent does the same. In a very general way, we allow historical effects (and knowledge) to depreciate at a geometric rate (1- θ) meaning that a teacher or peer's effect on test scores diminishes with time such that a good 4th grade

[3] Innate ability here simply refers to student differences set before entry to school that affect student learning growth and could arise from any combination of health, nutrition, genetic, or family factors.

teacher has a larger effect on 4th grade score than on 5th grade score.[4] If θ=1, the effects of prior experiences persist fully into the future, while, if θ=0, prior experiences and knowledge have no effect on current achievement. In the estimation, however, we do not constrain the knowledge depreciation rate to a specific value but instead directly estimate it.[5]

A value-added regression of achievement in grade G on achievement in grade G-1 along with contemporaneous family, school, and peer characteristics and a fixed effect for each teacher in each year provides a natural way to account for prior influences while estimating teacher effects on achievement.[6] Rewriting equations (1) and (2) for grade G-1 illustrates how the inclusion of A_{iG-1} as an explanatory variable with parameter θ in a regression with achievement in grade G as dependent variable potentially controls for the full set of historical factors.

$$(3) \qquad A_{iG} = \theta A_{iG-1} + \gamma_i + \beta X_{iG} + \lambda P_{iG} + \delta S_{iG} + \tau_{jy} + (\gamma_i + e_{iG})$$

The estimation presumes that there are at least two observations of achievement for each student and that there are multiple students with each teacher.

B. Estimation of Teacher Fixed Effects (τ_{jy})

Two central issues in the estimation of the teacher fixed effects are test measurement error and potential biases introduced by the purposeful sorting of students and teachers into schools and classrooms. The former inflates estimates of the variance in teacher quality and leads to less precise estimates of teacher value added. A number of studies highlight the issue test measurement error and consequent sampling variability and test measurement error (e.g., Gordon, Kane, and Staiger (2006), McCaffrey, Lockwood, Louis, and Hamilton (2004), Sass et al (forthcoming)). Purposeful sorting contaminates the estimates unless the model accounts for the determinants of school and classroom sorting that also affect achievement. We now outline the approaches used to mitigate omitted variables bias and account for test measurement error.

Whether the model generates unbiased estimates of the τ_{jy} depends importantly on whether the empirical specification accounts for relevant factors affecting schools that are also correlated with the teachers or other inputs. The inclusion of prior achievement mitigates bias from omitted family, neighborhood, and school influences. Yet this is generally insufficient because dynamic behavioral choices by families and school authorities introduce bias even to value added models.

For example, notable sources of "across school" unobserved heterogeneity include the quality of the principal, family background, the extent to which the curriculum for grade G comports with the state test, and the level of student disruption. Because available data typically have limited controls for differences in the quality of administration and other subtle aspects of schools, it is quite difficult to separate teacher and school effects in specifications that produce teacher fixed effects relative to all other teachers in the district. Therefore it is appealing to control for school or even school-by-grade-by-year fixed effects in order to account for both

---

[4] This does not exhaust the possibilities that have been used since using the difference in scores between grades G and G-1 as the dependent variable (i.e., imposing the assumption of θ=1) is sometimes accompanied by including student fixed effects. As Rivkin (2005) demonstrates, this gains specification will tend to bias downward differences among teachers in the absence of student fixed effects and bias upward differences among teachers if student fixed effects are included.

[5] For ease of estimation, we do constrain knowledge to depreciate at the same rate regardless of source. In other words, past knowledge has the same impact on the accumulation of new knowledge, regardless of the source of this past knowledge. Implicitly, this formulation does not allow for the "one special teacher" who has a lasting effect on student learning over and above the contemporaneous effect on achievement.

[6] See Hanushek (1979, (1986) for a discussion of value-added models.

observed and unobserved persistent differences among schools and districts, though as noted above this approach prohibits comparisons of teacher quality among schools.

Yet even estimates of teacher value added based solely on within school variation could suffer from omitted variables bias if classroom assignments are not random. Clotfelter, Ladd, and Vigdor (2006) and {Rothstein, 2008 #5302} document the existence of extensive within-school sorting on the basis of student characteristics and prior performance. Whether such sorting introduces substantial bias in commonly used value-added models is a topic of considerable debate. Rothstein argues that much of the sorting occurs on the basis of time varying student heterogeneity and therefore lagged test scores and even student fixed effects may fail to capture important determinants of classroom allocation, but specification issues raise some questions about the strength of this critique.[7] To provide an estimate of the potential importance of such selection, Kane and Staiger (2008) develop a specification test of the validity of nonexperimental estimates for a small sample of Los Angeles teachers and cannot reject unbiasedness of various standard estimators. However, a lack of power inherent in such specification tests and potentially select nature of the sample limit the strength of the findings.

Finally, compensatory behavior on the part of schools or parents can bias downward estimates of the variance in teacher quality and bias estimates of teacher value added toward the school mean. Principals may assign additional subject specialists, paraprofessionals or other support staff to classrooms in which a teacher is struggling. Moreover, parents may devote additional time to academic support if they believe instructional quality is below the standard expected.

Equation (4), in which the estimated fixed effect for teacher j is the sum of the persistent component of teacher quality and an error term, ν_{jy}, provides a simple framework for examination of these issues and the assumptions that must be satisfied to produce unbiased estimates of both individual teacher fixed effects and the variance in teacher value added.:

(4) $\hat{\tau}_{jy} = \tau_{jy} + \nu_{jy}$

In general terms, the error term incorporates:

$$(5) \quad \nu_{jy} = f(\textit{unmeasured student, classroom composition, and school factors,} \\ \textit{nonpersistent teacher effects, and test measurement error})$$

Note that the test measurement error incorporates both elements of test reliability (consistency across time) and test validity (accuracy of measurement of desired dimensions).

Consider first the case where the correlation between each of the error components and persistent teacher quality equals zero and there are no non-persistent teacher effects. In this case the estimates of teacher value added, τ_{jy}, are unbiased, while the sample variance of the teacher fixed effects equals the sum of the true variance, the variance in unmeasured student, classroom composition and school factors, and the variance of test error. Therefore the estimate of the variance in teacher value added must be shrunk by subtracting estimates of the variances of all error components other than the variance of true quality.

[7] The evidence that time varying classroom heterogeneity is important is based on findings from models with student fixed effects and test score gain as the dependent variable. If the assumption of no knowledge depreciation is incorrect, it could appear that much student heterogeneity was time varying even if that were not the case.

Many use an empirical Bayes shrinkage estimator to produce a consistent estimate of the true variance in teacher quality, while an alternative approach is to use the adjacent year correlation among the teacher-by-year fixed effects for the same teacher to estimate the true variance.[8] Under the assumption that the errors are orthogonal across years, the covariance between adjacent year fixed effects equals the variance of true value added and a direct estimate of the correlation ρ equals

(6) $r_{12} = \text{var}(\tau) / \text{var}(\hat{\tau})$

Therefore multiplication of the estimated sample variance of $\hat{\tau}$ by the year-to-year correlation produces an estimate of the variance in teacher quality. Importantly, this correction mitigates problems due to both random test error and to non-persistent differences in classroom average student quality, either purposeful or random.

Violation of the assumption of no non-persistent components of teacher quality would introduce downward bias in the estimates that attempt to purge variation from non-persistent sources as the variance of teacher quality would be the sum of the variances of the fixed and non-persistent components. Evidence suggests that teachers improve quite a bit early in their careers, and personal difficulties, the birth of children, and experimentation with new pedagogies all lead to variation over time in effectiveness.

Violation of the assumption that value added is orthogonal to the remaining error components may introduce positive or negative bias depending upon the nature of both student and teacher sorting among schools and classroom assignment. In this case the covariance across adjacent years equals the sum of the variance in quality, the variance in the persistent components of the error term such as student skill, plus the covariance terms, and the methods described above would not purge the estimates of these variance and covariance terms.

Given the difficulty of quantifying all relevant student and school variables related to the matching of students and teachers both within and between schools and the limitations of specification tests, we adopt an alternative approach that separates school, grade, and year observations of classrooms by the process used to sort students among classrooms. Specifically, we examine whether there are either 1) significant differences in mean prior test score among classrooms based on an F- test[9]; or 2) whether the allocation of students across classrooms in grade g is independent of the allocation in grade g-1 based on a chi-squared test of the transition matrix. A classrooms in a school are placed in the "purposefully sorted" category based on the particular test if the hypothesis of no significant differences (in the case of method 1) or independence (in the case of method 2) is rejected at the five percent level. Otherwise the classrooms in a school are placed in the "random" category.

These tests are weak in the sense that the failure to reject the hypotheses of independence or no significant difference at the five percent level does not provide strong evidence that a school actually randomly assigns students among classrooms. However, reinforced by a combination with falsification tests of the effects of future teachers on current achievement and specification tests related to assumptions about the rate of learning depreciation that we apply

[8] Empirical Bayes or shrinkage estimators move the separate estimates toward the mean according to the variance of the estimated parameter (e.g., Sanders and Horn (1994), Gordon, Kane, and Staiger (2006)). Aaronson, Barrow, and Sander (2007), Rockoff (2004), and others use estimates of the error variance for the teacher fixed effects to adjust raw fixed effect estimates.

[9] This test is similar in spirit to that used by Clotfelter, Ladd, and Vigdor (2006).

below, we believe it produces a sample of schools for which common selection mechanisms are highly unlikely to introduce significant biases in the estimation of teacher effects.

III. Texas Schools Project Data

The stacked panel datasets constructed by the Texas Schools Project contain administrative records on students and teachers collected by the Texas Education Agency (TEA) from the 1989-1990 through the 2001-2002 school years. The data permit the linkage of students over time and of students and teachers in the same school, grade, and year. The statewide data do not match students and classroom teachers, but those matches have been provided for a single large Texas urban district, known henceforth as "Lone Star" District. Typically this match identifies a subject specialist in middle school and a general teacher in elementary school. Only regular classroom teachers are included in the analysis.

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), classification as special needs, and classification as limited English proficient. 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. We concentrate on math performance in this analysis. These criterion referenced tests evaluate student mastery of grade-specific subject matter, and this paper presents results for mathematics. (Correspondingly, we include teachers matched to students who are identified as teaching mathematics if students are not in self-contained classrooms). Test scores are converted to z-scores using the mean and standard deviation for the entire state separately for each grade and year to account for the effects of test score inflation and other changes to the tests.

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 valid test scores or other missing data and classrooms with fewer than five students with non-missing data.

IV. Estimation of the variance in teacher quality

An important starting point for the consideration of teacher quality is simply how much variance is there in teacher effectiveness in the classroom. We construct estimates of the variation in teacher quality as measured by value added to student mathematics achievement. The estimates are follow equation (3) where the explanatory variables include lagged math achievement, student demographic characteristics, student mobility variables, and a full set of teacher-by-year fixed effects. We compare teachers to all teachers in the Lone Star district and, because of concerns about the selection of schools by both parents and teachers, to just the set of teachers in their own school, grade and year.

To investigate the implications of nonrandom sorting of students across classrooms, the estimates are compared across samples that differ according to empirical descriptions of the processes for assigning students to classrooms. Based on this, we also explore the impact of specification error stemming from imposing the assumption of no knowledge depreciation by producing full set of estimates for both the lagged achievement and test score gain specifications of student skill.

A. Basic Estimates of the Variance o Teacher Quality

Table 1 reports a series of estimates of the variance of teacher-by-year fixed effects, the adjacent year correlation of estimated fixed effects for the same teacher, and the measurement error adjusted estimate of the variance in teacher quality. The first and second columns use both within- and between-school variation (i.e., compare teachers across the entire district), while the third and fourth use only within-school-grade-year variation, thus restricting comparisons to colleagues in the same school, grade, and year. In addition, the second and fourth specifications regression-adjust for differences in observable student characteristics including eligibility for free or reduced lunch, gender, race and ethnicity, grade level, limited English proficiency, special education, student mobility status, and year dummy variables.

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 the variation within school, grade, and year noticeably reduces the estimated variance in teacher value-added from the less restrictive specification. As expected given that most sorting occurs among schools, the included student characteristics have a much larger effect in specifications not restricted to within-school-grade-year comparisons (columns 1 and 2 versus columns 3 and 4).

The second row reports the adjacent year correlations in estimated teacher value-added. The magnitudes range from 0.24 to 0.35, indicating that roughly a fourth of the overall variance and slightly more than one third of the within-grade variance is persistent. These correlations show considerable stability in the impact of teachers, particularly when comparing teachers just to other teachers in the same school. Again the controls for student heterogeneity reduce the correlations less in the within-school-and-year specifications. Note that this does not have to be the case since, some of the year-to-year variation in student gains comes from random differences in student characteristics, meaning that the inclusion of controls for student heterogeneity could potentially increase the adjacent year correlation.

The final two rows report estimates of the variance and standard deviation (σ_τ) of true teacher value-added (based on equation (6)). The estimate of the overall variance, even when regression adjusted, equals 0.027 and is more than twice as large as the regression adjusted within-school-grade-year variance estimate of 0.013. The variance estimates of 0.013 means that a one standard deviation difference in teacher quality equals 0.11 standard deviations in terms of student achievement score, a magnitude in line with existing estimates in the literature.[10] Note, however, that that if there are distinct quality differences in teachers across schools – say, because of a set of principals that is adept at staffing their schools with high quality teachers – the within-school estimator will neglect an important component of the teacher quality variance.[11]

B. Sorting Among Classrooms

Despite the elimination of any between school variation in teacher quality and changes over time in the quality of instruction for a given teacher and influences of random shocks or error, the within-school-grade-year estimates may be biased. On the one hand, if principals assign the more cooperative or more engaged students to the better teachers, the differential could conflate true differences in quality with any student influences for which the empirical model does not account fully. On the other hand, compensatory assignment of the better students

[10] Recent reports of estimated effects of a one standard deviation change in teacher quality on achievement include 0.1 (Rockoff (2004)), 0.15 (Aaronson, Barrow, and Sander (2007)), and 0.18-0.20 (Kane and Staiger (2008)).

[11] These estimates all come from a single urban district. If there are important teacher quality differences across districts, the overall variation in teacher quality would be commensurately larger.

to the less effective teachers would bias downward estimates of the within-school-grade-year variance.

We account more fully for any such biases introduced by sorting through the identification of two samples of school-grade-year combinations for which the hypotheses either of no significant differences among teachers' classes in average prior year achievement or of independence between prior year and current year teacher assignments cannot be rejected at the five percent level. The first approach follows in the spirit of Clotfelter, Ladd, and Vigdor (2006) and is based on an F-test of the equality of mean prior year test score, while the second approach uses a chi-square test to examine the transitions of students who remain in the same school from grade g-1 to grade g. The school observations where we reject the null hypothesis are considered observations affected by purposeful sorting ("sorted") and all others are classified as "not sorted".

The top panel of Table 2 reports estimates from the lagged achievement model and uses the same specifications as in Table 1 for two different samples determined by tests of differences among classrooms in mean pretest score. Estimates in Columns 2 and 4 are generated from the sample of school-grade-year combinations in which "random" allocation of students among teachers could not be rejected at the 5 percent level, while estimates in Columns 1 and 3 are generated from the sample of school-grade-year combinations for which the hypothesis of equal classroom pretest means is rejected (i.e., the sorted sample). For the "not-sorted" sample, the within-school-grade-year estimated variance of teacher quality equals 0.011 which means that a one standard deviation difference in teacher quality translates to a 0.10 standard deviation difference in achievement. This is roughly 10 percent smaller than the full sample estimate of 0.113 reported in Table 1.

The bottom panel repeats the estimation for samples determined by chi-square tests of the independence of the current and prior allocation of students to teachers within each school.[12] The estimated within school-grade-year standard deviation of teacher quality equals 0.098 standard deviations of achievement, just slightly smaller than the estimate reported in the top panel. Thus, these two alternative ways of defining samples where student sorting seems less important yield estimates of the within school-grade-year variance in teacher quality that are very similar to those found in the full sample.

The estimates of variations in teacher quality are some 22-40 percent higher in the sorted samples than in the not-sorted samples for the preferred within-school estimates. For the estimates across the entire district, the within-school sorting has virtually no effect on the estimated variance in teacher quality, although again the higher estimated variance suggests that there are remaining elements of between-school sorting of teachers.

The results provide a *prima facie* case for the existence of substantial within-school variation in teacher value-added that is not an artifact of classroom sorting. Another way to approach this issue is to construct and falsification tests similar to those suggested by Rothstein (2008). The underlying idea is to estimate basic teacher value-added models such as those in Table 2 except to apply information about the subsequent year's teacher for each student rather than the current year's teacher. The intuition is that a future teacher cannot affect current year performance, so finding a similar distribution of teacher value-added for future teachers would raise serious doubts about the estimation strategy.

---

[12] Doing this test requires three consecutive grades – years 1 and 2 for estimation of the teacher-by-year fixed effects and a prior year that is used to test for sorting of students. Therefore, the samples for estimation under this sample stratification are just 38-45 percent as large as those used in the top portion.

This approach is nonetheless inconclusive in the face of student sorting in schools, because the very impact of sorting that it is supposed to identify can itself affect the test. Presume that students in a school are sorted by initial achievement. An effective teacher in grade g will improve student achievement and make it more likely that the student is sorted into the high ability class in grade g+1. Thus, the teacher in grade g+1 will look like she is an effective teacher (for students prior to entry into her class), because students have been sorted on observable achievement at the end of grade g.

We pursue this approach by modifying the analysis in Table 2 with two crucial differences. First, we replace the indicators for each current year teacher with indicators for the subsequent year teacher, meaning that we are estimating mean differences in the value added to grade g achievement among grade g+1 teachers. Second, we use information on the allocation of students among classrooms in grade g+1 to determine whether a school-grade-year combination should be placed in the sorted or not sorted category. This second adjustment is designed to guard against concerns about the falsification approach.

Table 3 presents estimates of the variance and standard deviation of teacher value added for both the actual and subsequent year teacher. For both ways of defining the sample of "not-sorted schools", the within-school-grade-year variance estimates and adjacent year correlations for future teachers are much smaller than those for current teachers, despite the fact that the tests used to categorize schools as "not sorted" are quite weak. Based on the differences in lagged achievement sorting test (top panel), the estimated within school variance of future teacher quality is roughly half as large as the estimated variance of actual teacher quality (0.0026 versus 0.012). With the test based on year-to-year transitions, in the not-sorted schools there is a much smaller unadjusted variance for effectiveness based on future teachers compared to actual teachers and a negative correlation for adjacent year performance of these teachers, suggesting that zero effect based on future teachers is the best interpretation.

For sorted schools, there is a significant fall in estimated quality between those based on actual versus future, where the counterfactual estimates from the within-school variance estimates for future teachers range from 40 to 70 percent as large as the variance estimates using the actual teachers. [13] Such a pattern is consistent with sorting on the basis of grade g achievement and does not prove that the estimates based on grade g teachers are biased. Because differences in grade g teacher effectiveness influence classroom assignments in grade g+1, the estimates will show significant differences in value added among grade g+1 teachers.

The results from comparing all teachers in the district fail to find such sharp differences by within-school allocation practices, as might well be anticipated. Our sample division is based on sorting within schools, while the within district estimates include between-school choices by both parents and school personnel – things that we are not confident of modeling in a satisfactory manner. Behavioral choices of schools by students and by teachers introduce uncertainty about our ability to isolate teacher effectiveness from other determinants of achievement, leading us to focus on the within-school-grade-year estimates of the variance of teacher value added.

Interestingly, our preferred within-school estimates for the not-sorted sample are quite similar to those reported in {Rivkin, 2005 #4369}, a study that used grade level aggregate data to circumvent the problem of within-school sorting. This consistency across very different estimation methods reinforces the fact that there is significant variation in the true effectiveness

[13] The analysis in Rothstein (2008) was more dramatic, where the variance in "future teacher quality" looked very close to the variance of actual teacher quality. A significant portion of this, however, resulted from not correcting for measurement error in the value-added estimates.

of teachers within the typical school.

V. Teacher Transitions and Quality

The existing evidence – combining estimates in this paper with the variety of prior estimates discussed previously – reinforces the widely held opinion that high teacher quality is the most important element of a high quality school. No other measures of school factors – including class size, peer effects, and curriculum – have been shown to have effects approaching that of effective teachers. This finding suggests that the distribution of teachers by effectiveness can have significant effects on the distribution of student outcomes. And, this conclusion leads naturally into an investigation of how teacher behavior and the choices made by teachers determine the observed distribution of student achievement.

These issues seem particularly important for disadvantaged and minority students who might have more restricted options in their residential choices and thus implicitly in their schools given the limited resources for private education. Evidence suggests that new teachers gravitate toward the communities where they were raised, and one often hears that low income urban and rural districts lose many of their better teachers to suburban districts.[14] However, there remains little or no evidence regarding the crucial question of whether it is the more effective teachers who tend to leave urban and rural districts. We now make use of information on student achievement to describe differences in teacher effectiveness between teachers who remain in their initial school, teachers who transfer to another school within the large urban district, those who transition to another district, and those who leave the Texas Public Schools entirely. Essentially we re-estimate the previous models substituting transition indicators in place of the vector of teacher fixed effects in order to understand how teacher movements affect the distribution of education quality.

Three features of teacher mobility define the context for this work. First, teacher turnover is large (see, for example, Boyd, Lankford, Loeb, and Wyckoff (2002), Hanushek, Kain, and Rivkin (2004), and Podgursky, Monroe, and Watson (2004)). The turnover of inexperienced teachers is especially high; only 70 percent of teachers with fewer 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.

This analysis provides insight into how teacher mobility and teacher effectiveness interact to define the surface of school quality across schools and across subpopulations of students within an area. We focus on isolating patterns of implicit teacher quality through estimating a range of specifications that differ by comparison group, the classroom assignment process, and the timing of the measurement of teacher quality relative to the transition period. First, one dimension of specifications limits the comparison to differences among colleagues in the same school, grade, and year, and another broadens comparison to the district level. Second, another dimension contrasts results based on the entire sample of students with those that use subsamples determined by the same specification tests of "random" allocation used above. Third, most specifications estimate teacher effectiveness for the school year immediately prior to the transition, while a smaller number go back one more school year in order to distinguish between

[14] Boyd, Lankford, Loeb, and Wyckoff (2005) describe the importance of residential location in teacher job search.

temporary and longer term differences in teacher effectiveness. Fourth, some specifications characterize effectiveness when transition patterns are allowed to vary by teacher experience. Finally, the pattern of teacher transitions is permitted to differ by student achievement in the school and by the proportion of students who are black – factors previously identified in the literature on teacher mobility. (Note also that, in order to control for other achievement-related factors, all specifications include a full set of single year experience dummies, a full set of student race-ethnicity dummies, and indicators for female, eligible for a subsidized lunch, classification as special needs, classification as limited English proficient, a family move, and a transition to middle school).

Teacher movement is substantial within the Lone Star district. Among new teachers (0-1 years of experience), the annual exit rate from Texas public schools is 12 percent. Another 12 percent annual change campuses and 7 percent move to a new district in Texas. Even among more experienced teachers, however, some 18 percent leave their current school each year.

Table 4 reports estimates of mean differences in teacher value added to mathematics achievement by teacher transition type. These estimates provide little or no evidence that more effective teachers have higher probabilities of exiting the Lone Star District regardless of their destination. In fact those who exit the Texas public schools entirely are significantly less effective on average than those who stay regardless of whether they are compared to all stayers in the district or only those in the same school, grade, and year. In the school year immediately prior to leaving the Texas Public Schools, exiting teachers produced achievement gains that were 0.06 s.d. below the average teacher remaining in the school (or the district). Moreover, those who switch campuses within the same district were also significantly less effective, though the deficit is smaller than that observed for those exiting the Texas public schools. In contrast, those switching to another Texas school district were not significantly different on average from the stayers.[15]

Much of the attention to teacher movement focuses on initial years of experience, following the common observation that new teachers are initially assigned to the worst schools but then try to move quickly. The prior specifications included individual year of teacher experience controls, but Table 5 considers the possibility of separate patterns of effectiveness by transition type for teachers with one, two, and three years of experience. The small sample sizes lead to quite imprecise estimates, particularly for teachers with two and three years of experience. Nonetheless, the results suggest some marked differences between first year teachers who remain in teaching and those with more than one year of prior. The first year teachers who change schools are significantly more effective on average than stayers, regardless of whether the destination school is located in the same district or a different district. In contrast to school switchers, there appears to be little variation across these experience categories in the average effectiveness of teachers who exit the public schools. For experienced teachers (four or more years of experience), those staying in the school are uniformly more effective than those leaving.

Most of the concern about selective teacher attrition centers on lower performing schools serving disadvantaged students. As noted, past research has shown that schools serving disadvantaged populations have higher turnover and, by implication, more inexperienced teachers because teachers who leave tend to be replaced by new teachers. To address whether

[15] Another sensitivity check we examined was whether the relationship with prior achievement was linear. A series of indicator variables for different levels of prior achievement, which allow a very flexible relationship, produced virtually indistinguishable estimates of effectiveness for the different transition groups – leading us to continue with the linear specification in the analyses below.

the transition patterns differ by school characteristics, we divide the schools into two equal sized categories on the basis first of initial achievement and second of the proportion of students who are black.[16] We then examine quality differences by transition and these student characteristics for all teachers and for first year teachers.

The estimates in Table 6 provide little support for the view that the schools with lower performing students or higher black enrollment suffer larger losses of highly effective teachers. To the contrary, in disadvantaged schools defined either by low achievement or higher percentage black students there is a strong tendency for relatively ineffective teachers to depart. This finding holds regardless of whether the comparison group of teachers is defined by the district or the specific school, grade, and year. Moreover, the lesser effectiveness of leavers in the less advantaged schools is statistically significant for those who either switch campuses or exit the public schools. The average effectiveness of teachers moving to a new district (the middle panel) is not significantly different from the average for those who stay, regardless of the achievement or racial concentration of the origin school.

This rather desirable outcome is not the situation usually discussed.[17] While there is considerable variation in the effectiveness of teachers within each of these transition streams, the selection process is leading on average to favorable results in the more disadvantaged schools compared

The transition patterns for teachers in their first year of experience, reported in Table 7, show an interesting similarity from the prior aggregate results. Table 5 indicated that the first year teachers who leave a school tend to be more effective than those who stay. When we divide the schools by the underlying characteristics of the student body, however, we find that it is the more advantaged schools that are disproportionately the more effective teachers. Those leaving high achievement schools for other schools in the district and other districts are significantly more effective than remaining colleagues, while those who exit Texas public schools are not significantly different on average. Among teachers leaving from low achievement schools, the ineffectiveness of those exiting the public schools is particularly striking. The differences by student proportion black tend to be smaller and less pronounced but still follow the same pattern. Again, even for new teachers, the more disadvantaged schools by student achievement and race appear to suffer little disadvantage in terms of losing the more effective teachers.

The overall patterns in terms of effectiveness indicate little or no evidence that more effective teachers are the ones moving among Lone Star District schools, but the possibility still remains that the more effective teachers among those moving gravitate toward higher achieving schools within the district. If so, this would implicitly leave struggling schools with less effective teachers. To investigate this possibility, we concentrate on the within-district movers and add interactions between origin and destination school characteristics. The results in Table 8 reveal little or no evidence of significant differences by destination school type regardless of the student characteristics in the origin school. The bold coefficients indicate that teachers who make intradistrict moves from either schools with low achievement or with high concentrations of black students tend to be significantly less effective (compared to teachers who remain in

[16] Even though Texas and the Lone Star district have a substantial Hispanic population, our previous analysis of teacher mobility found that black concentrations and not overall minority concentrations were most salient for teacher moves (Hanushek, Kain, and Rivkin (2004)). Thus, we concentrate solely on black concentration throughout this work.

[17] Although test error could lead to some misclassification when schools are sorted by average achievement level, there is no reason for such error to influence estimated value-added relative to other teachers in the same school, grade, and year.

those schools. The marginal effect for the specific destination (either a high achieving school or a school with a high percentage of black students) are uniformly small and statistically insignificance. In other words, by tracing through destinations of the movers within the district there is again no sense that that teacher movement is systematically disadvantaging the schools with the greatest needs.

Nonrandom allocation of students among classrooms provides a potential impediment to identification of differences by transition type. For example, if teachers assigned the worst students are more likely to leave and if the regressions do not account fully for the within-school student heterogeneity, the quality of leavers may be biased downward by systematic errors, ν_{jy}, in the estimation of teacher effectiveness. To address this issue, we again divide the school, grade, year combinations into samples of schools on the basis of whether the hypothesis of no significant differences in classroom initial achievement can be rejected at the five percent level.[18]

The estimates in Table 9 show that coefficients from the "random" sample tend to be more negative, indicating that district switchers and particularly those exiting the public schools entirely are truly less effective relative to stayers. The only exception to this is the within-district movers who leave from the low achievement or high black concentration schools appear more ineffective in the sorted sample than in the not-sorted sample. But the conclusion remains that the most disadvantaged schools tend to lose ineffective teachers, and this does not appear to result from just biases arising from any purposeful sorting.

To this point estimates of teacher effectiveness are based on the academic year immediately prior to any transition, but this chronology potentially complicates interpretation of the results. Are movers less effective in their transition year because they are less skilled teachers? Or, because of a negative shock such as an unruly class or bad relationship with a new principal that both induces a transition and degrades instructional effectiveness? Or because they put forth less effort once they have decided to leave the school?

In order to isolate skill differences, we turn to measuring teacher quality by value added in the year prior to the transition year. For example, we describe the distribution of quality for transitions following the 1999 school year with average student achievement during the 1998 school year, implying that any shocks or change in effort related to the transition do not affect the estimates of teacher effectiveness. This approach does introduce analytical difficulties, however, because the sample size is significantly reduced by eliminating student performance information on the final year taught for each teacher and for all who teach only a single year in Lone Star district. Thus, the estimates will necessarily be more imprecise.

Table 10 reports within-school estimates of the effectiveness of leaving teachers that are based on achievement in both the transition year and the previous year (disaggregated by school demographic characteristics).[19] Two findings stand out in the comparison of performance in the exit year and the year prior. First, for those who leave the Texas public schools, the results for assessments based on teacher performance in the penultimate year are entirely consistent with those found above. Those exiting teaching are lower in average quality than those staying in their schools, and the differential effectiveness is, if anything, larger in the more disadvantaged schools (lower achievement or higher proportion black). In fact, the estimated average achievement deficit is actually slightly larger in the penultimate year: -0.094 vs. -0.061 in the

---

[18] As in the prior section, we duplicated the analysis based on a chi-square test of independence of assignment patters across grade and classroom. This produces similar but noisier results due to the much smaller sample.

[19] Note that, although the point estimates for the current scores in comparisons across the district (not shown) differ some from the comparable estimates in Table 8 that use the entire sample, the patterns are qualitatively the same.

low achievement schools and -0.084 vs -0.078 in the high proportion black schools. Second, however, for those who move to another school in the district, the estimated lower performance in comparison to stayers does not show up in the previous year performance. There appears to be little difference in average teacher effectiveness of within-district movers when assessed by earlier performance, suggesting the possibility of negative shocks or changes in effort account for the apparent lower average effectiveness seen before when based on effectiveness assessed in the transition year.

These overall findings would be consistent with the notion that self-recognition of not being a very effective teacher precipitates exit from the profession, while a temporary negative shock precipitates a transition to another school. But it is also consistent simply with principal pressure on ineffective teachers to leave, a possibility that has not been well-analyzed or documented. These alternative explanations clearly point to different potential policy actions, but within our current data it is impossible to distinguish between them.

If the difference in estimated effectiveness based for transition year and prior year assessments are the result of adverse shocks in the transition year, one might expect average improvements in effectiveness following a move to a new school. For example, in surveys teachers often cite school leader quality as an important determinant of working conditions; if teachers move within the district to find a better match with leadership, teacher effectiveness might be expected to rise following a move.

Table 11 provides estimates of the average change in estimated effectiveness for those staying in their prior school and those who switch schools within the Lone Star District. The pattern across all teachers suggests that those who move within the district are trivially different in performance change. The within-district difference is statistically significant, but the magnitude is only 0.003 s.d. gain in student achievement growth. The within-school assessments yield no significant differences in estimated effectiveness. Interestingly, there is also no differential for inexperienced teachers. As indicated, inexperienced teachers gain in performance with the first years of experience, but, if anything, the experience gain is less if it is accompanied by a move to another school in the district. Taken together, the table provides little evidence in support of the view that a change of school leads to substantial improvement for many teachers.

VI. Conclusion

Many policy discussions rightfully focus on the plight of disadvantaged students, particularly those in schools with concentrations of low income, low achieving, and heavily minority schools. On average these students reach significantly lower levels of achievement. Are these performance outcomes a result of the quality of schools and particularly the quality of teachers? Could the picture be altered by improving the effectiveness of teachers serving these populations?

We provide two pieces of evidence with substantial policy import that are directly related to these questions. First, we consider how much variation there is in teacher effectiveness, because this provides an indication of how much leverage could be attached to altering the distribution of teachers across schools. Second, we delve into a detailed consideration of how the distribution of teacher quality evolves over time with the mobility of teachers across schools and out of the profession. The latter issue arises from the commonly held belief that both school policy and teacher choices lead better teachers systematically to leave the most disadvantaged schools.

We can address both of these issues using rich student achievement data from a large urban district in Texas. The stacked panel data follow individual teachers and students over time and permit linking the growth in student math achievement to specific teachers and classrooms.

Ascertaining the variation in teacher effectiveness is complicated by several factors. First, we infer teacher effectiveness from scores on standardized mathematics tests, but these test measures are prone to error that will enter into estimates of the value-added of individual teachers. Second, a variety of both contemporaneous and historical factors affecting student achievement are difficult to measure adequately, implying that problems of omitted variables could bias estimates of teacher effectiveness. Third, value-added of teachers is inferred from the average achievement gains of students within a teacher's classroom, but students may not be randomly assigned to individual teachers, leading to potential difficulties in separating the teacher from the students or the collection of peers.

We deal with each of these issues in estimating the variance of true teacher effectiveness. The fundamental empirical specification employs models with lagged achievement that implicitly incorporate past influences on learning, while permitting direct estimation of the importance of depreciation of prior knowledge. We then use the time pattern of student placement in classrooms to separate a sample of schools where sorting appears significant versus those were it does not.[20] Importantly, the estimated variances in teacher quality from the "random" samples generally fall in the lower range of existing estimates. After correcting for measurement error (which accounts for somewhat over half of the observed classroom variation in achievement gains, our estimates are consistent with alternative approaches that are not subject to the same potential sorting difficulties and measurement problems. The efficacy of this approach is reinforced by falsification tests that compare actual teacher value-added to an estimate based on the students' future teachers. Future teachers have far less explanatory power in the sample of schools without significant sorting, but have much explanatory power for performance in schools that employ sorting (as should be the case).

The estimates show very significant heterogeneity in teachers. The magnitude implies substantial potential policy importance of policies directly related to improved teacher effectiveness.

Given the sizeable differences in effectiveness, the dynamics of the teacher-school matching process plays a substantial role in determining the pattern of school quality and implicitly of differences by race and income. Past analysis suggests considerable difficulty in ordering prospective teachers by quality during the hiring process (and before they are observed in the classroom). Therefore the subsequent choices of schools and particularly teachers play a primary role in the determination of the distribution of teacher quality, particularly in large urban districts that struggle to hire qualified teachers and that experience significant teacher mobility and turnover. An important element of many policy discussions is the possibility that teachers systematically choose schools with higher achieving and less disadvantaged students, reflecting an underlying belief that schools with more advantaged students are simply easier and more rewarding to teach in. More importantly, it is frequently asserted that it is the best teachers – those with better alternatives – that disproportionately leave the more difficult schools.

[20] We use two approaches to define schools with significant sorting. First, for each school, we test for equality of average entering achievement for students across the different classrooms and put schools with significant differences in the category of "sorted" and put all others in the "not-sorted" category. Second, we test for independence of classroom assignment of students from their classroom assignment in the prior year. Because the latter approach significantly reduces the samples of students, we rely on sorting by achievement through most of the analysis. However, when compared, the two methods yield qualitatively similar results.

Our estimates provide little to no support for the belief that those who transition out of Lone Star schools are more effective on average than stayers. This finding is clearest in looking at those who exit from teaching altogether, as leavers from low achievement and high proportion black schools are significantly less effective than stayers. Moreover, there is little evidence that teachers moving to high achievement or low proportion black schools systematically outperform stayers or movers to other schools.

These findings do not resolve all of the interpretive or policy questions, because we cannot distinguish among alternative underlying mechanisms that are consistent with these patterns of teacher transition. For example, our consistently low estimates of the average effectiveness of teachers exiting from Texas public schools could combine exits resulting from the identification and removal of poor performing teachers by principals with voluntary choices of teachers who recognize that they are not effective in the classroom. But, more importantly, the lower average effectiveness of leavers could combine the effects of some better than average teachers who choose to leave for other jobs with another group of ineffective teachers who are forced to leave. In the absence of information on the circumstances of the separation it is not possible to quantify the relative quality of voluntary leavers versus active policies. Our investigation of the pattern of teacher effectiveness leading up to a transition and following a transition provides suggestive information on these issues, but it is an insufficient basis for policy. Finally, the suggestive evidence that young district switchers outperform stayers is consistent with a concern that urban schools lose talented young teachers, though this problem does not seem to be confined to the more disadvantaged schools.

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).

———. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25,no.1 (January):95–135.

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, 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.

Boyd, Don, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2006. "How changes in entry requirements alter the teacher workforce and affect student achievement." *Education Finance and Policy* 1,no.2 (Spring):176-216.

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."

———. 2005. "The draw of home: How teachers' preferences for proximity disadvantage urban schools." *Journal of Policy Analysis and Management* 24,no.1 (Winter):113-132.

Clotfelter, Charles T., Helen F. Ladd, and Jacob l. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41,no.4 (Fall):778-820.

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.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger. 2006. "Identifying effective teachers using performance on the job." Washington, Hamilton Project Brookings Institution (April).

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.

———. 1986. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24,no.3 (September):1141-1177.

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

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39,no.2:326-354.

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

Kane, Thomas J., and Douglas O. Staiger. 2008. "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validation of Non-Experimental Estimates." Harvard University, (mimeo) March).

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher sorting and the plight of urban schools: A descriptive analysis." *Educational Evaluation and Policy Analysis* 24,no.1 (Spring):37-62.

McCaffrey, Daniel F., J.R. Lockwood, Thomas A. Louis, and Laura Hamilton. 2004. "Models for value-add modeling of teacher effects." *Journal of Educational and Behavioral Statistics* 29,no.1:67-101.

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 J., 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, Steve G. 2008. "Value Added Analysis and Education Policy." Washington, Calder Briefs Center for Analysis of Longitudinal Data in Education Research (

Rivkin, Steven G. 2005. "Cumulative nature of learning and specification bias in education research." mimeo, Amherst College (

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73,no.2 (March):417-458.

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.

Rothstein, Jesse. 2008. "Teacher quality in educational production: Tracking, decay, and student achievement." Cambridge, MA, NBER W14442, National Bureau of Economic Research (October).

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.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner. 2006. "Do Teachers Really Leave for Higher Paying Jobs in Alternative Occupations? ." *Advances in Economic Analysis & Policy* 6,no.1.

———. 2007. "Race, poverty, and teacher mobility." *Economics of Education Review* 26,no.2 (April):145-159.

Wainer, Howard. 2004. "Introduction to a Special Issue of the Journal of Educational and Behavioral Statistics on Value-Added Assessment." *Journal of Educational and Behavioral Statistics* 29,no.1 (January):1-3.

Table 1. Estimated Variance in Teacher Quality (254,046 observations)

	within district		within school	
	without demographic variables	with demographic variables	without demographic variables	with demographic variables
variance in fixed effects – $var(\hat{\tau})$	0.120	0.111	0.040	0.038
adjacent year correlation – r_{12}	0.269	0.244	0.348	0.339
variance in teacher quality – $var(\tau)$	0.032	0.027	0.014	0.013
stnd. dev. in teacher quality – σ_{τ}	0.180	0.165	0.118	0.113

Note: teacher fixed effects are produced from regressions of math score on lagged math score. Specifications with demographic characteristics also include indicators for female, race-ethnicity, low income, limited English proficient, special needs, first year in middle school, and family initiated school change.

Table 2. Estimated Variance in Teacher Quality by Alternative Tests of Significant Classroom Sorting

	within district		within school	
	sorted	not-sorted	sorted	not-sorted
1. Sorting Tested by Pretest Mean Achievement				
variance in fixed effects – $var(\hat{\tau})$	0.092	0.146	0.040	0.042
adjacent year correlation – r_{12}	0.425	0.283	0.496	0.254
variance in teacher quality – $var(\tau)$	0.039	0.041	0.020	0.011
stnd. dev. in teacher quality – σ_τ	0.198	0.203	0.141	0.103
Observations	161,990	79,047	161,990	79,047
2. Sorting Tested by Classroom Assignment Patterns				
variance in fixed effects – $var(\hat{\tau})$	0.081	0.116	0.034	0.036
adjacent year correlation – r_{12}	0.411	0.274	0.422	0.265
variance in teacher quality – $var(\tau)$	0.033	0.032	0.014	0.010
stnd. dev. in teacher quality – σ_τ	0.182	0.178	0.120	0.098
observations	56,656	24,010	56,656	24,010

Note: fixed effects derived from regressions that include lagged test score and the demographic variables listed in Table 1.

Table 3. Estimated Variance in Teacher Quality for Actual and Subsequent Grade Teachers and Alternative Tests of Classroom Sorting

	within district				within school			
	sorted		not-sorted		sorted		not-sorted	
	actual teacher	next teacher	actual teacher	next teacher	actual teacher	next teacher	actual teacher	next teacher
1. Sorting Tested by Pretest Mean Achievement (teachvartrackny.log)								
variance in fixed effects – $var(\hat{\tau})$	0.098	0.093	0.170	0.135	0.042	0.038	0.053	0.023
adjacent year correlation – r_{12}	0.264	0.180	0.151	0.078	0.440	0.174	0.219	0.111
variance in teacher quality – $var(\tau)$	0.026	0.017	0.026	0.011	0.018	0.0066	0.012	0.0026
stnd. dev. in teacher quality – σ_τ	0.161	0.129	0.160	0.103	0.136	0.081	0.108	0.051
Observations	56,051	56,051	23,172	23,172	56,051	56,051	23,172	23,172
2. Sorting Tested by Classroom Assignment Patterns (teachvartransition.log)								
variance in fixed effects – $var(\hat{\tau})$	0.096	0.088	0.169	0.146	0.044	0.036	0.049	0.025
adjacent year correlation – r_{12}	0.343	0.372	0.157	0.114	0.443	0.382	0.325	-0.083
variance in teacher quality – $var(\tau)$	0.033	0.033	0.027	0.017	0.019	0.014	0.016	
stnd. dev. in teacher quality – σ_τ	0.181	0.181	0.163	0.129	0.140	0.117	0.126	
Observations	55,418	55,418	23,485	23,485	55,418	55,418	23,485	23,485

Table 4. Average Differences in Teacher Quality by Transition

(no move is omitted category; 254,046 observations; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses)

	Within district	Within school
change campus	-0.048	-0.027
	(3.64)	(2.49)
change district	0.019	-0.019
	(0.99)	(1.26)
exit Texas public schools	-0.058	-0.061
	(3.79)	(5.24)

Note: Coefficients on teacher transition variables come from regressions of math score on the transition variables plus lagged score, indicators for female, race-ethnicity, low income, special needs, limited English proficient, first year in middle school, family initiated move, shares of students in campus, grade, and year who are black, Hispanic, asian, low income, special needs, limited English proficient, movers, peer average lagged achievement, a full set of teacher experience dummies, and a full set of year-by-grade dummies.

Table 5. Average Differences in Teacher Quality by Transition Type and Experience at Time of Move (within school comparisons; no move is omitted category; regressions from same sample and use same specifications as those in Table 4)

Experience at Move	change campus	change district	exit public schools
following 1st year	0.062	0.067	-0.049
	(1.97)	(1.79)	(1.66)
following 2nd year	-0.019	-0.057	-0.020
	(0.45)	(1.46)	(0.58)
following 3rd year	-0.048	-0.032	-0.057
	(1.57)	(0.89)	(1.68)
following 4th+ years	-0.040	-0.031	-0.070
	(3.23)	(1.66)	(4.89)

Table 6. Average Differences in Teacher Quality by Transition Type and Student Characteristics in School

(schools divided by being above or below district average for student achievement or percent black; no move is omitted category; no move is omitted category; 251,943 observations; same variables as in Table 4 specifications)

student characteristic	average achievement		proportion of students black	
	within district comparisons	within school comparisons	within district comparisons	within school comparisons
change campus				
from low value school	-0.081	-0.062	-0.027	-0.003
	(4.51)	(4.35)	(1.48)	(0.22)
from high value school	-0.013	0.006	-0.070	-0.055
	(0.72)	(0.40)	(3.89)	(3.40)
change district				
from low value school	-0.003	-0.033	0.028	-0.015
	(0.09)	(1.53)	(1.14)	(0.80)
from high value school	0.041	-0.008	0.010	-0.028
	(1.66)	(0.41)	(0.17)	(1.16)
exit public schools				
from low value school	-0.081	-0.086	-0.055	-0.051
	(3.76)	(5.25)	(3.05)	(3.75)
from high value school	-0.037	-0.043	-0.066	-0.082
	(1.80)	(2.79)	(2.62)	(4.16)

Table 7. Average Differences in First Year Teacher Quality by Transition Type and Student Characteristics in School

(schools divided by being above or below district average for student achievement or percent black; no move is omitted category; no move is omitted category; 251,943 observations; same variables as in Table 4 specifications plus a full set of teacher move indicators for all teachers not in their first year)

student characteristic	average achievement		proportion of students black	
	within district comparisons	within school comparisons	within district comparisons	within school comparisons
change campus				
from low value school	-0.038	0.018	-0.001	0.106
	(0.78)	(0.51)	(0.01)	(2.79)
from high value school	0.036	0.133	-0.033	0.008
	(0.42)	(2.28)	(0.56)	(0.19)
change district				
from low value school	0.107	0.034	0.098	0.051
	(2.06)	(0.67)	(1.93)	(1.73)
from high value school	0.064	0.114	0.082	0.073
	(1.00)	(2.37)	(1.39)	(1.23)
exit public schools				
from low value school	-0.127	-0.120	-0.077	-0.050
	(2.99)	(3.49)	(2.20)	(1.51)
from high value school	-0.010	0.033	-0.084	-0.061
	(0.22)	(0.97)	(1.58)	(1.47)

Table 8. Average Differences in Teacher Quality for Teachers Transferring within District by Student Characteristics in Origin and Destination School

	average achievement		proportion of students black	
	within district comparisons	within school comparisons	within district comparisons	within school comparisons
Exit low value school	**-0.077**	**-0.057**	-0.005	0.013
	(2.08)	**(1.90)**	(0.19)	(0.63)
-- transition to high value school	0.040	0.027	-0.037	-0.027
	(0.64)	(0.53)	(1.04)	(0.99)
Exit high value school	-0.22	0.063	**-0.096**	**-0.087**
	(0.53)	(1.61)	**(2.51)**	**(2.60)**
-- transition to high value school	0.044	-0.048	0.032	0.044
	(0.82)	(0.86)	(0.76)	(1.21)

Notes: Schools are divided by being above or below district median for achievement or percent black; no move is omitted transition category. All specifications include indicators for changing districts and for exiting public schools along with the covariates included in the models in Table 4.

Steve: It would be good to calculate t-stats for each of the "sums" – leave low/go to high and leave high/go to high. We can then put an a. not stat signif from 0 or b. stat signif from zero on those eight values.

Table 9. Differences in Teacher Quality by Transition and Distribution of Students Among Classrooms (within-school comparisons; no move is omitted category; classification of classroom allocation mechanism based on classroom differences in prior achievement)

	Average achievement		proportion black students	
	sorted	not-sorted	sorted	not-sorted
change campus				
from low value school	-0.076	-0.028	-0.010	0.021
	(4.36)	(1.12)	(0.56)	(0.87)
from high value school	0.012	-0.004	-0.064	-0.036
	(0.58)	(0.18)	(3.00)	(1.50)
change district				
from low value school	-0.022	-0.087	-0.012	-0.023
	(0.92)	(1.98)	(0.56)	(0.69)
from high value school	-0.011	0.003	-0.023	-0.057
	(0.51)	(0.08)	(0.86)	(0.96)
exit public schools				
from low value school	-0.065	-0.139	-0.035	-0.085
	(3.28)	(4.76)	(2.11)	(3.63)
from high value school	-0.035	-0.054	-0.072	-0.108
	(1.83)	(2.09)	(2.97)	(3.22)
Observations	159,569	77,939	159,569	77,939

Note: Same variables as regression specifications used in Table 4.

Table 10. Differences in Teacher Quality by Transition, School Characteristics, and timing of quality Estimate (within-school comparisons; no move is omitted category; 162,060 observations)

	Average achievement		proportion black students	
	Prior year estimate	Current year estimate	Prior year estimate	Current year estimate
change campus				
from low value school	0.002	-0.064	-0.010	-0.029
	(0.08)	(3.13)	(0.43)	(1.47)
from high value school	0.003	-0.026	0.014	-0.059
	(0.15)	(1.34)	(0.70)	(2.79)
change district				
from low value school	-0.014	-0.008	-0.010	-0.003
	(0.47)	(0.22)	(0.44)	(0.10)
from high value school	-0.001	0.012	-0.005	0.008
	(0.06)	(0.43)	(0.16)	(0.24)
exit public schools				
from low value school	-0.094	-0.061	-0.025	-0.043
	(3.76)	(2.61)	(1.34)	(2.35)
from high value school	-0.014	-0.054	-0.084	-0.078
	(0.75)	(2.90)	(3.34)	(3.03)

Note: Same variables as regression specifications used in Table 4.

Table 11. Average year to year change in value added over adjacent year for stayers and campus switchers within Lone Star District, by Transition and Experience (chqual.log)

	within-district comparisons	within-school comparisons	observations
all teachers			
same campus	-0.073	0.006	4,305
new campus	0.003*	-0.002	205
0 or 1 yr experience			
same campus	0.022	0.056	517
new campus	0.015	0.023	31

*reject hypothesis of no difference by transition status at 0.05 level