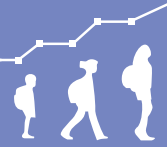


CALDER



NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute



A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

*Constrained
Job Matching*

Does Teacher Job Search
Harm Disadvantaged
Urban Schools?

ERIC A. HANUSHEK
AND STEVEN G. RIVKIN

Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?

Eric A. Hanushek
Stanford University & NBER

Steven G. Rivkin
Amherst College & NBER

Contents

Acknowledgements	ii
Abstract	iii
Introduction	1
Employee Turnover and Firm Matching	3
Measurement of Teacher Effectiveness	6
Cumulative Model of Learning	7
Estimation of Teacher Value-Added	9
Texas Schools Project Data	13
Estimation of the Variance in Teacher Quality	14
Basic Estimates of the Variance of Teacher Quality	14
Sorting Among Classrooms	16
Teacher Transitions and Productivity	20
Overall Differences by Transition	21
Differences by Student Demographic Composition	23
Sensitivity Analyses	26
Conclusion	29
Methodological Issues	29
Policy Implications	30
References	33
Tables and Figures	36

Acknowledgements

The authors thank Dan O’Brien for help with the data development and early analytical work. Greg Branch provided superb research assistance. This research received support from the Spencer Foundation, the Hewlett Foundation, and the Packard Humanities Institute, as well as the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) supported through Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education.

CALDER working papers have not gone through final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication.

The Urban Institute is a nonprofit, nonpartisan policy research and educational organization that examines the social, economic, and governance problems facing the nation. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or any of the funders. Any errors are attributable to the authors.

CALDER, The Urban Institute
2100 M Street N.W., Washington, D.C. 20037
202-261-5739 • www.caldercenter.org

Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?
Eric Hanushek and Steven Rivkin
CALDER Working Paper No. 42
May 2010

Abstract

Search theory suggests that early career job changes on balance lead to better matches that benefit both workers and firms, but this may not hold true in teacher labor markets characterized by salary rigidities, barriers to entry, and substantial differences in working conditions that are difficult for institutions to alter. Education policy makers are particularly concerned that teacher turnover may have adverse effects on the quality of instruction in schools serving predominantly disadvantaged children. Although these schools experience higher turnover, on average, than other schools, the impact on the quality of instruction depends on whether more productive teachers are more likely to depart. In Texas, the availability of matched panel data of students and teachers enables the isolation of teachers' contributions to achievement despite the intricacies of the choices of families, teachers, and administrators. Teachers who remain in their school tend to outperform those who leave, particularly those who exit Texas public schools entirely. This gap is larger for schools serving mainly low income students—evidence that high turnover is not nearly as damaging as many suggest.

Introduction

Policy discussions of teacher quality typically highlight the institution of tenure, barriers to entry into the profession, salary rigidities, seniority rules, and the importance of working conditions associated with school location and demographic composition to distinguish teacher labor markets from other occupations with fewer labor market rigidities. The unique structure of teacher labor markets is often thought to contribute to high early-career turnover in schools serving disadvantaged populations and is thought to impede efforts to raise the quality of education.

Quite generally, turnover is thought to be bad, and exiting teachers are thought to be the more productive. These concerns stand in stark contrast to much of labor search theory, where early-career turnover provides a mechanism for improving job matches that benefits both workers and firms. It may be that teacher labor markets lead to different outcomes, but it may also be the case that the concerns regarding turnover are misplaced. This paper provides direct evidence about the impacts of school job matching on productivity and student achievement.

Although teacher labor markets may be structured differently from other occupations, high turnover is certainly not a distinguishing feature. The data suggest that exit rates from teaching mirror those in non-teaching occupations (Stinebrickner (2002), Ballou and Podgursky (2002)). Consequently, disproportionate harm from the turnover of teachers must come from the character of that turnover rather than simply the level. Yet still, there remains little evidence on productivity differences between teachers who transition out of a school and those who remain.

In this paper we investigate the effects of turnover on the quality of instruction in a large urban district in Texas. The absence of direct measures of productivity typically hinders efforts to measure the effect of turnover on worker quality apart from the use of wage changes as an indirect proxy for changes in productivity. However, the availability of matched panel data of students and teachers enables the estimation of teacher quality based on the impact on student achievement. This outcome-based

productivity measure provides important information on the effects of turnover and the desirability of particular public policies designed to improve the quality of instruction in public schools broadly and more specifically in low achieving schools with high concentrations of disadvantaged and minority students.

The next section briefly reviews theories of job turnover and empirical evidence on turnover and wage changes and then considers ways in which the unique aspects of teacher labor markets may or may not lead teacher turnover to be particularly costly to schools. Section II develops the empirical model used to measure teacher productivity based on test scores that provide imprecise measures of achievement. This section highlights the difficulties introduced by the purposeful choices of families, teachers, and administrators and develops an approach to overcoming these impediments to identification.

The subsequent two sections present the results of the empirical analyses. Section III reports a series of estimates of the variation in productivity for a large urban school district in Texas based on specifications that progressively take more and more steps to protect against bias introduced by matching of students and teachers in classrooms.¹ The potentially unique aspects of teacher labor markets are easiest to see by tracking the job transitions of teachers, and Section IV investigates productivity differences between stayers and those in distinct mobility streams: moving to a new school in the same district, switching districts, and exiting the public schools entirely. This section also examines differences by average student achievement or demographic composition in the origin and destination schools, both because policy is concerned with the impacts of turnover on specific demographic groups and because the school population has implications for working conditions and the job demands on teachers. Finally, we also investigate whether teacher performance immediately prior to the transition

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 (2007), Rockoff (2004), Rivkin, Hanushek, and Kain (2005), Boyd et al. (2006), and Kane, Rockoff, and Staiger (2008).

could reflect temporary shocks in productivity rather than more permanent differences in quality.

Section V summarizes the findings and discusses implications for public policy given the institutional structures within which public schools currently operate. Even though the employer is a public agency with considerable latitude for action on its own, it is generally concluded that new public policy interventions are required to improve educational outcomes. The correct policy, however, requires a clearer understanding of teacher labor markets in general and the dynamics associated with teacher turnover.

I. Employee Turnover and Firm Matching

Economists have done considerable work on job search and mobility, and on first blush the public school teacher turnover situation appears to be similar to that of the typical young U.S. worker who changes jobs frequently early in the career. Most matching models posit that this turnover constitutes a productive component of the job search process. Topel and Ward (1992) show that job changes account for a substantial share of early-career wage growth, which is consistent with the belief that job changes tend to lead to better matches that enhance worker productivity.² Indeed, the flexibility and fluidity of American labor markets is often cited as one of the strengths of the U.S. economy as distinct from, say, European economies. These arguments raise the question of whether teacher turnover actually imposes substantial costs on schools in general and those serving disadvantaged children in particular.

Turnover can exact a cost to firms in the form of additional hiring expenses or the loss of job-specific human capital. But, it is not clear that elementary and secondary school teaching fits the description of a job characterized by particularly high turnover costs. Direct hiring costs for teachers appear to be modest. It could be that coordination among teachers in terms of curriculum and other factors requires substantial investment in school-specific knowledge, though that also seems unlikely.

² Burdett (1978) and Jovanovic (1979a, 1979b) offer alternative conceptions of the job search process.

Two structural differences between teacher labor markets and other markets for professionals may raise the costs of teacher turnover. First, rigidities in district salary and job promotion scales typically prevent efforts to reward more productive teachers. These rigidities would appear to be magnified by limited salary differences among districts. In contrast, private firms tend to use much more flexible job categorization and salary determination processes. Second, because of the early determination of tenure in many districts – sometimes by the end of the first or second year of employment, teachers exert substantial control over whether job matches persist.³ Third, collectively-bargained contracts for teachers frequently govern within-district mobility and employment rights, granting mobility options based on seniority.

These features of public schools may generate substantial turnover costs for schools serving disadvantaged populations. If such schools have difficulty attracting teachers with prior experience, they may also have difficulty replacing the general human capital acquired through initial experience. Contract provisions in many large urban districts specify that experience moves a teacher up the queue for vacant jobs, and this often leads to a concentration of rookie teachers in the highest-poverty schools. A high share of teachers with little or no experience may hinder efforts to improve and align curriculum, mentor new teachers, and advocate for the school. In addition, more appealing outside job opportunities for highly-skilled teachers may lead to the disproportionate departure of more effective teachers.

Existing evidence shows that teachers who switch schools tend to move to schools with higher-achieving, higher-income, and fewer minority students than their previous schools, and that teachers changing districts tend to get slightly higher wages on average once the wages are adjusted for changes

³ In 2008, ten states granted teacher tenure in one or two years of employment; only seven states had probationary periods greater than three years (National Council on Teacher Quality (2008)).

in student demographic composition.⁴ Since demographic composition is likely to be related to working conditions, these findings suggest that non-pecuniary considerations tend to outweigh salary in the job decisions of teachers. Boyd, Lankford, Loeb, and Wyckoff (2005) also find that teacher labor markets tend to be highly localized, which further disadvantages high-poverty, lower-achieving schools located in urban centers and rural areas that tend to produce few college graduates. Moreover, evidence that teacher exit probabilities are higher for those with better alternative earning opportunities or more education (Dolton and van der Klaauw (1995, 1999)) supports the belief that leavers are drawn from higher up in the quality distribution.

The generalizability of these findings on mobility patterns is unclear, however. Podgursky, Monroe, and Watson (2004) and Scafidi, Sjoquist, and Stinebrickner (2006) find that the majority of teachers exiting 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. This pattern is consistent with the observation by Stinebrickner (2002) that much of the occupational movement of teachers is related to family circumstances – something that bears no clear relationship to job effectiveness.

The overall implications of observed teacher movements on the distribution of teacher quality are nonetheless difficult to infer. Although salary changes provide indirect information on productivity in fluid labor markets, they likely provide little information in more rigid teacher labor markets. Moreover, individual salaries provide little information about differences in teacher productivity given the rigid salary structures, whose typical determinants – teacher education, teacher experience, or certification status – appear to explain little of the observed differences in effectiveness (Hanushek and Rivkin (2006)).

⁴ Lankford, Loeb, and Wyckoff (2002), Hanushek, Kain, and Rivkin (2004), and Scafidi, Sjoquist, and Stinebrickner (2007) find that working conditions, as measured by student demographic composition, tends to be a more important determinant of school choice than salaries. Average salary changes following a move adjusted for working conditions are positive but small.

Given the absence of direct information on the effectiveness of school leavers, the conflicting elements of existing research make it far from clear how high turnover affects the quality of instruction in high-poverty, low-achieving schools. In fact teaching jobs appear to fit the description of “experience goods” (Jovanovic (1979b)) in that neither teachers nor schools have good information about teacher quality prior to arrival at the school. If knowledge acquired about productivity in the first years at a school disproportionately leads to separations of less effective teachers – either because they are counseled out or because they find low performance personally unsatisfying, then negative selection out of high-poverty schools may partially, fully, or even more than fully offset any costs of turnover. Since the match response of principals and districts is muted, the decision making of individual teachers is likely to dominate the transition process, and the resulting distribution of teacher quality is uncertain.

II. Measurement of Teacher Effectiveness

By necessity, understanding the character and implications of teacher job mobility begins with the development of measures of teacher quality, a requirement sidestepped when competitive forces are presumed to make salaries a sufficient statistic for quality.

The primary analytical tasks at hand are the separation of teacher contributions to student achievement from other student, family, school, and community factors and the mitigation of sampling error introduced by the imperfections of the achievement measures. In terms of the separation of teacher contributions to achievement from those of other factors, we focus on accounting for student heterogeneity and differences in school factors in a specification built on a cumulative model of learning. Within this context, we discuss the advantages and drawbacks of specific methods of accounting for student and school differences.

A number of studies highlight the fact that random test measurement error inflates estimates of the variance of teacher value-added in addition to introducing noise into the estimation of teacher

contributions to learning (e.g., Gordon, Kane, and Staiger (2006), McCaffrey, Lockwood, Louis, and Hamilton (2004), and McCaffrey, Sass, Lockwood, and Mihaly (2009)). Other work focuses on the effects of nonrandom aspects of measurement error including problems introduced by attempts to compare students in different parts of the achievement distribution. Hanushek, Kain, O'Brien, and Rivkin (2005) show that the transformation of the value-added measure such that it captures value-added relative to other students from the same portion of the achievement distribution has little impact on estimates of the variance in teacher quality; we therefore focus on the mitigation of random test error.

Cumulative Model of Learning

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), teacher quality (τ), and of a random error (e).

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

Teacher productivity may vary over time and is estimated in a semi-parametric manner by netting out the contributions of α_{iG} , X , P , and S from the achievement of students taught by a teacher in a given year.

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 including: principal assignment of better teachers to classrooms with less disruptive students (or more disruptive students, if seeking to equalize the quality of education across classes); the gravitation of better teachers toward higher-resourced schools; the movement of families with strong educational investment motivation to school districts with the best teachers and

with the most extensive school programs. All complicate the estimation of teacher value-added to achievement, as teacher quality becomes intertwined with characteristics of students or schools.

Accounting for student heterogeneity is a crucial component of the identification of teacher quality. We model α in a one dimensional framework in which differences in cognitive skills are assumed to evolve over time with experiences at home, in school, and in the community in a manner consistent with Equation (1). Equation (2) describes a reasonably general formulation of this cumulative process,

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

where γ_i is “innate ability”, which is assumed to affect learning in each period.⁵

A good teacher likely raises achievement in the current year and subsequent years by increasing the stock of knowledge, and a supportive parent does the same. In a very general way, we allow historical effects (and knowledge) to depreciate at a geometric rate $(1 - \theta)$ meaning that a teacher or peer’s effect on test scores diminishes with time such that a good 4th grade teacher has a larger effect on 4th grade score than on 5th grade score.⁶ If $\theta=1$, the effects of prior experiences persist fully into the future, while, if $\theta=0$, 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.⁷

A value-added regression of achievement in grade G on achievement in grade G-1 along with

⁵ 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.

⁶ 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 $\theta=1$) is sometimes accompanied by including student fixed effects. As Rivkin (2005) demonstrates, when $\theta < 1$, the 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.

⁷ 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. Jacob, Lefgren, and Sims (2008) consider alternative specifications of the dynamics with attention to varying persistence of teacher effects.

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.⁸ 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.

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

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

Estimation of Teacher Value-Added

Whether the model generates unbiased estimates of the τ_{jy} and the variance in teacher quality 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 the dynamic behavioral choices of families, teachers, and school authorities may induce correlations with teacher quality and thus may introduce bias even to value-added models. Moreover, when achievement is captured by standardized tests, test measurement error will inflate estimates of the teacher quality variance, even in the case where the teacher quality estimates are unbiased.

Equation (4), in which the estimated fixed effect for teacher j is interpreted as the sum of the persistent component of teacher quality and an error term, v_{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.

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

⁸ See Hanushek (1979, 1986) for a discussion of value-added models.

In general terms, the error term incorporates:

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

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

Consider first the case where the correlation between persistent teacher quality and the error components (including nonpersistent teacher effects) equals zero.⁹ 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 and the variance of the other terms. Therefore an estimate of the variance in teacher value-added must be adjusted for these error components.

Many researchers have used 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.¹⁰ Equation 6 is the simple correlation of estimated teacher fixed effects for adjacent years y and $y-1$.

$$(6) \quad r_{12} = \text{cov}(\hat{\tau}_y, \hat{\tau}_{y-1}) / \left(\text{var}(\hat{\tau}_y) \text{var}(\hat{\tau}_{y-1}) \right)^{1/2}$$

If the errors in equation 4 are orthogonal across years, the covariance between adjacent year fixed effects equals the variance of true value-added. In addition, if the variance in estimated value-

⁹ We generally focus on persistent differences in teacher quality, but for some purposes it is appropriate to consider both persistent and nonpersistent quality. The sum is, for example, a reflection of the quality faced by students in any particular year. Evidence suggests that teachers improve substantially in the first few years of their careers, and personal difficulties, the birth of children, and experimentation with new pedagogies all lead to variation over time in effectiveness.

¹⁰ 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.

added is constant over time, the expected correlation, $E(r_{12})$, reduces to:

$$(7) \quad E(r_{12}) = \frac{\text{var}(\tau)}{\text{var}(\tau) + \text{var}(\nu)}$$

Therefore, multiplication of the estimated sample variance of $\hat{\tau}$ by the year-to-year correlation produces a direct estimate of the variance in true teacher quality that accounts for test error and nonpersistent differences in classroom average student quality, either purposeful or random.

Now consider the case where the correlation between the error components and persistent teacher quality does not equal zero. 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 unobserved peer quality, principal effectiveness, or student skill, plus the relevant covariance terms. The methods described above would not purge the estimates of these latter variance and covariance terms.

Because available data typically have limited controls for differences in the quality of administration and other subtle aspects of schools such as the degree to which the curriculum and state standards align, 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 include school or even school-by-grade-by-year fixed effects in order to account for both observed and unobserved grade-by-year specific differences among schools and districts, recognizing though that this approach prohibits comparisons of teacher quality among schools.

Yet even value-added estimates of teacher effectiveness 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 (2010) document the existence of extensive within-school sorting on the

basis of student characteristics and prior performance in North Carolina. Whether such sorting introduces substantial bias in commonly used value-added models is a topic of considerable debate. In an influential article, Rothstein (2010) 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.¹¹ To provide an estimate of the potential importance of such selection, Kane and Staiger (2008) develop a specification test of the validity of non-experimental 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 introduce questions about the generalizability of the findings.

Given the difficulty of quantifying all relevant student and school variables related to the matching of students and teachers both within as well as between schools and the limitations of the specification tests employed, we adopt an alternative approach to circumvent bias introduced by purposeful sorting into classrooms. We begin by dividing school, grade, and year cells into “sorted” and “not-sorted” categories on the basis of two different dimensions of the classroom allocation process. Specifically, we examine whether there are either 1) significant differences in mean prior test score among classrooms based on an F- test¹²; 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. All classrooms in the school-grade-year cell 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 they are placed in the “not-sorted” category.

¹¹ Some specification issues do raise questions about the strength of this critique. 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.

¹² This test is similar in spirit to that used by Clotfelter, Ladd, and Vigdor (2006).

These tests are weak in the sense that the failure to reject the hypotheses of independence or no significant difference in mean prior test score at the five percent level does not provide strong evidence that a school randomly assigns students. Therefore, we augment these tests 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 below. We believe that the combined evidence supports the conclusion that the sample of schools for which common selection mechanisms are not obviously employed provides estimates of teacher effects that are not significantly contaminated by classroom composition.

III. Texas Schools Project Data

The stacked panel datasets constructed by the Texas Schools Project contain administrative records collected by the Texas Education Agency (TEA) from the 1989-1990 through the 2001-2002 school years.¹³ 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 such 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 (a mathematics teacher here) 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

¹³ The underlying data were developed at the Texas Schools Project at the University of Texas at Dallas. For more detail, see <http://www.utdallas.edu/research/tsp-erc/>.

(TAAS), which was administered each spring to eligible students enrolled in grades three through eight. These criterion referenced tests evaluate student mastery of grade-specific subject matter, and this paper presents results for mathematics. 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 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

A basic starting point for the analysis of teacher transitions and related policies is the magnitude of variation in teacher quality. If variation is limited, concerns about turnover and its impact on the distribution of teacher quality become less relevant.

Our estimates of the variation in teacher quality are based on alternative specifications of equation (3). 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, also to just the set of teachers in their own school, grade and year. To investigate the implications of any nonrandom sorting of students across classrooms, we compute separate estimates for “sorted” and “not-sorted” samples based on the statistical tests described above.

Basic Estimates of the Variance of 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 least 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 year-to-year correlations are similar to those found by others.¹⁴ 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 (7)). Because the student test scores are normalized to mean zero and

¹⁴ McCaffrey, Sass, Lockwood, and Mihaly (2009) analyze the intertemporal stability of teacher effects in detail and show that the estimates become even more consistent if additional years are added.

standard deviation one, the standard deviation in teacher quality has the natural interpretation of the bonus in annual student growth obtained for a teacher at the 84th percentile of teachers compared to an average teacher (or the detriment of having a teacher at the 16th percentile compared to the average). The district wide estimate of the overall standard deviation in teacher effectiveness, even when regression adjusted for student demographic characteristics, equals 0.17. Note also that this is fifty percent larger than the regression adjusted within-school-grade-year standard deviation estimate of 0.11. The estimate of 0.11 is in line with existing estimates in the literature of within-school differences in effectiveness.¹⁵

We tend to emphasize the more conservative within-school estimates, because they guard against unmeasured selection into schools by students, teachers, and administrators. But, 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.¹⁶ It is interesting that the measured selection factors found in the inclusion of student demographic characteristics leads to relatively small changes in the estimated variation of teacher effectiveness.

Sorting Among Classrooms

Despite the elimination of any between school variation in teacher quality and of changes over time in the quality of instruction for a given teacher and of influences of random shocks or error, the within-school-grade-year estimates may be biased. On the one hand, if principals assign more cooperative or more engaged students to better teachers, the differential could conflate true differences in quality with

¹⁵ Recent reports of estimated effects of a one standard deviation change in teacher quality on achievement include 0.11 (Rockoff (2004)), 0.13 (Aaronson, Barrow, and Sander (2007)), and 0.22 (Kane and Staiger (2008)). See the summary and interpretation of results in Hanushek and Rivkin (2010).

¹⁶ 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.

any student influences for which the empirical model does not account fully. On the other hand, compensatory assignment of better students to 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; 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 schools observations where we reject the null hypothesis are considered observations affected by purposeful sorting ("sorted",) and all other schools are classified as "not sorted".¹⁷

The top panel of Table 2 reports estimates 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 "not-sorted" allocation of students among teachers was not be rejected at the 5 percent level; 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. All estimates include the demographic measures listed in Table 1. 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 0.103 standard deviation gain in achievement. This is roughly 10 percent smaller than the full sample estimate of 0.113

¹⁷ Appendix Table a1 describes the joint distribution of the distribution of schools into "sorted" and "not sorted" categories on the basis of these two criteria and illustrates that there is substantial but incomplete agreement in the divisions. Approximately three-quarters of the common schools fall into the same category of sorted or not-sorted, but, because of the sample loss from the tests of placements across years, only one-quarter of the schools are found at the intersection of the two criteria.

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.¹⁸ The estimated within school-grade-year standard deviation of teacher quality for the “not sorted” sample equals 0.098 standard deviations of achievement, essentially identical to 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 “not sorted” sample 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. But, at the same time they do not rule out the possibility that unobserved factors confound the estimates. In order to provide additional evidence we conduct falsification tests similar to those suggested by Rothstein (2010). 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.¹⁹

Importantly, evidence of systematic differences among future teachers in prior year value-added is evidence of sorting on the basis of prior grade achievement but not evidence that the fixed effect estimates are biased since they explicitly hold prior year achievement constant. Nevertheless, if

¹⁸ 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.

¹⁹ Finding that future teachers are correlated with current performance does not establish bias, because a policy of sorting on entry achievement would produce exactly this. Assume that each year students are placed into classes based on their beginning of the year achievement. Then student in an effective grade G class would tend to be placed together in grade G+1, and it would look like the grade G+1 teacher is effective. The same would hold for an ineffective grade G class.

we divide the sample into “sorted” and “not sorted” schools on the basis of classroom allocation in the subsequent year, we would expect the subsequent year teacher to explain little variation in current year achievement for schools in the “not sorted” category.

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. 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 small negative correlation for adjacent year performance of these teachers. Not surprisingly, in the “sorted” sample the within-school variance estimates for future teachers range from 40 to 70 percent as large as the estimates using the actual teachers.²⁰

Interestingly, our preferred within-school estimates for the not-sorted sample are quite similar to those reported in Rivkin, Hanushek, and Kain (2005), a study that used grade level aggregate data to circumvent the problem of within-school sorting. This consistency across very different estimation methods despite the challenges of directly measuring teacher productivity provides strong support for the belief that there is significant variation in productivity.

²⁰ The analysis in Rothstein (2010) 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, results from not correcting for measurement error in the value-added estimates.

V. Teacher Transitions and Productivity

The finding of sizeable variation in teacher effectiveness is consistent with the belief that teacher transitions substantially alter the distribution of teacher quality from what it would be in the absence of such turnover. Given that teachers initiate the vast majority of job separations, teacher movement would arguably improve the well-being of most school leavers, even those who choose to leave the profession entirely. Yet the implications for schools are uncertain and depend upon the factors that induce departures and ultimately on the productivity of school leavers and their replacements.

As noted earlier, the constrained labor markets for teachers – with strict district salary schedules that vary only modestly across districts – may lead the dynamics of the teacher labor market to diverge sharply from those of less fettered markets. The fact that much of the movement involves changes across the “establishments” of a single firm (district) in a context in which teachers typically maintain significant control over assignment to open positions introduces another dimension through which the choice process can lead to substantial inequality in teacher effectiveness among both districts and schools.

In this section we describe differences in teacher effectiveness by transition status and other school and teacher-related factors. We rely on similar specifications to those estimated in the prior section except that transition dummies replace the teacher fixed effects. Again we estimate a range of specifications in order to address potential problems introduced by purposeful sorting. In addition, we compare estimates based on achievement in the year immediately preceding the move with estimates based on achievement in the previous year in order to distinguish between temporary and longer term differences in teacher effectiveness. Finally, the pattern of teacher transitions is permitted to differ by average student achievement in the school and by the proportion of students who are black – factors previously identified as related to the likelihood of teacher exits and the degree of harm caused by

teacher departures.

Teacher movement is substantial within the Lone Star district. As shown in Table 4, among new teachers (0-1 years of experience), the annual exit rate from Texas public schools is 12 percent. Another 11 percent change campuses, and 7 percent move to another Texas district. Even among more experienced teachers, however, some 18 percent still leave their current school each year. The key issue is not, however, the magnitude of the movement *per se* but instead how teacher quality is related to these transitions.²¹

Overall Differences by Transition

Table 5 reports regression estimates of mean differences in value-added to mathematics achievement by transition (compared to teachers who stay in the same school). 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, the average value-added of a teacher who exited the Texas public schools entirely was 0.06 standard deviations below the average for a teacher remaining in the same school. Moreover, those who switched campuses within the same district were also significantly less effective than stayers, though the deficit is smaller than that observed for those exiting the Texas public schools. In contrast, those switching to another Texas school

²¹ It might be plausible to think that the quality of a school is linked to the stability of its teaching force. The limited direct evidence on this, however, does not suggest this is the case after allowing for other factors correlated with the rate of teacher turnover; see Rivkin, Hanushek, and Kain (2005).

district were not significantly different on average from teachers who remained in the same school.²²

The mean differences, however, offer a limited view of the character of transitions, because there is substantial quality variation in the all streams. Figures 1 and 2 provide kernel density plots of teacher value-added that illustrate both the mean differences and dispersion of each of the streams in terms of teacher quality. Although non-persistent factors certainly inflate the dispersion for all streams, the magnitude of the observed variation clearly indicates substantial productivity differences among stayers, school changers and those who exit the public schools.

The transition estimates presented do not allow for any possible interaction with student sorting (and thus potential biases in the teacher quality estimates). Table 6 divides schools into sorted and not-sorted based on the two approaches described earlier. In general, the quality differences of teachers switching to a different campus and teachers exiting from Texas public schools are clearer and more significant in the not-sorted samples where any possible biases are lessened. The lower effectiveness of those exiting teaching is particularly noticeable in the not-sorted sample. These teachers have a value-added of 0.1-0.15 standard deviations of student achievement below those who stay. These estimates are statistically significant regardless of the comparison group or of the method of defining sorted schools. It is also interesting to compare these estimates to the average lower performance of first-year teachers, typically estimated at about 0.10 (Rivkin, Hanushek, and Kain (2005)). If these exiting teachers are replaced with rookie teachers, the expected net effect is close to zero.

The high transition rate of teachers early in their careers magnifies the importance of the pattern of movement for this group. We now focus on teachers with fewer than four years of experience. Table 7 reports mean value-added differences by transition type and single year of experience for teachers with one, two, and three years of experience. The much smaller numbers of

²² 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.

teachers at each experience level 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 experience. Specifically, 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, those with more experience who switch schools or districts are uniformly less effective than those who stay in the school, although many of the point estimates are not statistically different from zero and the hypothesis of no difference between the coefficients for teachers in their first year and others is not rejected at conventional significance levels for any of the three transition types.

In contrast, there appears to be relatively little variation across all experience categories, including new teachers, in the average effectiveness of teachers who exit the public schools: on average they are less effective than those staying. Note that we cannot ascertain which of these exits result from the decision to pursue other opportunities, which result from the realization that one is ineffective, or which result from administrator pressure. Nonetheless, the evidence does not support the belief that such turnover adversely affects the composition of teachers.

Differences by Student Demographic Composition

Despite the absence of evidence that more productive teachers are more likely to leave schools on average, it remains possible that such job adjustments adversely affect lower-performing schools serving high concentrations of disadvantaged students. As noted, research to date has shown that schools serving disadvantaged populations have higher turnover than other schools. By implication, these schools will have more inexperienced teachers because teachers who leave tend to be replaced by new teachers. That by itself would potentially signal lower quality. Beyond that, differential reactions to uncompensated working conditions could also affect the overall quality distribution of teachers in

schools serving disadvantaged populations.

To address whether 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 nonblack.²³ We then examine quality differences by transition and student characteristics for all teachers and separately by initial years of experience.

The estimates in Table 8 provide little support for the view that turnover exerts a higher cost on schools with lower-performing students or higher black enrollment. To the contrary, the estimates indicate that teachers who depart low-achieving or low proportion nonblack schools are less effective on average relative to stayers than those who depart higher-achieving or higher proportion nonblack schools. This pattern holds for all transitions, though the difference across school categories is statistically significant only for those moving to another school within the district. Notice that this pattern holds regardless of whether the comparison group of teachers is defined by the district or the specific school, grade, and year.

Table 9 reports the same coefficients for first-year teachers, and, if anything, the estimates provide even less evidence of disproportionate harm to low-achieving or high-proportion black schools. (Estimates for second, third, and fourth-year teachers show little pattern and are rarely significant). The value-added differential between campus switchers and stayers relative to other teachers in the same grade, year, and school equals 0.133 for higher achievement schools and 0.018 for lower achievement schools; the same pattern holds if schools are divided by proportion non-black.

In the case of exits there appears to be little average difference in productivity between those who exit and stayers in higher-achieving schools, while in lower-achieving schools those who exit are significantly less effective teachers on average (the coefficient is -0.12 in both specifications). A

²³ Information on eligibility for a subsidized lunch provides a crude measure of income, and our earlier work on teacher mobility suggested that black concentration and not overall minority concentration was most salient for teacher moves (Hanushek, Kain, and Rivkin (2004)). Therefore, we concentrate on proportion black, and classify schools by proportion nonblack in order to align this and the achievement variables.

qualitatively similar but far smaller and insignificant differential appears if schools are divided by proportion black.

In contrast to the other transitions, little consistent pattern emerges for district switchers. If effectiveness is measured relative to teachers in the same school, grade, and year, there appears to be stronger positive selection out of higher-achieving schools. However, the differential by school achievement reverses if productivity is measured relative to teachers throughout the district, and none of these differences are significant at any conventional level. Finally, if schools are ranked by proportion non-black, the differences are small and insignificant.

A related question considers the destination of within-district campus switchers: are more effective teachers more likely to gravitate toward higher-achieving schools? Classifying transitions by both origin and destination school characteristics (not shown) reveals little or no evidence of significant differences by destination school type regardless of the characteristics of the origin school and the method of identifying any sorting into classrooms.

The basic results about the relative effectiveness of movers paint a clear picture that schools with low achievement or by high black student concentrations – often the schools that evoke particular concern about the adverse consequences of turnover – are not systematically losing their best teachers. This holds for new teachers as well as experienced teachers despite some differences by experience in the productivity distributions of school leavers. Perhaps most important, teachers who leave the Texas public schools tend to be noticeably less effective than the average teacher who does not move, and this holds independent of being a new or experienced teacher. Nonetheless, the average quality of movers masks the very significant heterogeneity of quality among stayers and among all streams of job and occupational movers.

Sensitivity Analyses

The final issue revolves around questions of quality measurement and specifically, whether the observed flows are an artifact of measurement problems with teacher quality. We focus first on the possibility of selection bias due to nonrandom allocation to classrooms and then consider the question of whether the measured quality differences appear to be persistent or transitory. Table 10 reports differences in teacher effectiveness relative to others in the same school, grade, and year by transition, student achievement and racial composition, and student assignment mechanism; we report results when schools are divided into “sorted” and “not sorted” groups on the basis of the anova test of differences by mean classroom pretest scores.²⁴ These results indicate that student sorting on unobservables does not produce the observed transition patterns. For both the divisions by achievement and proportion nonblack, all coefficients are negative for the “not sorted” samples. As above, the results are stronger for the lower-achievement schools, where teachers exiting the Texas public schools, and those moving to another district, are significantly less productive than stayers.

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 persistent productivity differences, we generate estimates based on value-added in the year prior to the transition year and compare those with estimates based on the transition year (immediately preceding the move). For example, we describe the distribution of quality for transitions following the 1999 school year with value-added based on average student achievement

²⁴ Division by the Chi-squared test of independent transitions produces a qualitatively similar pattern by much noisier estimates, not surprising given the much smaller sample.

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.

Table 11 reports within-school estimates of the effectiveness of departing teachers that are based on achievement in both the transition year and the previous year (disaggregated by school demographic characteristics).²⁵ 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 tend to exceed in magnitude estimates based on the transition year though they follow a similar pattern in that the differentials are larger for teachers in low-achievement and low-proportion nonblack schools. In the case of lower achievement schools, the exit-stayer differential is -0.094 based on the penultimate year and -0.061 based on the transition year; in the lower proportion nonblack schools the corresponding differences are -0.084 and -0.078.

Second, for those who move to another school in the district, the estimated lower performance in comparison to stayers does not show up in specifications based on the value-added in the penultimate year. There appears to be little difference in average teacher effectiveness between within-district movers and stayers when assessed by earlier performance, suggesting the possibility that negative shocks or changes in effort account for the apparent lower average effectiveness observed 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

²⁵ 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.

precipitates a transition to another school. But it is also consistent 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 12 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 indicates that the change in effectiveness of switchers relative to others in the district or to others in the same school-grade-year is roughly zero. The effectiveness of stayers declines by -0.073 standard deviations relative to others in the district (not significantly different from zero) but remains stable relative to others in the same school. Moreover, the pattern across new teachers reveals little difference between switchers and stayers and no evidence that switchers improve more than stayers relative to either the district or teachers in the same school-grade-year. Taken together, the table provides little evidence in support of the view that a change of school leads to substantial improvement for many teachers and suggests that job switches do not typically lead to more productive matches.

VI. Conclusion

Schools with high concentrations of low-income, low-achieving, and minority populations often experience a great deal of teacher turnover, but the evidence in this paper does not support the view that teacher exits adversely affect the quality of education in such schools. Rather the pattern of estimates supports the conception of jobs as experience goods in which workers learn about their productivity. Less productive teachers appear to be more likely to leave lower-achieving, urban schools, either by transitioning to another school in the district or exiting the Texas public schools entirely.

Methodological Issues

The results depend fundamentally on the validity of the measurement of productivity. The constraints of teacher contracts imply that price information is uninformative about quality movements, leading us to estimate teacher effectiveness directly based on growth in achievement of each teacher's students. This estimation is complicated, however, by the myriad choices made by families, teachers, and administrators that determine the matching of teachers and students in classrooms. In addition, the tests provide noisy measures of achievement, and the measurement error can inflate estimates of the variance in teacher effectiveness.

The fundamental empirical specification for productivity measurement employs models with lagged achievement that implicitly incorporate past influences on learning, while also permitting estimation of the importance of depreciation of prior knowledge. We then use the time pattern of student placement across classrooms to separate a sample of schools where sorting appears significant versus those where it does not. Although measurement error from the underlying achievement tests will propagate into the value-added measures of teachers, we focus on the persistent impacts across years to capture the stable part of teacher quality differences. Importantly, the estimates of teacher quality from the "not-sorted" samples show significant variations in the quality of instruction, which are quite

consistent with estimates obtained from a very different approach not susceptible to bias introduced by sorting into classrooms on unobservables, or bias introduced by test measurement error. Falsification tests that compare actual teacher value-added to an estimate based on the students' future teachers reinforce the validity of the estimates.

The measures of teacher effectiveness are employed to trace the impacts of mobility on school quality, but the impact could be misstated for reasons related to the move. A school transition could itself be the result of a particularly difficult class, or the performance of the teacher could be atypically poor because of a lack of effort by the teacher after deciding to move. In either case, classroom outcomes in the year prior to a move may not be indicative of the true underlying quality of the teacher. Average quality estimates for teachers exiting Texas schools are, however, unaffected by basing the estimate quality from the year before the year of a move decision, an estimate that is less prone to reflecting such temporary circumstances.

Policy Implications

In much of policy debate surrounding teacher retention, there is a prevalent belief that low-achieving schools lose many of their best teachers to other schools and professions. In combination with a high level of turnover, which leaves a large numbers of inexperienced teachers in such schools, these teacher transitions would imply that the job matching process hinders the improvement of low-achieving schools.

Our estimates, however, provide little or no support for the belief that those who transition out of Lone Star schools in general, and out of the lower-achieving subset of schools in particular, are more effective on average than stayers. In fact, teachers in low-achievement schools who exit teaching altogether tend to be far less effective on average than stayers, and the average quality difference is quite similar to the return to the first year of experience. Therefore, the increase in average quality of

instruction resulting from the departure of less effective teachers roughly offsets the decrease resulting from disproportionate reliance on inexperienced teachers. Moreover, these conclusions do not appear to be driven by bias in teacher quality estimates caused by the pattern of student sorting among classrooms.

The entire market appears to be one where employer choices are highly constrained. Salaries are effectively not used to induce desirable moves; there are generally weak performance evaluations before tenure, which is often awarded very quickly in a career; employers have often bargained away their ability to assign teachers to specific schools; and mobility does not on average lead to more productive job matches. These elements of the teacher labor market suggest that teacher preferences for job locations are key to determining the outcomes of mobility. This set of observations and conclusions suggests that the mobility that occurs in teacher labor markets does not lead to efficiency improvements that parallel those in more general labor markets within the economy. The opposite is also true. Simply lessening the aggregate flow of teachers, as is frequently suggested, would not yield large improvements in the quality of teachers in schools predominantly serving disadvantaged populations.

This leads back to the wide variation in effectiveness of teachers both staying and leaving schools. The relatively small average differences in average teacher effectiveness by mobility stream compared to the wide differences in quality found among both stayers and leavers indicates that changing the actual effectiveness of teachers in schools serving disadvantaged populations will likely require more direct actions related to the effectiveness of teachers, such as providing quality professional development opportunities, particularly in such schools.

These findings do not resolve all interpretive or policy questions, because we cannot distinguish among alternative underlying mechanisms that are consistent with these patterns of teacher transition. For example, our estimates of the average effectiveness of teachers exiting from Texas public schools

could combine exits resulting from the identification and removal of poorly performing teachers by principals with the 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 existing active policies.

References

- Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1):95–135.
- 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 Corporation.
- Ballou, Dale, and Michael J. Podgursky. 2002. "Returns to Seniority Among Public School Teachers." *Journal of Human Resources* 37(4):892–912.
- Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2006. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy* 1(2):176–216.
- Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2005. "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools." *Journal of Policy Analysis and Management* 24(1):113–32.
- Burdett, Kenneth. 1978. "A Theory of Employee Job Search and Quit Rates." *The American Economic Review* 68(1):212–20.
- 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(4):778–820.
- Dolton, Peter J., and Wilbert van der Klaauw. 1995. "Leaving Teaching in the UK: A Duration Analysis." *The Economic Journal* 105(429):431–44.
- . 1999. "The Turnover of Teachers: A Competing Risks Explanation." *Review of Economics and Statistics* 81(3):543–52.
- Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger. 2006. "Identifying Effective Teachers Using Performance on the Job." Washington, D.C.: Hamilton Project, Brookings Institution (April).
- Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 60(2):280–88.
- . 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." *Journal of Human Resources* 14(3):351–88.
- . 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24(3):1141–77.
- . 1992. "The Trade-off Between Child Quantity and Quality." *Journal of Political Economy* 100(1):84–117.
- Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steve G. Rivkin. 2005. "The Market for Teacher Quality." Working Paper No. 11154. Cambridge, MA: National Bureau of Economic Research.

- Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Why Public Schools Lose Teachers." *Journal of Human Resources* 39(2):326–54.
- Hanushek, Eric A., and Steven G. Rivkin. 2006. "Teacher Quality." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch (1051–78). Amsterdam: North Holland.
- . 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100(2): 267–71.
- Jacob, Brian A., Lars Lefgren, and David P. Sims. 2008. "The Persistence of Teacher-Induced Learning Gains." Working Paper No 14065. Cambridge, MA: National Bureau of Economic Research.
- Jovanovic, Boyan. 1979a. "Firm-specific Capital and Turnover." *The Journal of Political Economy* 87(6):1246–60.
- . 1979b. "Job Matching and the Theory of Turnover." *The Journal of Political Economy* 87(5):972–90.
- Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6): 615–31.
- Kane, Thomas J., and Douglas O. Staiger. 2008. "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validation of Non-Experimental Estimates." Mimeo. Harvard University.
- Lankford, Hamilton, Susanna Loeb, and James H. Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24(1):37–62.
- McCaffrey, Daniel F., J.R. Lockwood, Thomas A. Louis, and Laura Hamilton. 2004. "Models for Value-Added Modeling of Teacher Effects." *Journal of Educational and Behavioral Statistics* 29(1):67–101.
- McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. "The Intertemporal Variability of Teacher Effect Estimates." *Education Finance and Policy* 4(4):572–606.
- Murnane, Richard J. 1975. *Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger Publishing Company.
- Murnane, Richard J., and Barbara Phillips. 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10(1):83–100.
- National Council on Teacher Quality. 2008. *State Teacher Policy Yearbook, 2008*. Washington, D.C.: National Council on Teacher Quality.
- Podgursky, Michael J., Ryan Monroe, and Donald Watson. 2004. "The Academic Quality of Public School Teachers: An Analysis of Entry and Exit Behavior." *Economics of Education Review* 23(5):507–18.
- 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(2):417–58.
- Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94(2):247–52.
- Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125(1): 175–214.
- 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(3):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 and Policy* 6(1): Article 8.
- . 2007. "Race, Poverty, and Teacher Mobility." *Economics of Education Review* 26(2): 145–59.
- Stinebrickner, Todd R. 2002. "An Analysis of Occupational Change and Departure from the Labor Force." *Journal of Human Resources* 37(1):192–216.
- Topel, Robert H., and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics* 107(2):439–79.

Tables and Figures

Table 1. Estimated Variance in Teacher Quality

	within district		within school and grade	
	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
stdn. 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. Number of observations is 254,046.

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

	within district		within school and grade	
	sorted	not-sorted	sorted	not-sorted
A. 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
std. dev. in teacher quality – σ_{τ}	0.198	0.203	0.141	0.103
Observations	161,990	79,047	161,990	79,047
B. 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
std. 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 and grade			
	Sorted		not-sorted		sorted		not-sorted	
	actual teacher	subsequent teacher	actual teacher	subsequent teacher	actual teacher	subsequent teacher	actual teacher	subsequent teacher
A. Sorting Tested by Pretest Mean Achievement								
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
std. 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
B. Sorting Tested by Classroom Assignment Patterns								
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	
std. 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 . Transition Shares for Teacher Movements

	All	New Teachers
Stay at same campus	0.81	0.70
Change campus	0.09	0.11
Change district	0.04	0.07
Exit Texas public schools	0.06	0.12
Observations (Teacher/Years)	7589	801

Table 5. Average Differences in Teacher Quality by Transition Compared to Teachers Who Do Not Move

Teacher mobility	within district	within school and grade
change campus	-0.048 (3.64)	-0.027 (2.49)
change district	0.019 (0.99)	-0.019 (1.26)
exit Texas public schools	-0.058 (3.79)	-0.061 (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. No move is omitted category; 254,046 observations; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.

Table 6. Average Differences in Teacher Quality by Transition Compared to Teachers Who Do Not Move by Sorting of Students across Classrooms

	within district		within school and grade	
	sorted	not-sorted	sorted	not-sorted
A. Sorting Tested by Pretest Mean Achievement				
change campus	-0.043 (2.60)	-0.06 (2.64)	-0.031 (2.25)	-0.015 (0.88)
change district	0.006 (0.26)	0.005 (0.14)	-0.015 (0.88)	-0.034 (1.11)
exit	-0.053 (3.05)	-0.101 (3.84)	-0.045 (3.16)	-0.098 (4.93)
Observations	160,939	78,498	160,939	78,498
B. Sorting Tested by Classroom Assignment Patterns				
change campus	-0.031 (1.17)	-0.044 (1.01)	-0.014 (0.65)	-0.085 (2.50)
change district	0.06 (1.43)	0.018 (0.27)	0.024 (0.82)	-0.133 (3.97)
exit	-0.052 (2.26)	-0.154 (3.44)	-0.033 (1.61)	-0.136 (3.83)
Observations	56,166	24,461	56,166	24,461

Table 7. Average Differences in Teacher Quality by Transition Type and Experience at Time of Move

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

Note: Within school comparisons; no move is omitted category; regressions from same sample and use same specifications as those in Table 4; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.

Table 8. Average Differences in Teacher Quality by Transition Type and Student Characteristics in School^a

	School demographics based on average achievement		School demographics based on proportion of students nonblack	
	within district comparisons	within school/grade comparisons	within district comparisons	within school/grade comparisons
a. change campus				
from low value school	-0.081 (4.51)	-0.062 (4.35)	-0.070 (3.89)	-0.055 (3.40)
from high value school	-0.013 (0.72)	0.006 (0.40)	-0.027 (1.48)	-0.003 (0.22)
Difference (F) ^b	7.24	10.63	3.01	5.38
b. change district				
from low value school	-0.003 (0.09)	-0.033 (1.53)	0.010 (0.17)	-0.028 (1.16)
from high value school	0.041 (1.66)	-0.008 (0.41)	0.028 (1.14)	-0.015 (0.80)
Difference (F) ^b	1.38	0.76	0.34	0.19
c. exit public schools				
from low value school	-0.081 (3.76)	-0.086 (5.25)	-0.066 (2.62)	-0.082 (4.16)
from high value school	-0.037 (1.80)	-0.043 (2.79)	-0.055 (3.05)	-0.051 (3.75)
Difference (F) ^b	2.18	3.68	0.12	1.69

Notes:

a. 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; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.

b. F-statistic for difference between high and low value school. Degrees of freedom for achievement are F(1,7588) and for Proportion Black are F(1,7588). Difference that are significant at the 5 percent level in bold.

Table 9. Average Differences in First Year Teacher Quality by Transition Type and Student Characteristics in School^a

	School demographics based on average achievement		School demographics based on proportion of students nonblack	
	within district comparisons	within school/grade comparisons	within district comparisons	within school/grade comparisons
a. change campus				
from low value school	-0.038 (0.78)	0.018 (0.51)	-0.033 (0.56)	0.008 (0.19)
from high value school	0.036 (0.42)	0.133 (2.28)	-0.001 (0.01)	0.106 (2.79)
Difference (F) ^b	0.59	2.95	0.17	2.94
b. change district				
from low value school	0.107 (2.06)	0.034 (0.67)	0.082 (1.39)	0.073 (1.23)
from high value school	0.064 (1.00)	0.114 (2.37)	0.098 (1.93)	0.051 (1.73)
Difference (F) ^b	0.29	1.35	0.05	0.11
c. exit public schools				
from low value school	-0.127 (2.99)	-0.120 (3.49)	-0.084 (1.58)	-0.061 (1.47)
from high value school	-0.010 (0.22)	0.033 (0.97)	-0.077 (2.20)	-0.050 (1.51)
Difference (F) ^b	4.00	10.87	0.01	0.05

Notes:

- a. Schools divided by being above or below district average for student achievement or percent black; no move is omitted category; 251,943 observations; same variables as in Table 4 specifications; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.
- b. F-statistic for difference between high and low value schools for first year teachers. Degrees of freedom for achievement are F(1,7588) and for Proportion Black are F(1,7588). Difference that are significant at the 5 percent level in bold.

Table 10. Differences in Teacher Quality by Transition and Distribution of Students Among Classrooms

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

Note: Within-school comparisons; no move is omitted category; classification of classroom allocation mechanism based on classroom differences in prior achievement; same variables as regression specifications used in Table 4; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.

Table 11. Differences in Teacher Quality by Transition, School Characteristics, and Timing of Quality Estimate

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

Note: Within-school comparisons; no move is omitted category; classification of classroom allocation mechanism based on classroom differences in prior achievement; same variables as regression specifications used in Table 4; 162,060 observations; absolute value of t-statistics based on robust standard errors clustered by teacher-year in parentheses.

Table 12. Average Year to Year Change in Value-Added over Adjacent Year for Stayers and Campus Switchers Within Lone Star District, by Transition and Experience

	within district comparisons	within school/grade 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

Figure 1

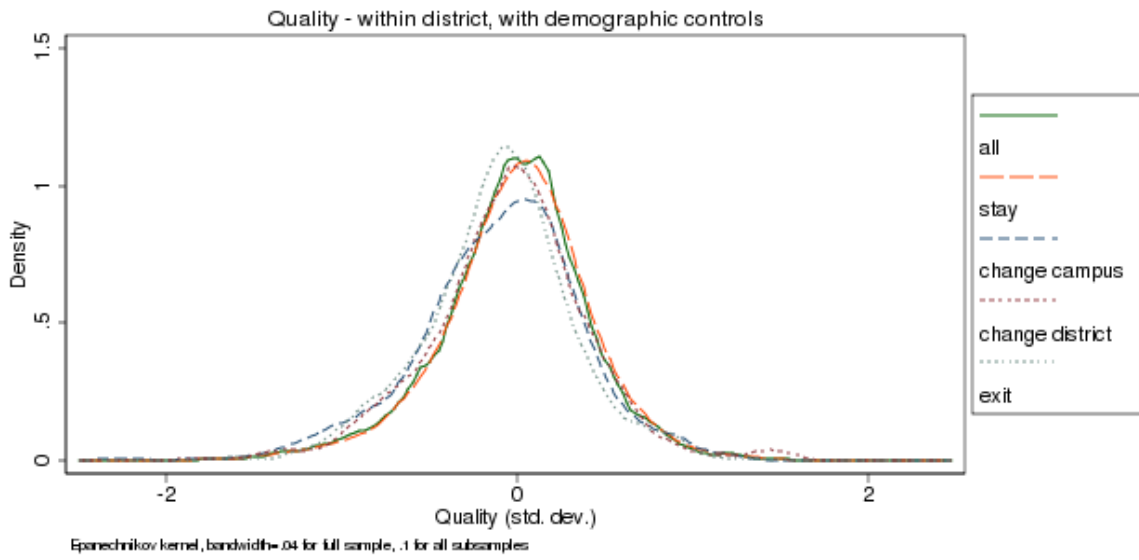
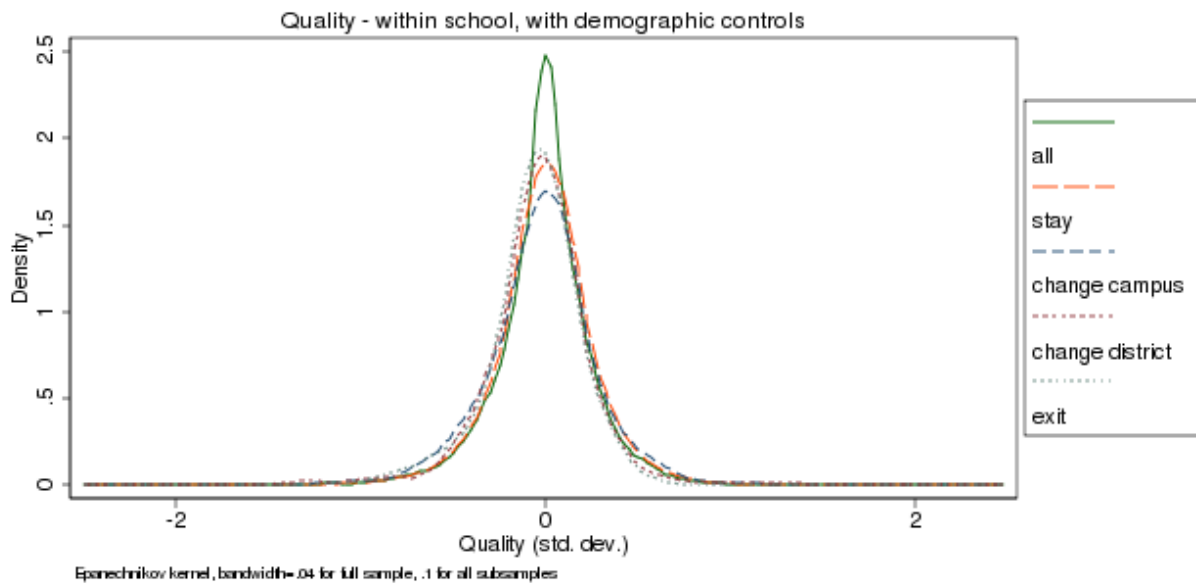


Figure 2



Appendix Table a1. Correspondence Between Assignment into Sorted and Not-sorted categories by Test Statistic

<i>F Test of Mean Pretest Score Difference</i>				
<i>Chi Squared Test of Transitions</i>	sorted	unsorted	missing	total
a. Proportions				
Sorted	0.71	0.29	0.00	0.13
Unsorted	0.27	0.73	0.00	0.10
missing	0.37	0.38	0.25	0.77
Total	0.40	0.40	0.19	1.0
b. Numbers (School/grade/year)				
Sorted	290	116	0	406
Unsorted	86	235	0	321
missing	879	899	600	2378
Total	1255	1250	600	3105

Note: The Chi Square test of the independence of transitions across grades requires an additional year of data with at least two teachers in the grade, leading to the classification of missing for a number of observations.

