



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness?

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters.

Citation	Chingos, Martin M., and Martin R. West. 2011. Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? <i>Economics of Education Review</i> 30, no. 3: 419-433.
Published Version	doi:10.1016/j.econedurev.2010.12.011
Accessed	December 16, 2017 12:30:50 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:4889479
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Program on Education Policy and Governance Working Papers Series

**Promotion and Reassignment in Public School Districts:
How Do Schools Respond to Differences in Teacher Effectiveness?**

Matthew M. Chingos
Program on Education Policy and Governance
Harvard University

and

Martin R. West
Harvard Graduate School of Education

PEPG 10-21

**Program on Education Policy and Governance
Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg/**

Promotion and Reassignment in Public School Districts: How Do Schools Respond to Differences in Teacher Effectiveness?

Matthew M. Chingos
Program on Education Policy and Governance
Harvard University
chingos@fas.harvard.edu

Martin R. West
Harvard Graduate School of Education
martin_west@gse.harvard.edu

December 2010

Abstract: We use a unique administrative database from the state of Florida to provide the first evidence that promotion and other job reassignments within school districts are systematically related to differences in teacher effectiveness in raising student achievement. We follow the career paths of a cohort of almost 25,000 classroom teachers during the 2001–02 school year for seven subsequent years. Our results confirm that effective teachers are more likely to become assistant principals or principals and less likely to be reassigned to a low-stakes teaching position. The tendency of highly effective teachers to continue teaching in high-stakes grades and subjects is strongest in schools receiving low ratings from the state’s school accountability system. Teachers entering the principal track experience a large increase in annual earnings, but the share of teachers promoted in this way is small enough that future compensation remains largely unrelated to effectiveness for teachers as a whole.

Acknowledgements – We are grateful to Tammy Duncan, Teresa Miller, and Jeff Sellers at the Florida Department of Education for supplying the data for this analysis and to Eric Hanushek, Paul Peterson, Ludger Woessmann, and anonymous reviewers for useful discussions and comments. Administrative and financial support was provided by the Program on Education Policy and Governance at Harvard University. This is a revised version of a paper originally prepared for the Conference on “Merit Pay: Will it Work? Is it Politically Viable?” sponsored by Harvard’s Program on Education Policy and Governance, Taubman Center on State and Local Government, Harvard’s Kennedy School, 2010.

1. Introduction

A growing body of research using administrative datasets to estimate the impact of individual teachers on student achievement has documented the existence of wide variation in the effectiveness of teachers employed by American school districts (see, e.g., Nye et al. 2004, Rockoff 2004, Rivkin et al. 2005). Moreover, evidence from low-stakes surveys (Jacob and Lefgren 2008, Harris and Sass 2009) and formal evaluation programs (Tyler et al. 2010, Rockoff and Speroni 2010) indicates that principals and other administrators can identify their most and least effective teachers. Yet there is little direct evidence that administrators' ability to recognize teacher effectiveness influences their personnel decisions.

Policy rigidities often constrain the ability of district and school administrators to respond to differences in teacher effectiveness. Despite recent experimentation with alternative pay structures, the dominant teacher compensation systems base salaries entirely on experience and education levels. Similarly, it is widely acknowledged that due process protections and the threat of litigation make it difficult and costly to dismiss tenured teachers. Although administrators often claim that they counsel ineffective teachers out of the profession, very few teachers are dismissed from their jobs due to poor performance (Weisberg 2009) and recent research suggests that ineffective early career teachers are only marginally more likely to leave the profession than their more effective counterparts (Goldhaber et al. 2009, Hanushek and Rivkin 2010, West and Chingos 2009).

Administrators may exercise substantially more control over the job assignments of teachers within school districts. For example, promoting teachers into positions as school leaders provides an opportunity to reward highly effective teachers with additional compensation and professional responsibility. Evidence on principal effects on student achievement is limited,

but recent studies suggest that it is an important determinant of school quality (Beteille et al. 2009, Branch et al. 2009). To the extent that the characteristics associated with classroom effectiveness predict performance as a school leader, promoting effective teachers should enhance leadership quality. It may also be useful for retaining effective teachers, who typically have stronger earnings potential outside of public education than their less effective peers (Chingos and West 2010).

The assignment of teachers to specific grades and subjects or to non-teaching positions also provides opportunities to act on information about teacher effectiveness. The high-stakes accountability systems now in place throughout American public education necessarily focus on performance in certain grades and subjects while ignoring others. In this environment, administrators have clear incentives to keep their most effective teachers in tested grades and subjects while reassigning their less effective teachers to positions in which they will less directly (or less immediately) influence a school's performance rating. Moreover, the rapid growth of non-teaching jobs in public school districts (Hanushek and Rivkin 1997) has created another means short of dismissal for limiting an ineffective teacher's influence on student achievement.

This study uses a unique administrative database from the state of Florida to provide the first evidence on the extent to which promotion and other job assignment decisions within school districts are systematically related to differences in teacher effectiveness. In particular, we follow the career paths of a cohort of almost 25,000 teachers in high-stakes classrooms during the 2001–02 school year for seven subsequent years. We limit our sample to math and reading teachers in grades four to eight for whom we are able to estimate value-added measures of classroom effectiveness. These measures enable us to examine whether more effective teachers

are more likely to be promoted into positions of school and instructional leadership and, conversely, whether less effective teachers are more likely to be reassigned to low-stakes grades and subjects or to other jobs within school districts. We also exploit earnings data from the state's unemployment insurance system to consider the degree to which promotions and reassignments are associated with changes in the total annual earnings (from all sources) of individual teachers.

Our results confirm that the career paths of individual teachers in Florida public school districts are systematically related to their classroom effectiveness. A teacher in the top effectiveness quartile is 32–36 percent more likely to become an assistant principal, principal, or instructional coach than a teacher in the bottom quartile. The same difference in effectiveness is associated with a 40 percent reduction in the (much greater) probability of having been reassigned to a low-stakes classroom and an eight percent reduction in the probability of no longer being the teacher of record in any classroom (despite remaining employed by a Florida school district). Consistent with previous research using administrative data to analyze teacher attrition, we find that more effective teachers are modestly less likely to leave the public schools altogether. Although teachers entering the principal track experience a substantial increase in annual earnings, the share of teachers promoted in this way is small enough that annual compensation across all teachers in our sample remains largely unrelated to effectiveness.

Our reliance on statewide administrative data, while necessary to observe a sufficient number of job transitions to examine their relationship with classroom effectiveness, means that we do not directly observe who within school districts is making decisions concerning job assignments. Although administrators have formal control over teachers' job assignments, promotions and reassignments within school districts are also likely influenced by teacher

preferences. Many of the patterns we observe, however, are consistent with the incentives Florida school administrators face to increase student achievement in tested grades and subjects as a result of the state's accountability system. We also show that top-quartile teachers are most likely to remain in high-stakes classrooms in schools facing strong pressure to improve their performance as measured by the state's accountability system. Finally, we note that our finding that teachers in the top effectiveness quartile will remain in a high-stakes classroom is driven entirely by teachers who remain in the same school, where building-level administrators have had the opportunity to observe the teacher's performance.

Our results therefore suggest that administrator decisions concerning teacher job placements reflect classroom effectiveness even in the absence of objective data on teachers' performance. At the same time, we also find that the promotion and reassignment of teachers is systematically related to the characteristics of the students to which they are assigned. Specifically, teachers who are consistently assigned students with higher prior achievement than other teachers in the same grade at their school are more likely to enter the principal track and to remain in a high-stakes classroom. This evidence may indicate that administrators do not adjust sufficiently when evaluating their teachers for the initial ability of the students to which they are assigned.¹ Providing administrators with objective reports on the effectiveness of individual teachers, an intervention evaluated in a sample of New York City schools by Rockoff et al (2010) and now being considered in several states attempting to incorporate information on student achievement into teacher evaluation systems, could therefore strengthen the relationships between effectiveness and job transitions documented in our study.

¹ An alternative explanation which we cannot rule out is that administrators offer preferential treatment to certain teachers for reasons unrelated to their effectiveness, so that teachers who receive better students in terms of prior achievement are also more likely to be promoted to the principal track and teachers who receive worse students are more likely to be involuntarily reassigned or counseled out of the profession.

2. Data

The primary data source for our study is the Florida Department of Education’s K–20 Education Data Warehouse (EDW). Our EDW extract contains observations of every student in Florida who took state tests from 1999 to 2009, with each student linked to his or her courses (and corresponding teachers) from 2001 to 2009.² These data include test scores from the Florida Comprehensive Assessment Test (FCAT), the state accountability system’s “high-stakes” test, and the Stanford Achievement Test (SAT), a nationally norm-referenced test that was administered to students alongside the FCAT until 2008 but is not used for accountability purposes. Beginning in 2001, students in grades three through 10 took both tests each year in math and reading. Annual gain scores can therefore be calculated for virtually all students in grades four through 10 beginning in 2002. The data also contain information on the demographic and educational characteristics of each student, including gender, race/ethnicity, limited English proficiency status, free or reduced-price lunch eligibility, special education status, days in attendance, and age.

Detailed information on individual teachers, including their demographic characteristics and experience, is also available from the EDW. We limit our analysis of job transitions to the cohort of teachers employed in 2002 in grades four to eight for whom value-added estimates can be calculated.³ We construct an employment file based on course enrollment data (that matches students and teachers) in order to track teachers’ classroom positions in subsequent years (2003

² Henceforth we refer to all school years by the spring date (when tests are administered).

³ We chose 2002 because it is the first year after Florida introduced annual testing in math and reading for grades 3–10. Although Florida also tests students annually in grades nine and 10, the wide variety of math and reading course offerings in these grades makes it difficult to construct reliable value-added measures.

through 2009). A separate file from the EDW enables us to identify the specific jobs of former classroom teachers who remained employed in the public schools. The teacher experience variable we construct reflects all years the teacher has spent in the profession, including both public and private schools in Florida and other states. Given our interest in the movement of teachers into different positions, we restrict our analysis to observations of teachers who were 55 years of age or younger in order to exclude likely retirees.⁴ The number of teachers in our analysis cohort therefore declines modestly over the study period.

Our analysis also draws on earnings records from the Florida Education and Training Placement Information Program (FETPIP). The FETPIP data consist of state Unemployment Compensation records that include the earnings of current public school teachers and former teachers working in Florida from the first quarter of 2001 through the third quarter of 2008. Only earnings received in Florida and reported to the state Unemployment Compensation system are included in these records. We inflate all quarterly earnings to 2008 dollars using the average of the monthly Consumer Price Indexes (CPI) for each quarter. We then calculate each current and former teacher's earnings in each school year from 2002 through 2008 by combining the earnings from the last quarter of the calendar year in which the school year began with the earnings from the first three quarters of the calendar year in which the school year ended. For example, earnings for the 2002 school year are calculated as earnings from October 2001 through September 2002.

We divide the 2002 cohort of teachers in reading or math classroom positions in grades four to eight, in each subsequent year, into six mutually exclusive categories based on the positions they held that year: 1) those on the "principal track" (assistant principals and principals,

⁴ Most Florida teachers become eligible for retirement with full pension benefits at age 62 or after 30 years of service. Assuming an entry age of 25, the earliest possible retirement age with full benefits is therefore 55.

including “interim” or “intern” placements); 2) those designated as reading or math coaches (but not on the principal track); 3) those remaining in high-stakes classroom positions (reading or math in grades three to 10 or science in grades five, eight, or 11) but who were not principals or coaches;⁵ 4) those teaching in low-stakes classroom positions (but not in any of the prior categories); 5) those working in the Florida public school districts but not as teachers of record for students in any subject (and who are not in any of the prior categories); and 6) those no longer employed by a Florida public school district. School district employees who are no longer teachers of record consist of individuals designated as teachers but not associated with individual students in the course files (e.g., teachers assigned to particular student populations such as special education or bilingual students) and individuals in non-teaching positions (e.g. librarians, clerical workers, and even bus drivers). The former make up 86 percent of the total, with the latter comprising the remaining 14 percent.

3. Analytic Strategy

Our primary aim is to examine how teachers’ career paths vary with their effectiveness in raising student achievement. We first explain our approach to measuring teacher effectiveness and then describe the models used to relate these measures to transitions from high-stakes classroom positions to other jobs.

Measuring Teacher Effectiveness

To measure teacher effectiveness we use student test score data from 2002 through 2009 to construct value-added measures for teachers of math and reading courses in grades four to

⁵ A teacher is only considered to be in a high-stakes classroom position if at least half of the teacher’s student-weighted classroom time was spent in high-stakes subject-grade combinations. Science classrooms in grades five, eight, and 11 are defined as high-stakes because the results of science tests administered in these grades are included in the calculation of school grades under Florida’s A+ Accountability Plan.

eight. Value-added measures attempt to isolate the causal effect that teachers have on their students' test scores. We use the course files to match fourth- and fifth-grade students (most of whom are enrolled in self-contained classrooms) to their primary teacher of record and sixth-through eighth- grade students to their math and reading/English Language Arts teachers.⁶ A small number of fourth- and fifth-grade students who were in classrooms of fewer than five or more than forty students are dropped from the analysis.

To generate value-added estimates for each teacher, we regress students' math and reading test scores separately on their prior-year test scores in both subjects (including squared and cubed terms); vectors of student, classroom, and school characteristics; dummy variables for teacher experience; and grade-by-year fixed effects.⁷ Student-level control variables in addition to prior achievement include the number of days absent the previous year and dummy variables for race, gender, special education status, limited English proficiency status, free or reduced-price lunch eligibility, whether the student was repeating the grade, and whether the student made a structural or non-structural move to a new school that year.⁸ Classroom- and school-level

⁶ For fourth- and fifth-grade students, the course files do not always clearly identify the student's regular classroom teacher. In order to match the maximum number of students to their teachers, we examined students' general (e.g., self-contained classroom), math, and reading teachers and matched them to the one or two teachers with whom they spent at least 40 percent of their academic (general, math, and reading) time. We then dropped students who were matched to two teachers and students who were not matched to any teachers (32 percent of all students in these grades). A large and increasing number of fourth- and fifth-grade students in Florida appear to have more than one regular classroom teacher, perhaps reflecting an increase in team teaching. We match sixth- through eighth-grade students to their primary reading and math teachers in a similar fashion. For each subject, we linked each student to the teacher(s) with whom they spent at least 40 percent of their time in that subject and dropped students who were matched to two teachers in a given subject as well as those who were not matched to any teachers in that subject (16 percent of students in math and 38 percent of students in reading).

⁷ We control for teacher experience using a dummy variable for each of the first twenty years of experience, so the omitted category includes all teachers with more than twenty years of experience.

⁸ Students are identified as having made a non-structural move from the previous year if they are in a different school than in the previous year and are in the lowest grade offered in their new school, and as having made a structural move if they are in a grade higher than the minimum grade of the new school.

control variables include all of the student-level characteristics (except prior-year test scores) aggregated to the relevant level and class size.

The model, then, is

$$(1) \quad A_{it} = \omega A_{i,t-1} + \beta X_{it} + \gamma C_{it} + \phi S_{it} + \delta W_{it} + \pi_{it} + \epsilon_{it},$$

where A_{it} is the test score of student i in year t (standardized by grade and year to have a mean of zero and standard deviation of one); $A_{i,t-1}$ includes the student's prior-year test scores in both subjects (and their squared and cubed terms); X , C , and S are student-, classroom-, and school-level characteristics; W is a vector of teacher experience dummy variables; π is a vector of grade-by-year fixed effects, and ϵ is a standard zero-mean error term. We estimate this equation separately by test (FCAT and SAT), subject (reading and math), and grade-level (fourth and fifth and sixth through eighth), and average the residuals by teacher and year to construct a value-added measure for each teacher in each year. Finally, we use the Bayesian (shrinkage) estimator described by Kane et al. (2007) to isolate the persistent component of each teacher's value added (using data from all available years).⁹ This persistent component forms the effectiveness measure used throughout our analysis. Consistent with previous research, we find that teacher effectiveness varies more for math than for reading and for elementary than for middle school teachers.¹⁰

⁹ One key difference is that for each year and teacher Kane et al. (2007) compute average residuals by class, whereas we compute average residuals by teacher (which is identical to class for fourth- and fifth-grade teachers because each teacher only has one class, but sixth- through eighth-grade teachers often teach multiple classes). We take this approach because the EDW course records make it difficult to definitively assign middle school students to a specific math classroom even though we can confidently match them to a specific math teacher.

¹⁰ For all teachers for whom we are able to estimate effectiveness measures using FCAT scores, the standard deviations of these measures are (in standard deviations of student test scores): 0.11 and 0.05 for fourth- and fifth-grade math and reading, respectively, and 0.07 and 0.03 for sixth- through eighth-grade math and reading, respectively. The standard deviations of measures calculated using SAT scores are 0.08 and 0.05 for fourth- and fifth-grade math and reading, and 0.06 and 0.03 for sixth- through eighth- grade math and reading.

Relating Effectiveness and Job Transitions

We use job assignment data at the teacher-year level to estimate the relationship between classroom effectiveness and movement out of high-stakes classrooms into other positions from 2003 through 2009. Our value-added measures of teacher effectiveness are standardized separately by grade level (grades four and five and grades six to eight), test (FCAT and SAT), and subject area (math and reading) to have a mean of zero and a standard deviation of 1. We average the value-added measure (based on data from all available years, 2002–2009) in math and reading for any teachers that taught both subjects in 2002 (almost all elementary teachers and 16 percent of middle school teachers). We also average the standardized value-added measures from both the high-stakes FCAT and low-stakes SAT tests to reduce measurement error.¹¹ Finally, for most of our analysis we split our final value-added measure into quartiles in order to capture non-linearities in the relationship between effectiveness and job transitions.

Combining elementary and middle school teachers, we estimate the following multinomial logit equation:

$$(2) \quad \Pr(y_{it} = j) = \frac{\exp(X_i\beta_j)}{1 + \sum_{j=1}^J \exp(X_i\beta_j)},$$

where y_i is the outcome j (one of the six job categories described above) of teacher i in year t and X_i is a vector of explanatory variables (with coefficient vectors β_j) that includes a set of dummy variables identifying the teacher's value-added quartile and a set of dummy variables corresponding to the teacher's grade level and subject (middle school math, middle school English/reading, or both subjects in middle school, with elementary school teachers making up

¹¹ Because Florida stopped administering the SAT in 2009, our SAT-based value-added estimates are based on 2002 to 2008 only. The main findings reported below concerning the relationship between this summary effectiveness measures and job transitions are qualitatively similar to those obtained using effectiveness measures based only on the FCAT or only on the SAT.

the omitted category). Standard errors are adjusted for clustering at the teacher level to account for the fact that most teachers are observed in more than one year. We also estimate specifications of equation (2) that include the following additional teacher-level variables: educational attainment, experience, race/ethnicity, and gender.¹² These models indicate the extent to which any relationship between effectiveness and job placements could be explained by correlations with these variables, as well as their independent predictive power.

4. Results

We begin our analysis by describing the job transitions made by the 2002 cohort of fourth- to eight-grade teachers in high-stakes classrooms and the characteristics of those teachers. Table 1 shows that, of the 24,475 teachers in the 2002 cohort, 84 percent were still primarily employed in high-stakes classrooms the following year. The remaining 16 percent were mainly in low-stakes classrooms (7 percent), not teachers of record (5 percent), or not employed in the public schools (3 percent). A very small number (0.3 percent) were assistant principals or principals. None were yet designated as reading or math coaches, as these positions were first created in Florida public schools in 2005. Six years later, only 52 percent were still teaching high-stakes grades and subjects, and sizable proportions were in low-stakes classrooms (15 percent), not teachers of record (9 percent), or not in the public schools (19 percent). A small but non-trivial number were on the principal track (4 percent) or reading or math coaches (2 percent).

¹² Experience in each year is calculated as the teacher's reported experience in 2002 plus the additional years spent in the public schools after 2002 (e.g., if a teacher had 10 years of experience in 2002 and worked in the public schools in 2003 but not in 2004, she would be coded as having 11 years of experience in both 2003 and 2004).

Teachers who transitioned into these positions differ noticeably in their observable characteristics, particularly their effectiveness in the classroom. Table 2 shows that teachers who were “promoted” in the sense that they were given positions of instructional leadership (as principals, assistant principals, or reading/math coaches) were more effective in the classroom—0.13 standard deviations better than the average teacher. Those who were “demoted” in the sense that they were no longer in high-stakes classrooms but were not promoted were consistently less effective classroom teachers: those who moved into low-stakes classrooms were 0.13 standard deviations worse than average, those who were no longer teachers of record were 0.01 standard deviations below average, and those who left the public schools entirely were 0.03 standard deviations below the mean. The majority of teachers who remained in high-stakes classrooms were 0.05 standard deviations above average. Figure 1 presents density plots illustrating the differences in the effectiveness distributions of the groups of former teachers with the largest differences in average effectiveness: those on the principal track and reading or math coaches (combined into a single group) and those in low-stakes classrooms. Although the distribution for the former group is clearly to the right of the distribution for the latter, the distributions are almost fully overlapping and there is considerable variation in effectiveness within each group.

Some of this additional variation in effectiveness among teachers reassigned to low-stakes classrooms appears to be related to the nature of their new assignment. More specifically, 40 percent of the observations in our dataset of teachers in low-stakes classrooms are of reading and math (primarily self-contained) classrooms in grades prekindergarten to two. As shown in Figure 2, these teachers were 0.20 standard deviations less effective than the average teacher when they were in high-stakes classrooms, as compared to the average deficit of 0.13 standard

deviations for the broader group. In contrast, teachers moving to math and reading classrooms in grades 11-12 were 0.01 standard deviations more effective than the average teacher.¹³

Table 2 also indicates differences across job categories in other teacher characteristics. Teachers who became principals or assistant principals were much more likely to be men, members of minority groups, and holders of advanced degrees than other teachers. Teachers who became math or reading coaches were predominantly female. But comparing the more numerous groups of teachers—those in high- or low-stakes classrooms, those not teachers of record, and those no longer working in the public schools—differences in characteristics other than value-added are more modest. Former teachers on the principal track also earned \$14,748 (32 percent) more annually than teachers remaining in high-stakes classrooms, while those not teachers of record and those no longer working in the public schools earned substantially less. We consider the implications of these differences in average annual earnings across positions below.

Teacher Effectiveness and Job Transitions

The descriptive statistics presented in Table 2 indicate that teachers who are promoted tend to be more effective and those who are reassigned to low-stakes positions tend to be less effective, but this mode of analysis does not directly address the question of whether more effective teachers are more likely to move to certain positions and less effective teachers are more likely to move to other positions. In particular, the magnitudes of the relationships between classroom effectiveness and the likelihood that a teacher will switch to various different positions are unclear. Multinomial logit analysis allows us to quantify these relationships while

¹³ Teachers are assigned to categories for this analysis based on the type of class they spent the majority of their student-weighted time teaching.

controlling for other teacher characteristics and to estimate the power of such characteristics to predict movement from high-stakes classrooms to other positions.

Figure 3 shows the (unadjusted) relationship between a teacher's value-added quartile and the probability that they will switch to various jobs outside of high-stakes classrooms. For many of the job transitions shown, the relationship between value added and the probability of switching to a position is non-monotonic. For example, second quartile teachers are more likely than bottom quartile teachers to switch to low-stakes classrooms (13.9 percent vs. 12.4 percent), but third and top quartile teachers are the least likely to transition to such jobs (11.5 percent and 7.8 percent, respectively). We therefore focus our discussion on models that compare the likelihood of each job transition across effectiveness quartiles.¹⁴

The first set of multinomial logit results are presented in Table 3.¹⁵ The coefficients are presented as marginal effects that indicate the difference in the probability that a teacher in a given value-added quartile will switch to (or remain in) a given type of job relative to a teacher in the bottom value-added quartile. For example, the coefficient in the upper-left corner of Table 3 indicates a second quartile teacher is 0.5 percentage points more likely than a bottom quartile teacher to be a principal or assistant principal in a given year (a difference that is statistically significant at the 5 percent level). The overall share of teachers in our initial cohort serving as principals or assistant principals in a given year is only two percent, so a 0.5 percentage point increase translates into a 25 percent increase in the probability of having entered the principal

¹⁴ Results using a continuous (standardized) measure of teacher effectiveness are reported in Appendix Table 1 and confirm the descriptive patterns reported above. For example, more effective teachers are consistently more likely to be promoted into positions as principals and instructional coaches and less likely to move into low-stakes or non-teaching positions. These results, however, mask the non-linearity and non-monotonicity in the relationship between effectiveness and job transitions evident in Table 3.

¹⁵ The analyses presented in Table 3 combine elementary and middle school teachers, but we obtain qualitatively similar results for the two sets of grade levels when we examine them separately. Results disaggregated by grade level are available from the authors upon request.

track. The point estimates suggest that third and top quartile teachers are even more likely to enter the principal track, although the differences in the coefficients estimated for each quartile are not statistically significant. The second column of Table 3 indicates that more effective teachers are also more likely to become reading or math coaches, with magnitudes similar to those estimated for the principal track (with the exception of the third quartile).

The most effective teachers are not only more likely to be promoted; they are less likely to be reassigned to low-stakes positions or to leave the public schools entirely. The fourth through sixth columns of Table 3 indicate that the differences between top and bottom quartile teachers in the probability of being in a low-stakes classroom, not a teacher of record, or not in the public schools are 40 percent, eight percent, and 11 percent, respectively.¹⁶ However, the least effective teachers are not the most likely to be in these types of positions—second and third quartile teachers generally are.¹⁷

A supplementary analysis (not shown) reveals that the relationships between value-added quartile and the probability of staying in a high-stakes classroom are driven entirely by teachers who remained in a high-stakes classroom in their original school, not by those who moved to a high-stakes classroom position in a different school. In other words, a teacher's value-added quartile is related to the probability that she remains in a high-stakes classroom in her original school, but not to the probability that she moves to a high-stakes classroom (as opposed to a different job assignment) in a different school. This result suggests that the patterns observed in

¹⁶ A supplementary analysis (not shown) shows that the results for staying in a high-stakes classroom (third column of Table 3) are driven entirely by teachers who remained in a high-stakes classroom in their original school, not by those who moved to a high-stakes classroom position in a different school. In other words, value added quartile is related to the probability that a teacher remains in a high-stakes classroom in her original school, but not to the probability that she moves to a high-stakes classroom in a different school.

¹⁷ The analyses presented in Table 3 combine elementary and middle school teachers, but we obtain qualitatively similar results for the two sets of grade levels when we examine them separately. Results disaggregated by grade level are available from the authors upon request.

Table 3 administrators are in fact driven by the ability of building-level administrators to base job assignment decisions on their observations of teachers' effectiveness, as administrators hiring teachers from other schools do not have this opportunity.

Given our primary interest in whether job transitions vary with effectiveness, the results without control variables (beyond the grade-subject dummies) represent our preferred estimates. Nonetheless, the bottom half of Table 3 demonstrates that the direction of the estimated relationship between value added and movement into other positions is not sensitive to controlling for other teacher characteristics, including educational attainment, experience, race/ethnicity, and gender. While the inclusion of controls generally reduces the magnitudes on the value-added quartile coefficients, they remain statistically distinguishable from the bottom quartile in all cases.

Additional Predictors of Job Transitions

Although not our focus in this paper, the coefficients on the control variables (reported in Appendix Table 2) also provide what is to our knowledge the first evidence from a statewide database on the correlates of entry into positions of school leadership. Black, Hispanic, and male teachers are substantially more likely to become principals or assistant principals than are white and female teachers. Teachers without advanced degrees are highly unlikely to enter the principal track, which typically requires additional certification obtained in the context of a master's degree program. Teachers with 6–15 years of experience are 44–49 percent more likely to become principals or assistant principals than the least experienced teachers, but after 15 years the relationship between experience and the odds of entering the principal track turns negative.

Experience is also a strong predictor of transitions into other positions. More experienced teachers are far more likely than early-career teachers to become reading or math

coaches. Teachers with 6–15 years of experience are more likely to move to a low-stakes classroom than those with 5 or fewer years of experience. As expected, more experienced teachers are also generally less likely to become non-teachers of record or to leave the public schools entirely.

Student Sorting, Job Transitions, and Additional Robustness Checks

Our measure of teacher effectiveness is based on observational data and thus could be biased by the non-random matching of teachers and students on their unobservable characteristics in a way that is related to teachers' future movement across positions (Rothstein 2010). For example, one might expect teachers that receive favors from administrators in the form of unobservably better students might also be likely to receive favors in the form of promotions. On the other hand, teachers who are successful with more challenging classrooms might also be rewarded with promotions.

To provide evidence on these issues, we estimate a measure of the extent to which teachers systematically receive students with higher prior-year test scores (as compared to other teachers in the same school, grade, and year) and add this sorting measure to the multinomial logit models estimated in Table 3.¹⁸ The first column of Table 4 indicates that more effective teachers are more likely to become principals or assistant principals (although this pattern is driven primarily by the least effective teachers being the least likely to enter the principal track), but that those that systematically receive students with higher prior-year test scores are also more likely to experience such promotions. In other words, while our original conclusion that more effective teachers are more likely to be promoted is robust to controlling for the sorting measure,

¹⁸ Specifically, we regress students' prior-year test scores on school-grade-year fixed effects, aggregate the residuals to the teacher-year level, then apply the same Bayesian shrinkage algorithm applied to the value-added estimates. Finally, we standardize the sorting measure to have a mean of zero and standard deviation of one, averaging across subjects and tests.

we also find suggestive evidence that teachers that receive preference in promotion to administrative positions also receive preference in classroom assignments. This relationship could result from administrators promoting teachers based on their perceived effectiveness if administrators fail to adjust sufficiently for differences in students' prior achievement.

Alternatively, it could be the case that administrators offer preferential treatment to certain teachers for idiosyncratic reasons unrelated to their effectiveness, so that teachers who receive better students in terms of prior achievement are also more likely to be promoted in this way. Regardless of the explanation, the same pattern is not evident in the selection of reading and math coaches; the most effective teachers are more likely to move to such positions, but those who are assigned better students are not.

Our finding that the most effective teachers are the least likely to be reassigned to low-stakes positions is also robust to controlling for the sorting measure, although once again the sorting measure is itself a strong predictor of job transitions. In the case of movement to low-stakes classrooms and positions not as teachers of record, the relationships are actually stronger for the sorting measure than for the value-added measure. For example, moving from the bottom to top value-added quartile is associated with a 31 percent decrease in the likelihood of moving to a low-stakes classroom, whereas the corresponding association for the sorting measure is 59 percent. Again, this pattern suggests that administrators may reassign teachers they perceive as less effective while failing to adjust sufficiently for the prior achievement of students assigned to them.

We also probe the robustness of our main results to two alternative value-added measures. First, we use value added estimates calculated based on data from 2002, 2003, and 2004 for all teachers who were employed in high-stakes classrooms for all three of those years

(recall that our preferred value-added measure uses data from all available years). As a result, the amount of measurement error in the value-added measures should not be correlated with transitions to other jobs because the same amount of data is used for all teachers. This restriction forces us to examine transitions to other jobs only from 2005 to 2009, rather than beginning in 2003. Second, we construct a value-added measure that conditions on school fixed effects rather than observable school characteristics. This measure controls for any time-invariant school characteristics (such as students' socioeconomic status), including those we cannot observe. As shown in Appendix Table 3, both alternative value-added measures produce similar estimates of the relationship between effectiveness and job transitions to those reported in Table 2.

School Accountability Ratings and Job Transitions

Although the nature of our data makes it difficult to identify the factors responsible for the patterns of job transitions reported above, the assignment of ineffective teachers to low-stakes and non-teaching positions is consistent with the incentives facing administrators under Florida's school accountability system. Because schools are rewarded and sanctioned based on their performance in raising student achievement, administrators have incentives to move their less effective teachers to classrooms where they will have a less direct (or immediate) impact on the outcomes used for accountability purposes. Florida's accountability policy was in place throughout the entire period covered by our data, so we are unable to examine whether the relationships documented above changed as this policy was implemented. However, we can provide suggestive evidence that accountability pressures influence promotion and reassignment decisions by testing whether the movement of less effective teachers out of high-stakes classrooms is most pronounced in schools that are under the greatest pressure from Florida's accountability policy.

Table 5 shows results separately by the “A”-“F” grade awarded to the school by Florida’s accountability system in 2002. For expositional convenience, we report only the marginal effects for the outcome of remaining in a high-stakes classroom.¹⁹ These results confirm that schools with lower grades (and thus greater pressure to improve) are the most likely to retain their best teachers in high-stakes classrooms. The differences in the top value-added quartile coefficients estimated for the “A” and “C” and the “A” and “D” schools are both statistically significant at the 1 percent level. However, the relationship between effectiveness and retention in a high-stakes classroom remains non-monotonic for all groups of schools, with the second quartile teachers consistently less likely to remain in a high-stakes classroom than the bottom quartile teachers.

Job Transitions and Compensation

Table 6 considers the implications of promotions and reassignments for the relationship between effectiveness and compensation across our entire cohort of teachers in high-stakes classrooms. The first column confirms that principals earn about 25 percent more than teachers on average, but reading and math coaches earn about the same and teachers in low-stakes classrooms earn slightly (5 percent) less. Non-teachers of record and especially those who leave the public schools earn substantially less, although these differences likely reflect movement into part-time work (Chingos and West 2010). Including our standard set of teacher-level controls does not alter this pattern of results (column 2), but adding teacher fixed effects—which allows us to see how much the average teacher’s earnings changed upon moving from a high-stakes classroom to another position—does induce some differences (column 3). The coefficient for the principal track falls by about half to 0.11, suggesting that future principals earned more, on

¹⁹ Note that the sum of the marginal effects for low-stakes classroom, not teacher of record, and not in public schools will have a similar magnitude to the high-stakes classroom marginal effect, given the small shares of teachers that are promoted.

average, when they were classroom teachers than their colleagues (perhaps by taking on additional responsibilities in their school or working outside of public education). Reading and math coaches saw a decline in their wages of about 14 percent on average, suggesting that many of those entering these roles were transitioning into part-time work.

Column 4 shows that, despite these differences in average annual earnings across positions, classroom effectiveness is only weakly correlated with total compensation for the cohort as a whole. More specifically, a one standard deviation increase in value added is associated with a 1.5 percent increase in wages. Column 5 shows that the correlation between value added and position transitions documented above explains about 40 percent of this already negligible relationship. In other words, neither the relationship between value added and job assignments nor the share of teachers experiencing job transitions is substantial enough to tie compensation closely to classroom effectiveness for teachers as a whole.

5. Discussion

The results presented above represent the first systematic evidence on the relationship between teacher effectiveness and job transitions within public school districts. Promotion and reassignment among Florida classroom teachers appear to be consistent with the incentives facing school administrators as a result of the state's high-stakes accountability system: More effective teachers are more likely to be promoted to positions of school and instructional leadership, while less effective teachers are more likely to be assigned to low-stakes positions. These patterns suggest that administrators' knowledge of teachers' ability to raise student test scores—knowledge which has been documented in low-stakes surveys—does influence their personnel decisions even in the absence of formal value-added data.

At the same time, it is important to acknowledge that the job transitions we observe in our dataset are not fully controlled by administrators. In the case of promotions, we can only identify teachers who apply for a position of school leadership, receive an offer, *and* accept it. Although administrators clearly control who is ultimately selected for positions among eligible applicants, those who apply may not reflect their preferred choices. Research conducted in the Miami Dade Public Schools during the same time period as our study indicates that the informal “tapping” of teachers by their principals to consider school leadership opportunities is common and strongly correlated with both interest in becoming a principal and the likelihood of entering the principal track (Myung et al. 2010). This evidence suggests that the inability to observe the application decision may not be an important limitation of our study. Even so, what we have interpreted as recognition of effectiveness could be biased downward if better teachers are more inclined to continue teaching (perhaps because they derive more intrinsic pleasure from their work). Alternatively, effective teachers could be more ambitious in seeking out promotions, in which case apparent recognition of effectiveness could be driven by teacher rather than administrator preferences.

Nor do we know whether reassignments to low-stakes positions are voluntary or involuntary. What we have loosely described as demotions could simply reflect a preference on the part of less effective teachers to try something different. Due to the non-tested nature of low-stakes subjects and grades we are unable to determine whether these teachers are better suited to their new jobs.

More generally, it is difficult to determine whether the patterns of promotion and reassignment we observe represent effective strategies for raising student achievement. If high-quality teachers make for high-quality principals, then promoting them into these positions may

well be. But if teacher quality is only weakly related to principal quality—or if principals have only a limited effect on school quality—then students may be best served by keeping the most effective teachers in the classroom. And moving less effective teachers to low-stakes positions may undermine performance in other subjects and even in the subjects on the basis of which schools will ultimately be evaluated. In particular, our finding that ineffective teachers are reassigned to early-grade math and reading classrooms, although arguably consistent with administrators’ short-term incentives, may well undermine achievement in those subjects in the long run.

The patterns we have documented thus provide grounds for both optimism and concern. On one hand, a profession that is largely unresponsive to quality in terms of compensation, formal evaluation, and dismissals does appear to factor effectiveness into important staffing decisions. Given that administrators also appear to be influenced by the prior achievement of students assigned to a teacher, providing them with formal performance reports could strengthen these relationships.²⁰ On the other hand, even if one believes that high-stakes subjects are most important, one should still be concerned about the movement of less effective teachers into classrooms in these subjects in low-stakes grades. Schools and students would be better served by dismissing their weakest teachers entirely, something that does not appear to occur in more than a trivial way in the average school.

²⁰ Rockoff et al (2010) show that providing value-added reports on individual teachers to New York City principals increased the probability that low-performing teachers exited their schools.

References

- Beteille, Tara, Demetra Kalogrides, and Susanna Loeb. 2009. "Effective Schools: Managing the Recruitment, Development, and Retention of High-Quality Teachers." CALDER Working Paper No. 37. Washington DC: Urban Institute.
- Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2009. "Estimating Principal Effectiveness." CALDER Working Paper No. 32. Washington DC: Urban Institute.
- Chingos, Matthew M. and Martin R. West. 2010. "Do Effective Teachers Earn More Outside of the Classroom?" Program on Education Policy and Governance Working Paper No. 10-02. Harvard University.
- Goldhaber, Dan, Betheny Gross, and Daniel Player. 2009. "Teacher Career Paths, Teacher Mobility, and Persistence in Teaching." CALDER Working Paper No. 29. Washington DC: Urban Institute.
- Hanushek, Eric and Steven Rivkin. 1997. "Understanding the Twentieth-Century Growth in U.S. School Spending." *Journal of Human Resources* 32(1): 35–68.
- Hanushek, Eric and Steven Rivkin. 2010. "Constrained Job Matching: Does Teacher Search Harm Disadvantaged Urban Schools?" NBER Working Paper No. 15816. Cambridge, Mass: National Bureau for Economic Research.
- Harris, Douglas, and Tim Sass. 2009. "What Makes for a Good Teacher and Who Can Tell?" CALDER Working Paper No. 30. Washington DC: Urban Institute.
- Jacob, Brian L. and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26(1): 101–136.
- Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2007. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6): 615–631.
- Myung, Jeannie, Susanna Loeb, and Eileen Lai Horng. 2010. "Tapping the Principal Pipeline: Identifying Talent for Future School Leadership in the Absence of Formal Succession Management Programs." Institute for Research on Education Policy and Practice Report No. 10-1. Stanford University.
- Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26(3): 237–57.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73: 417–458.

Rockoff, Jonah. 2004. "The impact of individual teachers on student achievement: evidence from panel data." *American Economic Review* 94(2): 236–240.

Rockoff, Jonah E. and Cecilia Speroni. 2010. "Subjective and Objective Evaluations of Teacher Effectiveness." *American Economic Review* 100(2): 261–266.

Rockoff, Jonah E, Douglas O. Staiger, Thomas J. Kane, and Eric S. Taylor. 2010. "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools." NBER Working Paper No. 16240. Cambridge, Mass: National Bureau of Economic Research.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125(1): 175–214.

Tyler, John H., Eric S. Taylor, Thomas J. Kane, and Amy L. Wooten. 2010. "Using Student Performance Data to Identify Effective Classroom Practices." *American Economic Review* 100(2): 256–60.

Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. 2009. "The Widget Effect." *Education Digest* 75(2): 31–35.

West, Martin R. and Matthew M. Chingos. 2009. "Teacher Effectiveness, Mobility, and Attrition in Florida," in Matthew G. Springer, ed., *Performance Incentives: Their Growing Impact on American K-12 Education*. Washington DC: Brookings Institution Press, pp. 251–271.

Figure 1. Teacher Effectiveness by Job Type
Teachers Leaving High-Stakes Classrooms

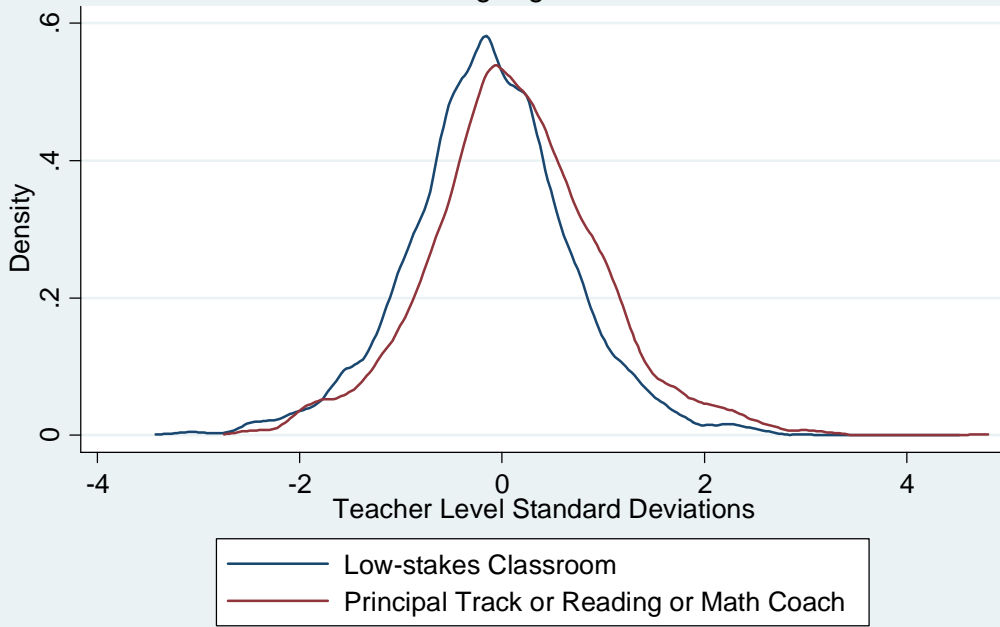
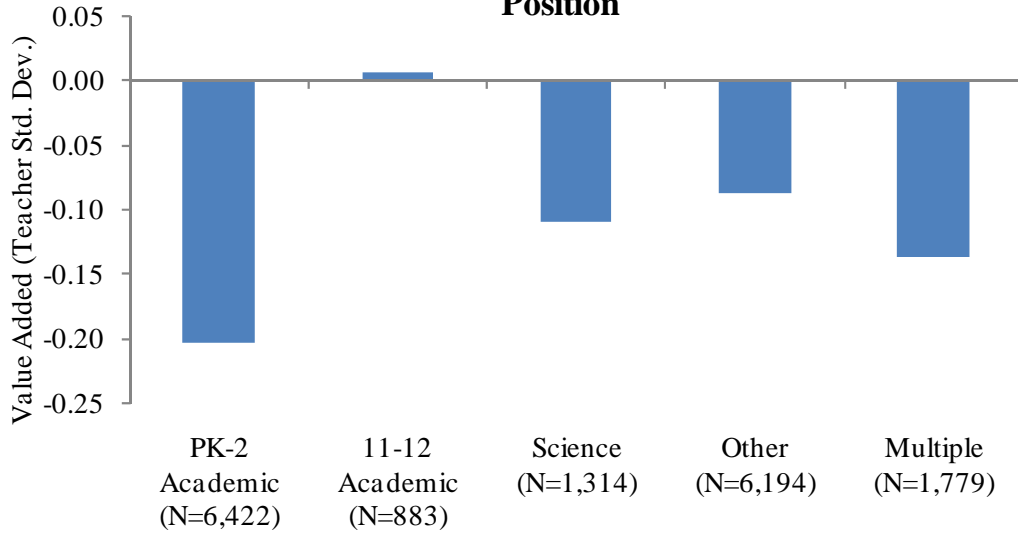


Figure 2. Average Value Added, by Low-Stakes Position



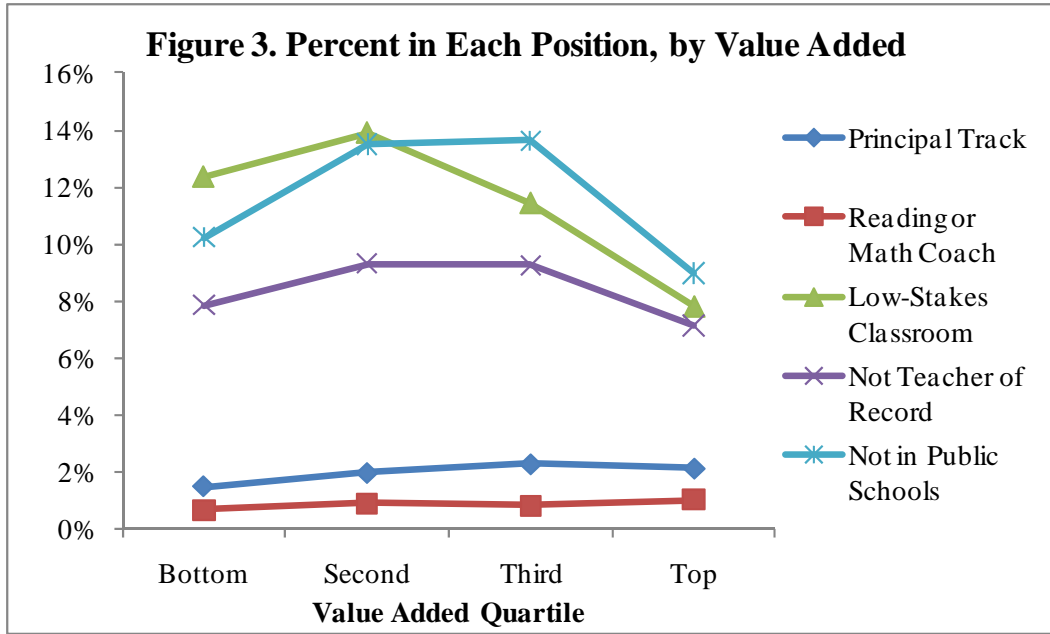


Table 1

Positions of 2002 Cohort of 4th- to 8th-Grade Teachers in High-Stakes Classrooms, 2002-2009

	2002	2003	2004	2005	2006	2007	2008	2009
Principal Track	0.0%	0.3%	0.7%	1.2%	2.2%	2.9%	3.5%	3.8%
Reading or Math Coach	0.0%	0.0%	0.0%	0.2%	0.5%	1.6%	2.0%	2.3%
High-Stakes Classroom	100.0%	84.2%	73.8%	68.0%	64.0%	58.6%	54.3%	51.6%
Low-Stakes Classroom	0.0%	7.1%	10.7%	11.2%	11.3%	12.8%	13.0%	14.7%
Not Teacher of Record	0.0%	5.3%	7.8%	9.6%	9.4%	9.1%	9.4%	8.5%
Not in Public Schools	0.0%	3.1%	6.9%	9.7%	12.6%	15.1%	17.8%	19.1%
Number of teachers	24,475	23,618	22,702	21,771	20,852	19,886	19,001	18,132

Notes: The number of teachers decreases over time because teacher*year observations when the teacher was older than 55 are excluded.

Table 2

Summary Statistics, 2002 Cohort of 4th- to 8th-Grade Teachers in High-Stakes Classrooms, 2003-2009

	All	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-added in math and/or reading	0.02	0.13	0.13	0.05	-0.13	-0.01	-0.03
Mean Earnings	\$44,799	\$60,921	\$48,713	\$46,173	\$44,659	\$38,313	\$25,897
Mean Years of Experience	10.7	9.9	12.9	11.3	10.4	10.0	7.1
Mean Age	41.8	39.4	42.8	42.4	42.2	40.7	39.0
Percent Black	18%	30%	26%	17%	18%	20%	15%
Percent Hispanic	9%	12%	11%	9%	10%	11%	9%
Percent Male	15%	30%	6%	14%	17%	14%	16%
Percent with Master's	41%	87%	58%	40%	40%	50%	30%
Percent with Doctorate	2%	5%	3%	2%	2%	3%	2%
Percent Middle School	39%	39%	46%	38%	42%	38%	44%
Number of Teacher*Year Observations	145,962	2,878	1,272	96,110	16,592	12,232	16,878
Number of Teachers	23,618	830	648	21,588	6,762	6,138	4,824

Notes: Teacher*year observations when the teacher was older than 55 are excluded. Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher.

Table 3

Relationship between Effectiveness Quartile (Relative to Bottom Quartile) and Movement from High-Stakes Classrooms to Other Positions, Multinomial Logit Models, Marginal Effects

Without Controls						
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.005 (0.002)* 25%	0.002 (0.001)* 23%	-0.067 (0.007)** -10%	0.013 (0.004)** 11%	0.014 (0.003)** 17%	0.032 (0.005)** 28%
Value-Added in Third Quartile	0.008 (0.002)** 41%	0.001 (0.001) 14%	-0.046 (0.007)** -7%	-0.011 (0.004)* -9%	0.014 (0.003)** 17%	0.033 (0.005)** 28%
Value-Added in Top Quartile	0.006 (0.002)** 32%	0.003 (0.001)** 36%	0.056 (0.006)** 8%	-0.045 (0.004)** -40%	-0.007 (0.003)* -8%	-0.013 (0.004)** -11%
Overall Share in Position	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Teacher*Year Observations	145,962					
Number of Teachers	23,618					
Pseudo R-squared	0.01					
With Controls						
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.002 (0.001)* 10%	0.002 (0.001)* 20%	-0.059 (0.007)** -9%	0.015 (0.005)** 13%	0.015 (0.003)** 17%	0.026 (0.004)** 22%
Value-Added in Third Quartile	0.003 (0.001)** 15%	0.001 (0.001) 12%	-0.037 (0.007)** -6%	-0.009 (0.004)* -8%	0.014 (0.003)** 17%	0.027 (0.004)** 23%
Value-Added in Top Quartile	0.002 (0.001)** 11%	0.003 (0.001)** 28%	0.062 (0.006)** 9%	-0.046 (0.004)** -41%	-0.009 (0.003)** -10%	-0.012 (0.004)** -10%
Overall Share in Position	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Teacher*Year Observations	145,962					
Number of Teachers	23,618					
Pseudo R-squared	0.05					

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of the listed outcome, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Controls include teacher race/ethnicity, gender, educational attainment, and experience. Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.

Table 4

Relationship between Effectiveness Quartile (Relative to Bottom Quartile) and Movement from High-Stakes Classrooms to Other Positions, Multinomial Logit Models, Marginal Effects, Controlling for Sorting Measure

	Without Controls					
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.005 (0.002)* 24%	0.002 (0.001)* 23%	-0.073 (0.007)** -11%	0.018 (0.004)** 15%	0.016 (0.003)** 19%	0.033 (0.005)** 28%
Value-Added in Third Quartile	0.008 (0.002)** 38%	0.001 (0.001) 14%	-0.059 (0.007)** -9%	-0.002 (0.004) -2%	0.018 (0.003)** 21%	0.035 (0.005)** 30%
Value-Added in Top Quartile	0.006 (0.002)** 29%	0.003 (0.001)** 36%	0.039 (0.006)** 6%	-0.035 (0.004)** -31%	-0.003 (0.003) -4%	-0.010 (0.004)* -9%
Sorting Measure in Second Quartile	0.002 (0.002) 8%	0.001 (0.001) 9%	0.039 (0.007)** 6%	-0.040 (0.005)** -35%	-0.016 (0.004)** -19%	0.014 (0.005)** 12%
Sorting Measure in Third Quartile	0.005 (0.002)* 24%	0.001 (0.001) 7%	0.067 (0.007)** 10%	-0.052 (0.004)** -45%	-0.022 (0.004)** -26%	0.001 (0.005) 1%
Sorting Measure in Top Quartile	0.004 (0.002)+ 22%	0.000 (0.001) 4%	0.110 (0.007)** 17%	-0.067 (0.004)** -59%	-0.030 (0.004)** -35%	-0.018 (0.005)** -16%
Overall Share in Position	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Teacher*Year Observations	145,962					
Number of Teachers	23,618					
Pseudo R-squared	0.02					

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of the listed outcome, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.

Table 5

Relationship between Effectiveness Quartile (Relative to Bottom Quartile) and Retention in High-Stakes Classrooms, Multinomial Logit Models, Marginal Effects, Accountability Grade

	Accountability Grade in 2002 of School Worked at in 2002				
	A	B	C	D	F
Value-Added in Second Quartile	-0.066 (0.011)** -10%	-0.069 (0.014)** -10%	-0.069 (0.013)** -11%	-0.047 (0.025)+ -8%	-0.106 (0.054)+ -19%
Value-Added in Third Quartile	-0.066 (0.011)** -10%	-0.062 (0.014)** -9%	-0.028 (0.013)* -4%	0.037 (0.028) 6%	0.012 (0.053) 2%
Value-Added in Top Quartile	0.035 (0.010)** 5%	0.032 (0.013)* 5%	0.073 (0.012)** 11%	0.119 (0.026)** 20%	0.098 (0.052)+ 17%
Share Stayed in High-Stakes Classroom	68.5%	66.1%	64.4%	60.0%	56.6%
Number of Teachers	9,512	5,618	5,946	1,369	331

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of remaining in a high-stakes classroom, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.

Table 6

Differences in Log(Earnings) by Job (Compared to High-Stakes Classroom Teachers), 2002-2008

	(1)	(2)	(3)	(4)	(5)
Value-Added in math and/or reading (standardized)				0.015 (0.003)**	0.009 (0.002)**
Principal Track	0.254 (0.011)**	0.212 (0.011)**	0.112 (0.016)**		0.253 (0.011)**
Reading or Math Coach	-0.001 (0.016)	-0.023 (0.014)	-0.136 (0.017)**		-0.002 (0.016)
Low-Stakes Classroom	-0.054 (0.005)**	-0.037 (0.004)**	-0.037 (0.006)**		-0.052 (0.005)**
Not Teacher of Record	-0.557 (0.017)**	-0.547 (0.016)**	-0.610 (0.017)**		-0.557 (0.017)**
Not in Public Schools	-1.343 (0.034)**	-1.243 (0.034)**	-1.376 (0.035)**		-1.342 (0.034)**
Controls?	No	Yes	Yes	No	No
Teacher Fixed Effects?	No	No	Yes	No	No
Teacher*Year Observations	142,966	142,966	142,966	142,966	142,966
Number of Teachers	24,475	24,475	24,475	24,475	24,475
R-squared	0.19	0.26	0.54	0.01	0.19

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Omitted category is High-Stakes Classroom Teachers. Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. All regressions including year dummies as well as dummy variables identifying middle school math, reading, and math/reading teachers. Controls include teacher educational attainment, experience, race, and gender. Data are based on cohort of teachers in 2002, and cover period from 2002 to 2008. Teacher*year observations when the teacher was older than 55 are excluded.

Appendix Table 1

Relationship between Effectiveness and Movement from High-Stakes Classrooms to Other Positions, Multinomial Logit Models, Marginal Effects, Linear Models

	Without Controls					
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in math and/or reading (standardized)	0.003 (0.001)**	0.001 (0.000)**	0.025 (0.002)**	-0.020 (0.001)**	-0.003 (0.001)**	-0.006 (0.002)**
Overall Share in Position	13%	12%	4%	-18%	-3%	-5%
Teacher*Year Observations	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Number of Teachers			145,962			
Pseudo R-squared			23,618			
			0.01			
	With Controls					
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in math and/or reading (standardized)	0.001 (0.000)**	0.001 (0.000)**	0.028 (0.002)**	-0.021 (0.001)**	-0.004 (0.001)**	-0.006 (0.001)**
Overall Share in Position	4%	10%	4%	-18%	-5%	-5%
Teacher*Year Observations	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Number of Teachers			145,962			
Pseudo R-squared			23,618			
			0.05			

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of the listed outcome, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Controls include teacher race/ethnicity, gender, educational attainment, and experience. Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.

Appendix Table 2

Relationship between Effectiveness Quartile (Relative to Bottom Quartile) and Movement from High-Stakes Classrooms to Other Positions, Multinomial Logit Models, Coefficients on Control Variables

	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.002 (0.001)* 10%	0.002 (0.001)* 20%	-0.059 (0.007)** -9%	0.015 (0.005)** 13%	0.015 (0.003)** 17%	0.026 (0.004)** 22%
Value-Added in Third Quartile	0.003 (0.001)** 15%	0.001 (0.001) 12%	-0.037 (0.007)** -6%	-0.009 (0.004)* -8%	0.014 (0.003)** 17%	0.027 (0.004)** 23%
Value-Added in Top Quartile	0.002 (0.001)** 11%	0.003 (0.001)** 28%	0.062 (0.006)** 9%	-0.046 (0.004)** -41%	-0.009 (0.003)** -10%	-0.012 (0.004)** -10%
Middle school reading	0.000 (0.001) -2%	0.008 (0.001)** 87%	-0.028 (0.006)** -4%	0.001 (0.004) 1%	-0.006 (0.003)+ -7%	0.025 (0.004)** 22%
Middle school math	-0.001 (0.001) -5%	-0.003 (0.001)** -30%	0.016 (0.006)* 2%	-0.005 (0.004) -5%	-0.014 (0.003)** -17%	0.007 (0.004) 6%
Middle school reading/math	-0.003 (0.001)** -14%	-0.002 (0.001)+ -20%	-0.136 (0.010)** -21%	0.106 (0.008)** 93%	0.015 (0.005)** 18%	0.019 (0.007)** 16%
Master's degree	0.030 (0.002)** 152%	0.004 (0.001)** 44%	-0.027 (0.005)** -4%	-0.003 (0.003) -3%	0.032 (0.003)** 39%	-0.037 (0.003)** -32%
Doctoral degree	0.028 (0.005)** 142%	0.003 (0.002)+ 37%	-0.060 (0.016)** -9%	0.001 (0.009) 1%	0.034 (0.009)** 40%	-0.006 (0.011) -5%
6-10 years experience	0.009 (0.001)** 44%	0.008 (0.001)** 88%	0.013 (0.005)* 2%	0.027 (0.003)** 23%	0.021 (0.002)** 25%	-0.078 (0.005)** -67%
11-15 years experience	0.010 (0.001)** 49%	0.006 (0.001)** 69%	0.063 (0.007)** 10%	0.026 (0.004)** 23%	0.014 (0.003)** 16%	-0.118 (0.005)** -102%
16-20 years experience	0.006 (0.001)** 28%	0.007 (0.001)** 74%	0.116 (0.007)** 18%	0.016 (0.004)** 14%	0.002 (0.004) 2%	-0.145 (0.005)** -125%
21+ years experience	0.000 (0.001) -1%	0.007 (0.001)** 73%	0.157 (0.007)** 24%	0.011 (0.004)** 10%	-0.013 (0.003)** -15%	-0.162 (0.005)** -139%

Black	0.007 (0.001)**	0.004 (0.001)**	-0.007 (0.006)	0.002 (0.004)	0.015 (0.003)**	-0.020 (0.004)**
	33%	48%	-1%	2%	17%	-17%
Hispanic	0.003 (0.001)**	0.004 (0.001)**	-0.026 (0.008)**	0.021 (0.006)**	0.018 (0.004)**	-0.020 (0.005)**
	17%	39%	-4%	19%	22%	-17%
Male	0.011 (0.001)**	-0.004 (0.001)**	-0.022 (0.007)**	0.018 (0.005)**	-0.004 (0.003)	0.002 (0.004)
	55%	-44%	-3%	16%	-5%	1%
Overall Share in Position	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Teacher*Year Observations	145,962					
Number of Teachers	23,618					
Pseudo R-squared	0.05					

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of the listed outcome, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Controls include teacher race/ethnicity, gender, educational attainment, and experience. Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.

Appendix Table 3

Relationship between Effectiveness Quartile (Relative to Bottom Quartile) and Movement from High-Stakes Classrooms to Other Positions, Multinomial Logit Models, Marginal Effects, Alternative Specifications

Value Added Calculated Using 3 Years of Data for All Teachers						
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.005 (0.002)* 26%	0.002 (0.001)* 22%	-0.038 (0.007)** -6%	0.002 (0.005) 2%	0.012 (0.004)** 13%	0.017 (0.005)** 14%
Value-Added in Third Quartile	0.009 (0.002)** 43%	0.001 (0.001) 12%	-0.022 (0.007)** -3%	-0.017 (0.005)** -15%	0.011 (0.004)** 13%	0.018 (0.005)** 15%
Value-Added in Top Quartile	0.008 (0.002)** 41%	0.003 (0.001)** 38%	0.053 (0.007)** 8%	-0.044 (0.004)** -39%	-0.004 (0.003) -5%	-0.016 (0.005)** -13%
Overall Share in Position	2.0%	0.9%	64.8%	11.4%	8.7%	12.2%
Teacher*Year Observations	130,123					
Number of Teachers	21,107					
Pseudo R-squared	0.01					
Value Added Estimates Calculated Conditional on School Fixed Effects						
	Principal Track	Reading or Math Coach	High-Stakes Classroom	Low-Stakes Classroom	Not Teacher of Record	Not in Public Schools
Value-Added in Second Quartile	0.007 (0.002)** 33%	0.002 (0.001)* 22%	-0.065 (0.007)** -10%	0.008 (0.004)+ 7%	0.014 (0.003)** 17%	0.035 (0.005)** 30%
Value-Added in Third Quartile	0.009 (0.002)** 44%	0.003 (0.001)** 34%	-0.050 (0.007)** -8%	-0.011 (0.004)** -10%	0.018 (0.003)** 22%	0.032 (0.005)** 27%
Value-Added in Top Quartile	0.009 (0.002)** 43%	0.004 (0.001)** 48%	0.053 (0.006)** 8%	-0.047 (0.004)** -41%	-0.006 (0.003)+ -7%	-0.013 (0.004)** -11%
Overall Share in Position	2.0%	0.9%	65.8%	11.4%	8.4%	11.6%
Teacher*Year Observations	145,962					
Number of Teachers	23,618					
Pseudo R-squared	0.01					

Notes: + significant at 10%; * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in parentheses. Coefficients indicate predicted change in probability of the listed outcome, holding all control variables at their means. All regressions include dummies identifying teachers in the following grade-subject combinations: middle reading, middle math, and middle reading-math (elementary is the omitted category). Standardized (mean zero, standard deviation 1) value-added measures were estimated separately by grade-level and subject (4-5 math, 4-5 reading, 6-8 math, and 6-8 reading) and test (FCAT and Stanford Achievement Test) and averaged (when multiple measures were available) to form a single effectiveness measure for each teacher. Data are based on cohort of teachers in 2002, and cover period from 2003 to 2009. Teacher*year observations when the teacher was older than 55 are excluded.