



NATIONAL
CENTER *for* ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington



**Testing, Teacher
Turnover and the
Distribution of
Teachers Across Grades
and Schools**

**Dillon Fuchsman
Tim R. Sass
Gema Zamarro**

Testing, Teacher Turnover and the Distribution of Teachers Across Grades and Schools

Dillon Fuchsman
University of Arkansas

Tim R. Sass
Georgia State University/CALDER

Gema Zamarro
University of Arkansas

Contents

Contents	i
Acknowledgments	ii
Abstract	iii
I. Introduction	1
II. Literature Review	3
III. Background and Data	6
IV. Econometric Methods	7
V. Empirical Results	14
VI. Robustness Checks	18
VII. Discussion and Conclusions	21
References	23
Figures and Tables	25

Acknowledgments

The contents of this report were developed using data provided by Georgia’s Academic and Workforce Analysis and Research Data System (GA•AWARDS). However, those contents do not necessarily represent the policy of GA•AWARDS or any of its participating organizations, and you should not assume endorsement by GA•AWARDS or any of its participating organizations. This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated. We are thankful for insightful comments and feedback by participants at the AEFP 45th Annual Conference, the Department of Education Reform Brownbag Seminar, and the NC State Microeconomics Workshop Series.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street NW, Washington, DC 20007
202-403-5796 • www.caldercenter.org

Testing, Teacher Turnover and the Distribution of Teachers Across Grades and Schools

Dillon Fuchsman, Tim R. Sass, Gema Zamarro
CALDER Working Paper No. 229-0220
February 2020

Abstract

Teacher turnover has adverse consequences for student achievement and imposes large financial costs for schools. Some have argued that high-stakes testing may lower teachers' satisfaction with their jobs and could be a major contributor to teacher attrition. In this paper, we exploit changes in the tested grades and subjects in Georgia to study the effects of eliminating high-stakes testing on teacher turnover and the distribution of teachers across grades and schools. To measure the effect of testing pressures on teacher mobility choices we use a "difference-in-differences" approach, comparing changes in mobility over time in grades/subjects that discontinue testing vis-à-vis grades/subjects that are always tested. Our results show that eliminating testing did not have an impact on the likelihood of leaving teaching, changing schools within a district, or moving between districts. We only uncover small negative effects on the likelihood of grade switching. However, we do find relevant positive effects on retention of beginning teachers in the profession. In particular, the average probability of exit for teachers with 0-4 years of experience fell from 14 to 13 percentage points for teachers in grades 1 and 2 and from 14 to 11 percentage points in grades 6 and 7.

Keywords: *Teacher turnover; High-stakes Testing; Accountability Pressure; Difference-in-differences approach*

I. Introduction

Teacher turnover is found to have adverse consequences for student achievement (Ronfledt et al., 2013) and to impose large financial costs for schools (Barnes, Crowe, & Schaefer, 2007; Birkeland & Curtis, 2006; Milanowski & Odden, 2007). Turnover is especially problematic among early-career teachers, with 17 percent leaving the profession within their first five years (Gray & Taie, 2015). Some have argued that high-stakes testing lowers teachers' job satisfaction and is thus a major contributor to teacher attrition.¹ This in turn has fueled calls for a reduction in the amount of mandatory statewide testing.

High-stakes testing could also affect the distribution of teachers across grades and across schools. Within a school, principals might allocate their more experienced or higher-performing teachers to tested grades in response to accountability pressure (Grissom, Kalogrides and Loeb, 2017). Similarly, teachers may have a greater incentive to flee schools with low-performing students if their students are subject to high-stakes testing.² If high-performing teachers switch to schools with more advantaged students this in turn could exacerbate differences in school quality.

Changes in Georgia's testing practices present a unique opportunity to study the effect of testing on teacher turnover, the distribution of teachers across grades within schools and the variation in quality across schools. Over the last decade, Georgia has experienced four different testing regimes with variation in both subjects tested and grade levels in which testing occurred. This includes dropping all statewide achievement tests in some grades, eliminating some subject

¹ In a recent survey of Georgia teachers, "the number and emphasis of mandated tests" was ranked as the most prominent among eight often-cited reasons that teachers leave the profession. See Owens (2015).

² Teachers would have an incentive to leave if either accountability is based on student achievement (rather than on growth), or if the evaluation system does not fully account for student/family characteristics, resulting in biased estimates of productivity for schools and teachers serving high proportions of disadvantaged students.

areas from testing in all grades and reducing the number of grades in which certain subject areas are tested.

Given concerns over testing and its effects on teachers' labor market decisions, we seek to derive causal estimates of the impact of testing on teacher mobility by analyzing the effects of changes in mandated testing in Georgia. Previous studies have looked at the effect of establishing a statewide accountability system or the impact of accountability pressure within an established system on teacher mobility. However, our focus is somewhat different. Driven by federal mandates in the No Child Left Behind Act and incentives associated with the Race to the Top competition, most states have established test-score based metrics of school performance and associated accountability systems. However, the recent Every Student Succeeds Act (ESSA) grants more flexibility over testing requirements to states. Thus, the debate has shifted toward the extent of testing within pre-existing accountability structures. We therefore strive to understand the implications of altering the grades and subjects tested on teachers' labor market choices. In particular, we seek to answer the following research questions:

1. Does reducing the number of tested subjects (or eliminating testing entirely) affect the likelihood teachers will switch schools or exit the profession?
2. Does reducing the number of tested subjects (or eliminating testing entirely) affect the likelihood that teachers will change grade assignments?
3. Does reducing the number of tested subjects (or eliminating testing entirely) alter the distribution of school quality?

II. Literature Review

Most prior research focuses on the effects of establishing test-based accountability systems or differential pressure faced by schools within those systems. Early work by Clotfelter, et al. (2004) and a subsequent study by Fuller and Ladd (2013) study the impact on teacher labor markets of the introduction of a statewide accountability system in North Carolina in 1996/97. The system not only established a test-based grading system for schools, it also created a bonus system whereby teachers in the highest-rated (“exemplary”) category received bonuses of \$1500. Clotfelter, et al. compare retention rates of new teachers in low-performing elementary schools in North Carolina two years prior to the introduction the school accountability system with attrition of new teachers in the first year the accountability system was in place. Estimating hazard models of teacher exit, they find that post reform, the probability that first-year teachers in low-performing school leave increased.

Fuller and Ladd extend the analysis period to 1994/95 through 2008/09, which allows them to investigate both the effects of the initial state-based school accountability system established in 1996/97 and the later federal No Child Left Behind Act, which began in 2002/03. They compare elementary school teachers in kindergarten through second grade, where testing did not occur, to teachers in grades three through five, where statewide testing occurred. Estimating logit regressions of the probability of moving up from untested lower elementary grades to tested upper elementary grades and vice-versa. Consistent with the notion that principals seek to maximize test scores in response to accountability pressure, the authors find that movement of teachers from lower to upper-elementary grades increases in both the state-based and federal NCLB accountability regimes. They also show that in the NCLB era higher quality teachers (as measured

by licensure test scores) were more likely to move from the lower elementary to upper elementary grades. Further, in the NCLB era higher-value-added teachers were less likely to move down from the upper-elementary to lower-elementary grades.

Perhaps the work that is closest to the present study is an analysis of the institution of statewide testing in selected grades in New York State by Boyd et al. (2008). In 1998/99, New York established new statewide exams in grades four and eight and produced public report cards of individual school performance based on the test score results. A year later, the state adopted a school accountability system that established performance standards. Under this system, failure to meet standards triggered a planning process and schools that continued to not meet the standards were labeled “in need of improvement” and could eventually be dissolved. However, there were no direct consequences for teachers until the 2002/03 school year. Boyd et al. utilize a logit analysis to compare rates of teacher exit from grade four to the departure-from-grade rates of teachers in grades one, two, three, five and six. They find that the turnover rate of fourth-grade teachers decreased relative to teachers in other elementary grades after testing began. They further distinguish between inter-grade movement and exit from public school teaching by estimating multinomial logit models of the choice between remaining in the same grade, switching grades and exiting public school teaching. First-year teachers were 25 percent less likely to leave the fourth grade after testing began, but there was no statistically significant impact on the likelihood that fourth grade teachers left the public school system after testing began.

Sun, Saultz and Ye (2017) analyze the effects of school accountability on teacher turnover by comparing the pre- and post-NCLB periods in states with pre-existing state accountability systems to the changes over time in states that did not have school accountability systems in place

prior to the institution of NCLB. To conduct the analysis they utilize data from the nationally representative Schools and Staffing Surveys and associated Teacher Follow-Up Surveys from 1993–2009. Neither the states with pre-existing systems nor the states without school accountability prior to NCLB experienced any significant change in voluntary between-school mobility or attrition from the profession. However, they find that with the introduction of NCLB, involuntary teacher inter-school mobility and attrition increased in states lacking prior accountability systems relative to those that had state accountability systems in place before the federally mandated accountability system under NCLB.

In contrast to prior work that exploits the initiation of accountability systems to identify the impact of test-based measures of school performance on teacher mobility, two recent papers, Feng, Figlio & Sass (2018) and Dizon-Ross (2018) utilize variation in accountability pressure within an established accountability system to identify impacts on teacher labor-market decisions. Using data from Florida, Feng, Figlio & Sass exploit a change in the way that school grades were assigned to measure the impact of an unanticipated change in the grade a school received on teacher turnover. Using both difference-in-differences and regression discontinuity strategies, they find that school grading shocks do not affect teacher mobility decisions in most schools, but do have significant impacts in schools that unexpectedly receive a failing grade. Teachers in schools that receive an “F” rating are significantly more likely to leave the school and the turnover increase is disproportionately higher among the best teachers (as measured by estimated impacts on student test scores). Employing a regression discontinuity design, Dizon-Ross obtains rather different results for teachers and schools in New York City. For schools rated C and below, those just below the C/D or D/F threshold have lower rates of teacher turnover than schools just above

the threshold. Dizon-Ross argues that the local decrease in turnover is likely due to accountability pressure driving principals to make improvements in the school environment that make the school more desirable to teachers and this outweighs negative stigma effects from a lower school grade.

III. Background and Data

In Georgia, statewide testing was first mandated by the A+ Education Reform Act of 2000.³ The Act required that students in grades 1-8 be tested in reading, English/language arts, and math. Students in grades 3-8 were also required to be tested annually in science and social studies. Initiation of the testing regime was left to the discretion of the State Board of Education. Beginning in spring 2000, all students in grades four, six, and eight were required to take the Criterion-Referenced Competency Tests (CRCT) in reading, English/language arts, and math. In spring 2002, testing in these three subjects was expanded to include students in all grades 1 through 8. Also in spring 2002, science and social studies tests were added for students in grades 3-8. Testing of reading, English/language arts and math competencies in grades 1-8 and assessments in science and social studies in grades 3-8 continued through spring 2010. Due to budget constraints associated with the Great Recession, the state stopped administering the CRCT in grades one and two; beginning in school year 2010/11 (henceforth 2011), end-of-grade exams were only given to students in grades 3-8.⁴ In 2015 the state adopted a new test, Georgia Milestones. In 2015 and 2016, students in grades 3-8 were tested in ELA, math, science, and social studies. In 2017, science

³ O.C.G.A. §20-2-281. The Act also required end-of-course assessments in grades 9-12.

⁴ On April 29, 2010, the Georgia Legislature adopted a budget that instructed the Georgia Department of Education (GaDOE) to eliminate CRCT testing in grades 1 and 2. Governor Perdue subsequently vetoed that provision and gave the GaDOE the authority to continue testing in June 2010. However, the associated budget cuts remained and the GaDOE choose not to administer CRCT tests in grades 1 and 2 in the 2010/11 school year “due to budgetary constraints.” See Perdue and Dlugolenski (2010).

and social studies testing was scaled back to grades 5 and 8, while students in grades 3-8 were still tested in ELA and math. The patterns of testing over our sample period are summarized in Table 1. The plausibly exogenous changes in tested grades and subjects over time provide us with a unique opportunity to estimate the effects of eliminating mandatory testing on teacher mobility decisions, the allocation of teachers across grades within schools and the distribution of school quality.

We employ data from Georgia's statewide longitudinal database, Georgia's Academic and Workforce Analysis and Research Data System (GA•AWARDS). GA•AWARDS combines data from all educational agencies in Georgia as well as unemployment insurance (UI) records from the Georgia Department of Labor. Thus, individual students can be tracked from pre-K through post-secondary institutions and into the labor force. The database includes data from the 2007 school year forward; thus our sample includes students in grades K-12 for the years 2007 through 2018. We limit our analytical sample to teachers in grades one through eight.⁵

IV. Econometric Methods

To measure the effect of reducing accountability pressures on teacher mobility choices we will use a "difference-in-differences" approach, comparing changes in mobility over time in grades/subjects that discontinue testing vis-à-vis grades/subjects that are always tested. A difference-in-differences strategy will allow us to make causal inferences under the assumption

⁵Given that kindergarten students have not been subject to statewide testing, we do not include kindergarten teachers in our analytical sample. Their inclusion, however, does not affect the results.

that teachers in always-tested and (eventually) non-tested grades and subjects follow the same trend in outcome variables prior to changes in test coverage.

To model turnover, we estimate multinomial logit hazard models, as in Feng and Sass (2017). Five choices are studied: (i) remain in the present school and grade, (ii) remain in the present school but change grade, (iii) transfer to another school within district, (iii) transfer out of the district and (iv) exit teaching in K-12 public schools in Georgia. Our models control for teacher characteristics such as gender, race/ethnicity, teaching experience, educational attainment, pathway to teaching and the institution attended for traditional teacher preparation program completers. The models also control for school-level student body demographics such as school enrollments, the proportion of students eligible for free-and-reduced-price lunch, the proportion of students with disabilities, the proportion classified as English language learners and shares from different races/ethnicities.

To further study the impact of eliminating testing requirements on the allocation of teachers across grades, we also estimate a binary choice logit model of whether or not a teacher changes grades from one year to the next as a function of the extent of state-mandated testing in the relevant grade level. This model also includes controls for teacher and school characteristics as described above.

In order to gauge the impact of test-based accountability pressure on the distribution of school quality, we first estimate linear regression models of student achievement in both math and ELA, where current achievement is a function of prior achievement, student characteristics and peer characteristics. The achievement models also include school indicators (school fixed effects) which yield so-called “value added” estimates of school effectiveness. We then study changes in

the distribution of school effectiveness as a function of changes in testing policies using a Kolgomorov-Smirnov equality-of-distribution test. If school testing policies affect the distribution of teachers across schools, and hence the distribution of school quality, we would expect greater changes in the distribution of school quality when grade or subject-area testing changes. We would also anticipate larger differences in grade groups where there are more substantial changes (e.g. elementary) than in grade groups where testing regimes were more stable (e.g. high school).

A. Teacher Turnover

We model a teacher’s decision about job change as an individual utility maximizing problem over a number of job choices.⁶ A teacher will select among a group of jobs based on her individual preferences and the characteristics of the job, including testing requirements. A teacher will compare the available options and select the job that yields the highest present value of expected utility. The decision facing a teacher during each time period t is represented by:

$$Max[PV(U_S, U_g, U_c, U_D, U_L,)] \text{ where } U = f(\mathbf{D}_{kgmrt}) \quad (1)$$

where the subscript $kgmrt$ indicates teacher k , teaching grade g , in school m and district r at time t . \mathbf{D}_{kgmrt} , represents a vector of control variables that have been shown in the literature to be important in influencing a teacher’s mobility decision. These include teacher demographics and professional credentials, as well as school characteristics. A teacher’s race and gender are included to account for teacher preferences. A teacher’s experience and education level, professional certification status, and subject-specific certification are all included to reflect human capital

⁶ Although we formally model the problem as a one-sided decision, we recognize that some (particularly early-career) teachers may be subject to involuntary terminations and principals may respond to accountability pressure by re-assigning high-performing teachers to tested grades.

investments. The effects of non-monetary working conditions on teacher mobility decisions are taken into account by including a variety of school characteristics such as enrollment, student racial/ethnic composition, the percent of students receiving free or reduced-price lunch (a proxy for family income), English language learners and students with disabilities.

A teacher maximizes his utility by selecting the best of five options. The five possible choices are: stay at the present school and grade (S), move to a different grade within the school (G), change school within the school district (C), move between districts to a new school in a different school district (D), and leave public school teaching in Georgia (L). All moves are assumed to result from utility-maximizing choices. While this assumption may not be correct in cases of involuntary separation due to poor performance or workforce reduction, such cases are relatively rare.

For teachers, most moves and exits occur at the end of the school year. In addition, information on schools is typically only available at yearly intervals. Given this discreteness in the data, we employ a discrete multinomial logit hazard model with both time-varying and time-invariant variables. The discrete-time hazard function models the probability that any of the five events—staying in the same school and grade, moving grades within a school, moving within the district, moving between districts, or leaving—happens to teacher k during period $t+1$, which is conditional on staying in the teaching profession until that time. The discrete-time hazard function can be interpreted as the probability of transition at discrete time $t+1$ given survival up to time $t+1$:

$$h_{kgmr,t+1} = \Pr[T_{kgmr,t+1} = t + 1 \mid T_{kgmr,t+1} \geq t + 1, D_{kmrt}] \quad (2)$$

Assuming independence of irrelevant alternatives and error terms that are independently and identically extreme value distributed, a multinomial logit hazard model specifies the probability of choosing each alternative as a function of teacher and school characteristics.

To capture the effect on teachers' moves of testing requirements we use a difference-in-differences approach following this specification for each potential type of move:

$$\text{Logit}(h_{kmr,t+1}(j)) = \beta_j + \beta_{j1}D_{kgmrt} + \beta_{j2}E + \beta_{j3}Z_t + \beta_{j4}(E \times Z_t); \quad j = S, G, C, D, L \quad (3)$$

where E is an indicator for teachers in grades that were affected by the testing requirements change (i.e. teachers in grades one and two and science or social science teachers in grades six or seven). Z_t is an indicator that signifies the period after the testing requirement change (2011 for the grades one and two testing change and 2017 for the science and social science change in testing for grades six and seven) and $E \times Z_t$ is the interaction between the two. Our main coefficient of interest is the coefficient of this interaction, β_{j4} , which represents the extent that the probability of a certain teacher move changed because of the changes in testing requirements.⁷ We also explore heterogeneous effects by teacher experience including an interaction term with an indicator variable for early-career teachers, i.e. teachers with less than five years of experience⁸. We present separate estimates for elementary school teachers in grades one to five, who are affected by the

⁷ By defining the no-testing period for grades one and two as 2011 and beyond, we are implicitly assuming that teachers did not know when making future employment decisions in spring 2010 that there would not be testing in grades one and two in school year 2011. As described in footnote 4 above, there was uncertainty about future testing at least as late as June 2010 when Governor Perdue signed the FY 2011 budget. Though very unlikely, if teachers knew that testing in grades one and two would be dropped they might have been more likely to continue teaching in 2010/11 (and hence be categorized as a stayer in 2010). We study the potential for these anticipatory effects in the robustness section of this paper.

⁸ We also estimated models allowing for differential effects for later career teachers but did not find any significant differences in the estimated effects in this case.

testing change in 2011 if teaching in grades one and two, using data up to school year 2017, before the second change took place. We also present separate estimates for middle school teachers in grades six to eight, who were affected by the second testing change in 2017 if teaching in science or social sciences in grades six and seven, using all data available until school year 2018. We believe estimating the models separately for elementary and middle school teachers is the best strategy as testing changes affected these two groups of teachers differently. This is so because most teachers in elementary school teach all subjects and so will only be affected by the eliminating of testing in 2011. Third grade teachers would also in theory be affected by the eliminating of testing in science and social science in 2017 but, as they continued to be tested in other subjects, they would not be able to avoid testing entirely. Middle school teachers typically teach one subject and for them the change in testing in 2017 was subject and grade specific. Then, teachers could avoid testing entirely by changing grades in this case.

B. Teacher Mobility Between Grades

To further study the effect of testing on movement between grades, we define an indicator for the teacher moving grades independently of doing so within or across different schools. Similar to our difference-in-differences approach explained above, we estimate simple logit hazard models of the following form:

$$\text{logit}[\lambda(t_i)] = \beta_0 + \beta_1 D_{kgmrt} + \beta_2 E + \beta_3 Z_t + \beta_4 (E \times Z_t) \quad (4)$$

where all the variables are described as above.

C. Changes in School Quality

To study changes in the distribution of school quality due to the testing policy change, we model the contribution of school m to the test score (Y) of student i at time t and in grade g following this value-added model for each year separately:

$$Y_{igmt} = \delta_0 + \delta_1 Y_{imt-1} + \delta_2 X_{igmt} + \delta_3 G_{gmt} + \gamma_g + \alpha_m + \varepsilon_{igmt} \quad (5)$$

Test scores are standardized within subject (ELA and math), school year, and grade-level; X_{igst} contains student-level indicator variables for gender, race/ethnicity (Hispanic, Asian, Black, or other race), free or reduced price lunch status, and English language learner status; G_{gst} are grade-level covariates for the proportions of female students, Hispanic students, Asian, Black, or other race students, free or reduced price lunch eligible students, English language learner students, and students with disabilities in a given grade, as well as grade-level enrollment; and γ_g are grade-level fixed effects. α_m represents a series of school fixed effects capturing our estimates of school quality.

For this analysis we focus on elementary school grades and estimate equation (5) for grades 4 and 5 for each year of our data separately, starting with the 2008 school year and concluding with the 2018 school year. We could estimate equation (5) for grades 2 and 3 in the pre-policy change period, but opt to omit them to increase comparability with the post-policy change period where we can only estimate the model for grades 4 and 5.⁹ We test for changes in the distribution of school quality resulting from the testing policy change using a non-parametric Kolmogorov-

⁹ Results including grades 2 and 3 in the pre-policy period were, however, similar and are available from the authors upon request.

Smirnov equality-of-distributions test over the sample period of 2008 through 2018, focusing on the first testing change in 2011.

V. Empirical Results

A. Descriptive Analyses

In Figures 1A-1D we present the proportion of teachers who experienced a given move in a certain year. If testing is driving teacher mobility decisions, we would expect the greatest impacts on attrition from the profession and within-school grade level changes, as intra-district or inter-district moves between schools would not necessarily reduce exposure to testing. Visually, we do not observe any differences in attrition, within-school grade changes or inter-school mobility trends across grades before and after the testing requirements changed for grades 1 and 2 in 2011. We also do not observe significant differences in trends after the change in testing requirements for science and social-science teachers in grades 6 and 7.

Table 2 presents descriptive statistics for our sample of first to eight-grade teachers across the years of analysis, from 2007 to the 2018 academic year. We present descriptive statistics for all teachers together, as well as for early-career teachers (i.e. those with less than five years of experience) separately. Twelve and one-half percent of the observations in our sample represent moves to leave the teaching profession, 2.6 percent moves to change districts, 4.5 percent moves to different schools and 12.6 percent changes of grades taught within the same school¹⁰. About 20 percent of the teachers in our data are early-career teachers with less than five years of experience and the majority of teachers are white (75 percent). Early-career teachers present relatively higher

¹⁰ We do not find significant differences in these statistics depending on the teacher being an elementary school teacher or a middle school teacher.

rates of mobility as 14 percent of the observations among early-career teachers represent moves to leave the teaching profession, five percent moves across districts, six percent moves within district across schools and 14 percent moves within school across grades. Schools have on average about half of their student body that is white, 37 percent African American, 14 percent Hispanic and about three percent Asian or other ethnicities. Schools on average have about 58 percent of their student body that qualifies for Free-Reduced lunch, about eight percent are English Language Learners and about 13 percent are students with disabilities.

B. Quantitative Analyses – Teacher Turnover

Table 3 presents average marginal effects from a multinomial logit hazard model of teacher mobility, as described in (2). Panel A presents estimated effects of the removal of testing requirements for first and second grade that took place in 2011 using the sample of elementary school teachers teaching in grades one to five. Panel B presents estimated effects for the removal of testing requirements for science and social science teachers in grades six and seven in 2017 including data from teachers in grades six to eight¹¹. Estimates control for teacher and school characteristics. In particular, the following variables are also included as controls: an indicator for female teachers, an indicator for early-career teachers (less than five years of experience), four indicators for a teacher's race/ethnicity (Hispanic, Asian, African-American and other), five indicators for a teacher's level of educational attainment (high school, associate's degree, master's degree, education specialist and doctorate), three indicators for teachers with special certification (special education, English as a second language, gifted), an indicator teachers who received an

¹¹ In the multinomial choice analyses, we restrict the sample to teachers in grades one to five and grades six to eight, respectively, but moves include all grade levels. Thus, for example, if a teacher starts in grade five and transfers to a middle school in the same district they would be classified as a within-district mover, but in the following years their observations are censored as they would no longer be part of the analytical sample.

education degree from the state of Georgia during our data time frame, the proportion of students in the school from different ethnicities (Hispanic, Asian, African American, other), the proportion of students in the school that qualify for free or reduced-priced lunch, the proportion of students in the school that are English Language Learners, the percentage of students with disabilities in the school and number of students enrolled in the school. Finally, standard errors are clustered at the teacher level.

As can be seen in the estimates presented in the columns labeled as (1), in Panel A, overall, we do not observe significant effects of eliminating testing in grades one and two on the probability of leaving the teaching profession, changing districts or changing schools within a district. We do observe, however, a small but statistically significant negative effect of eliminating testing on the probability of changing grades within schools. After the policy change, the probability of changing grades within the same school for first and second-grade teachers fell by 0.5 percentage points (relative to a mean of 12.6 percent). In general, turnover is lower among first and second grade teachers, compared to teachers in grades three through five. After the testing changes in 2011, we observe higher overall probabilities of changing districts and changing grades, lower probabilities of changing schools and no change in the likelihood of leaving the profession. Early-career teachers are observed to have overall higher levels of turnover. They are about two percentage points more likely to change district, change schools within the district, and move grades.

In the estimates presented in the columns labeled as (2), we allow for different effects of the testing changes on the mobility decisions of early-career teachers. Looking at the changes in testing in grades one and two we observe that early-career first and second grade teachers are 0.9 percentage points less likely to leave the profession (a 6 percent reduction relative to a mean of 14

percentage points) after the elimination of testing and about 0.4 percentage points more likely to change districts.

Panel B of Table 3 shows estimates of the impact of the discontinuation of testing for science and social science teachers in grades six and seven that took place in 2017. Similarly to the results presented in panel A for eliminating testing in grades one and two, we observe no general effect of this test change on either mobility across grades, schools and districts, or exit from public-school teaching. Early-career science and social science teachers of grades six and seven, however, do present a reduction of 2.8 percentage points in the probability of leaving the profession after the change in testing requirements. This is a 20 percent reduction in the probability of exit.

C. Quantitative Analyses – Teacher Overall Mobility Between Grades

To further study the effects of the testing policy change on the probability of changing teaching grades, we also estimate logit hazard models, as described in (4) above. The dependent variable in these models takes a value of one if the teacher changed grades from one year to the next, whether or not they changed schools. Our estimates control for teacher and school characteristics as described above and again, standard errors are clustered at the teacher level.

Table 4, panel A, presents the estimated average marginal effects for the discontinuation of testing in grades one and two. In this case, we fail to find any effect of the testing change in the probability of first and second grade teachers changing grades. We also find no differential effect of the testing policy change for early-career teachers. Results are similar for the discontinuation of testing of science and social science teachers as presented in Panel B.

D. Quantitative Analyses – Changes in School Quality

Figures 2a and 2b display the kernel densities of estimated school value-added contributions for ELA and math, respectively, using our model presented in (5) above. In both plots, the solid line represents the distribution of school quality prior to the testing policy change while the dashed line represents the distribution of school quality after the policy change. In both math and reading, we reject the null hypothesis of the Kolmogorov-Smirnov equality-of-distributions test that the distributions pre- and post-policy change are the same at any level of significance. However, changes in the distribution appear small. Table 5 provides summary statistics of our value-added estimates for 4th and 5th grade pre- and post-policy testing change in 2011 for ELA and math. The mean of the distributions for ELA is lower in the post-policy period, but is higher for math in the post-policy period. However, the differences between pre-policy school quality and post-policy school quality in grades four and five are generally small. It is, however, not surprising that these effects in school quality appear to be small, given the limited overall estimated effects of the testing policy change on teacher mobility.

VI. Robustness Checks

A key assumption for our difference-in-differences empirical strategy is that teachers in always-tested and (eventually) non-tested grades and subjects follow the same time trend in outcome variables prior to changes in test coverage. To further study the plausibility of this assumption, we expanded the model presented in (3) to allow for differential leads and lag effects for the different years of data following this model:

$$\text{Logit}(h_{kmr,t+1}(j)) = \beta_j + \beta_{j1}D_{kgmrt} + \gamma_g + \lambda_t + \sum_{t=-q}^{-1} (E \times \lambda_t) + \sum_{t=1}^m (E \times \lambda_t); \quad j = S, G, C, D, L \quad (6)$$

The model then includes teacher grade and time indicators for each grade and year in the sample dataset and allows for differential lead effects before treatment and lag effects after treatment per year, using the year before the policy change as reference. For the common trends assumption to be satisfied lead effects should not be statistically significant.

Figures 3.A and 3. B show the estimated lead and lag effects for the change in testing grades one and two in 2011 and the change in testing science and social science teachers in 2017, respectively. Overall, leads effects are close to zero and not statistically significant with a few exceptions (marked in black if statistically significant at the 95% level), for teachers in grades one to five affected by the 2011 policy change. Some more lead effects appear significant, however, for the case of teachers affected by the change in testing in science in social science in 2017, especially as it concerns the probability of changing schools. Estimates for this second testing change should be then taken more cautiously as there is less supporting evidence in this case for the common trends assumption to be satisfied. However, our general results were robust to limiting the analysis to years where the common trend assumption appears to be satisfied, years 2012 onwards.¹²

As a robustness check we also estimated binary logit duration models for the decision of leaving the profession alone, as well as for the decisions of changing districts and changing schools but without changing grades in these latter moves. For the binary logit duration model for the decision of leaving the profession, we expected to find similar results to those of the multinomial logit hazard models discussed above. As reported in Appendix Table A.2 overall we

¹² Results of this analysis can be found in the Appendix Table A.1.

found comparable results in this case. In contrast, the models of the decisions of changing districts or schools but no changing grades, act like placebo outcomes and, in the absence of bias, we should find null effects on these. Results are presented in the Appendix Tables A.3 and A.4. Overall, we do not find an effect on the probability of changing districts or schools without changing grades. We only find a marginally significant effect of eliminating testing on the probability of changing districts in the specification that allows for differential treatment effects for beginning teachers. These effects are also an order of magnitude smaller than the estimated significant effects on the probability of leaving the profession. Therefore, based on these results, we believe that the evidence of bias, if any, appears to be weak.

Finally, as the second test change that took place in 2017 only affected science and social-science teachers in grades six and seven, we could also perform a difference-in-differences-in-differences analysis, first comparing turnover rates of science and social-science teachers in grades six and seven with turnover rates of teachers in other subjects in the same grades and subtract the differences in turnover rates across subjects using information for teachers in eight grade. In particular, we follow this difference-in-differences-in differences specification:

$$\begin{aligned} \text{Logit}(h_{kmr,t+1}(j)) = & \beta_j + \beta_{j1}D_{kgmrt} + \beta_{j2}S + \beta_{j3}E + \beta_{j4}Z_t + \\ & \beta_{j5}(S \times E) + \beta_{j6}(S \times Z_t) + \beta_{j7}(E \times Z_t) + \beta_{j7}(S \times E \times Z_t) \end{aligned} \quad ; \quad j = S, G, C, D, L \quad (7)$$

Where S is an indicator for a teacher teaching science or social-science, E is an indicator for teachers in grades that were affected by the testing requirements change (i.e. six or seven). Z_t is

an indicator that signifies the period after the testing requirement change (2017). Our main coefficient of interest in this case is the coefficient of the triple interaction, β_{j7} .

In the absence of violations of the common trends assumption that differences in turnover rates across subjects observed for eight grade teachers are comparable to the differences we would have observed across subjects for six and seven grade teaches in the absence of the change of testing, we expect to find similar results than those presented above using a difference-in-differences approach and using both teachers in other subjects and teachers in eighth grade together as control. Our results are presented in Table 6. Using a triple-differences approach, we still find a significant negative effect of eliminating testing on the probability of leaving the profession for beginning teachers of about three percentage points. However, using this approach we also now find significant negative effects on the probability of changing schools within district for all teachers of about two percentage points. These estimated effects for changing schools within district, however, disappear if we consider binary logit specifications that do not allow changes in grades as a result of the school change.¹³

VII. Discussion and Conclusions

There is growing concern over shortages of teachers and some have argued that testing is contributing to the shortage by driving teachers out of the profession. There is little direct evidence, however, on whether this is indeed the case. Some previous studies have investigated the relationship between testing and teacher turnover, but they are based on situations where testing was simultaneously instituted along with other changes to school accountability. In contrast, we

¹³ Results of these binary logit difference-in-difference-in-differences models are presented in the Appendix Tables A.5, A.6 and A.7.

consider how eliminating testing in certain grades within an established accountability system affects teacher labor market decisions.

Exploiting the elimination of testing in grades one and two and for science and social science teachers in grades six and seven in Georgia, we find that the removal of statewide achievement tests had no effect on the likelihood of leaving teaching in general, nor did it impact the probability of changing schools within a district or the probability of moving between districts for teachers as a whole. We did find that eliminating testing in the lower elementary grades reduced the likelihood of teachers switching grades, but the effect was small. Thus, overall, we find no support for the notion that testing alone is having a major impact on overall teacher attrition or, conversely, that reducing the amount of testing will reduce teacher shortages. However, we did find a meaningful reduction in the probability of early-career teachers leaving the profession when testing requirements were relaxed. In particular, the average probability of exit for teachers with 0-4 years of experience fell from 14 to 13 percentage points for teachers in grades one and two and from 14 to 11 percentage points in grades six and seven.

We also investigated the net impact of eliminating testing in grades one and two on the distribution of school quality in elementary school grades four and five. If eliminating testing in the early elementary grades reduces the incentive for high quality teachers to leave low-performing schools, it could reduce inter-school variation in quality. However, consistent with the mobility analysis, we find no meaningful change in the distribution of school quality when testing in grades one and two was eliminated.

References

- Barnes, G., Crowe, E., & Schaefer, B. (2007). *The Cost of Teacher Turnover in Five School Districts: A Pilot Study*. Washington, DC: National Commission on Teaching and America's Future.
- Birkeland, S. E., & Curtis, R. (2006). *Ensuring the support and development of new teachers in the Boston Public Schools: A proposal to improve teacher quality and retention*. Boston: Boston Public Schools.
- Boyd, D., Lankford, H., & Loeb, S. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, 36(1): 88-111.
- Clotfelter, C., Ladd, H., Vigdor, J. and Aliaga, R. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management* 23(2): 251-71.
- Dizon-Ross, R. (2018). *How does School Accountability Affect Teachers? Evidence from New York City*. (NBER Working Paper No. 24658). Retrieved from National Bureau of Economic Research.
- Feng, L., Figlio, D. N., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics* 103: 1-17.
- Feng, L., & Sass, T. (2017). Teacher Quality and Teacher Mobility. *Education Finance and Policy* 12(3): 396-418.
- Fuller, S. C. & Ladd, H. F. (2013). School-Based Accountability and the Distribution of Teacher Quality Across Grades in Elementary School. *Education Finance and Policy* 8(4): 528-559.
- Gray, L., & Taie, S. (2015). *Public School Teacher Attrition and Mobility in the First Five Years: Results from the First Through Fifth Waves of the 2007–08 Beginning Teacher Longitudinal Study (NCES 2015-337)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Grissom, J. A., Kalogrides, D. & Loeb, S. (2017). Strategic Staffing? How Performance Pressures Affect the Distribution of Teachers Within Schools and Resulting Student Achievement. *American Educational Research Journal* 54(6): 1079-1116.
- Milanowski, A. T., & Odden, A. R. (2007). *A new approach to the cost of teacher turnover*. Washington, DC: School Finance Redesign Project, Center on Reinventing Public Education.

- Owens, S. J. (2015). Georgia's Teacher Dropout Crisis: A Look at Why Nearly Half of Georgia Public School Teachers are Leaving the Profession. Atlanta, GA: Georgia Department of Education.
- Perdue, S. & D. Dlugolenski (2010). Budget in Brief: Amended Fiscal Year 2010 and Fiscal Year 2011. Atlanta, GA: Governor's Office of Planning and Budget.
- Ronfledt M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal* 50(1): 4-36.
- Sun, M., Saultz, A., & Ye, Y. (2017). Federal policy and the teacher labor market: Exploring the effects of NCLB school accountability on teacher turnover. *School Effectiveness and School Improvement* 28(1): 102-122.

Figures and Tables

Figure 1.A: Percentage of Teachers Leaving the Profession By Grade and Year

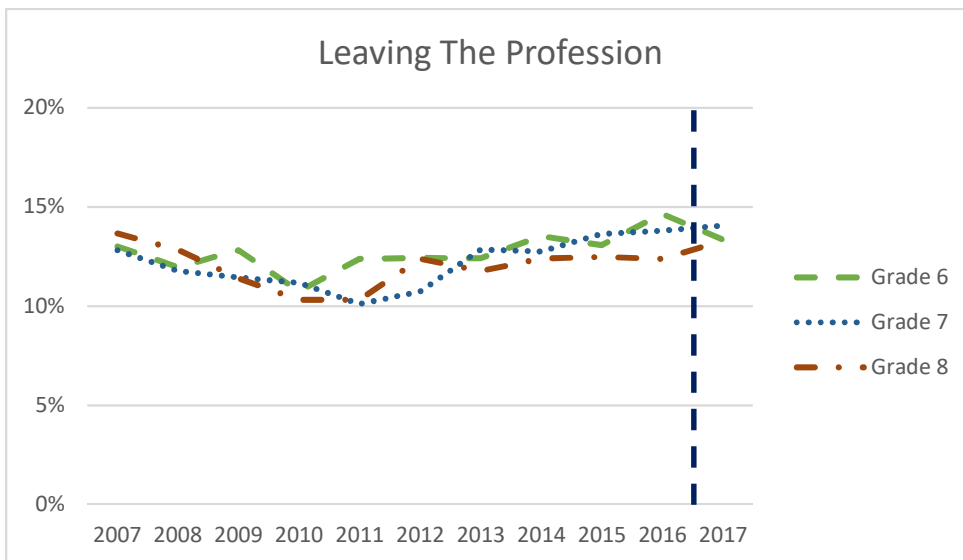
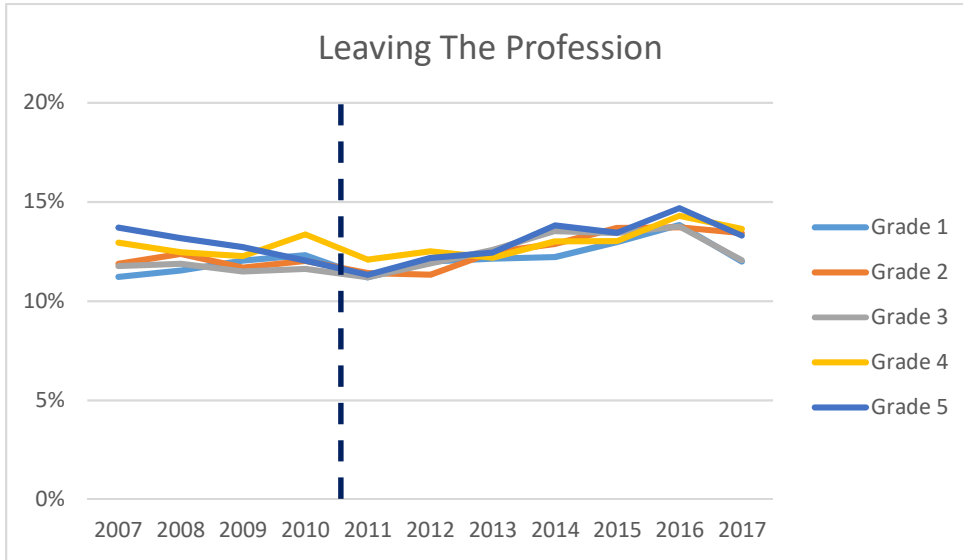


Figure 1.B: Percentage of Teachers Changing Districts By Grade and Year

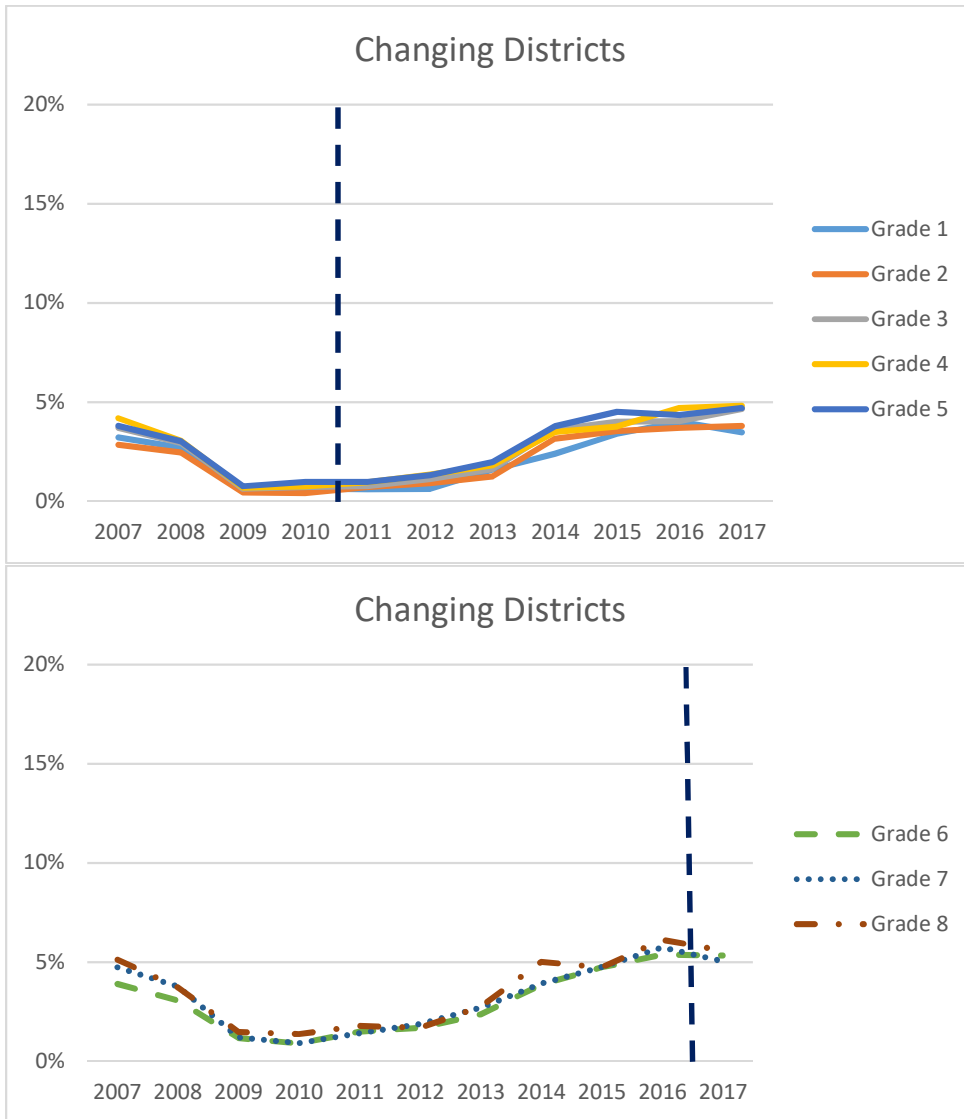


Figure 1.C: Percentage of Teachers Changing Schools Within District By Grade and Year

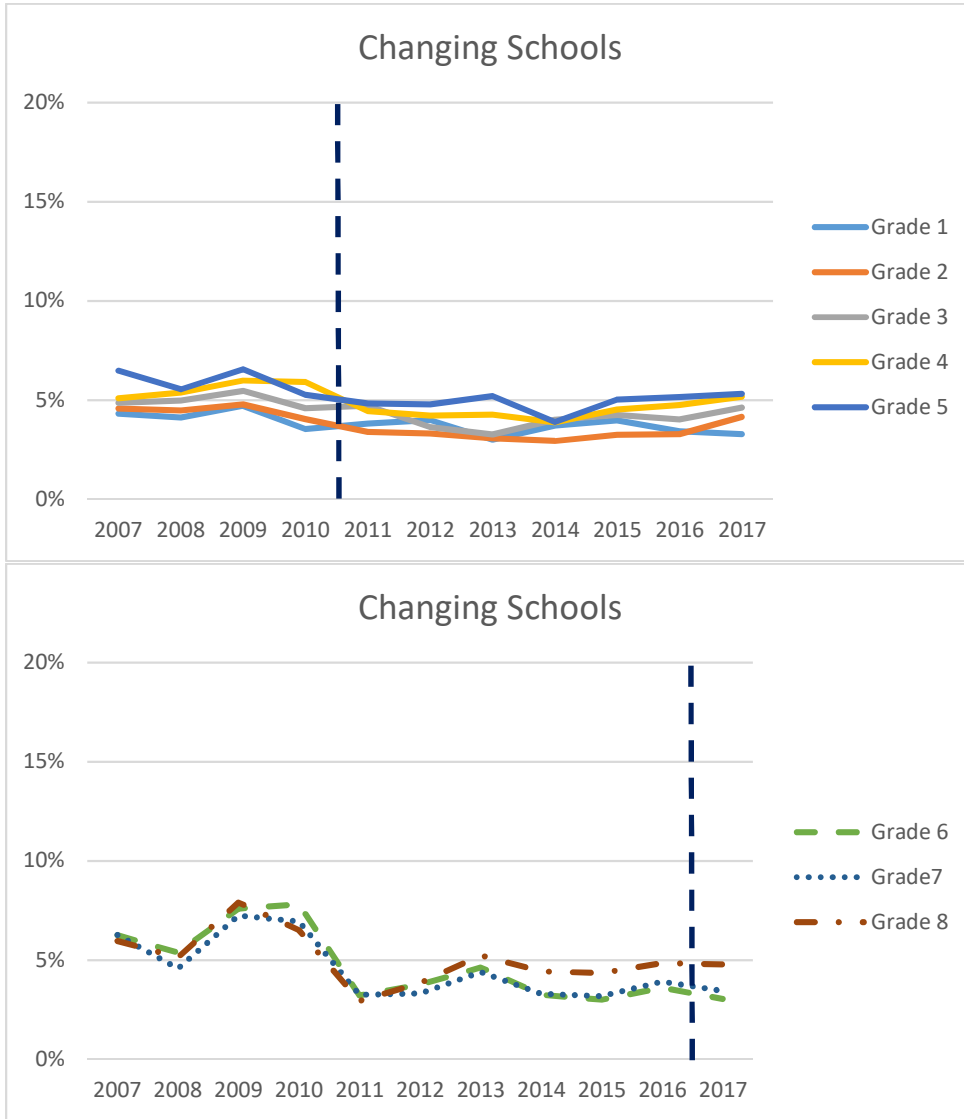


Figure 1.D: Percentage of Teachers Changing Grades Within School By Grade and Year

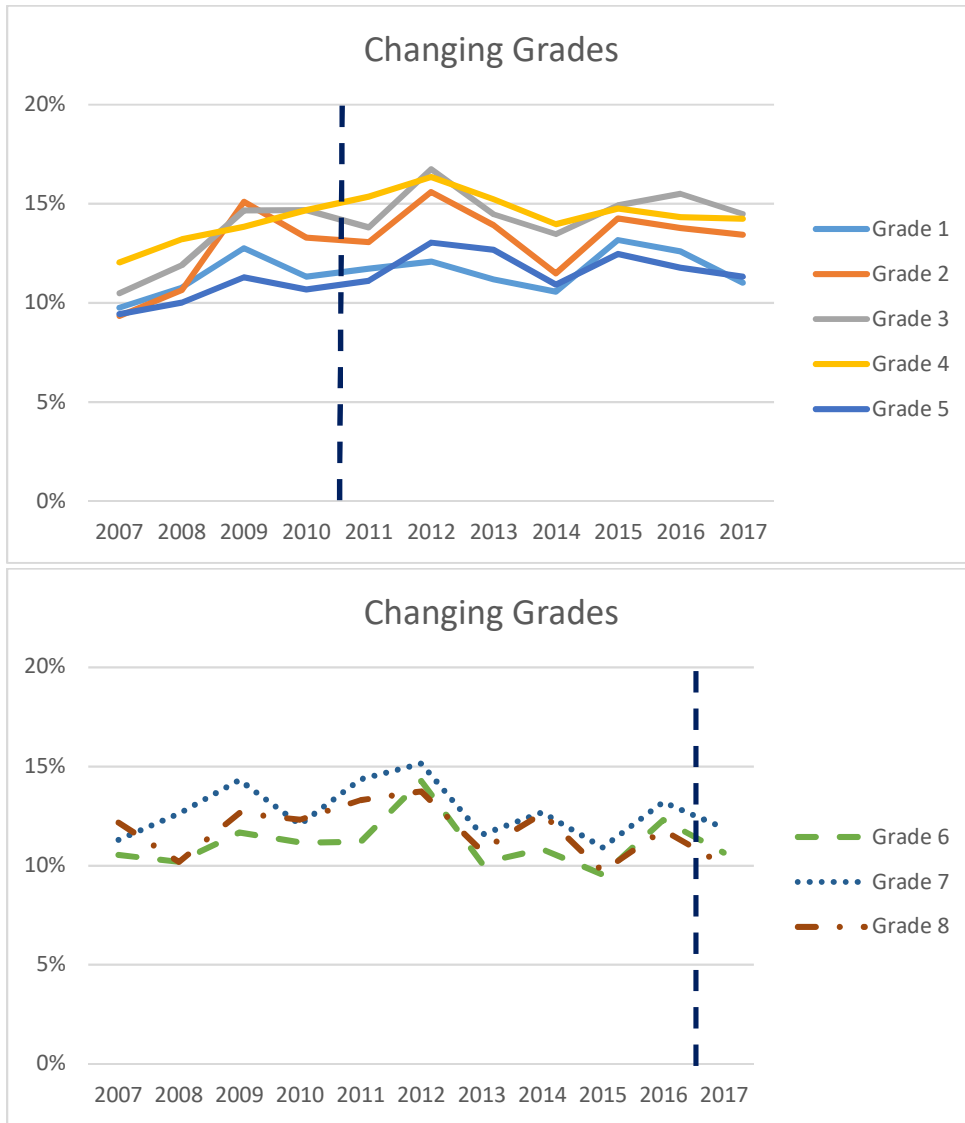


Figure 2A: Distribution ELA School Value-Added

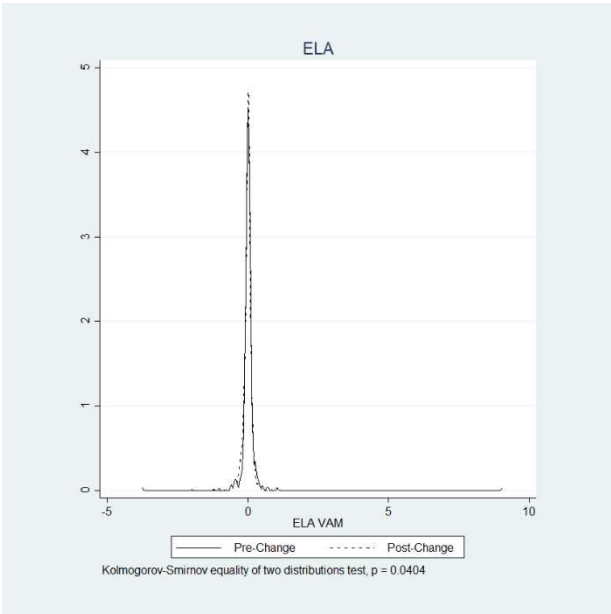


Figure 2B: Distribution of Math School Value-Added

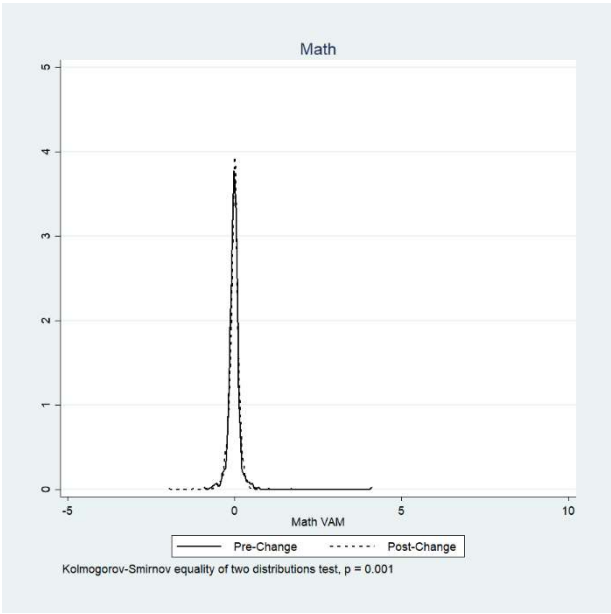


Figure 3A: Leads and Lags Estimated Effects of the 2011 Testing Change

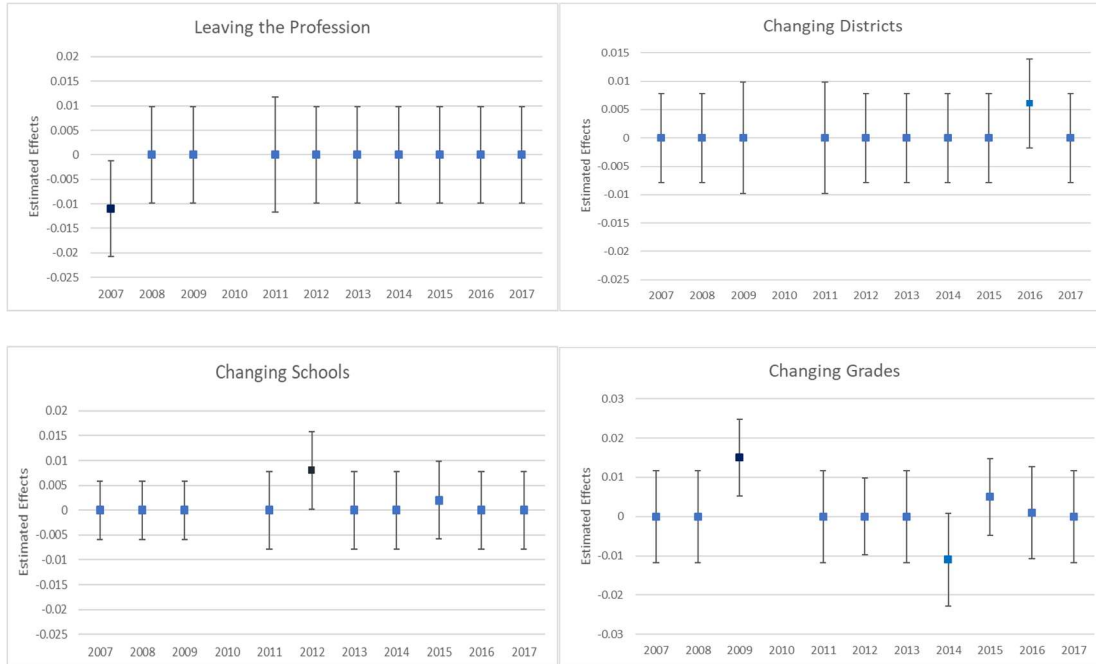


Figure 3B: Leads and Lags Estimated Effects of the 2017 Testing Change

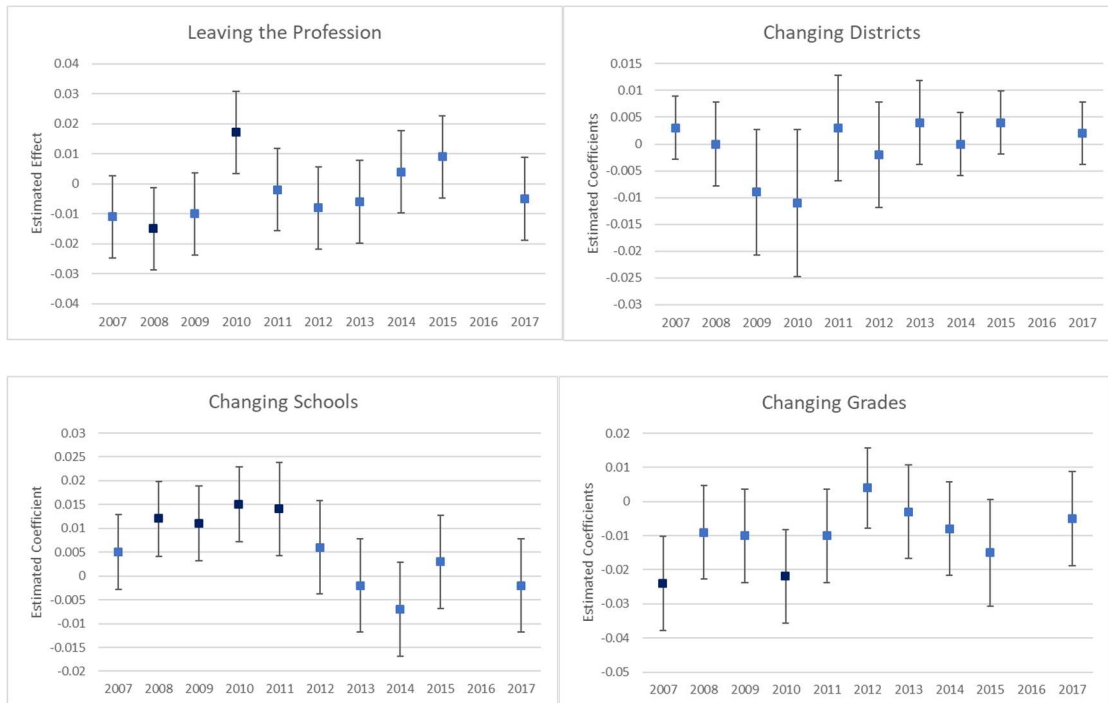


Table 1. Tested Grades and Subjects, 2007-2018

Grade	Subject	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	ELA	x	x	x	x								
	Math	x	x	x	x								
	Reading	x	x	x	x								
	Science												
	Social Studies												
2	ELA	x	x	x	x								
	Math	x	x	x	x								
	Reading	x	x	x	x								
	Science												
	Social Studies												
3	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x		
	Social Studies	x	x	x	x	x	x	x	x	x	x		
4	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x		
	Social Studies	x	x	x	x	x	x	x	x	x	x		
5	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x	x	x
	Social Studies	x	x	x	x	x	x	x	x	x	x	x	x
6	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x		
	Social Studies	x	x	x	x	x	x	x	x	x	x		
7	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x		
	Social Studies	x	x	x	x	x	x	x	x	x	x		
8	ELA	x	x	x	x	x	x	x	x	x	x	x	x
	Math	x	x	x	x	x	x	x	x	x	x	x	x
	Reading	x	x	x	x	x	x	x	x				
	Science	x	x	x	x	x	x	x	x	x	x	x	x
	Social Studies	x	x	x	x	x	x	x	x	x	x	x	x

Table 2: Teacher Sample Descriptive Statistics

	<i>All teachers</i>		<i>Early-Career Teachers</i>	
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Standard Deviation</i>
<i>Turnover</i>				
Leaving Teaching	0.125	0.330	0.140	0.347
Changing District	0.026	0.160	0.049	0.216
Changing School	0.045	0.208	0.059	0.235
Changing Grades	0.126	0.331	0.143	0.350
<i>Demographics</i>				
Beginning Teacher	0.199	0.399		
Hispanich Teacher	0.011	0.106	0.018	0.132
Asian Teacher	0.008	0.088	0.014	0.117
African American	0.238	0.426	0.279	0.449
White Teacher	0.753	0.431	0.703	0.457
Other ethnicity	0.001	0.029	0.001	0.036
<i>Education/ Certification</i>				
Associates Degree	0.000	0.002	0.000	0.005
Bachelors Degree	0.252	0.434	0.392	0.488
Master Degree	0.445	0.497	0.430	0.495
Education Specialist	0.266	0.442	0.146	0.353
Doctorate	0.028	0.166	0.013	0.113
Special Education	0.074	0.262	0.099	0.299
ESL certified	0.078	0.269	0.085	0.279
Gifted certified	0.224	0.417	0.105	0.306
Georgia Education Degree	0.001	0.035	0.002	0.046
<i>School Characteristics</i>				
Percentage Hispanic	0.143	0.160	0.165	0.187
Percentage Asian	0.033	0.056	0.035	0.061
Percentage African American	0.373	0.292	0.436	0.309
Percentage White	0.504	0.287	0.421	0.291
Percentage Other Ethnicity	0.003	0.004	0.003	0.005
Percentage Free-Reduced Lunch	0.585	0.258	0.623	0.265
Percentage English Language Learners	0.081	0.124	0.098	0.145
Percentage with disabilities	0.127	0.040	0.125	0.040
Enrollment	815.945	611.188	816.409	530.115

Note: Number of observations = 453,955 teachers in total with 91,529 early-career teachers

Table 3: Average Marginal Effect Estimates from Multinomial Logit Hazard Models

Panel A: Stop Testing for Grades 1 and 2 in 2011-Teachers from Grades 1 to 5

	<i>Leaving</i>		<i>Changing District</i>		<i>Changing School</i>		<i>Changing Grade</i>	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>Treated Grades</i>	-0.006*** (0.002)	-0.006*** (0.002)	-0.005*** (0.001)	-0.005*** (0.001)	-0.009*** (0.001)	-0.009*** (0.001)	-0.006*** (0.002)	-0.006*** (0.002)
<i>After Change</i>	-0.001 (0.002)	-0.000 (0.002)	0.005*** (0.001)	0.005*** (0.001)	-0.011*** (0.001)	-0.011*** (0.001)	0.014*** (0.002)	0.014*** (0.002)
<i>Treated*After Change</i>	0.001 (0.002)	0.003 (0.003)	0.000 (0.001)	-0.001 (0.001)	-0.000 (0.002)	-0.001 (0.002)	-0.005* (0.003)	-0.005** (0.003)
<i>Beginning Teacher</i>	-0.001 (0.002)	0.001 (0.002)	0.018*** (0.001)	0.017*** (0.001)	0.015*** (0.001)	0.015*** (0.001)	0.022*** (0.002)	0.021*** (0.002)
<i>Beginning Teacher *Treated* After Change</i>		-0.009** (0.004)		0.004*** (0.001)		0.002 (0.002)		0.005 (0.004)

Note: Number of observations = 295,564. Estimates include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Panel B: Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Leaving</i>		<i>Changing District</i>		<i>Changing School Within District</i>		<i>Changing Grade Within School</i>	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>Treated Grades</i>	-0.007*** (0.002)	-0.007*** (0.002)	-0.003** (0.001)	-0.003** (0.001)	-0.002* (0.001)	-0.002* (0.001)	-0.009*** (0.002)	-0.009*** (0.002)
<i>After Change</i>	0.002 (0.004)	0.002 (0.004)	0.016*** (0.002)	0.016*** (0.002)	-0.010*** (0.003)	-0.010*** (0.003)	-0.017*** (0.004)	-0.017*** (0.004)
<i>Treated × After Change</i>	-0.000 (0.007)	0.006 (0.007)	0.000 (0.003)	0.002 (0.004)	-0.008 (0.005)	-0.009 (0.006)	0.006 (0.007)	0.003 (0.008)
<i>Beginning Teacher</i>	0.002 (0.002)	0.003 (0.002)	0.025*** (0.001)	0.025*** (0.001)	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)
<i>Beginning Teacher × Treated × After Change</i>		-0.028** (0.013)		-0.006 (0.005)		0.004 (0.009)		0.008 (0.013)

Note: Number of observations = 129,033. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Table 4: Average Marginal Effects from a Logit Hazard Model for the Probability of Changing Grades

Panel A: Stop Testing for Grades 1 and 2 in 2011-Teachers from Grades 1 to 5

	<i>Changing Grades</i>	
<i>Treated Grades</i>	-0.111*** (0.014)	-0.111*** (0.014)
<i>After Change</i>	0.085*** (0.012)	0.085*** (0.012)
<i>Treated*After Change</i>	-0.003 (0.017)	-0.005 (0.018)
<i>Beginning Teacher</i>	0.220*** (0.011)	0.217*** (0.012)
<i>Beginning Teacher × Treated × After Change</i>		0.013 (0.025)

Note: Number of observations = 295,563. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Panel B: Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Changing Grades</i>	
<i>Treated Grades</i>	-0.115*** (0.016)	-0.115*** (0.016)
<i>After Change</i>	0.058** (0.027)	0.058** (0.027)
<i>Treated × After Change</i>	-0.007 (0.046)	0.016 (0.050)
<i>Beginning Teacher</i>	0.211*** (0.017)	0.214*** (0.017)
<i>Beginning Teacher × Treated × After Change</i>		-0.099 (0.087)

Note: Nobs=129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table 5: Summary Statistics of School Quality

	ELA		Math	
	Pre-Policy Change 2008-2010	Post-Policy Change 2011-2016	Pre-Policy Change 2008-2010	Post-Policy Change 2011-2016
5 th Percentile	-0.1791	-0.2622	-0.2517	-0.2652
10 th Percentile	-0.1178	-0.1691	-0.1760	-0.1846
25 th Percentile	-0.0484	-0.0684	-0.0965	-0.0783
Median	0.0027	-0.0005	-0.0244	-0.0009
75 th Percentile	0.0653	0.0530	0.0455	0.0654
90 th Percentile	0.1392	0.1250	0.1232	0.1499
95 th Percentile	0.2322	0.1698	0.1823	0.2099
Mean	0.0158	-0.0174	-0.0195	-0.0114
Std. Dev	0.4338	0.1557	0.2303	0.1660
Number of Observations	579	1845	581	1858

Table 6: Average Marginal Effect Estimates from Multinomial Logit Hazard Models- Difference-in-Differences-in-Differences Approach. Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 5 to 8

	<i>Leaving</i>		<i>Changing District</i>		<i>Changing School</i>		<i>Changing Grade</i>	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>Treated Grades</i>	0.003 (0.003)	0.003 (0.003)	-0.006*** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	0.006* (0.003)	0.006* (0.003)
<i>Science and Social Science Teachers</i>	-0.011*** (0.004)	-0.011*** (0.004)	-0.003 (0.002)	-0.003 (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	-0.000 (0.004)	-0.000 (0.004)
<i>Treated Grades *Science and Social Science</i>	0.000 (0.004)	0.000 (0.004)	0.003 (0.002)	0.003 (0.002)	0.006** (0.003)	0.006** (0.003)	-0.012*** (0.004)	-0.012*** (0.004)
<i>After Change</i>	-0.002 (0.008)	-0.002 (0.008)	0.012*** (0.004)	0.012*** (0.004)	-0.014** (0.006)	-0.014** (0.006)	-0.022** (0.009)	-0.022** (0.009)
<i>After Change*Treated Grades</i>	0.007 (0.009)	0.007 (0.009)	0.006 (0.004)	0.006 (0.004)	-0.005 (0.007)	-0.005 (0.007)	0.010 (0.010)	0.010 (0.010)
<i>After Change*Science and Social Science</i>	0.005 (0.011)	0.005 (0.011)	0.007 (0.005)	0.007 (0.005)	0.022*** (0.008)	0.022*** (0.008)	-0.000 (0.012)	-0.000 (0.012)
<i>After Change*Treated Grade* Science and Social Science</i>	-0.007 (0.014)	-0.002 (0.014)	-0.008 (0.006)	-0.006 (0.006)	-0.021** (0.010)	-0.022** (0.010)	0.001 (0.015)	-0.001 (0.015)
<i>Beginning Teacher</i>	0.002 (0.002)	0.002 (0.002)	0.025*** (0.001)	0.025*** (0.001)	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	0.011*** (0.002)
<i>Beginning Teacher *After*Treated Grade* Science and Social Science</i>		-0.028** (0.013)		-0.006 (0.005)		0.003 (0.009)		0.008 (0.013)

Note: Number of observations = 129,033. Estimates include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

APPENDIX TABLES

Table A.1: Average Marginal Effect Estimates from Multinomial Logit Models Restricting to Years of Common Support (2012 onwards)-Grades 6 to 8

	<i>Leaving</i>		<i>Changing District</i>		<i>Changing School</i>		Changing Grade	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>Treated Grades</i>	-0.004 (0.003)	-0.004 (0.003)	-0.002 (0.002)	-0.002 (0.002)	-0.008*** (0.002)	-0.008*** (0.002)	-0.004 (0.003)	-0.004 (0.003)
<i>After Change</i>	0.002 (0.004)	0.002 (0.004)	0.011*** (0.002)	0.011*** (0.002)	-0.004 (0.002)	-0.004 (0.002)	-0.016*** (0.004)	-0.016*** (0.004)
<i>Treated*After Change</i>	-0.004 (0.007)	0.001 (0.008)	-0.001 (0.004)	0.003 (0.004)	-0.001 (0.004)	-0.002 (0.005)	-0.000 (0.007)	-0.002 (0.008)
<i>Beginning Teacher</i>	-0.003 (0.003)	-0.002 (0.004)	0.033*** (0.002)	0.033*** (0.002)	0.007*** (0.002)	0.007*** (0.002)	0.013*** (0.003)	0.013*** (0.004)
<i>Beginning Teacher *Treated* After Change</i>		-0.023* (0.014)		-0.010 (0.006)		0.005 (0.008)		0.006 (0.013)

Note: Number of observations = 68,244. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Table A.2: Average Marginal Effects from a Logit Hazard Model for the Probability of Leaving Teaching

Panel A: Stop Testing for Grades 1 and 2 in 2011-Teachers from Grades 1 to 5

	<i>Leaving</i>	
<i>Treated Grades</i>	-0.006*** (0.002)	-0.006*** (0.002)
<i>After Change</i>	-0.0007 (0.002)	-0.0006 (0.002)
<i>Treated*After Change</i>	0.0009 (0.002)	0.002 (0.002)
<i>Beginning Teacher</i>	-0.003* (0.001)	-0.0006 (0.002)
<i>Beginning Teacher *Treated* After Change</i>		-0.009** (0.004)

Note: Number of observations = 295,563. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Panel B: Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Leaving</i>	
<i>Treated Grades</i>	-0.007*** (0.002)	-0.007*** (0.002)
<i>After Change</i>	0.0015 (0.004)	0.0015 (0.004)
<i>Treated*After Change</i>	-0.0003 (0.007)	0.006 (0.007)
<i>Beginning Teacher</i>	0.0004 (0.002)	0.001 (0.002)
<i>Beginning Teacher *Treated* After Change</i>		-0.028** (0.013)

Note: Nobs=129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table A.3: Average Marginal Effects from a Logit Hazard Model for the Probability of Changing Districts Without Changing Grades

Panel A: Stop Testing for Grades 1 and 2 in 2011-Teachers from Grades 1 to 5

	<i>Changing District</i>	
<i>Treated Grades</i>	-0.002*** (0.0005)	-0.002*** (0.0005)
<i>After Change</i>	0.001*** (0.0004)	0.001*** (0.0004)
<i>Treated*After Change</i>	-0.0009 (0.0007)	-0.001** (0.0007)
<i>Beginning Teacher</i>	0.006*** (0.0004)	0.005*** (0.0004)
<i>Beginning Teacher *Treated* After Change</i>		0.002** (0.0008)

Note: Number of observations = 295,563. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Panel B: Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Changing District</i>	
<i>Treated Grades</i>	-0.0007 (0.0006)	-0.0007 (0.0006)
<i>After Change</i>	0.003*** (0.0009)	0.003*** (0.0009)
<i>Treated*After Change</i>	0.002 (0.001)	0.003* (0.002)
<i>Beginning Teacher</i>	0.006*** (0.0006)	0.006*** (0.0006)
<i>Beginning Teacher *Treated* After Change</i>		-0.003 (0.002)

Note: Nobs=129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table A.4: Average Marginal Effects from a Logit Hazard Model for the Probability of Changing Schools Without Changing Grades

Panel A: Stop Testing for Grades 1 and 2 in 2011-Teachers from Grades 1 to 5

	<i>Changing School</i>	
<i>Treated Grades</i>	-0.003*** (0.0007)	-0.003*** (0.0007)
<i>After Change</i>	-0.007*** (0.0007)	-0.007*** (0.0007)
<i>Treated*After Change</i>	0.00002 (0.001)	0.0002 (0.001)
<i>Beginning Teacher</i>	0.006*** (0.0006)	0.006*** (0.0007)
<i>Beginning Teacher *Treated* After Change</i>		-0.001 (0.002)

Note: Number of observations = 295,563. Models include controls for teacher and school characteristics. Standard errors clustered by teacher are presented in parentheses.

Panel B: Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Changing School</i>	
<i>Treated Grades</i>	0.002*** (0.008)	0.002*** (0.008)
<i>After Change</i>	-0.015*** (0.002)	-0.015*** (0.002)
<i>Treated*After Change</i>	-0.004 (0.004)	-0.006 (0.005)
<i>Beginning Teacher</i>	0.004*** (0.0009)	0.004*** (0.0009)
<i>Beginning Teacher *Treated* After Change</i>		0.007 (0.007)

Note: Nobs=129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table A.5: Average Marginal Effects from a Logit Hazard Model for the Probability of Leaving Teaching-Difference-in-Difference-in-Differences

Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Leaving</i>	
<i>Treated Grades</i>	0.003 (0.003)	0.003 (0.003)
<i>Science and Social Science Teachers</i>	-0.011*** (0.003)	-0.011*** (0.003)
<i>Treated Grades *Science and Social Science</i>	-0.000 (0.004)	-0.000 (0.004)
<i>After Change</i>	-0.003 (0.008)	-0.003 (0.008)
<i>After Change*Treated Grades</i>	0.007 (0.009)	0.007 (0.009)
<i>After Change*Science and Social Science</i>	0.004 (0.011)	0.004 (0.011)
<i>After Change*Treated Grade* Science and Social Science</i>	-0.007 (0.013)	-0.008 (0.014)
<i>Beginning Teacher</i>	0.0002 (0.002)	0.001 (0.002)
<i>Beginning Teacher *After*Treated Grade* Science and Social Science</i>		-0.028** (0.013)

Note: Nobs= 129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table A.6: Average Marginal Effects from a Logit Hazard Model for the Probability of Changing Districts Without Changing Grades-Difference-in-Difference-in-Differences Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Changing District</i>	
<i>Treated Grades</i>	-0.001** (0.0008)	-0.001** (0.0008)
<i>Science and Social Science Teachers</i>	-0.0007 (0.0009)	-0.0007 (0.0009)
<i>Treated Grades *Science and Social Science</i>	0.0006 (0.001)	0.0006 (0.001)
<i>After Change</i>	0.0002 (0.002)	0.0002 (0.002)
<i>After Change*Treated Grades</i>	0.003 (0.002)	0.003 (0.002)
<i>After Change*Science and Social Science</i>	0.004 (0.003)	0.004 (0.003)
<i>After Change*Treated Grade* Science and Social Science</i>	-0.002 (0.003)	-0.001 (0.003)
<i>Beginning Teacher</i>	0.006*** (0.0006)	0.006*** (0.0006)
<i>Beginning Teacher *After*Treated Grade* Science and Social Science</i>		-0.003 (0.002)

Note: Nobs= 129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis

Table A.7: Average Marginal Effects from a Logit Hazard Model for the Probability of Changing Schools Without Changing Grades--Difference-in-Difference-in-Differences Stop Testing for Science/Social Science Teachers Grades 6 and 7 in 2017-Teachers from Grades 6 to 8

	<i>Changing School</i>	
<i>Treated Grades</i>	0.001 (0.001)	0.001 (0.001)
<i>Science and Social Science Teachers</i>	0.0005 (0.001)	0.0005 (0.001)
<i>Treated Grades *Science and Social Science</i>	0.001 (0.002)	0.001 (0.002)
<i>After Change</i>	-0.010** (0.005)	-0.010** (0.005)
<i>After Change*Treated Grades</i>	-0.015** (0.006)	-0.015** (0.006)
<i>After Change*Science and Social Science</i>	0.006 (0.006)	0.006 (0.006)
<i>After Change*Treated Grade* Science and Social Science</i>	0.0004 (0.008)	-0.002 (0.009)
<i>Beginning Teacher</i>	0.004*** (0.001)	0.004*** (0.001)
<i>Beginning Teacher *After*Treated Grade* Science and Social Science</i>		0.007 (0.007)

Note: Nobs= 129,032; Estimates include controls for teacher and school characteristics; Clustered standard errors by teacher presented in parenthesis