Disparate Impacts of Teacher Certification Exams

[Preliminary draft — Please do not cite or distribute]

Christa Deneault*

Evan Riehl†

Jian Zou‡

September 25, 2025

Abstract

We use Texas administrative data to assess the long-standing claim that teacher certification exams have a disparate impact on prospective teachers from underrepresented minority (URM) groups. In a regression discontinuity design, we find that failing a certification exam delays entry into teaching, ultimately costing the average candidate who marginally fails $10,000 in forgone earnings. These costs fall disproportionately on URM teaching candidates because they are both more likely to fail and have more negative causal effects of failing. To assess the predictive validity of the exams for teaching quality, we exploit a reform that significantly increased the difficulty of Texas' elementary certification exam. We find that the distribution of teacher value-added among those who passed the easier pre-reform exam is similar to that for those who passed the harder post-reform exam, showing that the harder exam was not any better at screening out low-quality teachers. Taken together, our findings show that certification exams have a disparate impact on URMs in the sense that they impose disproportionate economic costs that are not justified by improvements in teaching quality.

*Deneault: Federal Reserve Bank of Dallas (email: christa.deneault@dal.frb.org)

†Riehl: Department of Economics and ILR School, Cornell University (email: eriehl@cornell.edu)

‡Zou: ILR School, Cornell University (email: jz2326@cornell.edu)

The views expressed in this article are those of the authors and do not necessarily reflect the opinions or official position of the Federal Reserve Bank of Dallas, Federal Reserve System, Texas Education Agency, the Texas Higher Education Coordinating Board, Texas Workforce Commission, or the state of Texas. This research was supported by a grant from the Spencer Foundation. We are grateful to Holly Kosiewicz, Zhixiang Lu, and members of the Texas Education Research Center Advisory Board for help in accessing and working with the data. All errors are our own.

1 Introduction

Economists have extensively examined labor market discrimination from the perspective of individual agents, such as firms engaging in taste-based or statistical discrimination.[1] By contrast, relatively little attention has been paid to how systemic or institutional features of labor markets may interact to generate unequal outcomes across groups of prospective workers (Bohren et al., 2025b; Small and Pager, 2020). Occupational licensing provides a salient example: certification exams are a prerequisite for entry into many licensed occupations, yet differences in pre-market test-taking abilities between underrepresented minorities (URMs) and white candidates could lead to disparities in access.[2] Even if such disparities exist, it is difficult to conclude that they constitute discrimination, since exams may reflect skills relevant to productivity. Moreover, quantifying the effects of barriers relative to productivity gains remains an empirical challenge in part because outcomes, such as earnings, are not always available based on *intent* to become licensed but rather exist by current licensing status.

We address these challenges to provide new evidence of the disparate impacts of teacher license exams, a requirement of certification, on URMs (Blacks and Hispanics). We utilize rich administrative data from the state of Texas that allow us to examine outcomes not previously considered. We leverage multiple quasi-experimental designs, including a certification exam regime change in Texas that made some licensing exams relatively more difficult and reduced license exam pass rates. This certification reform motivates our development of a *policy-relevant* disparate impacts measure which quantifies the difference in pass rates between URM and white candidates of equal productivity—where productivity is defined by teacher value added. By combining causal evidence on the costs of failing exams with evidence on productivity, we provide the best evidence to date that certification exams place a disproportionate burden on URM prospective teachers relative to white prospective teachers of comparable productivity.

We first begin by providing new evidence on the costs of failing license exams for prospective teachers. Using a regression discontinuity (RD) design, we find that narrowly failing a license exam delays entry into the teaching profession for both whites and URMs, with larger effects for URMs. In the year following their first certification exam attempt, whites are 9 percentage points (pp) less likely to be working in a public school while URMs are

[1] For recent examples see Bohren et al. (2025a), Kline et al. (2022), Benson et al. (2024), Lepage (2024), among others.

[2] Several papers document differences in test scores in k-12 or college standardized exams. For example, Fryer Jr and Levitt (2006) demonstrates that black-white test score gaps begin to diverge as soon as third grade. Recently Chetty et al. (2023) show SAT scores are correlated with parental income.

15pp less likely if they narrowly fail. Although these employment effects decline over time as candidates retake exams, URMs remain 3pp less likely to be employed in Texas public schools ten years later. Because of these delaying and discouraging effects, we find that narrowly failing an exam imposes a cost to long-term cumulative earnings: whites earn $7,800 less and URMs earn $12,200 less over a ten year period.[3] Importantly, URM candidates are also more likely to experience these earnings losses since they have much lower first-attempt pass rates than White candidates (57 percent vs. 83 percent).

Next, we develop a research design that shows how these individual effects aggregate into compositional changes in the teacher workforce. We utilize a 2010 certification exam reform in Texas that increased the difficulty of passing a certification exam for prospective elementary teachers relative to middle school using differences-in-differences (DiD). We combine this with a shock to schools' labor supply through teacher departures (RD) to isolate how demand for new teachers affects student achievement under easier or harder certification exam regimes. Our RD-DiD analysis shows when schools have a vacancy under stricter licensing regime, they employ 2.5pp fewer new URM teachers, offsetting them with white teachers. This compositional shift also results in a redistribution of income: white teachers collectively earn, on average, $6,000 more under the stricter exam regime, while URM teachers earn collectively less.

Although certification is costly, its expense could be justified if it is a key predictor of quality—yet we find no evidence of gains in student achievement from increased license exam stringency. Using our RD-DiD strategy, which incorporates supply-side compositional effects, we find essentially no improvement in student achievement when licensing exams become harder. We additionally show that the distributions of value added among teachers who were certified under differentially difficult certification exam regimes are nearly identical. Thus, certification exams are not necessarily pushing out relatively low-performing teachers. We complement this finding by demonstrating that value added is weakly correlated with certification exam scores and narrowly failing their first certification exam is essentially unrelated to value added 2-10 years later.

To quantify the effects of the disproportionate costs of licensing for equally productive prospective teachers, we create a *policy-relevant* disparate impacts measure. We adapt Arnold et al. (2020)'s estimate of disparate impacts to utilize the quasi-variation provided by our certification exam reform. Our policy-relevant disparate impacts measure relies on two assumptions that we believe are plausible in our context. First, it requires anyone who

[3]These earnings are the present value of cumulative earnings over a 10 year period. This is estimated as the sum of discounted nominal earnings through for the 10 years following taking the exam using a 5% discount rate.

passed a harder certification exam could have also passed an easier certification exam. Second, latent potential value-added (i.e., value-added for all individuals regardless of whether we observe it) does not change systematically with the timing of the certification reform. Our policy-relevant disparate impacts measure indicates that among those who would have passed their certification exams during the easier exam regime, an URM prospective teacher is 13pp (10pp) less likely to pass the harder exam than white exam takers with the same math (ELA) value-added.

We make three contributions to the existing literature. First, we estimate both the costs to failed access to the teaching profession and the heterogeneous costs across race and ethnicity. Some work with respect to teachers has explicitly studied how certification stringency changes heterogeneously affect subpopulations' supply but have found mixed outcomes. Law and Marks (2009) find an increase in Black participation following certification, Larsen et al. (2020) and Law et al. (2023) find no impact of changes to licensing stringency on diversity of the teacher workforce, and Angrist and Guryan (2008) and Chung and Zou (2025) find it leads to fewer diverse teachers. Other work has examined the wage gains for licensed workers by race and finds that the earnings gap shrinks for licensed workers (e.g., Blair and Chung, 2025; Yan, 2024; Xia, 2021). However, these papers lack information about the costs of licensing barriers, and comparing wages of licensed to unlicensed workers (or even an increase post-license) has a selection issue: those who want to become licensed may be different than those who do not. Our RD design circumvents selection issues by examining teachers who narrowly fail the exam to show certification exams are a costly barrier in terms of employment and earnings, particularly for URMs. We also additionally add to the literature on effects of certification stringency on supply by showing the composition of teachers changes in favor of whites when license exams become more difficult.

Second, we provide new and more convincing evidence that certification exams are unrelated to value added and that changes to certification exam stringency have little impact on student achievement on net (including supply side effects). Previous work has correlated value added with certification exams in an attempt to address whether it is predictive of teacher productivity and has found that it is weakly correlated (Goldhaber, 2007; Aaronson et al., 2007; Kane et al., 2008; Clotfelter et al., 2010; Goldhaber and Hansen, 2010; Shuls and Trivitt, 2015; Hendricks, 2016; Goldhaber et al., 2017; Cowan et al., 2020).[4] But this

[4]Only a subset of these papers consider that certification exams may be differentially predictive across race/ethnicity. In all cases the finding is a weak correlation between value added and certification. But there is mixed evidence about the predictiveness across race/ethnicity. For example, Goldhaber and Hansen (2010) finds license exams are more predictive of student test achievement for white teachers than Black teachers, while Cowan et al. (2020) finds the opposite. We find that certification is weakly predictive for whites but not predictive for URMs.

approach falls short of observing all teachers who initially fail and are discouraged from the profession, does not take into consideration potential supply-side impacts of changes to exam stringency, and may simply represent a lack of correlation among those who passed. In a different approach using RD, Orellana and Winters (2023) find failing a certification pushes out those who would have been *better* teachers.[5] Their method requires strong assumptions about value added for those who never enter the teaching profession, and they do not have information on race. We add to this literature by examining a certification exam reform, which allows us to rule out meaningful differences in the value added across teachers who pass differentially difficult exams. Our RD-DiD measure also incorporates potential supply-side adjustments and points to similar evidence: harder certification exams did not improve student achievement yet resulted in a less diverse workforce.

Finally, we contribute to a burgeoning literature on quantifying disparate impacts by defining a policy-relevant disparate impacts measure. Recent work has measured disparate treatment in bail decisions, foster care, and policing, among others (e.g., Arnold et al., 2020; Baron et al., 2024; Goncalves and Mello, 2021). We contribute to this literature by modifying and extending the work from Arnold et al. (2020), which we believe can be used as a rubric for future work where population-level quasi-experimental variation is not available. Instead, our measure relies on changes in subpopulations affected by a policy change, a situation common in other settings.

2 Background and Data

2.1 TExES certification exams. The Texas Education Agency (TEA) has implemented state-specific requirements for selecting prospective teachers. Since the 2002–2003 academic year, the primary certification requirement has been the Texas Examinations of Educator Standards (TExES), which is the focus of our study.[6] The TExES consists of a series of standardized assessments, administered by Pearson under the oversight of the TEA, designed to evaluate teacher candidates' subject-matter knowledge and pedagogical skills. These assessments include both content-specific exams and Pedagogy and Professional Responsibilities (PPR) exams. To become a certified teacher in Texas, a candidate must pass at least two TExES exams, including one content test aligned with the candidate's intended teaching subject and grade level and the PPR exam.[7] The TExES exams employ a scaled scoring

[5]Other work has examined quasi-experimental changes to licensing on teacher quality but cannot connect teachers to students. See Angrist and Guryan (2008), Larsen et al. (2020), and Chung and Zou (2025).

[6]Prior to 2003, Texas used the Examination for the Certification of Educators in Texas (ExCET).

[7]In addition to passing the certification exams, the candidate need to hold a bachelor's degree, complete an approved Educator Preparation Program, and complete the background check (Agency, 2022).

system, with a passing threshold set at 240. Finally, we leverage a TExES elementary exam reform between 2002 and 2015 to measure changes in licensing stringency, which we describe in detail in Section 4.1.

2.2 Texas administrative data. We use administrative data from four Texas agencies:

- **Texas Education Agency (TEA)**. TEA provides data on all public school teachers in Texas from 1996–2019. These data include each teacher's demographics, salary, years of teaching experience, school of employment, and full-time equivalent (FTE) years associated with their teaching grade(s) and subject(s). We also use TEA data on grade 3–8 state standardized tests from 1994–2019, which include student test scores and demographics.[8] We can connect student test scores to groups of teachers at the school/grade/subject level for all years from 1996–2019. From 2012–2019, we also observe classroom identifiers that allow us to connect students to individual teachers and compute teacher value-added estimates.

- **Texas State Board for Educator Certification (SBEC)**. SBEC provides data on all teacher certification exams in Texas between 1986 and 2019. This data includes the exam date, test type, test description, test score, test result, certification route, the in-state educator preparation program that approved the educator to test, and the state in which the educator was prepared. For our analysis, we restrict the sample to content tests of the TExES, the primary certification exam type since 2002, and exclude PPR exams. We use the exam date to identify the first exam taken by each candidate, ensuring that we focus on initial exam performance. The test description enables us to distinguish between specific exams, allowing us to compare certification performance and examinees' outcomes within the same exams.

- **Texas Higher Education Coordinating Board (THECB)**. We use THECB data on all bachelor's degree graduates from public (1992–2019) and private colleges (2003–2019) in Texas. We observe each graduate's college, major, and graduation year.

- **Texas Workforce Commission (TWC).** TWC provides administrative earnings records for all individuals working at registered firms in Texas from 1992–2021. We use this data to measure individuals' annual and cumulative earnings in the years before and after they transition into teaching careers. All earnings are expressed in constant 2019 dollars, deflated using the U.S. Consumer Price Index from the Bureau

[8]Texas had three testing regimes during this time period: Texas Assessment of Academic Skills (TAAS, 1994–2002), Texas Assessment of Academic Skills (TAKS, 2003–2011), and State of Texas Assessments of Academic Readiness (STAAR, 2012–2019).

of Labor Statistics. We calculate the present value (PV) of cumulative earnings as the sum of discounted nominal earnings over 1, 5, or 10 years after taking the exam, using a 5% discount rate. This measure better captures the economic cost of failing the exam by expressing the earnings loss in dollar terms at the time of the exam.

We accessed the data via the University of Texas at Dallas Education Research Center (ERC), which contains unique personal identifiers that allow us to connect all datasets at the individual level. With these links, we can follow teachers from their teacher certification exams into teaching jobs and observe their earnings regardless of whether they become teachers. Appendix B.1 provides details on the variables that are derived from the administrative data.

Table 1 presents the means for demographics, TExES exam performance, and earnings among first-time certification exam takers and comparison populations, including first-year teachers, bachelor's degree (BA) recipients, and K-12 students between 2002 and 2019. Panel A shows that the share of females is higher among exam takers and new teachers than among BA graduates and K-12 students. In terms of race composition, there are no notable differences across exam takers, new teachers, and BA graduates. Regarding high school achievement, BA graduates tend to have the highest test scores, followed by first-year teachers and then exam takers. Notably, the relatively low average high school test scores of exam takers are mainly driven by URMs. Panels B and C present results for TExES content and PPR exam performance, respectively. As expected, first-year teachers hold better exam performance (test scores and passing rates) than first-time exam takers. The comparison between race groups among exam takers is more informative: URM exam takers, on average, perform worse than their White counterparts, especially on the content exams. Panel D reveals two insights. First, when examining cumulative earnings five years later, teachers earn less than the average BA graduate. Second, certification exam takers do suffer from passing the exams and earn less earnings than new teachers, and this earning gaps persist over the following five years.

3 Certification exams and prospective teachers' careers

3.1 Regression model. We begin our analysis by using a local linear regression discontinuity (RD) model to estimate the causal impacts of failing a teacher certification exam on prospective teachers' careers:

$$Y_{ie} = \beta D_{ie} + \alpha x_{ie} + \psi D_{ie} x_{ie} + \gamma_{t(i)e} + \varepsilon_{ie} \quad \text{if } |x_{ie}| \leq h^Y. \tag{1}$$

Y_{ie} is an outcome for individual i who took certification exam e. The variable of interest, D_{ie}, is an indicator for failing the certification exam on the first attempt. We include an interaction between D_{ie} and the running variable, x_{ie}, which is individual i's certification exam score on the first attempt normalized to equal zero for the lowest passing score (240). To restrict identification to individuals taking similar exams, we include fixed effects, $\gamma_{t(i)e}$ for exams, e, interacted with the year in which individual i took the exam, $t(i)$. Our regression samples include only applicants whose admission scores are within h^Y standard deviations of the admission threshold. Our results use the Calonico et al. (2019) bandwidth computed separately for each outcome Y. We weight observations in equation (1) using a triangular kernel following the default options in the `rdrobust` package of Calonico et al. (2019). We cluster standard errors at the individual level, as roughly 0.8 percent of individuals in our sample appear in the data multiple times because they took more than one certification exam on the same day.

We estimate equation (1) for all certification exam takers and separately for two groups: white and URM (Hispanic and Black) candidates. The separate estimates for these groups show how failing a teacher certification exam might differently affect teacher candidates by race. We also jointly estimate the model with interaction terms and report p-values to indicate whether differences in effects across the two racial groups are statistically significant.

3.2 Identification assumptions and balance tests. The main RD identification assumption is that candidates' exam scores are effectively randomly assigned near the thresholds. This assumption is supported by balance tests that show that exam taker characteristics do not change discontinuously at the passing threshold. Appendix Table A1 presents estimates from RD regressions that use demographic characteristics, K-12 enrolled schools and programs, high school achievement, and prior year earnings as dependent variables. Most RD estimates are statistically insignificant, and we cannot reject the hypothesis that these coefficients are jointly equal to zero for the full sample ($p = 0.242$), white ($p = 0.737$), and URM candidates ($p = 0.524$).

There is a discontinuous change in the density of certification exam scores at the passing threshold, but this likely reflects how TExES scores are scaled rather than score manipulation. According to a TExES technical manual, the formula that converts raw scores (proportion of correct answer) to scale scores differs depending on whether the raw score is above or below the passing standard.[9] This means that the relationship between underlying

[9]TExES scores are scaled from 100–300 with 240 representing the minimum passing score. The scaling formula is as follows. "For raw scores greater than or equal to the minimum passing (cut) score obtained by a standard-setting study: Scaled score = 240 + [60*(raw score − raw cut score) / (max raw score − raw cut score)]. For raw scores less than the minimum passing score: Scaled score = 100 + 140*(raw score)/(raw cut score)." See "Texas Educator Certification Program Technical Manual, Valid through Au-

ability and scale scores is likely to differ above and below the threshold for any particular exam, and thus scale scores may be more "spaced out" on one side of the threshold. Indeed, we find that the relationship between exam takers' high school math scores and their certification exam scores tends to be flatter below the passing threshold than above it, consistent with the lower density of scores below the threshold. Appendix Figure A1 shows that if we rescale failing scores to have the same linear relationship with high school math scores as passing scores in the vicinity of the threshold, we no longer find a discontinuous change in the density of scores using the test recommended by Cattaneo et al. (2020). Appendix Table A4 shows that our main RD results are essentially unchanged when we use these rescaled scores as our running variable (i.e., the rescaling mainly effects the slope coefficient ψ but not the RD coefficient β). Further, we are not aware of any mechanism through which exam takers could manipulate their scores near the passing threshold.

3.3 Career effects of failing certification exams. Table 2 presents RD estimates of the impacts of failing a certification exam on candidates' teaching careers. In Panel A, candidates who marginally fail, on average, are 11.8% (column C) less likely to be employed as a teacher one year later. URM candidates experience larger negative effects (15%, a 28.6% decrease relative to the mean of the dependent variable) compared to white candidates (8.7%, a 17.5% decrease), and the differential effects are statistically significant ($p = 0.000$). In addition, marginally-failed candidates have 12.2% less years of teaching experience, where URM candidates experience larger effects (15.2%, a 25.5% reduction relative to the dependent variable mean) than white candidates (9.4%, a 16.6% reduction). The RD effect differences on teaching experience between URM and white candidates are statistically significant ($p = 0.000$).[10] Figure 1 graphically illustrates these RD results, displaying plots of teaching outcomes across certification exam scores separately for white (black) and URM (red) candidate groups. These plots visually show outcome disparities by race and ethnicity across the passing threshold.

For earning effects, on average, marginally failed candidates receive $2,831 less one year after the exam. The negative effects are more prominent for URM candidates ($3,942, a 9.5% reduction relative to the dependent variable mean) than white candidates ($1,919, a 5% reduction). The unequal effects between the URM and white candidates are statistically

gust 31, 2018," downloaded in October 2024 at: https://tea.texas.gov/texas-educators/preparation-and-continuing-education/program-provider-resources/texastechnicalmanual8.31.18.pdf

[10] As a supplement to results on teaching outcomes, Table A2 shows RD effects of failing a certification exam on exam outcomes, which includes indicators on whether pass the certification by now, whether pass any certification exam by now, the number of days to pass this exam, and the number of days to pass any exam, by white and URM groups. Results in Table A2 confirm that marginally-failed candidates are less likely to pass both current and any certification exam and take more days to pass both current and any exam. Again, URM candidates experience statistically larger impacts on exam outcomes than white candidates.

different from zero ($p = 0.000$). Similarly, marginally-failed candidates experience \$4,062 less cumulative earnings (in present value) one year after the exam, and the effects are significantly larger ($p = 0.000$) for URM candidates (\$5,672, a 10.41% reduction relative to the dependent variable mean) than white candidates (\$2,532, a 5.16% reduction).[11] Figure 2 graphs the associated RD plots of earning outcomes across exam scores separately for white (black) and URM (red) groups, illustrating outcome disparities by race beyond the passing threshold.

Results in Panels B and C examine longer-run RD effects. Five years after the certification exam (Panel B), significant negative career impacts persist for marginally failing candidates overall and within URM and white subgroups. Over this extended period, RD effects diminish slightly for the likelihood of employment as a teacher, annual earnings, and cumulative earnings, while the effects on teaching experience grow larger. In addition, the differences between URM and white candidates in these RD effects remain statistically significant for teaching and earnings outcomes. Ten years post-exam (Panel C), most RD effects continue to be significant, but differences between URM and white candidates become statistically insignificant. Figure 3 provides a graphical illustration of these lasting RD effects by white and URM groups, over each of the post-period up to ten years after failing the exam, using a balanced panel sample.

These findings indicate that failing a certification exam substantially delays or deters entry into teaching, resulting in losses in both teaching experience and earnings. On average, marginally failed candidates lose around \$9,700 in cumulative earnings over the ten years following the exam, with significantly larger negative impacts for URM candidates. These disparities remain pronounced ten years after the exam, while racial gaps in longer-run outcomes eventually narrow.

3.4 Heterogeneity. Table A5 examines heterogeneity within URM groups by further separating the URM group into Black and Hispanic subsamples. Results indicate that marginally-failed Black teacher candidates are less likely to become teachers, have less teaching experience, and have less annual and cumulative earnings, compared to their Hispanic counterparts. The RD effect differences on teaching outcomes are statistically significant in one and five years after the certification exam.

In addition to heterogeneity across race groups, Table A6 investigates RD effects of failing a certification exam by SES, where candidates who participated in the free or reduced-price

[11] Table A3 details the effects of failing a certification exam on earnings outcomes, including zero-earnings samples and teaching-specific earnings. The results show a consistent pattern as found in Table 2, reinforcing the robustness of RD effects on earning outcomes. Additionally, for earnings from teaching, statistically significant racial differences in the impact on teaching earnings are evident in the short term (one year after the exam) but become statistically insignificant over the longer run (five or ten years after the exam).

lunch when they were students are defined as low SES. Results show that, compared to high SES teacher candidates, low SES candidates who marginally failed the exam experience larger negative effects — are less likely to become a teacher, have less teaching experience, and have less annual and cumulative earnings. The unequal effects on teacher employment indicator, teaching experience, and cumulative earnings one year after the exam are statistically different from zero.

Lastly, we show that SES does not explain the RD differential effects between white and URM groups. Table A7 further separates the group by race and SES, revealing racially heterogeneous effects within both the high- and low- SES groups. These findings confirm that the RD differential effects between white and URM groups are not driven by differences in socioeconomic status.

3.5 Extrapolated effects. While RD estimates indicate the impacts of failing a certification exam on outcomes around the passing score, the results are less informative on the impacts for candidates away from the cutoff. We thus conduct an extended exercise to examine the extrapolated effects. To do so, we use the observed relationship between the scores and outcomes using samples with an exam score range of 240 and 260, and employ a linear extrapolation to project the outcomes for candidates who are below the passing score. Figure A2 visually illustrates the scatter plots and the extrapolated linear lines for the baseline outcomes between white and URM candidates.

Figure 4 summarizes the RD effects and extrapolated effects of the three samples — exam takers with a score within ten points of the passing score, all exam takers with a score below the 240 cutoff, and all exam takers — for teacher employment and cumulative earnings. The results show that, while failing the exam is costly for candidates just at the margin (the RD effects), failing the exam is more costly for people far below the cutoff (the extrapolated effects), especially for the URM group. In addition, when comparing the extrapolated effects between exam takers who failed by 10 points with all exam takers who failed, the more pronounced effects on the latter indicate that more URM candidates are affected when we look at candidates far from the passing cutoff.[12]

[12]Table A8 provides the extrapolated effect estimates of failing certification exam on the baseline outcomes in one, five, and ten years after the exam. Table A9 replicates the exercise but computes the extrapolated effect estimates using the sample that generates the lower bounds on the white-URM gap for the extrapolation, rather than the sample with a score ranging between 240 and 260. Our results still hold when using the lower bounds on the white-URM gap for extrapolation, as in Table A9.

4 Certification exam reform

Our results so far show that certification exams impose significant costs on prospective teachers who fail, but these findings do not capture all of the potential costs and benefits of certification exams. The RD results in Section 3 reflect the impacts of marginal changes in the passing threshold for certification exam, but more significant changes in exam difficulty may generate additional supply side responses. For example, harder exams may reduce the supply of certified teachers, making it harder for schools to find teachers to fill open vacancies. On the other hand, the burdens of certification exams for prospective teachers may be worth it if the exams are effective at screening out low quality teachers. Yet it is hard to know how whether exams are an effective screen because teaching quality is not observed for prospective teachers who cannot pass the exams.

In the rest of this paper, we exploit a reform of the TExES certification exam for elementary school teachers to estimate the impacts of certification exam difficulty on teacher supply, teacher quality, and student achievement. This section describes the TExES reform and its impact on the supply of exams takers and exam passers. In Section 5, we show how this reform can be used to overcome the typical challenge inherent in evaluating whether certification exams are an effective screen for teaching quality. Section 6 examines the overall impacts of the TExES reform on teacher composition and student achievement.

4.1 TExES elementary exam reform. Our analysis going forward takes advantage of a reform that increased the difficulty of the TExES certification exams for elementary school teachers. Panel A of Table 3 shows the TExES exams that led to elementary teaching certificates in core subjects between 2002 and 2015. From 2002–2010, Texas offered a Generalist EC-4 exam for certification in early childhood (EC) education through fourth grade. In 2010, the state replaced this exam with the Generalist EC-6 exam, which expanded the certification grade range up to sixth grade. The EC-6 exam was harder than the EC-4 exam because the EC-6 exam covered both more material and more advanced content. Column (E) of Table 3 shows that 82.7 percent of Generalist EC-4 exam takers passed on their first try, while only 65.5 percent of Generalist EC-6 exam takers did so. At the same time, Texas also replaced the Bilingual Generalist EC-4 exam and the English as a Second Language (ESL)/Generalist EC-4 exam with EC-6 versions of these exams, which also increased their difficulty.[13]

[13]Column (F) of Table 3 shows that the first-attempt pass rates on the Generalist EC-6 exams remain 18–27pp lower in regressions that control for exam takers' high school math and ELA scores, showing that the lower pass rates are not driven by a time trend in the average ability of exam takers. In 2015, the Generalist EC-6 exams were replaced with the Core Subjects EC-6 exam. While the Generalist EC-4 and EC-6 exams produced a single score that determined the pass/fail decision, individuals who took the Core Subjects EC-6

During the same time period, the certification exams for middle school teachers remained constant. Panel B of Table 3 shows that individuals could be certified to teach grades 4–8 by passing either a Generalist 4-8 exam or by passing subject-specific exams (e.g., Mathematics 4-8). This suite of exams did not change from 2002–2015, which allows us to use middle school grades as a control group in our analysis of student impacts.

4.2 Effects of exam difficulty on exam takers. We begin by showing how the TExES reform affected the composition of exam takers and exam passers. Table 4 shows estimates from DiD regressions that compare Generalist EC-4/EC-6 exam takers (treated group) to individuals who took all other TExES teacher certification exams that were offered in 2004–2015 (control group). Columns (A)–(B) show means for white and URM individuals who took a Generalist EC-4 exam (2004–2009). Columns (C)–(E) show DiD coefficients estimated separately for all, white, and URM exam takers; these are coefficients on an indicator for taking a Generalist exam interacted with a dummy for years after the switch to the EC-6 exam (2010–2015). Figure 5 shows event study graphs for this DiD specification.[14]

The TExES reform significantly increased the difficulty of passing elementary certification exams, particularly for URM candidates. Panel A of Table 4 shows that the switch from the EC-4 to EC-6 exam reduced first-attempt passing rates by 13.6pp for white candidates and by 21.0pp for URM candidates. Although some individuals retook the exams, the reform reduced the proportion of white and URM exam takers who ever passed the Generalist exams by 5.4pp and 14.3pp, respectively. As a result, the URM share of exam takers who passed on the first try fell by 3.2pp, and the URM share of candidates who ever passed fell by 1.5pp (Panel B of Table 4). Panels A–D of Figure 5 show that passing rates declined immediately with the introduction of the Generalist EC-6 exam in 2010 and remained lower in subsequent years. The reform also increased the average number of exam attempts and the average length of time that candidates took to pass the exam (Panel A of Table 4).

---

were required to pass each subject separately (Math, ELA, Natural Science, and Social Science), and exam takers could retake individual subjects if they failed. We focus our analysis on the switch from the Generalist EC-4 to EC-6 exam because the changes in exam structure are more substantial for the Core Subjects exam.

[14] Specifically, our DiD specification for Table 4 is:

$$Y_i = \gamma_{e(i)} + \gamma_{t(i)} + \beta \text{Generalist}_{e(i)} \text{Post}_{t(i)} + \varepsilon_i, \tag{2}$$

where Y_i is an outcome for exam taker i. The sample includes individuals who took one of the elementary certification exams in Panel A of Table 3 (Generalist EC-4/EC-6, Bilingual Generalist EC-4/EC-6, or ESL/-Generalist EC-4/EC-6) or any of the 62 other TExES teacher certification exams that were offered in every year from 2004–2015. We include fixed effects for exam years, $\gamma_{t(i)}$, and TExES exams, $\gamma_{e(i)}$. We treat each EC-4/EC-6 pair as a single exam for the $\gamma_{e(i)}$ fixed effects. $\text{Generalist}_{e(i)}$ is an indicator for taking one of the Generalist exams. $\text{Post}_{t(i)}$ is an indicator for exam years $t(i) \geq 2010$, when the EC-6 exams were offered. Columns (C)–(E) of Table 4 show the DiD coefficients β estimated separately for all, white, and URM exam takers. Figure 5 shows event study estimates from a version of equation (2) that computes separate $\beta_{t(i)}$ coefficients for each exam year $t(i)$ (omitting 2009).

As a result of this decline in passing rates, the *number* of individuals who passed the elementary certification exams also fell. For Panel C of Table 4, we collapse our data to the exam × year level, and our outcome variables are the log number of exam takers and exam passers. The switch from the EC-4 to EC-6 exam had essentially no impact on the number of first-time exam takers (0.005 log points), which is notable given that, all else equal, the EC-6 certificate should have been more popular since it allowed individuals to teach a wider range of grades. The total number of exams taken increased by 24 percent (0.22 log points), reflecting a higher retake rate. Yet the number of individuals who passed on the first attempt declined by 20 percent (−0.23 log points), and the number of exam takers who ever passed declined by 13 percent (−0.13 log points). Consistent with the results in Panels A–B, these reductions are more pronounced for URM teaching candidates.

A potential upside of this supply reduction is that the reform increased the average ability of exam passers. Panel D of Table 4 shows effects on teaching candidate ability as measured by the exam takers' high school math scores (see also Panels E–F of Figure 5).[15] We do not find evidence that TExES reform affected the average ability of individuals who attempted the Generalist exams. Yet the switch from the EC-4 to EC-6 exam led to a 0.10 SD increase in the average math ability of individuals who passed the exam on their first try and a 0.04 SD increase in the mean math ability of individuals who ever passed. Although this finding is consistent with the goal of harder certification exams, the magnitude of this effect is relatively modest. Assuming a normal distribution for math scores, the average candidate who passed the EC-4 exam was at the 67th percentile of the distribution of math ability among all high school students in Texas, while the average candidate who passed the EC-6 exam was at the 68th percentile of this distribution. The next section examines the relationship between certification exam difficulty and teaching quality in more detail.

5 Certification exams and teacher quality

This section asks whether the racial disparities in certification exam pass rates are justified by differences in teaching quality. In Section 5.1, we develop an econometric framework that yields an empirical test of whether harder teacher certification exams have a disparate impact on URM candidates relative to easier exams. We implement this test in Section 5.2 by combining estimates of teacher value-added on standardized exams with variation from the TExES reform.

[15]The results in Panel D of Table 4 are similar if we use exam takers' high school ELA scores.

5.1 Policy-relevant disparate impact. We adapt the notion of disparate impact in Arnold et al. (2022) to one that we can test using quasi-experimental variation from the TExES certification exam reform. Arnold et al. (2022) develop and implement an empirical test for disparate impact in the context of bail judge release decisions. While their empirical evidence is compelling, it relies on both the quasi-random assignment of defendants to judges and the fact that some judges are extremely lenient in their release decisions (so that econometricians can observe uncensored outcomes). These two features—random assignment and lenient decision makers—are difficult to replicate in other settings, including ours. Thus we modify Arnold et al. (2022)'s definition of disparate impact to one that we call *policy-relevant disparate impact*, which we can credibly estimate using the TExES reform.

We consider a population of teacher certification exam takers defined by their race, potential exam performance, and teaching quality. Let $R_i \in \{U, W\}$ denote the race of exam taker i, where U is URM and W is White. Let $T_i \in \{E, H\}$ denote whether individual i took an easy or hard certification exam. In our context, $T_i = E$ represents the easier EC-4 TExES exam and $T_i = H$ represents the harder EC-6 exam. Let D_i^E and D_i^H be indicators for whether individual i would pass the easy and hard exams on their first try. D_i^E and D_i^H are potential outcomes because individuals' first attempt can only be on one of the two exams. We let $D_i = \mathbb{1}\{T_i = E\}D_i^E + \mathbb{1}\{T_i = H\}D_i^H$ be the observed indicator of whether individual i passed their exam. Lastly, we let Y_i^* be individual i's teaching quality as measured by value-added if they were to become a teacher. Y_i^* is a censored outcome because value-added is not observed for some individuals who fail the exam.[16]

Given this notation, we define the *policy-relevant disparate impact* (PRDI) at a particular value of teacher value-added $Y_i^* = y$ as:

$$\begin{aligned} \text{PRDI}_y \equiv \quad & \Pr(D_i^H = 1 | R_i = U, T_i = H, D_i^E = 1, Y_i^* = y) \\ & - \Pr(D_i^H = 1 | R_i = W, T_i = H, D_i^E = 1, Y_i^* = y). \end{aligned} \tag{3}$$

Equation (3) says there is disparate impact if the pass rate on the hard exam, $\Pr(D_i^H = 1)$, differs between URM (U) and White (W) exam takers who have the same level of value-added, $Y_i^* = y$. Our measure of disparate impact is policy-relevant in the sense that it corresponds to a particular population that is affected by the switch from an easy to a hard exam: individuals who took the hard exam ($T_i = H$) and who *would have passed* the easy exam if they had taken it instead ($D_i^E = 1$). This differs from the population average definition of disparate impact in Arnold et al. (2022), which is equivalent to dropping the

[16]Below we discuss additional considerations that arise from the fact that Y_i^* is not observed even for many teachers who pass the exam, either because they choose not to become teachers or because they teach subjects without student exams. Further, we address the fact that Y_i^* may vary over a teacher's career.

$T_i = H$ and $D_i^E = 1$ conditions in equation (3). We focus on policy-relevant disparate impact because it allows for more credible identification assumptions in our context, and because it answers the specific policy question of whether the EC-6 exam (H) had a disparate impact relative to the EC-4 exam (E) for the population of exam takers affected by this reform.[17]

We transform equation (3) into a form that clarifies the assumptions needed to identify policy-relevant disparate impact. Ignoring race for the moment to simplify notation, the key object to be estimated in equation (3) is the hard exam pass rate conditional on a particular level of value-added $Y_i^* = y$:

$$\text{PassRate}_y \equiv \Pr(D_i^H = 1 | T_i = H, D_i^E = 1, Y_i^* = y).$$

By Bayes' Rule, we have:

$$\text{PassRate}_y = \Pr(D_i^H = 1 | T_i = H, D_i^E = 1) \times \frac{\Pr(Y_i^* = y | T_i = H, D_i^E = 1, D_i^H = 1)}{\Pr(Y_i^* = y | T_i = H, D_i^E = 1)}. \quad (4)$$

The first term on the right-hand side of equation (4) is the hard exam pass rate in our policy-relevant population ($T_i = H$ and $D_i^E = 1$). To transform this into the pass rate at a particular level $Y_i^* = y$, this term is multiplied by the ratio of two PDFs of teacher value-added at Y_i^*. The numerator is the PDF of value-added for the subset of our policy-relevant population who *actually passed* the hard exam ($D_i^H = 1$), while the denominator is the PDF of value-added for the full policy-relevant population. None of the terms in equation (4) is observed in the data, for two reasons. First, we do not observe which hard exam test takers ($T_i = H$) would have passed the easy exam if they had taken it ($D_i^E = 1$). Second, value-added (Y_i^*) is censored for some individuals who would have passed the easy exam but instead took and failed the hard exam.

To identify these unobservable quantities, we make the following two assumptions:

1. $D_i^H = 1 \implies D_i^E = 1$: Anyone who passed the hard exam would also have passed the easy exam.

2. $T_i \perp\!\!\!\perp D_i^E, D_i^H, Y_i^*$: Individuals are as good as randomly assigned to the easy or hard exams. More formally, individuals' potential pass rates and value-added are independent of the exam they take. We assume random assignment for simplicity of exposition, but this assumption may be violated in our context if, for example, there is a time trend in the characteristics of TExES exam takers. In our empirical analysis below, we relax this assumption to a DiD parallel trends assumption using middle school certification

[17] Our analysis will not speak to any potential disparate impact of certification exams on exam takers who would have failed the easy exam.

exams as a control group, and we use balance tests to examine the validity of this assumption.

Under these assumptions, the pass rate at $Y_i^* = y$ is equal to:[18]

$$\text{PassRate}_y = \frac{\Pr(D_i = 1|T_i = H)}{\Pr(D_i = 1|T_i = E)} \times \frac{\Pr(Y_i = y|T_i = H, D_i = 1)}{\Pr(Y_i = y|T_i = E, D_i = 1)} \tag{5}$$

Equation (5) shows that the pass rate at $Y_i^* = y$ among hard exam takers who would have passed the easy exam is a function of two ratios of observable quantities: 1) the ratio of pass rates for the hard and easy exams; and 2) the ratio of the PDFs of value-added for exam takers who passed each exam. Below we compute equation (5) separately for URM and White exam takers to estimate the policy-relevant disparate impact at $Y_i^* = y$ defined in equation (3). We then average across values of y to compute an average measure of disparate impact for our policy-relevant population.

5.2 Estimating disparate impact. We estimate policy-relevant disparate impact using teacher value-added on standardized exams as a measure of teaching quality (Kane and Staiger, 2008; Chetty et al., 2014c; Koedel and Rockoff, 2015). We follow the methodology in Chetty et al. (2014a) to compute teacher value-added for grade 4–8 math and ELA test scores using a large vector of student-, classroom-, and school-level controls, including lagged test scores. We can only compute value-added for teachers whose students take standardized exams and for the years in which our data includes classroom identifiers, so this analysis is limited to teachers who teach math or ELA in grades 4–8 during 2012–2019. Appendix B.2 provides details on our methods for calculating value-added.

As an initial descriptive fact, Table 5 shows that certification exam performance is only weakly related to value-added, particularly for URM teachers. This table regresses student math (Panel A) and ELA (Panel B) scores on indicators for teacher race, measures of teachers' performance on their first attempt at a certification exam, and the full vector of student-, classroom-, and school-level value-added controls. The certification exam variables are defined only for teachers whose first content exam was one of the tests listed in Table 3, which are the TExES exams that lead to math and ELA teaching in grades 4–8.[19] On

[18]The derivation of the ratio of pass rates in equation (5) use Bayes' rule combined with our two assumptions:

$$\Pr(D_i^H = 1|T_i = H, D_i^E = 1) = \frac{\Pr(D_i^E = 1|T_i = H, D_i^H = 1) \times \Pr(D_i^H = 1|T_i = H)}{\Pr(D_i^E = 1|T_i = H)} = \frac{\Pr(D_i^H = 1|T_i = H)}{\Pr(D_i^E = 1|T_i = E)}$$

[19]Teachers who took pre-TExES exams or other tests as their first content exam are still included in regressions in Table 5 so that they influence the distribution of teacher value-added, but the certification exam variables reflect comparisons only among those teachers who took one of the exams in Table 3 as their

average, URM teachers have slightly higher value-added in both math (0.007 SDs) and ELA (0.010 SDs) than White teachers. White teachers who passed the certification exam on their first attempt have math value-added that is 0.019 SDs higher than those who initially failed, but this difference is only 0.007 SDs for URM teachers and is statistically insignificant. This pattern is similar but smaller in magnitude for ELA scores. This pattern is also similar using certification exam scores as the measure of teacher performance (column C) rather than an indicator for passing the exam (column B). Further, these findings are robust to restricting the estimates to teachers who passed the exam on their first attempt (columns D–E).[20]

The modest correlations between certification exam performance and value-added in Table 5 mirror those in other papers in the literature (Goldhaber, 2007; Clotfelter et al., 2010; Chingos and Peterson, 2011; Shuls and Trivitt, 2015; Hendricks, 2016; Goldhaber et al., 2017), but a fundamental problem with all of these analyses is that value-added is not observed for all teachers who initially fail the exam. There is likely to be selection between those who ultimately manage to pass after initially failing and those who do not (e.g., in terms of their knowledge of the material). Further, the passing threshold is explicitly chosen to distinguish prepared and unprepared teachers; the exams are not designed to distinguish great teachers from average teachers. Thus, the results in Table 5 do not conclusively prove that certification exams are a bad screening mechanism for teaching quality.

To overcome some of these challenges, we exploit the TExES reform to estimate the policy-relevant disparate impact of the harder EC-6 exam relative to the easier EC-4 exam. Under the assumptions described in Section 5.1, the reform allows us to overcome the fundamental problem of unobserved teaching quality because we observe value-added for individuals who passed the easy exam but who would not have passed the hard exam. Following equation (5), we estimate the distributions of teacher value-added, $\Pr(Y_i = y)$, among both easy exam passers ($T_i = E, D_i = 1$) and hard exam passers ($T_i = H, D_i = 1$). Our population of easy exam passers includes individuals who took the EC-4 exam in 2002–2009 as their first certification exam, who passed on their first attempt, and who subsequently went on to teach fourth grade. Similarly, our population of hard exam passers includes individuals whose first certification exam was the EC-6 exam in 2010–2015, who passed on their first attempt, and who then taught fourth grade. We also compute distributions of value-added for grade 7–8 teachers who took the middle school certification exams in the pre-reform (2002–2009) and post-reform (2010–2015) years. Since these exams did not change over this time period, we use these estimates to compute a placebo measure of policy-relevant disparate impact that

first content exam.

[20]Appendix Table A11 similarly finds that race and certification exam performance are weakly correlated with measures of teacher value-added for non-cognitive outcomes.

allows us to control for any time trend in the characteristics of certification exam takers.

Panel A of Figure 6 shows that the distributions of math value-added are similar for EC-4 and EC-6 passers, suggesting that the harder EC-6 exam was not any more effective at screening out low-quality teachers.[21] In fact, the mean of the EC-4 distribution is 0.03 SDs *higher* than that for the EC-6 exam, as indicated by the vertical lines. Panel B of this figure shows that math value-added is also 0.02 SDs higher on average for individuals who passed middle school math certification exams in the pre-reform period than for those who passed the same exams in the post-reform years, suggesting a slight negative time trend in teacher quality over this period. But the similarity between the grade 4 and grade 7–8 value-added distributions shows that the harder EC-6 exam did not produce higher quality math teachers than the easier EC-4 exam. This pattern of findings is similar for ELA value-added (Appendix Figure A4) and if we compute the value-added distributions separately for URM and White teachers (Appendix Figures A5–A6).

We use these value-added distributions to compute our overall measure of policy-relevant disparate impact. We begin by using equation (5) to compute race-specific values of PassRate_y. The first term in equation (5) is simply the difference in first-attempt pass rates on the EC-4 and EC-6 exams. To compute the second term in equation (5), we group teacher value-added into deciles and compute the proportion of observations in each decile for each exam. We perform these calculations separately for each year of potential experience to ensure that EC-4 and EC-6 teachers are measured at the same number of years since passing the exams. We then plug the race-specific values of PassRate_y into equation (3) to compute the policy-relevant disparate impact at each value of teacher value-added $Y_i^* = y$, PRDI_y. Lastly, we average the PRDI_y estimates across values of y based on the distribution of value-added for URM exam takers who passed the easy exam. Thus our overall measure of disparate impact is equal to the observed URM pass rate on the hard exam among those who would have passed the easy exam minus an analogous pass rate for White exam takers, but where the White pass rate is reweighted to match the URM distribution of value-added. We perform the same calculations for grade 7–8 teachers to compute a placebo measure of disparate impact for the middle school exams.

Our main finding is that the harder EC-6 exam had a large disparate impact on URM exam takers relative to the easier EC-4 exam. Panel A of Figure 7 shows that the EC-6 pass rate among individuals who would have passed the EC-4 exam was 66 percent for URM test takers and 82 percent for White test takers. Reweighting the White pass rate to match the

[21] We observe EC-4 exam passers at higher levels of potential experience (i.e., years since the exam) than EC-6 exam passers, so we weight observations in this figure so that the average level of potential experience is the same in the two samples.

distribution of math value-added among URM teachers changes this pass rate only slightly to 83 percent. Thus our initial estimate of disparate impact is a 17pp URM/White gap in EC-6 exam pass rates among individuals who would have passed the EC-4 exam. Our placebo disparate impact calculation using middle school exams yields an URM/White pass rate gap of 4pp, which reflects a general decline in the ability of certification exam takers over time. Taking the difference between the grade 4 disparate impact estimate and the placebo estimate for grades 7–8 yields our preferred estimate of policy-relevant disparate impact: a 13pp URM/White gap in pass rates. Similarly, Panel B shows that the policy-relevant disparate impact is 10pp using ELA value-added as the measure of teaching quality.

In sum, we find that the harder EC-6 exam was not any more effective at screening out low-quality teachers than the EC-4 exam, which means that the EC-6 exam had a disparate impact on URM test takers relative to the EC-4 exam given the large disparity in pass rates. The similarity between the distributions of teacher value-added for EC-4 passers and EC-6 passers means that the observed difference in pass rates on the two exams is essentially equal to our measure of disparate impact that controls for value-added (after accounting for a slight time trend in certification pass rates). Taken together with our findings from Section 3 on the career consequences of failing a certification exam, this shows that certification exams impose disproportionate costs on URM exam takers without corresponding benefits for teaching quality.[22]

6 Overall effects of certification exams on teachers and students

Although our above results document both high costs and limited benefits of certification exams, reforms that significantly change the difficulty of certification exams may have additional impacts on teachers and students through supply side channels. Since K–12 education is mandatory and essentially free, the demand for public school teachers is roughly fixed re-

[22]In Appendix Tables A12–A14 and Appendix Figures A7–A9, we follow Orellana and Winters (2023) in asking whether teacher value-added changes discontinuously at the certification exam passing threshold using our RD specification (1). We find that math value-added measured one year after the certification exam is 0.048 SDs higher among teachers who marginally failed than those who marginally passed, suggesting that the certification exam initially delayed some lower-quality teachers from starting their careers. But there is no discontinuity in math value-added measured 2–10 years after the certification exam, which shows that, in the long run, individuals who marginally failed and went on to become teachers were neither better nor worse math teachers than individuals who initially passed. We find smaller and statistically insignificant RD coefficients for ELA value-added in both one year after the exam and 2–10 years after the exam. These results corroborate our finding from the disparate impact analysis that certification exam performance is not strongly related to teaching quality, although the RD results can only speak to the quality of individuals who marginally fail the exam relative to those who marginally passed.

gardless of the supply and quality of potential teachers. In other words, schools will still have to find individuals to fill open teaching vacancies even if harder certification exams reduce the size of the pool of potential teachers. Thus it is unclear how certification exam reforms will impact both the composition of the teaching workforce and student achievement.

This section develops a new identification strategy that allows us to examine the overall impacts of certification exam difficulty on teachers and students. We identify a set of open teaching vacancies by using the departure of experienced teachers as a shock to schools' labor supply (Section 6.1). Our main specification is an RD-DiD regression model that combines variation from teacher departures with variation from the TExES exam reform (Section 6.2). Together, these two sources of variation allow us to ask how harder certification exams affect the composition of the teaching workforce (Section 6.3) and student achievement (Section 6.4).

6.1 Teacher departures. To examine the supply-side impacts of the TExES reform, we bring in a second source of variation: the departure of experienced teachers. Intuitively, departures are a shock to schools' labor supply that create open teaching vacancies, which allows us to ask how demand for new teachers affects teacher composition under easier or harder certification exams. We define a *teacher departure* as an instance in which a math or ELA teacher with five or more years of experience leaves a given school. To isolate large changes in teacher composition, we restrict our analysis to cases in which the departing teacher taught one-third or more of the students in a given school, grade, and subject in the year before their departure.[23] This strategy is analogous to research that uses worker deaths as a shock to firm labor supply to ask how firms find substitutes for their workers (e.g., Jäger and Heining, 2022). Our strategy also follows Chetty et al. (2014b) in using teacher departures to identify impacts of teacher quality.

To use all possible variation from teacher departures, we create a "stacked" dataset for each teacher departure from a given school, grade, and subject. We first collapse our individual-level data to the school/grade/subject/year level. We let s denote school/grade/-subject triplets and use t to denote years, so the variables in our collapsed dataset are mean teacher characteristics at the st level. A given school/grade/subject may experience multiple teacher departures during the period of our data. Thus we let y denote the year of a teacher departure, and we "stack" our collapsed dataset so that st observations occur multiple times for each departure. (We drop st observations for which there is no teacher departure.) Lastly,

[23]For grades 7–8, we focus on two subjects: math or ELA (English + reading). For grades 3–4, our main specification uses a single core subject (the combination of math, ELA, science, social studies, and generic) since most elementary teachers teach all core subjects. We focus on departures of teachers with 5+ years of experience so that the departing teacher likely took a different certification exam than the exam that was currently offered to new teachers. See Appendix B.3.1 for details on our definition of teacher departures.

we let $\tau_{ty} = t - y$ denote years *relative* to the teacher departure, where $\tau_{ty} = 0$ is the first year in which the teacher is no longer in the school.

Since teacher departures may be caused in part by school-specific trends, we use an RD model to isolate sharp changes in teacher composition due to departures.[24] Specifically, we use our collapsed and stacked dataset to estimate a local linear RD regression:

$$Y_{st} = \beta \mathbf{1}\{\tau_{ty} \geq 0\} + \alpha \tau_{ty} + \psi \mathbf{1}\{\tau_{ty} \geq 0\}\tau_{ty} + \gamma_{sy} + \varepsilon_{sty} \quad \text{if } |\tau_{ty}| \leq h^Y. \tag{6}$$

The dependent variable, Y_{st} is an average teacher characteristic or student outcome at the school/grade/subject (s) and year (t) level. Our variable of interest is an indicator for years after the teacher departure, $\mathbf{1}\{\tau_{ty} \geq 0\}$. The running variable is years relative to the teacher departure, τ_{ty}, and we include an interaction between τ_{ty} and $\mathbf{1}\{\tau_{ty} \geq 0\}$. We include fixed effects for school/grade/subject/departure-year quadruplets, γ_{sy}, so that identification comes only from before and after variation within the same departure event. The regression includes years relative to departure, τ_{ty}, that are within the Calonico et al. (2019) RD bandwidth, h^Y, computed separately for each outcome Y.[25] Standard errors are clustered at the school level to allow for correlation in outcomes within the same school (e.g., if schools ask teachers to switch grades following a departure). The coefficient of interest, β, estimates the projected change in average outcomes in the first year after the teacher departure ($\tau_{ty} = 0$). For example, if departing teachers are less likely to be URMs than the teachers that replace them, we would find $\beta > 0$ for the outcome of the URM share of teachers.

6.2 RD-DiD specification. Our full specification combines our teacher departure RD specification with DiD variation in exposure to the TExES reform. Specifically, we estimate our RD regression (6) separately for grades that were more and less affected by the exam reform (treated/control groups) and for teacher departures that occurred before and after the reform (pre/post variation). Our treated group includes grades 3–4 because the vast majority of new teachers in these grades took the Generalist EC-4 or EC-6 exams during this period (see Appendix Table A10). We use grades 7–8 as our control group because neither the EC-4 exam nor the EC-6 exam qualifies one to teach these grades.[26] Our pre-period includes teacher departures that occurred in $y \in$ 2005–2010, when the Generalist EC-4 exam was in

[24] Jäger and Heining (2022) use a DiD specification because worker deaths are plausibly exogenous. As we show below, a DiD model does not work as well in our case because teachers may choose to leave a school if its performance is on a negative trend. The RD specification helps to address this issue by separating sharp changes in teacher composition from longer-term trends that may cause teacher turnover.

[25] We weight observations in equation (6) by the product of a triangular kernel (based on the RD bandwidth) and the number of individual teachers or students that were used to compute the outcome variable, Y_{st}.

[26] The effect of the TExES reform is less clear for grades 5–6 since the EC-6 exam is a pathway to teaching these grades but the EC-4 exam is not. Thus we omit grades 5–6 from this analysis.

place. Our post-period includes departures in $y \in$ 2011–2016, when the EC-6 exam was in place.[27] We estimate equation (6) separately for each pairwise combination of grade group $g \in \{3\text{–}4,\ 7\text{–}8\}$ and departure period $p \in \{2005\text{–}2010,\ 2011\text{–}2016\}$, which gives four RD coefficients β_{gp}.

Lastly, we use these RD coefficients as the dependent variable in a simple DiD regression:

$$\beta_{gp} = \phi \text{Treated}_g + \delta \text{Post}_p + \theta \text{Treated}_g \text{Post}_p + \varepsilon_{gp}, \tag{7}$$

where Treated_g is a dummy for grades 3–4 and Post_p is a dummy for teacher departures in $y \in$ 2011–2016. Equation (7) gives our RD-DiD specification. The coefficient of interest, θ, shows how the effects of teacher departures (as estimated by the RD coefficients) changed with the TExES reform in grades 3–4 relative to grades 7–8. For example, if the TExES reform made it more difficult for elementary schools to hire URM teachers relative to middle schools, then we would find $\theta < 0$ for the outcome of the URM share of teachers.

Although equations (6)–(7) present our RD-DiD specification in two steps to build intuition, in practice we estimate a single-step specification by plugging equation (7) into equation (6).[28] Our single-step RD-DiD specification is equivalent to estimating the RD equation (6) separately for each gp pair and then constructing the DiD estimator: $\theta = \Big(\beta_{\text{Treated,Post}} - \beta_{\text{Treated,Pre}}\Big) - \Big(\beta_{\text{Control,Post}} - \beta_{\text{Control,Pre}}\Big)$.

The key identification assumption in our RD-DiD strategy is parallel trends *in the RD coefficients*. Specifically, we assume that the effects of teacher departures on teacher composition (and student outcomes in Section 6.4) as estimated by the RD coefficients would have trended in the same manner for grades 3–4 and grades 7–8 in the absence of the TExES reform. One might be concerned that the standard RD assumption of "no threshold manipulation" may be violated in our case if, for example, teacher departures cause students to change schools. But our RD-DiD specification allows for this type of behavioral response provided that it does not change differentially in response to the TExES reform. This is analogous to the "difference-in-IV" estimator in Alsan et al. (2025), in which violations of

[27] The EC-6 exam was introduced in 2010, but we consider teacher departures in $y = 2011$ to be the first post-reform period since most individuals do not get a teaching job right away after passing the exam.

[28] In other words, our single-step RD-DiD specification is:

$$\begin{aligned} Y_{st} = \big(\phi \text{Treated}_g + \delta \text{Post}_p + \theta \text{Treated}_g \text{Post}_p\big)\mathbf{1}\{\tau_{ty} \geq 0\} + \\ \alpha_{gp}\tau_{ty} + \psi_{gp}\mathbf{1}\{\tau_{ty} \geq 0\}\tau_{ty} + \gamma_{sy} + \varepsilon_{sty} \quad \text{if } |\tau_{ty}| \leq h^Y_{gp}. \end{aligned} \tag{8}$$

Note that we estimate separate running variable coefficients α_{gp} and ψ_{gp} for each grade group g and departure period p pair. We also compute Calonico et al. (2019) RD bandwidths h^Y_{gp} separately for each gp pair. Grade groups g are a function of the school/grade/subject triplet s, i.e., $g = g(s)$. Departure periods p are a function of teacher departure years y, i.e., $p = p(y)$. Thus the fixed effects γ_{sy} subsume dummies for gp pairs.

the standard IV assumptions are permissible provided that they are the same in the pre- and post-periods. Below we test the validity of our parallel trends assumption by using a variety of student characteristics as outcome variables, Y_{st}.

6.3 Effects of exam difficulty on teacher composition. Table 6 shows our RD-DiD estimates for the effects of the TExES reform on teacher composition. Column (A)–(C) present results for grade 3–4 teacher departures. Column (A) shows the mean of each outcome in the year prior to the departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in 2011–2016. Columns (B)–(C) show RD coefficients β from equation (6) estimated separately for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Similarly, columns (D)–(E) show β coefficients β for grades 7–8 teacher departures in the post- and pre-reform periods, respectively. Column (F) shows our main object of interest, the RD-DiD coefficient θ from equation (7), which is equal to column (B) − column (C) − (column (D) − column (E)). Figure 8 shows RD graphs for a few of the outcomes in Table 6. The left graphs in each panel (red circles) depict grade 3–4 teacher departures, and the right graphs (black triangles) depict grade 7–8 departures. Hollow symbols represent pre-reform departures ($y \in$ 2005–2010), and solid symbols represent post-reform departures ($y \in$ 2011–2016). The x-axis denotes years relative to the teacher departure, τ_{ty}.

We first show that our RD-DiD approach identifies open teaching vacancies with differential exposure to the Generalist EC-4 and EC-6 exams, as intended. Panel A of Table 6 shows that grade 3–4 teachers who departed in 2005–2010 were often replaced with teachers who took the EC-4 exam, whereas elementary teachers who departed in 2011–2016 were more likely to be replaced with EC-6 exam takers. Specifically, the proportion of a school/-grade's teachers who took the EC-4 exam increases discontinuously from about 20pp to 35pp following a grade 3–4 teacher departure in 2005–2010 (Figure 8, Panel A), yielding an RD coefficient of 14.6pp (Table 7, column C). There are only small discontinuities in the proportion of teachers who took the EC-4 exam following grade 3–4 teacher departures in 2011–2016 (column B) and grade 7–8 teacher departures in either time period (columns D–E). Thus our RD-DiD estimate shows that the TExES reform reduced the proportion of newly-hired elementary teachers who took the EC-4 exam by 14.4pp (column F). Analogously, the TExES reform increased the proportion of newly-hired elementary teachers who took the EC-6 exam by 7.6pp, which is driven by a large discontinuity for grade 3–4 teacher departures in 2011–2016 (Figure 8, Panel B). The reform also increased the difficulty of newly-hired elementary teachers' exams as measured by the average pass rate and the teachers' own exam scores (Appendix Table A15).

Our main finding on teacher composition is that the TExES reform reduced the URM share of newly-hired elementary teachers. Panel B of Table 6 shows that teacher departures

tend to increase the URM share of teachers, with RD coefficients ranging from 1.4pp to 2.1pp in columns (C)–(E). In other words, newly-hired teachers were more likely to be URM than the departing teachers they replaced, which reflects the increasing diversity of the Texas teaching workforce over time. The one exception to this pattern is for grade 3–4 teacher departures in the post-reform period (column B), for which there is no significant discontinuity in the URM share (see also Figure 8, Panel C). This suggests that the more difficult EC-6 exam reduced the share of URMs in the pool of certified teachers. The RD-DiD estimate for the URM share of teachers ($\theta = -1.9$pp), which roughly matches our estimate for the URM share of exam passers from Table 4 (-1.5pp). Consistent with these exam results, we also find that the TExES increased the average high school math scores of newly-hired teachers by 0.08 SDs, although this effect is statistically insignificant (Appendix Table A15).

The reduction in the URM share of teachers came from a decline in new URM teachers and an increase in more experienced White teachers. In the other rows of Panel B of Table 6, our outcome variables include an indicator for new teachers (those within five years of receiving their first certification) and the interaction of this variable with teacher race. The TExES reform reduced the share of elementary teaching vacancies that were filled by new teachers by 3pp, driven primarily by a 2.5pp reduction in the share of new URM teachers. By contrast, the share of elementary teaching vacancies that were filled by White teachers who earned a certification more than five years ago increased by 1.7pp.

This change in teacher composition meant that total earnings declined for URM teachers in elementary schools and while total earnings for White teachers increased. Our outcome variables in Panel C of Table 6 are the sum of earnings to teachers in a given school/grade/-subject/year in cell. Teacher departures typically lead to a decline in total teacher pay since the newly-hired teachers tend to make lower salaries than the departing teachers. But relative to middle schools, the TExES reform led to a \$6,800 increase the total payment to teachers in elementary schools with departing teachers, consistent with the higher experience of the newly-hired teachers. This higher payment was driven by a \$6,300 increase in payment to White teachers, whereas the total payment to URM teachers fell by \$2,300 with the reform. We also find that the costs of harder certification exams fall on URM teachers if use teachers' year-to-year earnings gains as our outcome variable (Appendix Table A15). Thus harder certification exams not only have disproportionate costs on prospective teachers from URM groups, but they also reduce the share of teaching salaries that go to URMs as a whole.

6.4 Effects of exam difficulty on student achievement. We find no systematic evidence that the TExES reform affected the number of students in school/grades with de-

parting teachers or their demographic characteristics. In Panel A of Table 7, the outcome variables are the number of students in the school/grade who took student achievement exams and their demographic traits. We also combine all the demographic characteristics into indices of predicted math and ELA scores. Panel A of Figure 9 shows RD graphs for the math score demographic index, and Appendix Figure A10 shows the ELA score index. We find a small and statistically insignificant effect on the number of exam takers, with an RD-DiD coefficients of -1.4 students relative to a pre-departure mean of 74.3 students. We also find statistically insignificant effects on all demographic characteristics, with one exceptions: the proportion of students at risk of dropping out ($\theta = -2.0pp$). Our Math and ELA score demographic indices suggest that there were no significant differential trends in student characteristics between grades 3–4 and grades 7–8.

Turning to our main results on student achievement, we find that the harder elementary certification exams neither increased nor decreased student test scores. Panel B of Table 7 shows RD-DiD estimates on student math and ELA scores. Our outcome variables in these panels also include test score residuals, which are the residuals from a value-added-like regression that includes individual, school/grade mean, and school mean student characteristics and lagged test scores. The RD coefficients in columns (B)–(E) show that teacher departures have mostly small and insignificant effects on student achievement in either time period and for either grade level (see also Panels B–C of Figure 9 and Appendix Figure A10). Our RD-DiD estimates in column (F) also show that these RD coefficients did not change differentially between grades 3–4 and grades 7–8 with the TExES reform. We find positive estimates for ELA scores and ELA score residuals (0.007 SDs and 0.009 SDs) and negative estimates for math scores and math score residuals (-0.014 SDs and -0.000 SDs), but all estimates are statistically insignificant.

Table 8 shows test score effects using an alternative specification that exploits variation across math and ELA teacher departures within grades 3–4. Although the majority of grade 3–4 teachers teach all core subjects, some schools in our data have teachers that specialize in either math or ELA. In Table 8, we restrict our sample to school/grades with teachers who specialize in either subject and estimate our RD-DiD specification using the departure of one of these specialized teachers. Panel A asks how student math score changes relative to ELA test scores when the math teacher departs. Panel B asks how ELA scores change relative to math scores when the ELA teacher departs. Panel C asks how math scores change in school/grades with a departing math teacher relative to school/grades with a departing ELA teacher. Analogously, Panel D asks how ELA scores change in school/grades with a departing ELA teacher relative to school/grades with a departing math teacher.[29]

[29]Panels A–B exploit "within school" variation by comparing changes in test scores across the two subjects,

These alternative specifications show evidence that the TExES reform *reduced* student achievement in math. In Table 8, the RD-DiD coefficients for demographic indices are small and statistically insignificant effects, suggesting that our specifications that restrict identified to only grades 3–4 are balanced. In Panels A and C, we find negative and statistically significant RD-DiD coefficients for both math scores and math score residuals, with estimates ranging from −0.029 SDs to −0.055 SDs. In other words, the effect of math teacher departures on student math scores declined with the reform relative to ELA scores for the same students (Panel A) and relative to math scores at other schools with departing ELA teachers (Panel C). These results could be driven by the reduction in the supply for elementary teachers documented in Table 4, which may have made it harder for schools to find qualified math teachers. We do not find statistically significant effects on ELA test scores following the departure of ELA teachers in either Panel B or Panel D.

Our finding that the TExES reform did not improve student achievement is robust to other specifications. A caveat with our RD-DiD approach is that it only estimates effects in the first year of a teacher departure ($\tau_{ty} = 0$). In Appendix Table A16, we modify the RD regression (6) so that our RD-DiD specification estimates changes in outcomes from three years before to three years after the teacher departure ($\tau_{ty} = -3$ to $+3$). We continue to find small and statistically insignificant effects on student test scores in this specification, although there is imbalance on student demographics in some specifications (which is why we prefer the RD model that estimates effects at $\tau_{ty} = 0$). Our point estimates for math and ELA scores are relatively stable and statistically insignificant across a wide range of RD bandwidths, although we find demographic imbalance using the ELA score index with large bandwidth (Appendix Figure A12). We do not find systematic patterns of heterogeneity in the RD-DiD estimates for student test scores by race, although we find a statistically significant 0.028 SD decline in math scores for White students (Appendix Table A17).

In sum, we find that the switch from the easier EC-4 exam to the harder EC-6 exam reduced the share of teachers from URM groups and without any measurable improvements in student achievement, and some evidence of negative effects on student math performance. These results corroborate our findings that certification exams impose disproportionate costs on URM prospective teachers (Section 3) and are effective at screening out teachers with lower value-added (Section 5).

while Panels C–D exploit "across school" variation by comparing changes in scores in a given subject across departing teachers who teach different subjects.

7 Conclusion

Dating back to the 1980s, numerous lawsuits have been filed alleging that teacher certification exams have a disparate impact on prospective teachers from underrepresented minority (URM) groups, which violates Title VII of the U.S. Civil Rights Act. The legal argument for disparate impact relies on two claims: 1) the exams impose disproportionate costs on URM teaching candidates because they have lower pass rates; and 2) the exams are not sufficiently related to teaching performance to justify the pass rate disparities.

Both of these claims are difficult to prove, which explains the mixed success of these lawsuits. While New York City recently agreed to pay $1.8 billion to Black and Hispanic teachers to settle a more than 25-year lawsuit over its certification tests,[30] similar lawsuits were withdrawn in Texas and Arkansas after plaintiffs failed to show that the exams imposed a significant cost on individuals who failed.[31] In a lawsuit between the Association of Mexican-American Educators and the State of California, both District and Circuit Court judges ruled in favor of the State, arguing that it had successfully shown that the exam was a valid measure of job performance.[32]

The academic literature too provides little conclusive evidence on the costs and benefits of certification exams. No study to date has estimated the economic costs of failing certification exams for prospective teachers. While there is a large literature that correlates certification exam performance with teacher value-added (e.g., Goldhaber, 2007; Clotfelter et al., 2010), these analyses suffer from the fundamental problem that teaching quality is not observed for candidates who do not manage to pass the exams. As a result of this inconclusive evidence, the vast majority of U.S. states continue to require that individuals pass certification exams to receive a teaching license.

This paper presented new evidence on both the economic costs and predictive validity of certification exams. Using administrative data from Texas on certification exams, K–12 teachers, and students, we found that individuals who marginally fail certification exams are delayed from starting their teaching careers, ultimately costing them roughly $10,000 in forgone earnings. Not only are Black and Hispanic teaching candidates more likely to bear these costs given their lower pass rates, but they also have more negative causal effects of failing. Ultimately, we found that increases in the difficulty of certification exams reduce the share of teachers who are Black and Hispanic, thus contributing to their under-representation

[30] See Green (2023) and https://clearinghouse.net/case/11541/.

[31] The Texas and Arkansas lawsuits were withdrawn because most candidates who initially failed ended up passing after subsequent retaking the exam, and thus it was argued that the exams did not create a significant harm. See: https://www.edweek.org/education/arkansas-union-drops-suit-over-teacher-tests/1987/11.

[32] See: https://clearinghouse.net/case/10826/.

in the profession relative to the racial composition of students.

Our paper also showed that more difficult certification exams do not increase the quality of the teaching workforce. For this we exploited a reform that significantly increased the difficulty of Texas' certification exam for elementary teachers. This reform allowed us to observe the teaching quality of individuals who passed the easier pre-reform exam but would not have passed the harder post-reform exam, thus overcoming the overcoming the typical challenge inherent in measuring certification exam validity under plausible assumptions. We found that the distribution of teacher value-added was similar for teachers who went through the easier and harder exams, showing that the harder exam was not a more effective screen for teaching quality. A complementarity analysis showed that the switch to the harder elementary certification exam did not improve student achievement on standardized tests, and we found evidence of reductions in math scores in some specifications.

Taken together, our findings support the plaintiffs' claim that certification exams have a disparate impact on URM teaching candidates. Within the population of prospective teachers who would have passed the easier pre-reform exam, our preferred estimates suggest that URM teachers had 10–13pp lower pass rates on the harder exams than White teachers with comparable teaching quality. Combined with our evidence on the career consequences of failing exams, this suggests that certification exams impose disproportionate economic costs on URM teaching candidates.

An important caveat is that our findings pertain to the elementary teacher certification exams in Texas. It is possible that certification exams are more informative for teaching quality at higher grade levels, but we cannot credibly test for such effects given the nature of the Texas exam reform that underlies our analysis. More broadly, examinations are widely used to determine certification and licensing in other professions, and they are also a requirement for high school graduation in some states. These exams typically exhibit large disparities in pass rates between White and URM exam takers, but it is similarly challenging to measure their validity and economic consequences. We hope future research will shed light on both the costs and potential benefits of certification exams in other settings.

References

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the chicago public high schools. *Journal of labor Economics*, 25(1):95–135.

Agency, T. E. (2022). Becoming a classroom teacher in texas. https://tea.texas.gov/texas-educators/certification/initial-certification/becoming-a-classroom-teacher-in-texas. Accessed 2022.

Alsan, M., Barnett, A., Hull, P., and Yang, C. S. (2025). "Something works" in US jails: Misconduct and recidivism effects of the IGNITE program. *The Quarterly Journal of Economics*, 140(2):1367–1415.

Angrist, J. D. and Guryan, J. (2008). Does teacher testing raise teacher quality? evidence from state certification requirements. *Economics of Education Review*, 27(5):483–503.

Arnold, D., Dobbie, W., and Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038.

Arnold, D., Dobbie, W. S., and Hull, P. (2020). Measuring racial discrimination in bail decisions. National Bureau of Economic Research Working Paper No. 26999.

Baron, E. J., Doyle Jr, J. J., Emanuel, N., Hull, P., and Ryan, J. (2024). Discrimination in multiphase systems: Evidence from child protection. *The Quarterly Journal of Economics*, 139(3):1611–1664.

Benson, A., Board, S., and Meyer-ter Vehn, M. (2024). Discrimination in hiring: Evidence from retail sales. *Review of Economic Studies*, 91(4):1956–1987.

Blair, P. Q. and Chung, B. W. (2025). Job market signaling through occupational licensing. *Review of Economics and Statistics*, 107(2):338–354.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2025a). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, 107(3):605–620.

Bohren, J. A., Hull, P., and Imas, A. (2025b). Systemic discrimination: Theory and measurement. *The Quarterly Journal of Economics*, page qjaf022.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying society's leaders? the determinants and causal effects of admission to highly selective private colleges. Technical report, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014c). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–79.

Chingos, M. M. and Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3):449–465.

Chung, B. W. and Zou, J. (2025). Occupational licensing in us public schools: Nationwide implementation of teacher performance assessment. *Journal of Public Economics*, 244:105328.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3):655–681.

Cowan, J., Goldhaber, D., Jin, Z., and Theobald, R. (2020). Teacher licensure tests: Barrier or predictive tool? working paper no. 245-1020. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Fryer Jr, R. G. and Levitt, S. D. (2006). The black-white test score gap through third grade. *American law and economics review*, 8(2):249–281.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of human Resources*, 42(4):765–794.

Goldhaber, D., Gratz, T., and Theobald, R. (2017). What's in a teacher test? assessing the relationship between teacher licensure test scores and student stem achievement and course-taking. *Economics of Education Review*, 61:112–129.

Goldhaber, D. and Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47(1):218–251.

Goncalves, F. and Mello, S. (2021). A few bad apples? racial bias in policing. *American Economic Review*, 111(5):1406–41.

Green, E. (2023). The $1.8-billion lawsuit over a teacher test. *The New Yorker*.

Hendricks, M. (2016). Teacher characteristics and productivity: Quasi-experimental evidence from teacher mobility. *Available at SSRN 2822041*.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 126(5):2072–2107.

Jäger, S. and Heining, J. (2022). How substitutable are workers? evidence from worker deaths. NBER Working Paper No. 30629.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education review*, 27(6):615–631.

Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. National Bureau of Economic Research.

Kline, P., Rose, E. K., and Walters, C. R. (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics*, 137(4):1963–2036.

Koedel, C. and Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47:180–195.

Larsen, B., Ju, Z., Kapor, A., and Yu, C. (2020). The effect of occupational licensing stringency on the teacher quality distribution. Technical report, National Bureau of Economic Research.

Law, M. T., Marks, M., and Stern, T. (2023). Teacher testing standards and the new teacher pipeline. *Available at SSRN 4644356*.

Law, M. T. and Marks, M. S. (2009). Effects of occupational licensing laws on minorities: Evidence from the progressive era. *The Journal of Law and Economics*, 52(2):351–366.

Lepage, L.-P. (2024). Experience-based discrimination. *American Economic Journal: Applied Economics*, 16(4):288–321.

Orellana, A. and Winters, M. A. (2023). Licensure tests and teacher supply. Technical report, Working Paper. Retrieved from https://alexisorellana. github.io/assets/pdf ....

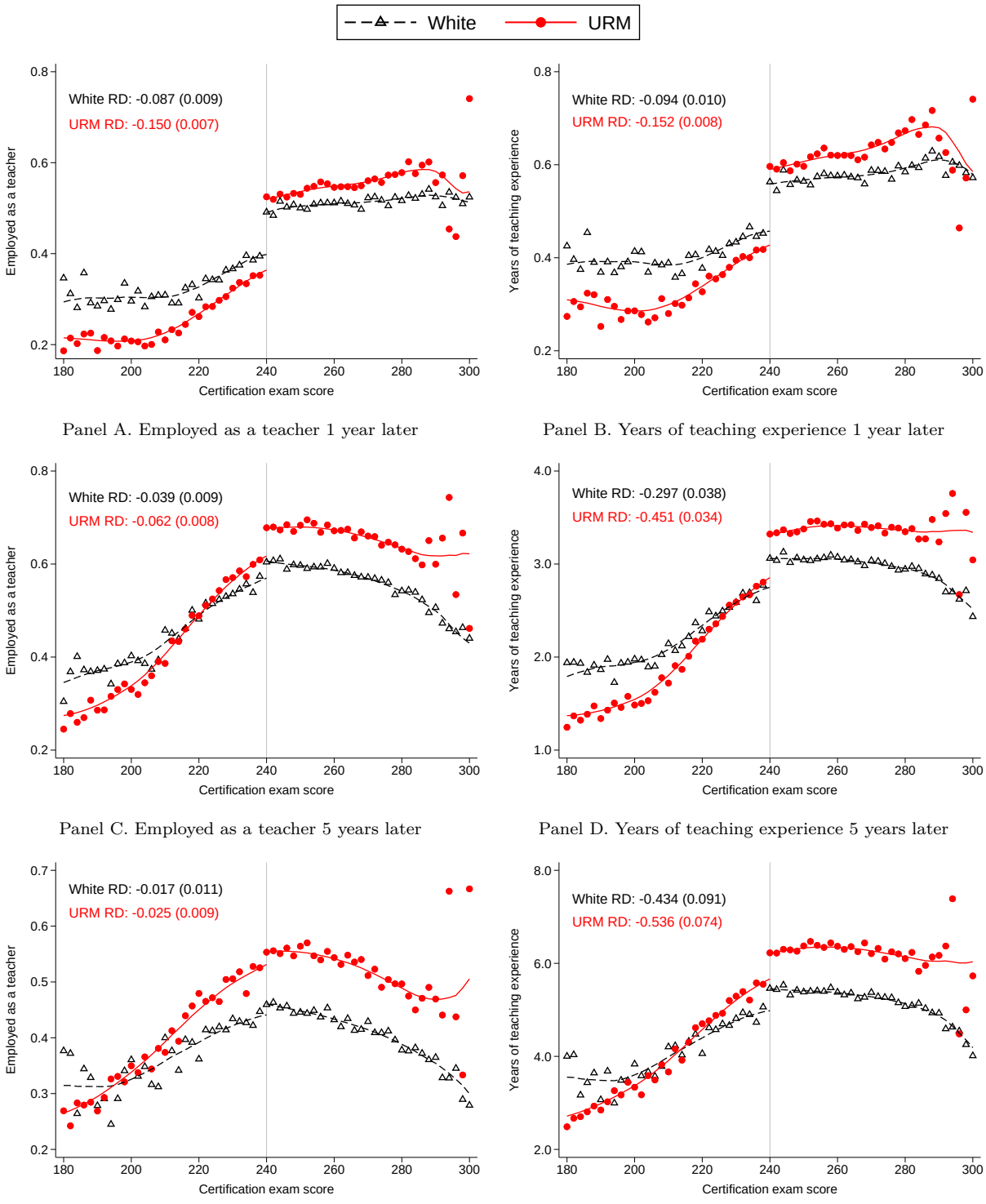
Shuls, J. V. and Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy*, 29(4):645–675.

Small, M. L. and Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34(2):49–67.

Xia, X. (2021). Barrier to entry or signal of quality? the effects of occupational licensing on minority dental assistants. *Labour Economics*, 71.

Yan, Z. (2024). The effect of occupational licensing on the gender wage gap. *Available at SSRN 5090345*.

Figures



Panel A. Employed as a teacher 1 year later

Panel B. Years of teaching experience 1 year later

Panel C. Employed as a teacher 5 years later

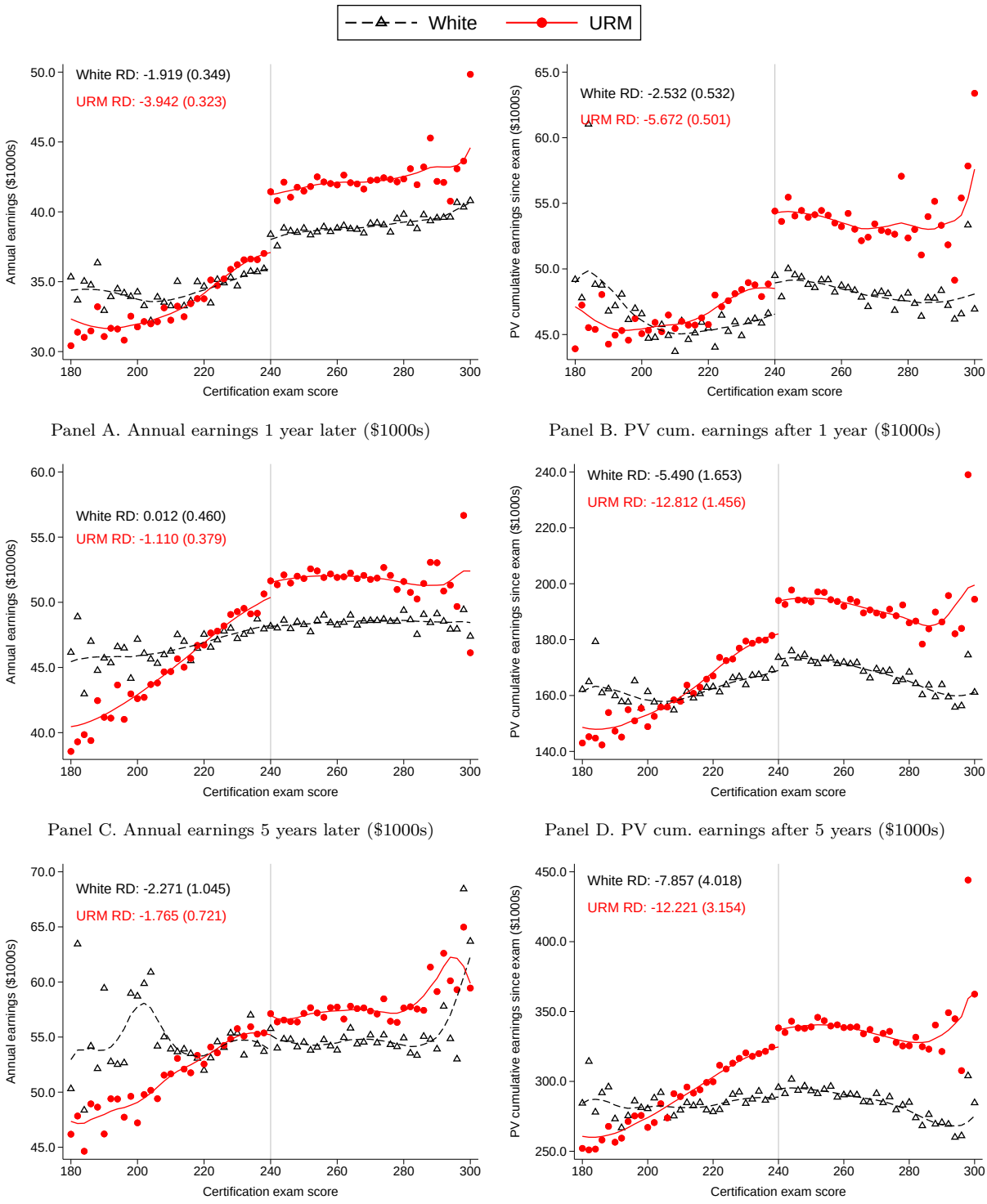
Panel D. Years of teaching experience 5 years later

Panel E. Employed as a teacher 10 years later

Panel F. Years of teaching experience 10 years later

Figure 1: RD plots: Teaching outcomes by race

Notes: The figure presents the RD effects of failing a certification exam on teaching outcomes by white and URM groups. Panels A-F display the RD effects on teacher employment and years of teaching experience one, five, and ten years after the exam, respectively. The RD estimates are shown at the upper-left corner, with standard deviations included in the parentheses.



Panel A. Annual earnings 1 year later ($1000s)

Panel B. PV cum. earnings after 1 year ($1000s)

Panel C. Annual earnings 5 years later ($1000s)

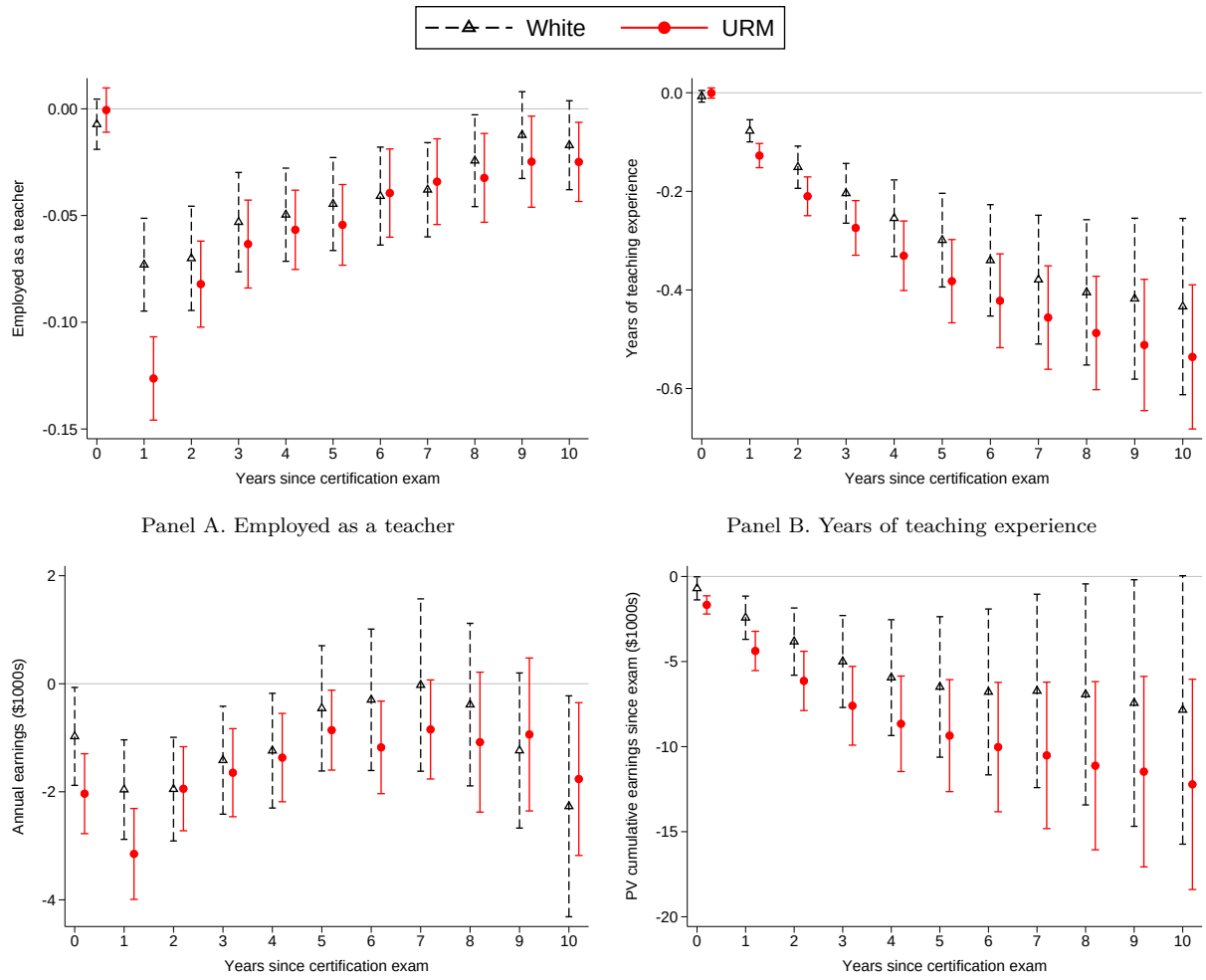
Panel D. PV cum. earnings after 5 years ($1000s)

Panel E. Annual earnings 10 year later ($1000s)

Panel F. PV cum. earnings after 10 years ($1000s)

Figure 2: RD plots: Earnings outcomes by race

Notes: The figure presents the RD effects of failing a certification exam on earnings outcomes by white and URM groups. Panels A-F display the RD effects on annual earnings and the present value of cumulative earnings one, five, and ten years after the exam, respectively. The RD estimates are shown at the upper-left corner, with standard deviations included in the parentheses.



Panel A. Employed as a teacher

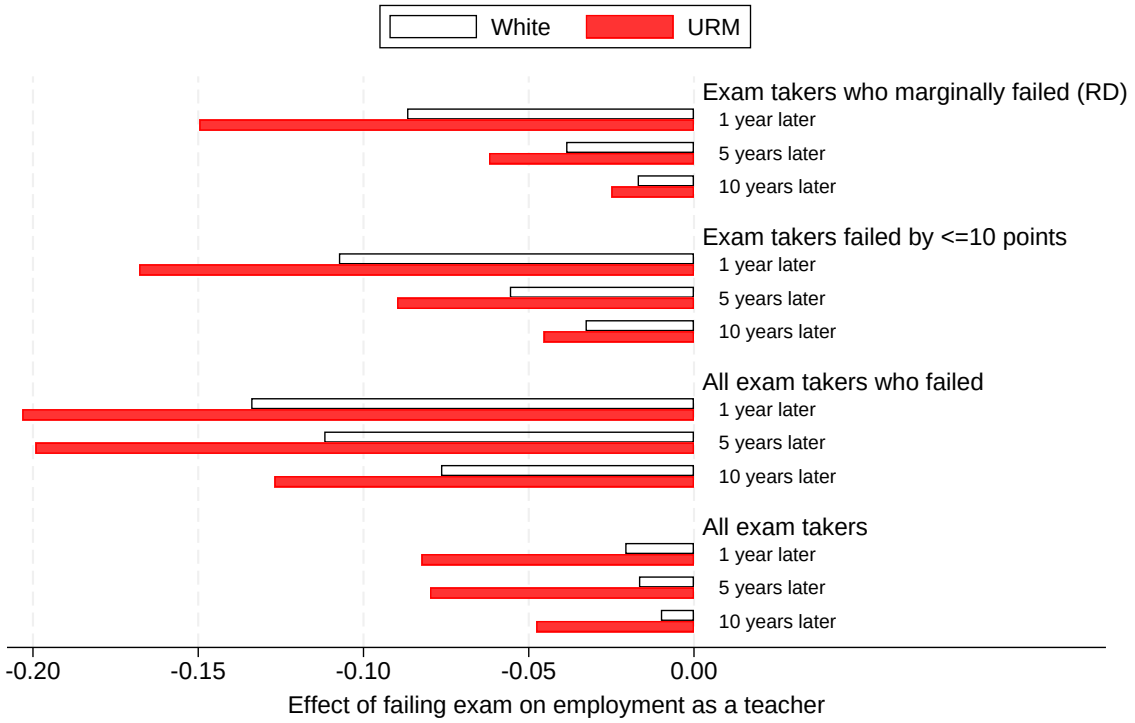
Panel B. Years of teaching experience

Panel C. Annual earnings ($1000s)

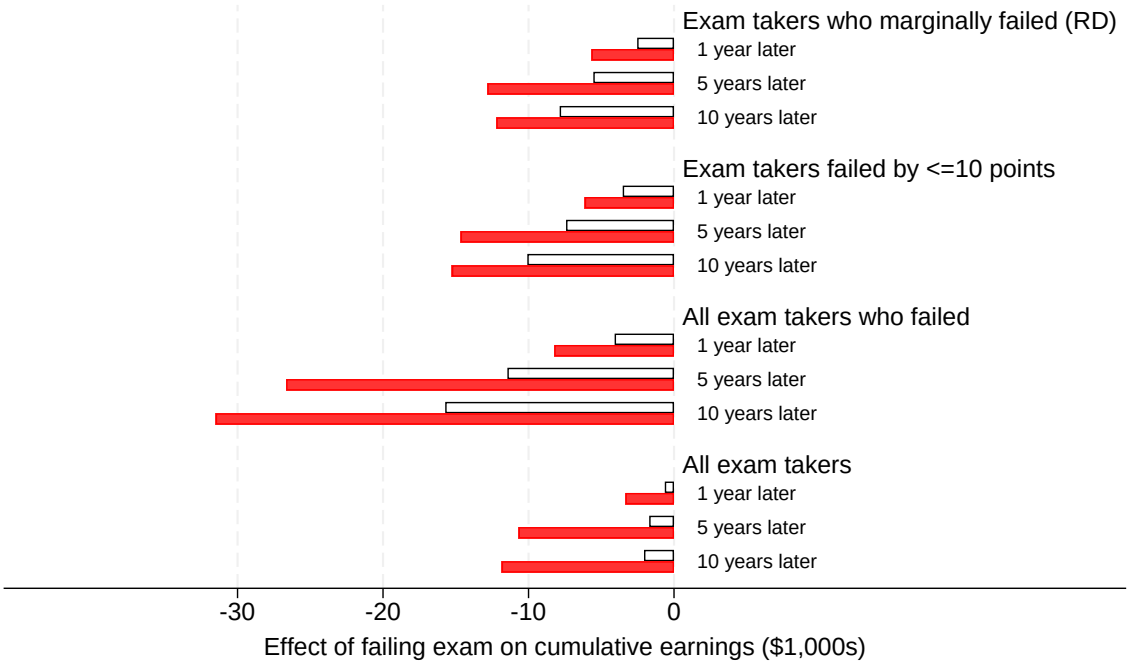
Panel D. PV cumulative earnings since exam ($1000s)

Figure 3: RD effects of failing a certification exam over time — Balanced panel

Notes: The figure presents the RD effects of failing a certification exam on teaching and earnings outcomes by white and URM groups. Panels A-F display the RD effects on teaching employment, years of teaching experience, annual earnings, and the present value of cumulative earnings, respectively. The figure shows lasting RD effects over each of the post-period up to ten years after failing the exam, using a balanced panel sample.



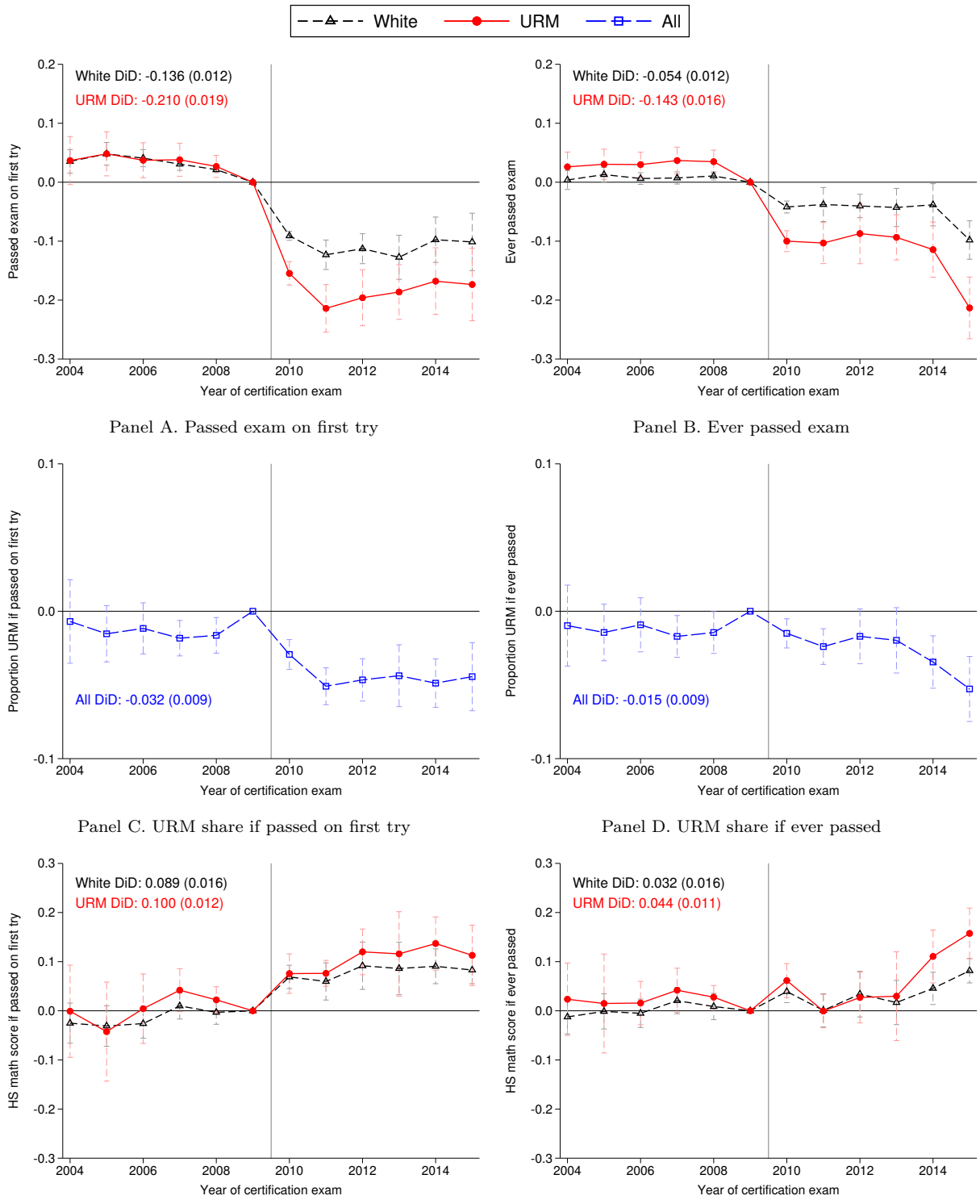
Panel A. Employed as a teacher



Panel B. PV cumulative earnings since exam ($1000s)

Figure 4: Extrapolated effects of failing certification exam on outcomes

Notes: The figure summarizes the baseline RD effects and extrapolated effects of the three samples — exam takers with a score within ten points of the passing score, all exam takers with a score below the 240 cutoff, and all exam takers — by white and URM groups in one, five, and ten years after the exam. Panels A and B display the effects on teaching employment and the present value of cumulative earnings, respectively.



Panel A. Passed exam on first try

Panel B. Ever passed exam

Panel C. URM share if passed on first try

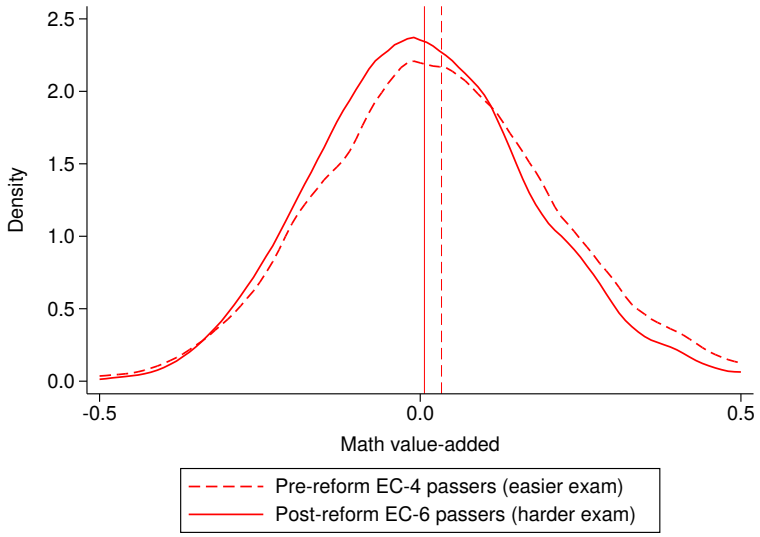
Panel D. URM share if ever passed

Panel E. HS math score if passed on first try

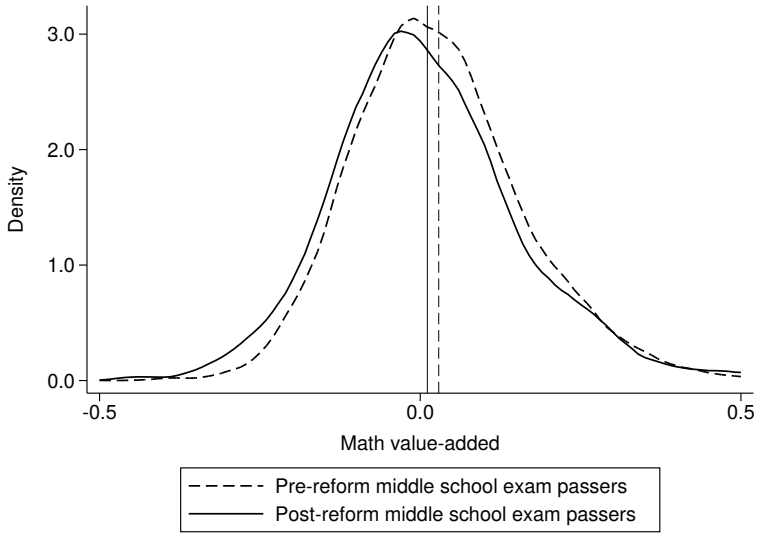
Panel F. HS math score if ever passed

Figure 5: Event studies: Effects of harder certification exams on exam outcomes

Notes: This figure shows the effects of the TExES reform on exam outcomes (Panels A–B) and the composition of individuals who took/passed the exams (Panels C–F). The graphs display estimates from an event-study version of equation (2) that computes separate $\beta_{t(i)}$ coefficients for each exam year $t(i)$ (omitting 2009). Each figure displays the β coefficient from equation (2) and with its standard error in parentheses. We estimate these regressions separately for White, URM, or all exam takers, as indicated by the legend. Dashed lines contain 95 percent confidence intervals using standard errors clustered at the exam level.



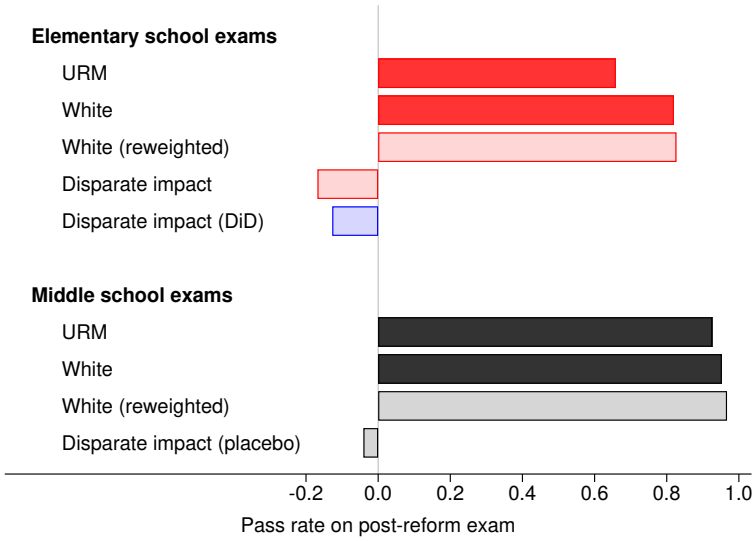
Panel A. Elementary school exams



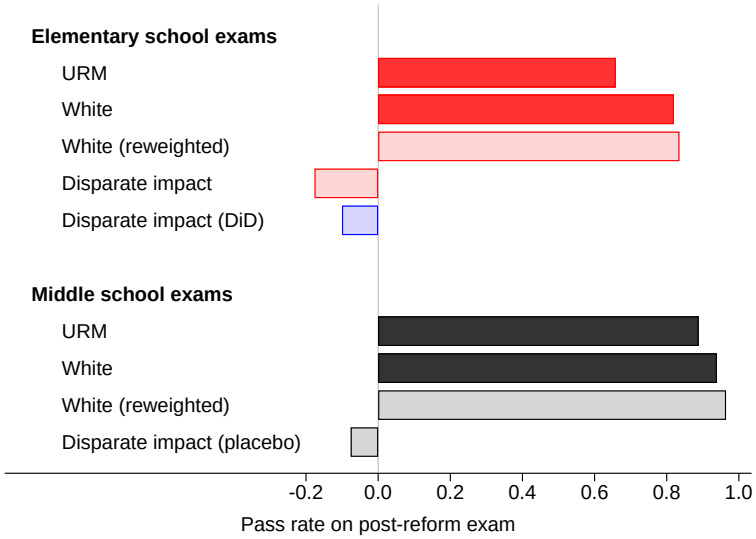
Panel B. Middle school exams

Figure 6: Distributions of math value-added for exam passers measured 3–9 years later

Notes: This figure shows kernel densities of math value-added for individuals who passed different TExES certification exams. In Panel A, the sample includes individuals who took the EC-4 or EC-6 exams (Panel A of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 4 teachers. The graph plots distributions of grade 4 math value-added for individuals who took the EC-4 exam in 2002–2009 (dashed line) or the EC-6 exam in 2010–2015 (solid line). In Panel B, the sample includes individuals who took middle school Generalist or math exams (Panel B of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 7–8 teachers. The graph plots distributions of grade 7–8 math value-added for individuals who took the middle school exams in 2002–2009 (dashed line) or in 2010–2015 (solid line). We measure value-added 3–9 years after taking the certification exam ("potential experience"), and weight observations so that the distribution of potential experience is the same for the two exams in each graph (using the exam with the minimum number of observations for each level of potential experience). Vertical lines denote the mean of each distribution.



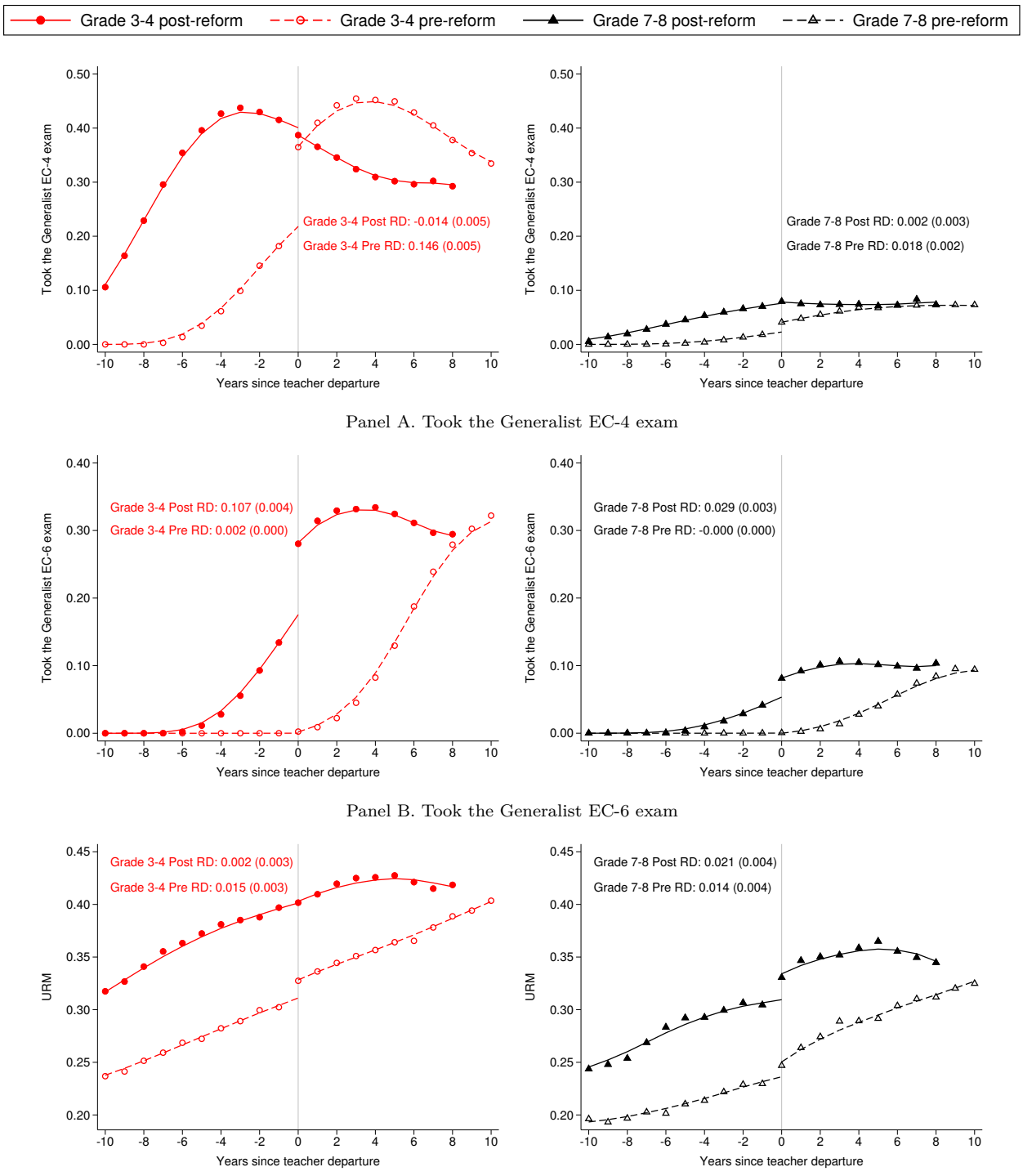
Panel A. Math value-added



Panel B. ELA value-added

Figure 7: Policy-relevant disparate impact of EC-6 exam relative to EC-4 exam

Notes: This figure shows estimates of the policy-relevant disparate impact of EC-6 exam relative to EC-4 exam using math value-added (Panel A) and ELA value-added (Panel B) as measures of teaching quality. For elementary teachers, the solid red bars display EC-6 pass rates divided by EC-4 pass rates estimated separately for URM and White exam takers (i.e., the first term on the righthand side in equation 5). The "White (reweighted)" bar shows the same ratio of pass rates for White exam takers after it has been reweighted to match the distribution of teacher value-added for URM exam takers. For this, we group value-added into deciles and compute the proportion of observations in each decile for each exam. We do this separately by race and potential experience (i.e., years since the exam), and reweight the cell-specific pass rates for White exam takers to match the proportion of observations in each of the URM cells. The "Disparate impact" bar shows the difference between the URM pass rate and the White (reweighted) pass rate. We do the same calculation for middle school certification exams using pre-reform (2002–2009) and post-reform (2010–2015) exams to compute a placebo measure of disparate impact. The "Disparate impact (DiD)" bar reports the difference between the "Disparate impact" bar for elementary exams and the "Disparate impact (placebo)" bar for middle school exams.



Panel A. Took the Generalist EC-4 exam

Panel B. Took the Generalist EC-6 exam

Panel C. URM

Figure 8: Effects of teacher departures on teacher composition

Notes: This figure contains RD graphs that show how teacher departures affect the composition of teachers in a school/grade. Red circles represent grade 3–4 teachers. Black triangles represent grade 7–8 teachers. Hollow symbols depict teacher departures that occurred in $y \in$ 2005–2010 (pre-reform). Solid symbols depict teacher departures that occurred in $y \in$ 2011–2016 (post-reform). The x-axis is years relative to the teacher departure, τ_{ty}. The y-axis depicts the average outcome at the school/-grade/year level. Each graph displays RD coefficients β from equation (6) with standard errors clustered at the school level in parentheses.

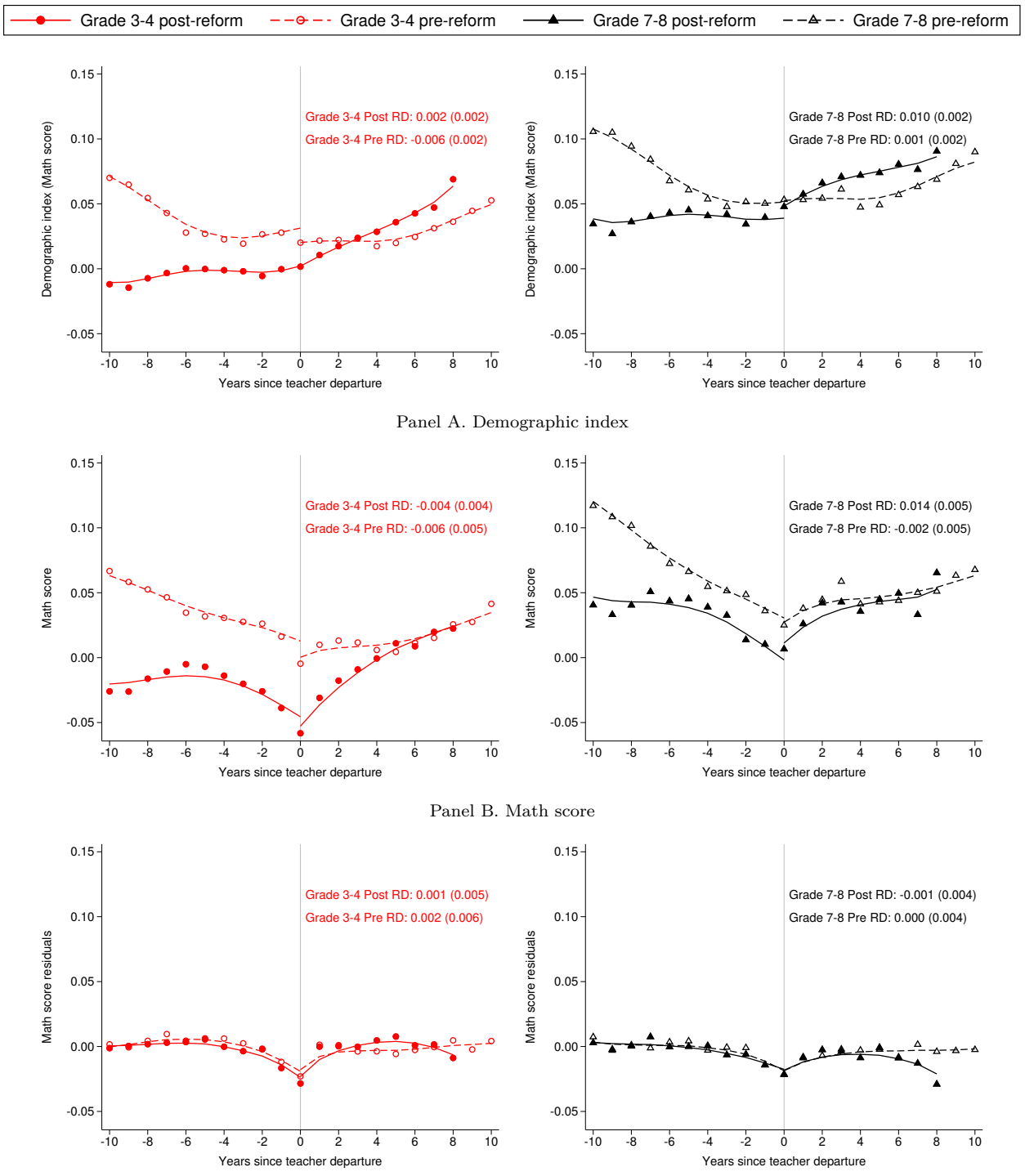


Figure 9: Effects of teacher departures on student math achievement

Notes: This figure contains RD graphs that show how teacher departures affect the math achievement of students in a school/-grade. Red circles represent grade 3–4 teachers. Black triangles represent grade 7–8 teachers. Hollow symbols depict teacher departures that occurred in $y \in$ 2005–2010 (pre-reform). Solid symbols depict teacher departures that occurred in $y \in$ 2011–2016 (post-reform). The x-axis is years relative to the teacher departure, τ_{ty}. The y-axis depicts the average outcome at the school/grade/year level. Each graph displays RD coefficients β from equation (6) with standard errors clustered at the school level in parentheses.

Tables

Table 1: Summary statistics for TExES exam takers and comparison populations in 2002–2019

	(A)	(B)	(C)	(D)	(E)	(F)
	First-time certification exam takers			First-year teachers	New BA graduates	K–12 students
	White	URM	All	All	All	All
Panel A. Demographics						
Female	0.777	0.742	0.762	0.747	0.581	0.487
White	1.000	0.000	0.588	0.590	0.524	0.326
Black	0.000	0.307	0.119	0.113	0.094	0.134
Hispanic	0.000	0.693	0.269	0.261	0.265	0.488
HS math score	0.626	0.250	0.483	0.552	0.676	0.043
HS ELA score	0.700	0.307	0.546	0.602	0.632	0.041
Panel B. TExES content exam performance						
Content exam score on first attempt	257.8	241.5	252.1	254.0		
Passed content exam on first attempt	0.833	0.573	0.739	0.787		
Ever passed first content exam	0.925	0.769	0.863	0.927		
Ever passed any content exam	0.959	0.851	0.913	0.979		
Panel C. TExES PPR exam performance						
PPR exam score on first attempt	266.6	254.6	262.5	263.3		
Passed PPR exam on first attempt	0.945	0.807	0.894	0.907		
Ever passed first PPR exam	0.992	0.955	0.976	0.987		
Ever passed any PPR exam	0.994	0.966	0.982	0.992		
Panel D. Annual earnings (2019 USD)						
Earnings in year of exam/first teaching/graduation	26,492	26,909	29,158	47,716	24,529	
Earnings 1 year later	37,195	37,481	39,286	48,491	38,482	
Earnings 5 years later	49,422	50,668	50,528	51,806	57,513	
N (# unique individuals)	292,785	193,094	662,332	348,800	1,472,457	13,313,603

Notes: The table presents the summary statistics (mean) for TExES exam takers in the first three columns and comparison populations in the last three columns (first-year teachers, bachelor degree recipients, and K-12 students, respectively) between 2002 and 2019. The sample first-time certification exam takers is further separated into Whie and URM groups. Panel A displays demographic variables, including gender, race, and high school achievement. Panel B presents TExES content exam performance, including exam scores and passing rates. Panel C reports the TExES PPR exam performance, including the exam scores and passing rate. Panel D shows earnings (deflated to constant 2019 US dollars), including earnings in the year of exam/first teaching/graduation, and cumulative earnings one and five years later (in present value). "N (# unique individuals)" indicates the unique number of observation for each of the populations.

Table 2: RD effects of failing a certification exam on outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Employed as a teacher	0.497	0.525	-0.118***	-0.087***	-0.150***	0.000
			(0.005)	(0.009)	(0.007)	
Years of teaching experience	0.565	0.597	-0.122***	-0.094***	-0.152***	0.000
			(0.007)	(0.010)	(0.008)	
Annual earnings ($1000s)	38.239	41.432	-2.831***	-1.919***	-3.942***	0.000
			(0.213)	(0.349)	(0.323)	
PV cumulative earnings since exam ($1000s)	49.084	54.467	-4.062***	-2.532***	-5.672***	0.000
			(0.321)	(0.532)	(0.501)	
N (Employed as a teacher)	16,683	16,896	506,345	227,740	156,499	384,239
Panel B. 5 years after certification exam						
Employed as a teacher	0.608	0.677	-0.048***	-0.039***	-0.062***	0.046
			(0.006)	(0.009)	(0.008)	
Years of teaching experience	3.076	3.342	-0.364***	-0.297***	-0.451***	0.003
			(0.026)	(0.038)	(0.034)	
Annual earnings ($1000s)	48.275	51.687	-0.552*	0.012	-1.110***	0.058
			(0.290)	(0.460)	(0.379)	
PV cumulative earnings since exam ($1000s)	173.604	194.717	-9.062***	-5.490***	-12.812***	0.001
			(1.039)	(1.653)	(1.456)	
N (Employed as a teacher)	14,528	14,425	442,510	201,864	133,272	335,136
Panel C. 10 years after certification exam						
Employed as a teacher	0.459	0.553	-0.020***	-0.017	-0.025***	0.587
			(0.007)	(0.011)	(0.009)	
Years of teaching experience	5.477	6.248	-0.464***	-0.434***	-0.536***	0.390
			(0.059)	(0.091)	(0.074)	
Annual earnings ($1000s)	54.772	56.658	-1.595***	-2.271**	-1.765**	0.697
			(0.559)	(1.045)	(0.721)	
PV cumulative earnings since exam ($1000s)	296.041	338.626	-9.668***	-7.857*	-12.221***	0.400
			(2.390)	(4.018)	(3.154)	
N (Employed as a teacher)	9,848	9,734	310,404	145,790	88,718	234,508

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teaching and earning outcomes. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether being employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the *p*-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table 3: TExES certification exams for math and ELA teachers, 2002–2015

(A) TExES Exam	(B) First offered	(C) Last offered	(D) # exam takers	(E) First try pass rate	(F) Pass rate w/ HS score controls
Panel A. Elementary school exams					
Generalist EC-4	2002	2010	137,375	0.827	
Generalist EC-6	2010	2015	140,609	0.655	-0.180
Bilingual Generalist EC-4	2002	2010	29,397	0.642	
Bilingual Generalist EC-6	2010	2015	17,290	0.442	-0.242
English as a Second Language (ESL)/Generalist EC-4	2003	2010	8,577	0.892	
English as a Second Language (ESL)/Generalist EC-6	2010	2015	10,937	0.655	-0.265
Panel B. Middle school exams					
Generalist 4-8	2002	2015	83,059	0.895	
Bilingual Generalist 4-8	2002	2015	4,038	0.383	-0.462
English as a Second Language (ESL)/Generalist 4-8	2003	2015	3,785	0.640	-0.245
Mathematics 4-8	2002	2022	49,905	0.646	-0.275
English Language Arts and Reading 4-8	2002	2022	37,785	0.852	0.004
Mathematics/Science 4-8	2002	2022	8,635	0.696	-0.238
English Language Arts and Reading/Social Studies 4-8	2002	2022	9,994	0.842	-0.044

Notes: This table shows the TExES exams for elementary (Panel A) and middle (Panel B) school math and ELA teachers that were offered between 2002 and 2015. Column (A) shows the name of each exam. Columns (B) and (C) show the first and the last year in which exam was offered (up through the end of our data in 2022). Column (D) shows the total number of exam takers in 2002–2015. Column (E) shows the proportion of exam takers who passed on their first attempt at the exam. For column (F), we regress an indicator for passing the exam on the first attempt on a cubic in teachers' high school math and ELA scores plus dummies for each exam. We estimate these regressions separately for each pair of EC-4/EC-6 exams in Panel A and for all exams in Panel B, omitting the dummy for the first exam in each group. Column (F) shows the coefficients on the exam dummies in these regressions.

Table 4: Effects of harder certification exams on exam outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Pre-reform means		DiD coefficients			*P*-value: White=URM
	White	URM	All	White	URM	
Panel A. Exam outcomes						
Passed exam on first try	0.930	0.660	-0.159***	-0.136***	-0.210***	0.000
			(0.014)	(0.012)	(0.019)	
Ever passed exam	0.974	0.827	-0.094***	-0.054***	-0.143***	0.000
			(0.014)	(0.012)	(0.016)	
Number of times taking this exam	1.092	1.530	0.338***	0.271***	0.538***	0.000
			(0.026)	(0.019)	(0.051)	
Days between first try and passing	8.836	50.638	33.059***	26.737***	57.369***	0.000
			(3.681)	(3.182)	(6.203)	
Panel B. Racial distribution of exam takers						
URM share of first-time exam takers		0.418	0.009			
			(0.012)			
URM share of exam takers who passed on first try		0.339	-0.032***			
			(0.009)			
URM share of exam takers who ever passed		0.380	-0.015*			
			(0.009)			
Panel C. Number of exam takers						
Log number of first-time exam takers	8.640	7.874	0.005	0.010	-0.059	0.664
			(0.082)	(0.067)	(0.145)	
Log total number of exams taken	8.717	8.295	0.219*	0.236***	0.226	0.957
			(0.117)	(0.068)	(0.186)	
Log number passed on first try	8.570	7.462	-0.228**	-0.153**	-0.464***	0.047
			(0.091)	(0.061)	(0.143)	
Log number ever passed	8.617	7.694	-0.134	-0.051	-0.298*	0.154
			(0.106)	(0.060)	(0.165)	
Panel D. HS test scores of exam takers						
HS math score of first-time exam takers	0.527	0.203	-0.015	0.001	-0.013	0.096
			(0.011)	(0.011)	(0.009)	
HS math score if passed on first try	0.580	0.371	0.099***	0.089***	0.100***	0.472
			(0.016)	(0.016)	(0.012)	
HS math score if ever passed	0.543	0.274	0.037**	0.032**	0.044***	0.314
			(0.015)	(0.016)	(0.011)	
N (# exam takers)	50,739	37,877	734,463	346,944	209,955	

Notes: This table shows the effects of the TExES reform on exam outcomes (Panel A) and the composition of individuals who took/passed the exams (Panels B–D). Columns (A)–(B) show means of each outcome in the pre-reform years (2004–2009) for White and URM exam takers. Columns (C)–(E) show β coefficients from the DiD regression (2) estimated separately for all, White, and URM exam takers. Column (F) shows P-values from a test of equality of the White and URM coefficients in columns (D)–(E). Standard errors in parentheses are clustered at the exam level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Teacher value added by certification exam performance and race

	(A)	(B)	(C)	(D)	(E)
	Estimates for all certification exam takers			Estimates for exam takers who passed	
Panel A. Math value added					
URM	0.007***			0.013***	
	(0.003)			(0.003)	
Passed certification exam * White		0.019***			
		(0.005)			
Passed certification exam * URM		0.007			
		(0.004)			
Certification exam score/10 * White			0.005***		0.005***
			(0.001)		(0.001)
Certification exam score/10 * URM			0.001		-0.002
			(0.001)		(0.002)
N (# students with Math scores)	9,067,841	9,067,841	9,067,841	9,067,841	9,067,841
P-value: White=URM	0.006	0.059	0.004	0.000	0.002
Panel B. ELA value added					
URM	0.010***			0.014***	
	(0.001)			(0.002)	
Passed certification exam * White		0.006**			
		(0.003)			
Passed certification exam * URM		-0.001			
		(0.003)			
Certification exam score/10 * White			0.002***		0.002***
			(0.001)		(0.001)
Certification exam score/10 * URM			-0.001		-0.002*
			(0.001)		(0.001)
N (# students with ELA scores)	9,489,789	9,489,789	9,489,789	9,489,789	9,489,789
P-value: White=URM	0.000	0.071	0.000	0.000	0.000
Fixed effects:					
Exam*Year	Y				
Exam*Year*Race		Y	Y		
Exam*Year*Passed exam				Y	
Exam*Year*Race*Passed exam					Y

Notes: The table presents the correlations between teacher effectiveness by race and certification performance using equations B3. Panels A and B examine the math and ELA value added, separately. Estimates in Columns (A)-(C) are obtained using all exam takers, while estimates in Columns (D)-(E) are based on exam takers who passed. "N (#students with Math-/ELA scores)" indicates the sample size, while "*P*-value: White = URM" shows *p*-value for estimate differences between white and URM groups.

Table 6: RD-DiD effects of harder certification exams on teacher composition

	(A)	(B)	(C)	(D)	(E)	(F)
	Grade 3–4 departures			Grade 7–8 departures		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Teachers' cert exams						
Took the Generalist EC-4 exam	0.415	-0.014***	0.146***	0.002	0.018***	-0.144***
		(0.005)	(0.005)	(0.003)	(0.002)	(0.008)
Took the Generalist EC-6 exam	0.134	0.107***	0.002***	0.029***	-0.000	0.076***
		(0.004)	(0.000)	(0.003)	(0.000)	(0.005)
Panel B. Teacher demographics						
URM	0.397	0.002	0.015***	0.021***	0.014***	-0.019**
		(0.003)	(0.003)	(0.004)	(0.004)	(0.008)
New teacher (within 5 years of cert)	0.223	0.190***	0.206***	0.250***	0.236***	-0.030***
		(0.005)	(0.006)	(0.007)	(0.006)	(0.012)
New URM teacher	0.097	0.073***	0.069***	0.087***	0.058***	-0.025***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
New White teacher	0.120	0.112***	0.126***	0.151***	0.156***	-0.008
		(0.004)	(0.005)	(0.005)	(0.006)	(0.010)
Experienced URM teacher	0.300	-0.070***	-0.054***	-0.064***	-0.044***	0.004
		(0.004)	(0.004)	(0.005)	(0.004)	(0.008)
Experienced White teacher	0.464	-0.113***	-0.142***	-0.169***	-0.180***	0.017*
		(0.004)	(0.005)	(0.006)	(0.006)	(0.010)
Panel C. Teacher pay						
Total salary for all teachers	197,199	-815	-7,997***	-2,246***	-2,611***	6,818***
		(811)	(1,094)	(523)	(547)	(1,584)
Total salary for URM teachers	79,801	30	603	2,002***	226	-2,349*
		(751)	(703)	(479)	(420)	(1,234)
Total salary for White teachers	113,641	-1,374*	-9,094***	-3,072***	-4,466***	6,327***
		(790)	(986)	(510)	(591)	(1,531)
N (# *sty* observations)	4,887	105,829	86,102	123,706	107,812	423,449

Notes: This table displays RD and RD-DiD estimates of the effects of the TExES reform on teacher composition as defined by their certification exams (Panel A), demographics (Panel B), and pay (Panel C). Column (A) shows the mean of each outcome in the year prior to the teacher departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in the post-reform years (2011–2016). Columns (B)–(E) show RD coefficients β from equation (6) estimated separately for grades 3–4 and 7–8, and for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Column (F) shows the RD-DiD coefficient θ from equation (7). Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: RD-DiD effects of harder certification exams on student achievement

	(A)	(B)	(C)	(D)	(E)	(F)
	Grade 3–4 departures			Grade 7–8 departures		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Student demographics (balance tests)						
Number of students with test scores	74.268	-0.107	-0.819**	-0.642	-2.799***	-1.445
		(0.253)	(0.347)	(0.675)	(0.668)	(1.063)
Male	0.498	-0.004***	0.001	-0.000	0.001	-0.004
		(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
White	0.262	0.003***	0.002	0.004***	0.002*	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Hispanic	0.537	-0.002**	0.001	-0.004***	-0.001	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Black	0.148	-0.000	-0.001	0.000	-0.002**	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Economically disadvantaged	0.667	-0.002	0.004**	-0.010***	-0.000	0.004
		(0.002)	(0.002)	(0.001)	(0.001)	(0.003)
In gifted education	0.096	-0.004***	-0.000	-0.000	0.002*	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
At risk of dropping out	0.485	-0.004	-0.000	-0.000	-0.017***	-0.020***
		(0.003)	(0.003)	(0.003)	(0.002)	(0.006)
Demographic index (Math score)	-0.000	0.002	-0.006***	0.010***	0.001	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Demographic index (ELA score)	-0.005	0.006***	-0.006***	0.010***	0.002	0.004
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Panel B. Student achievement						
Math score	-0.039	-0.004	-0.006	0.014**	-0.002	-0.014
		(0.004)	(0.005)	(0.005)	(0.005)	(0.010)
Math score residuals	-0.017	0.001	0.002	-0.001	0.000	-0.000
		(0.005)	(0.006)	(0.004)	(0.004)	(0.010)
ELA score	-0.035	0.000	-0.010**	0.009***	0.006	0.007
		(0.004)	(0.004)	(0.003)	(0.004)	(0.008)
ELA score residuals	-0.009	-0.004	-0.004	-0.004	0.006*	0.009
		(0.004)	(0.005)	(0.003)	(0.003)	(0.008)
N (# *sty* observations)	4,887	105,829	86,102	123,706	107,812	423,449

Notes: This table displays RD and RD-DiD estimates of the effects of the TExES reform on student composition (Panel A) and student achievement (Panel B). Column (A) shows the mean of each outcome in the year prior to the teacher departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in the post-reform years (2011–2016). Columns (B)–(E) show RD coefficients β from equation (6) estimated separately for grades 3–4 and 7–8, and for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Column (F) shows the RD-DiD coefficient θ from equation (7). Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

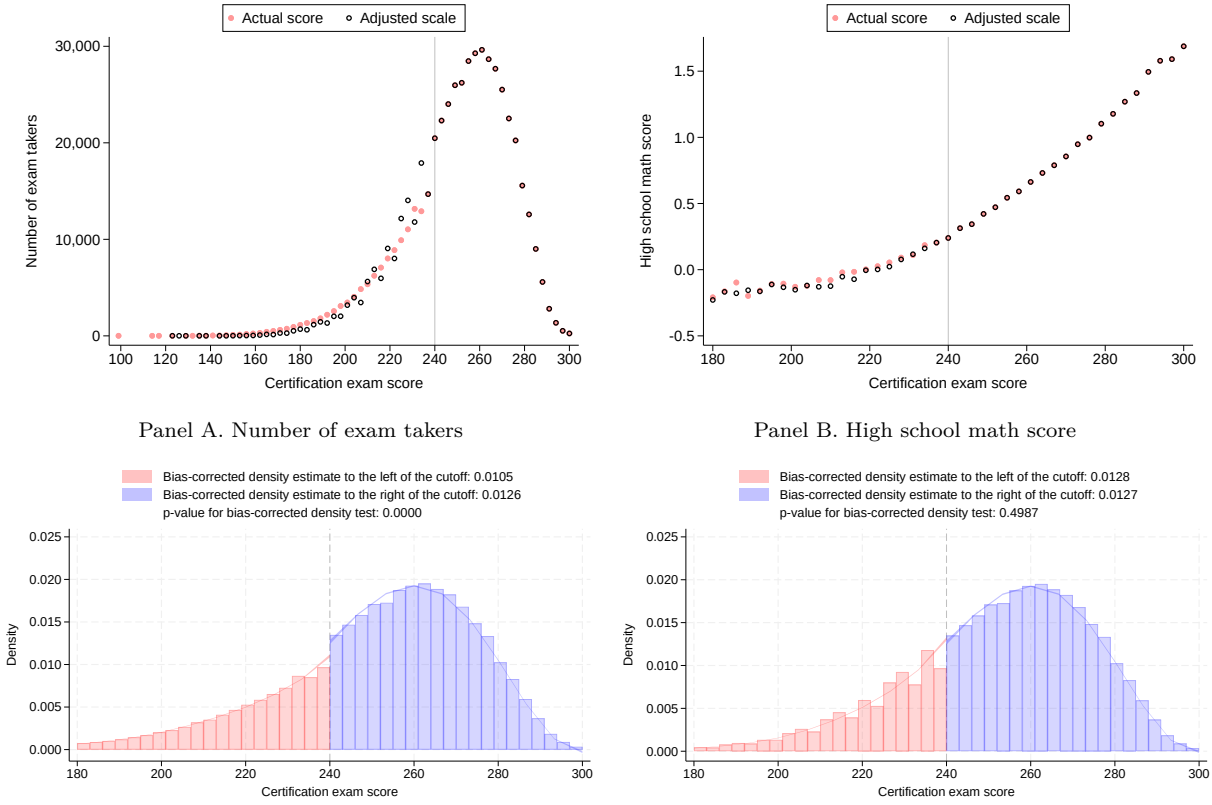
Table 8: RD-DiD effects of harder certification exams on student achievement — Grades 3–4 with various control groups

	(A)	(B)	(C)	(D)	(E)	(F)
	Departures in which teaching subject matches the exam subject			Departures in which subjects don't match		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Math vs. ELA scores when Math teacher departs (within school)						
Demographic index (Test score)	0.021	0.009***	-0.000	0.010**	0.002	0.001
		(0.004)	(0.005)	(0.004)	(0.005)	(0.002)
Test score	0.031	-0.012	0.022*	0.009	0.012	-0.030**
		(0.011)	(0.013)	(0.008)	(0.009)	(0.015)
Test score residuals	0.001	-0.021	0.049***	0.007	0.022*	-0.055***
		(0.013)	(0.015)	(0.009)	(0.011)	(0.018)
N (# *sty* observations)	740	14,776	9,578	14,775	9,577	48,706
Panel B. ELA vs. Math scores when ELA teacher departs (within school)						
Demographic index (Test score)	0.043	0.003	-0.002	0.004	-0.003	-0.002
		(0.003)	(0.004)	(0.003)	(0.004)	(0.002)
Test score	0.020	0.005	0.012*	0.004	0.011	-0.001
		(0.006)	(0.007)	(0.008)	(0.009)	(0.011)
Test score residuals	-0.015	0.015*	0.010	0.013	0.012	0.003
		(0.008)	(0.009)	(0.011)	(0.009)	(0.013)
N (# *sty* observations)	1,143	23,074	16,325	23,071	16,326	78,796
Panel C. Math teacher departs vs. ELA teacher departs (across school)						
Demographic index (Math score)	0.021	0.009***	-0.000	0.004	-0.003	0.007
		(0.004)	(0.005)	(0.003)	(0.004)	(0.006)
Math score	0.031	-0.012	0.022*	0.004	0.011	-0.029*
		(0.011)	(0.013)	(0.008)	(0.009)	(0.016)
Math score residuals	0.001	-0.021	0.049***	0.013	0.012	-0.039***
		(0.013)	(0.015)	(0.011)	(0.009)	(0.013)
N (# *sty* observations)	740	14,776	9,578	23,071	16,326	63,751
Panel D. ELA teacher departs vs. Math teacher departs (across school)						
Demographic index (ELA score)	0.043	0.003	-0.002	0.010**	0.002	-0.004
		(0.003)	(0.004)	(0.004)	(0.005)	(0.006)
ELA score	0.020	0.005	0.012*	0.009	0.012	-0.002
		(0.006)	(0.007)	(0.008)	(0.009)	(0.012)
ELA score residuals	-0.015	0.015*	0.010	0.007	0.022*	0.007
		(0.008)	(0.009)	(0.009)	(0.011)	(0.010)
N (# *sty* observations)	1,143	23,074	16,325	14,775	9,577	63,751

Notes: This table displays RD and RD-DiD estimates of the effects of the TExES reform on student composition and student achievement. The structure of this table is similar to Table 7, but in this table we restrict the sample to students in grades 3–4 who have a departing teacher that is identified in the data as specifically teaching math or ELA, and we use test scores or students in unaffected subjects as the control group. In Panels A–B, the regression sample includes school/grades with a departing math (Panel A) and ELA (Panel B) teacher, and the RD coefficients show how teacher departures affect exam scores in the same subject (columns B–C) relative to exam scores in the other subject (columns D–E). In Panels C–D, the outcome variables are math scores (Panel C) and ELA scores (Panel D), and the RD coefficients show how teacher departures affect test scores when the departing teacher teaches the same subject (columns B–C) relative to when the departing teacher teaches the other subject (columns D–E). In all panels, column (F) shows RD-DiD coefficient that equal column (B) − column (C) − (column D − column E). See Appendix B.3.3 for details on the sample and these RD-DiD specifications. Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix — For Online Publication Only

A Appendix Figures and Tables



Panel A. Number of exam takers

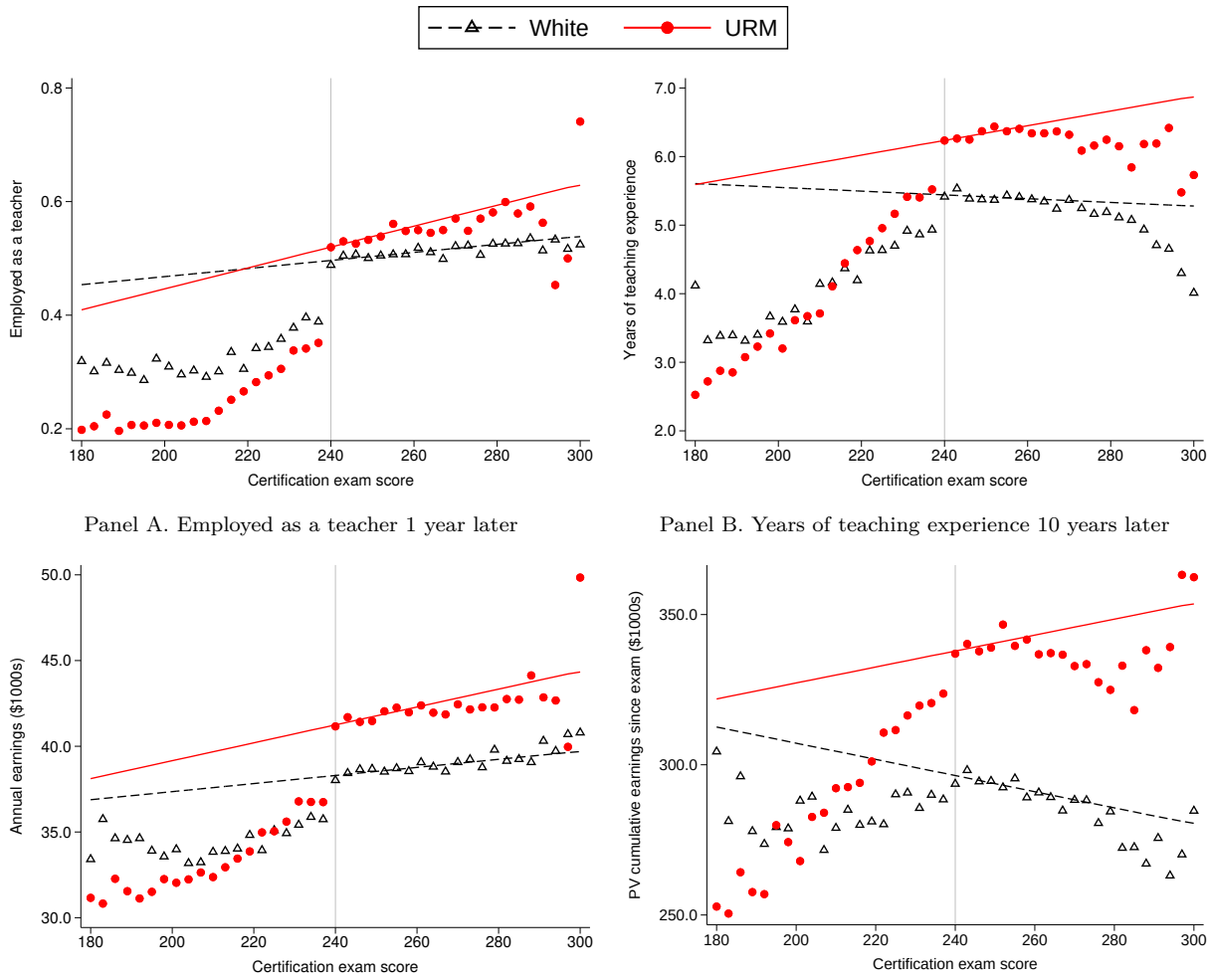
Panel B. High school math score

Panel C. rddensity test: Actual score

Panel D. rddensity test: Adjusted scale

Figure A1: Density tests

Notes: The figure presents density tests of certification exam scores using both actual and adjusted scores. Panel A shows scatter plots of the number of exam takers within score bins, while Panel B illustrates the relationship between high school math scores and certification exam scores. Panels C and D display the regression discontinuity density tests using the `rddensity` package of Cattaneo et al. (2020), based on the actual and adjusted scores, respectively. The actual TExES certification exam scores are calculated using a scaled formula, see details in Section 3.2. We rescale failing scores so that they follow the same linear relationship with high school math scores as passing scores in the vicinity of the threshold, which constructs the adjusted scores.



Panel A. Employed as a teacher 1 year later

Panel B. Years of teaching experience 10 years later

Panel C. Annual earnings 1 year later ($1000s)

Panel D. PV cumulative earnings after 10 years ($1000s)

Figure A2: Extrapolated effects of failing certification exam on outcomes

Notes: The figure presents the extrapolated effects of failing a certification exam on baseline outcomes by white and URM groups. Panels A-D display the extrapolated effects on teacher employment (1 year later), years of teaching experience (10 years later), annual earnings (1 year later), and the present value of cumulative earnings (10 years later), respectively. While the scatter plots indicate the observed relationship between exam scores and outcomes, the projected linear lines are extrapolated using the relationship for exam takers with a score range of 240 and 260.

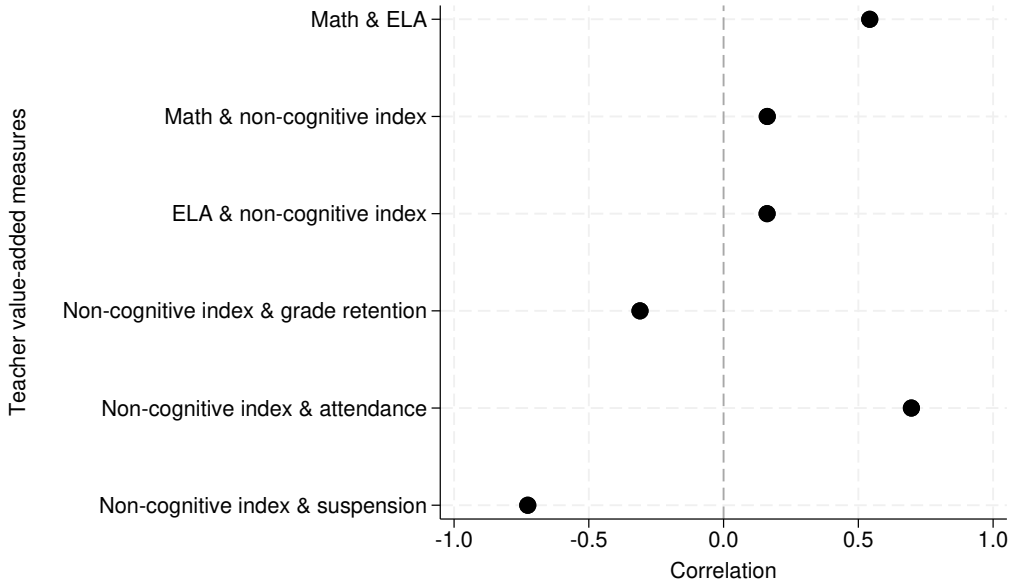
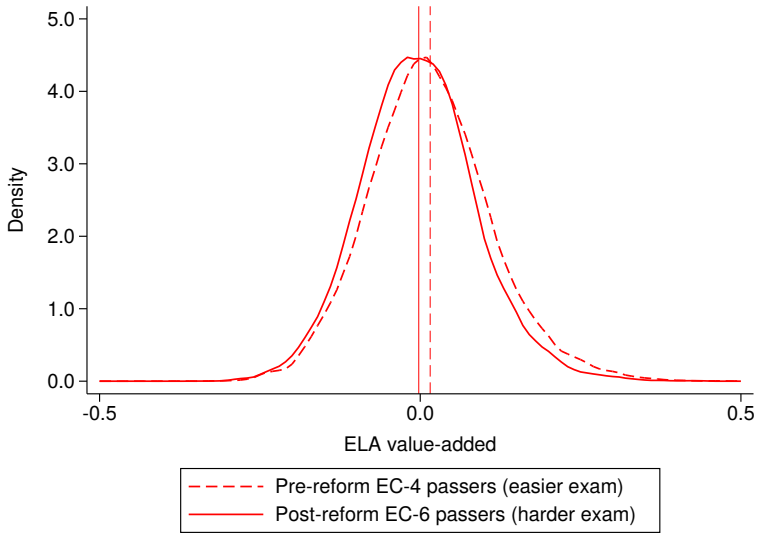
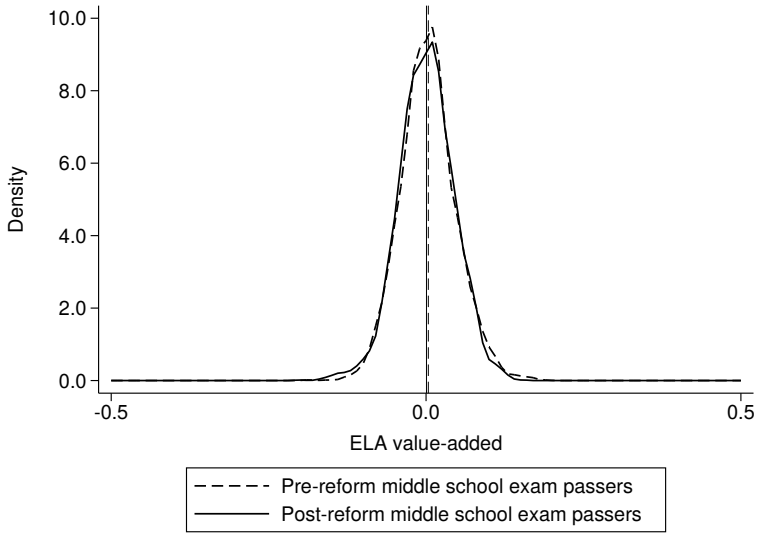


Figure A3: Correlations between teacher value-added measures

Notes: The figure displays correlations between teacher value-added measures. The academic (math and ELA) and non-cognitive value-added are estimated using equations B1 and B2, respectively. The non-cognitive index is constructed using students' non-achievement behaviors in the data: grade retention, attendance, and suspension. See Appendix B.2 for detailed information on the value-added estimation procedure.



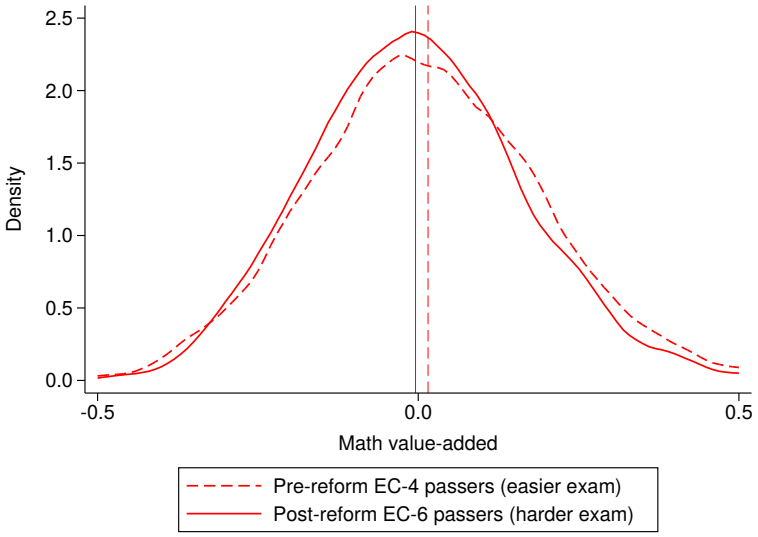
Panel A. Elementary school exams



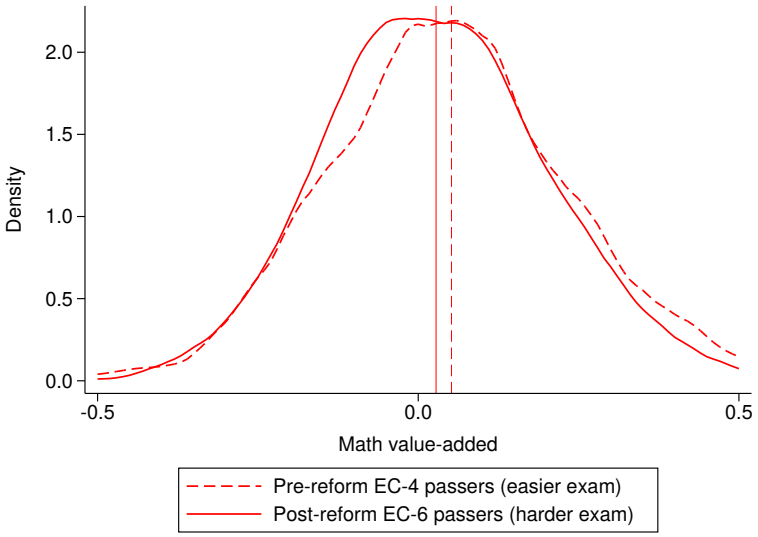
Panel B. Middle school exams

Figure A4: Distributions of ELA value-added for exam passers measured 3–9 years later

Notes: This figure shows kernel densities of ELA value-added for individuals who passed different TExES certification exams. In Panel A, the sample includes individuals who took the EC-4 or EC-6 exams (Panel A of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 4 teachers. The graph plots distributions of grade 4 ELA value-added for individuals who took the EC-4 exam in 2002–2009 (dashed line) or the EC-6 exam in 2010–2015 (solid line). In Panel B, the sample includes individuals who took middle school Generalist or ELA exams (Panel B of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 7–8 teachers. The graph plots distributions of grade 7–8 ELA value-added for individuals who took the middle school exams in 2002–2009 (dashed line) or in 2010–2015 (solid line). We measure value-added 3–9 years after taking the certification exam ("potential experience"), and weight observations so that the distribution of potential experience is the same for the two exams in each graph (using the exam with the minimum number of observations for each level of potential experience). Vertical lines denote the mean of each distribution.



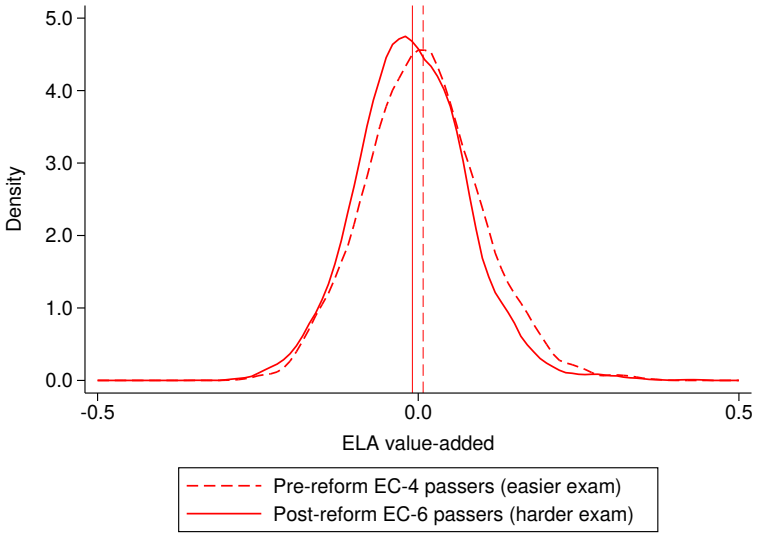
Panel A. White fourth grade teachers



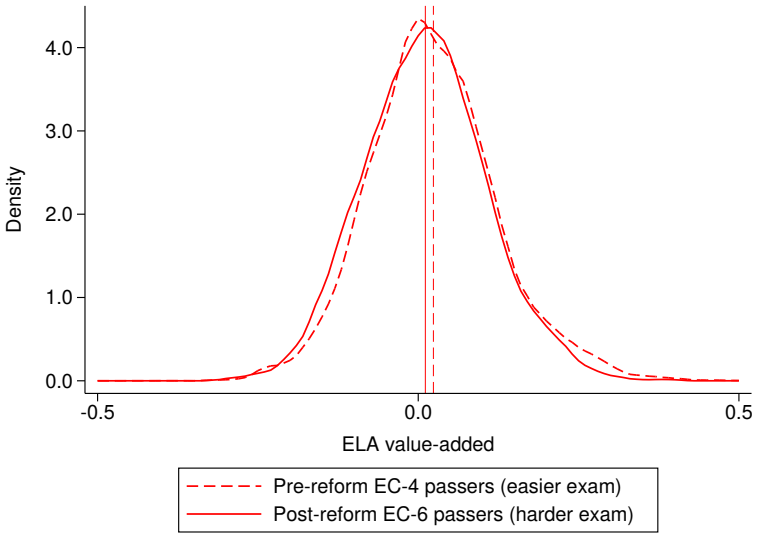
Panel B. URM fourth grade teachers

Figure A5: Race-specific distributions of math value-added for exam passers measured 3–9 years later

Notes: This figure shows kernel densities of math value-added for White and URM individuals who passed different TExES certification exams. The sample includes individuals who took the EC-4 or EC-6 exams (Panel A of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 4 teachers. The graphs plot distributions of grade 4 math value-added for individuals who took the EC-4 exam in 2002–2009 (dashed lines) or the EC-6 exam in 2010–2015 (solid lines). Panel A shows estimates for White teachers, and Panel B shows estimates for URM teachers. We measure value-added 3–9 years after taking the certification exam ("potential experience"), and weight observations so that the distribution of potential experience is the same for the two exams in each graph (using the exam with the minimum number of observations for each level of potential experience). Vertical lines denote the mean of each distribution.



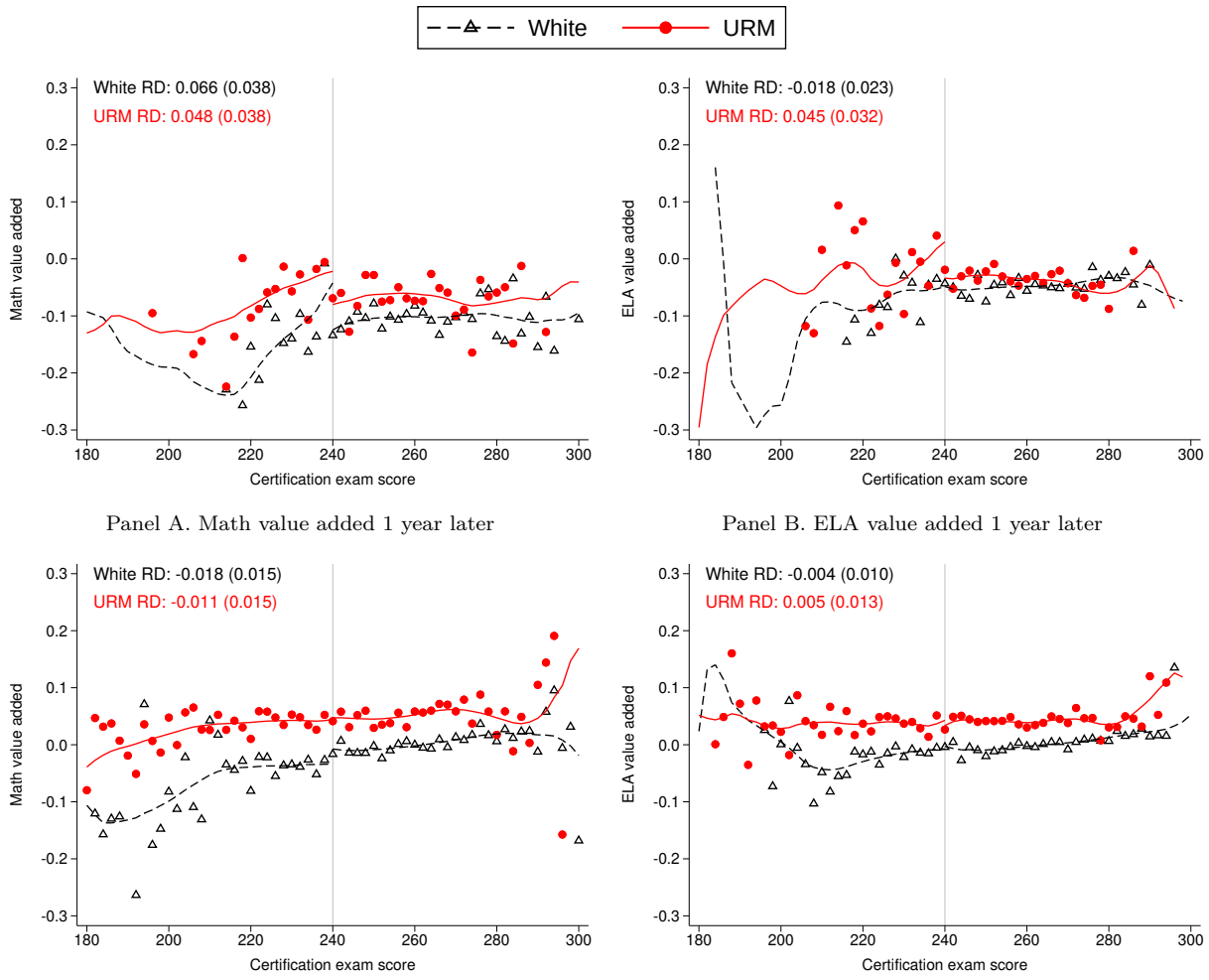
Panel A. White fourth grade teachers



Panel B. URM fourth grade teachers

Figure A6: Race-specific distributions of ELA value-added for exam passers measured 3–9 years later

Notes: This figure shows kernel densities of ELA value-added for White and URM individuals who passed different TExES certification exams. The sample includes individuals who took the EC-4 or EC-6 exams (Panel A of Table 3) as their first certification exam, who passed the exam, and who subsequently became grade 4 teachers. The graphs plot distributions of grade 4 ELA value-added for individuals who took the EC-4 exam in 2002–2009 (dashed lines) or the EC-6 exam in 2010–2015 (solid lines). Panel A shows estimates for White teachers, and Panel B shows estimates for URM teachers. We measure value-added 3–9 years after taking the certification exam ("potential experience"), and weight observations so that the distribution of potential experience is the same for the two exams in each graph (using the exam with the minimum number of observations for each level of potential experience). Vertical lines denote the mean of each distribution.



Panel A. Math value added 1 year later

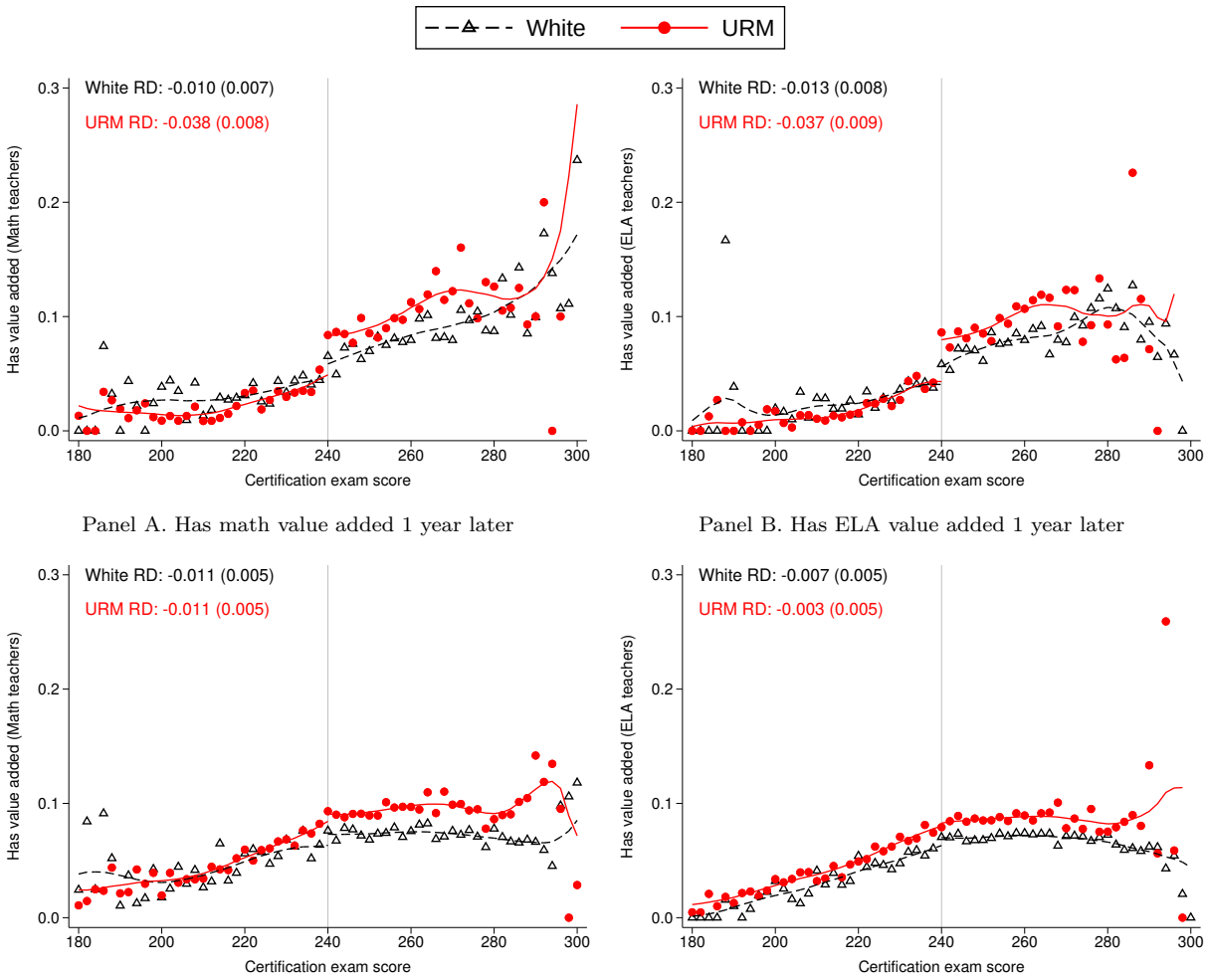
Panel B. ELA value added 1 year later

Panel C. Math value added 2-10 years later

Panel D. ELA value added 2-10 years later

Figure A7: RD plots: Teacher value added by race

Notes: The figure presents the RD effects of failing a certification exam on math and ELA value-added by white and URM groups. Panels A-B and C-D display the RD effects 1 and 2-10 year after the exam, respectively. The RD estimates are shown at the upper-left corner, with standard deviations included in the parentheses.



Panel A. Has math value added 1 year later

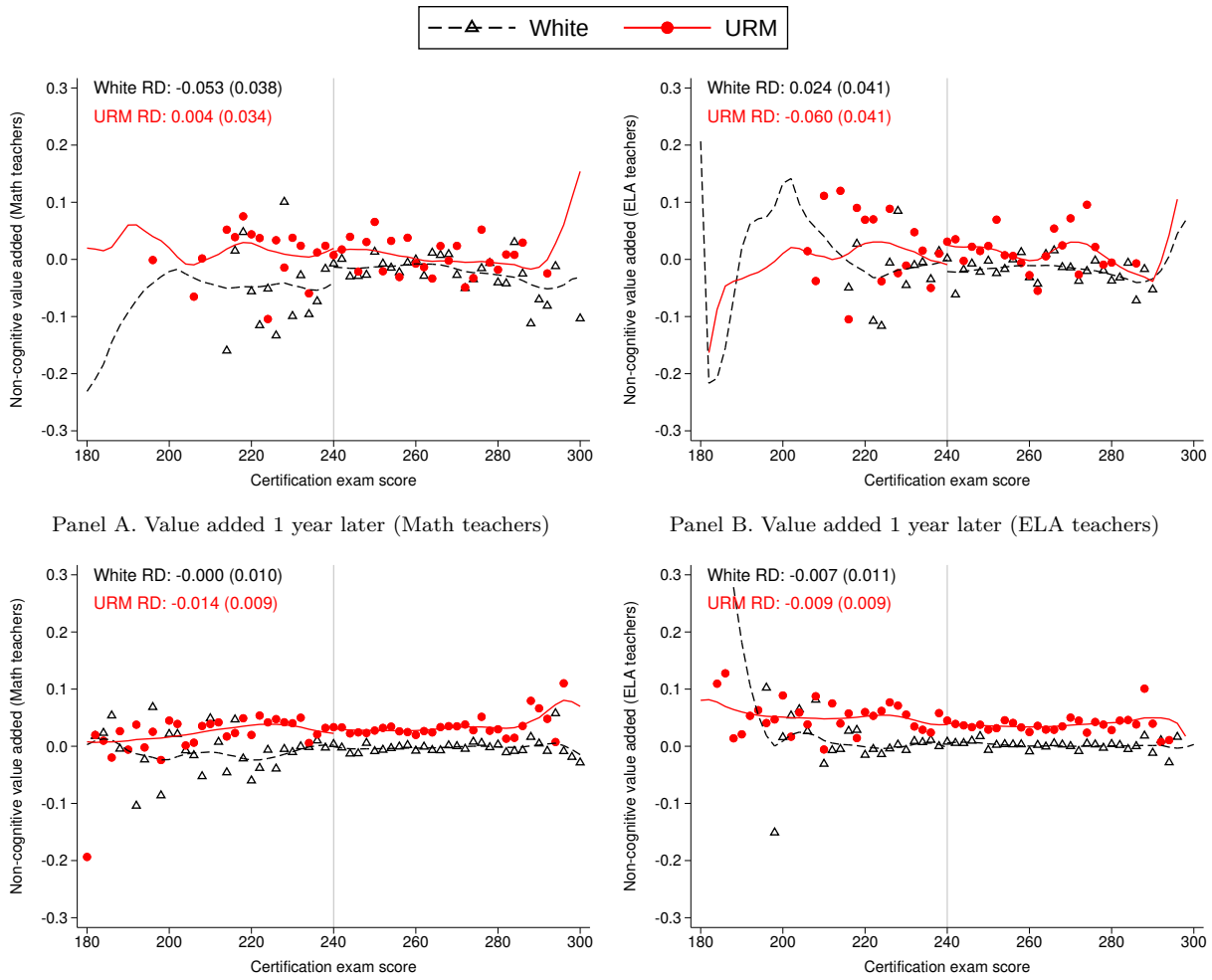
Panel B. Has ELA value added 1 year later

Panel C. Has math value added 2-10 years later

Panel D. Has ELA value added 2-10 years later

Figure A8: RD plots: Teacher has value added by race

Notes: The figure presents the RD effects of failing a certification exam on whether having math and ELA value-added by white and URM groups. Panels A-B and C-D display the RD effects 1 and 2-10 year after the exam, respectively. The RD estimates are shown at the upper-left corner, with standard deviations included in the parentheses.



Panel A. Value added 1 year later (Math teachers)

Panel B. Value added 1 year later (ELA teachers)

Panel C. Value added 2-10 years later (Math teachers)

Panel D. Value added 2-10 years later (ELA teachers)

Figure A9: RD plots: Teacher non-cognitive value added by race

Notes: The figure presents the RD effects of failing a certification exam on non-cognitive value added by white and URM groups. Panels A-B and C-D display the RD effects 1 and 2-10 year after the exam, respectively. The RD estimates are shown at the upper-left corner, with standard deviations included in the parentheses.

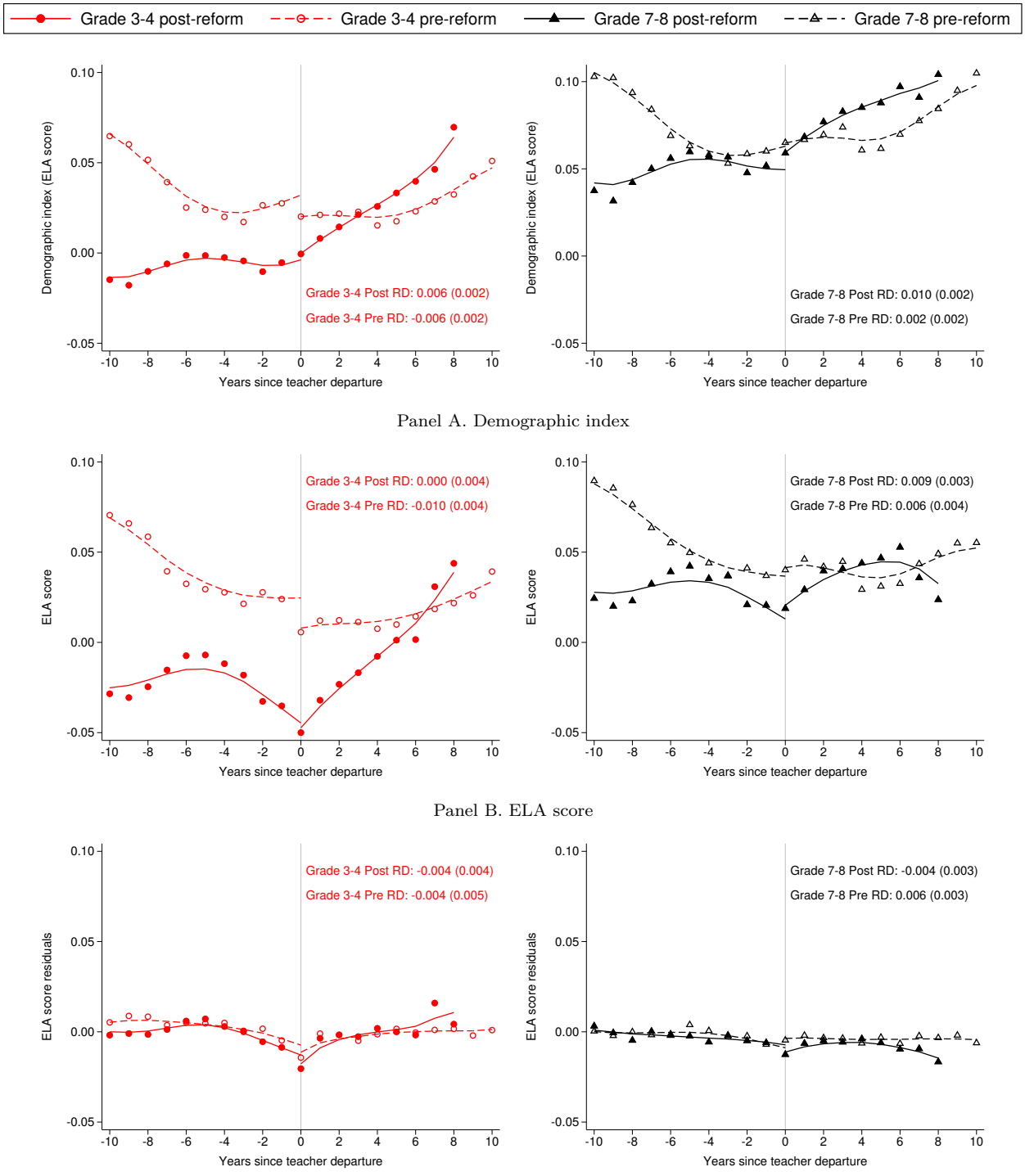
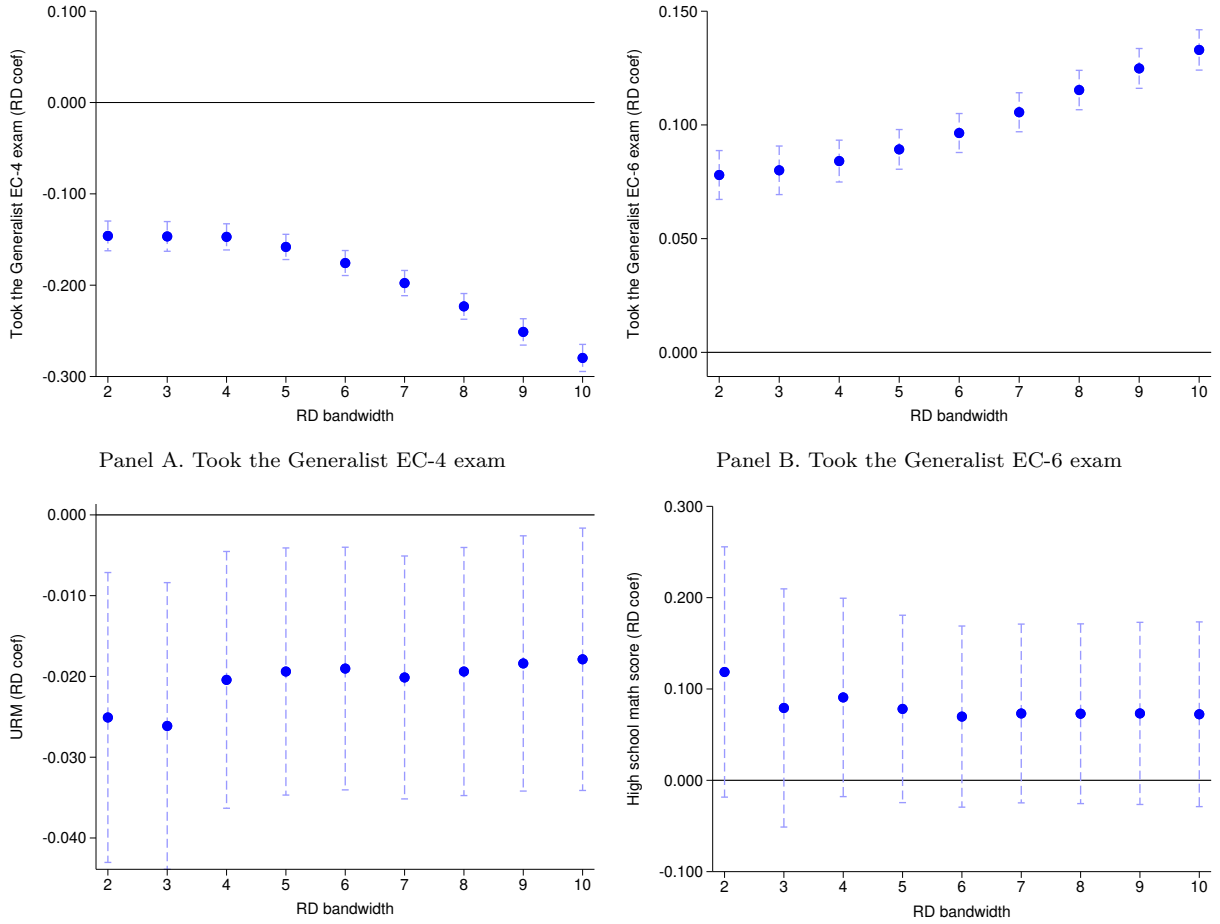


Figure A10: Effects of teacher departures on student ELA achievement

Notes: This figure contains RD graphs that show how teacher departures affect the ELA achievement of students in a school/-grade. Red circles represent grade 3–4 teachers. Black triangles represent grade 7–8 teachers. Hollow symbols depict teacher departures that occurred in $y \in$ 2005–2010 (pre-reform). Solid symbols depict teacher departures that occurred in $y \in$ 2011–2016 (post-reform). The x-axis is years relative to the teacher departure, τ_{ty}. The y-axis depicts the average outcome at the school/grade/year level. Each graph displays RD coefficients β from equation (6) with standard errors clustered at the school level in parentheses.



Panel A. Took the Generalist EC-4 exam

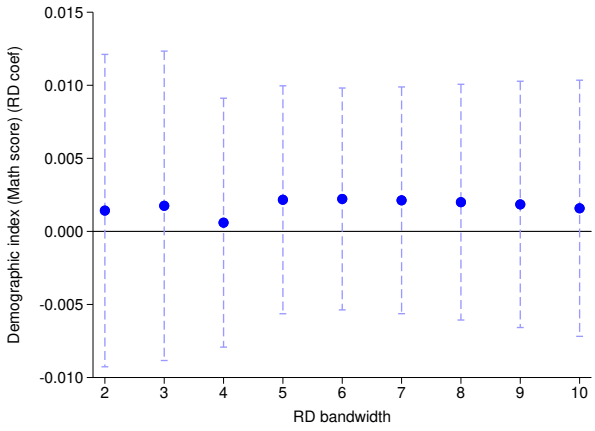
Panel B. Took the Generalist EC-6 exam

Panel C. URM

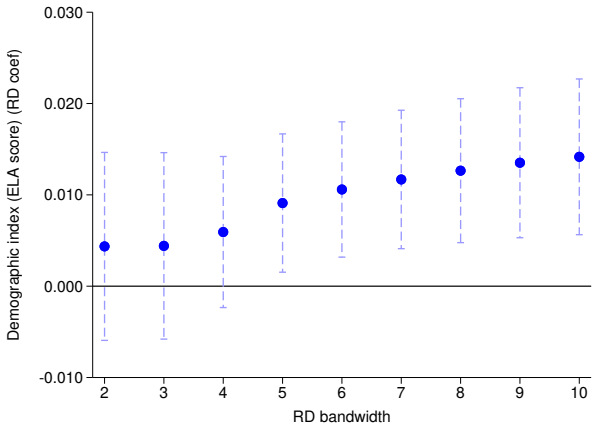
Panel D. HS math score

Figure A11: RD-DiD effects of harder certification exams on teacher composition — Robustness to RD bandwidth

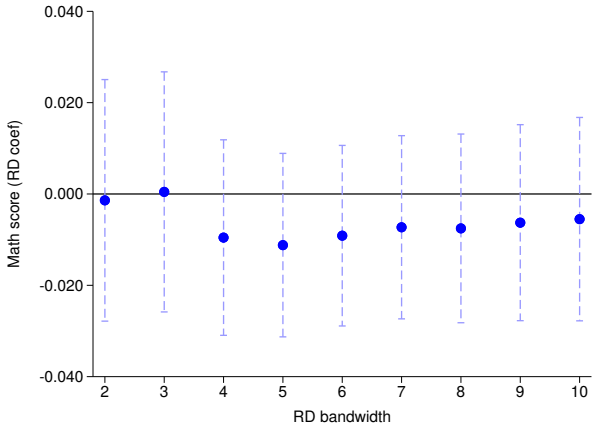
Notes: This figure displays the robustness of our RD-DiD estimates for teacher composition outcomes to the choice of RD bandwidth. The y-axis in each graph displays the θ from equation (7). The x-axis displays the bandwidth used for the RD regression (6), which we vary from $h^Y \in 2$–10 years relative to the year of the teacher departure. Dashed lines represent 95 percent confidence intervals using standard errors clustered at the school level.



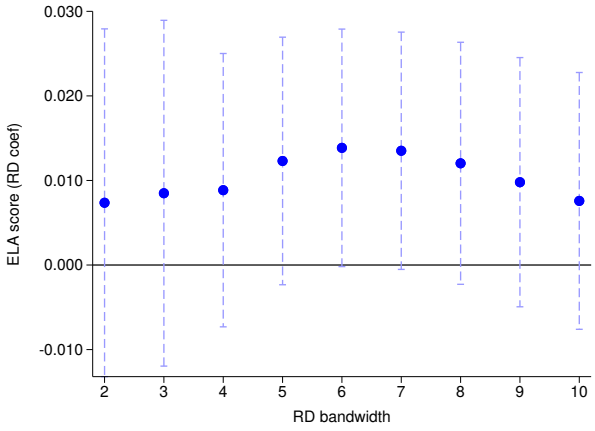
Panel A. Demographic index (Math score)



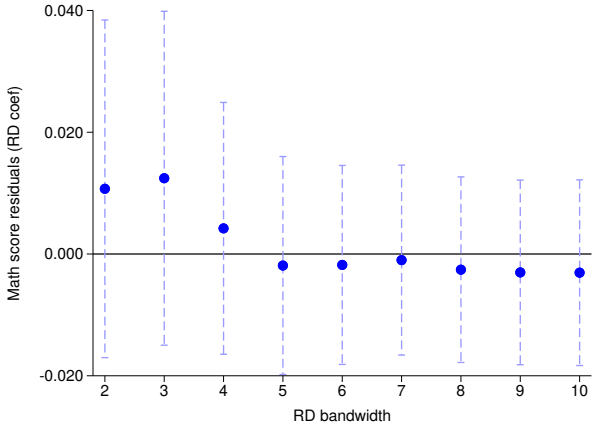
Panel B. Demographic index (ELA score)



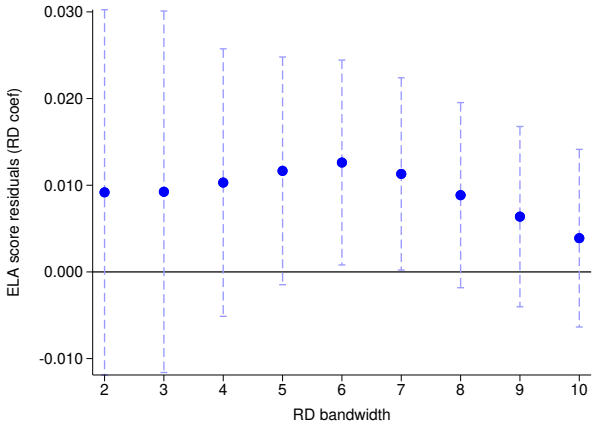
Panel C. Math score



Panel D. ELA score



Panel E. Math score residuals



Panel F. ELA score residuals

Figure A12: RD-DiD effects of harder certification exams on student achievement — Robustness to RD bandwidth

Notes: This figure displays the robustness of our RD-DiD estimates for student achievement outcomes to the choice of RD bandwidth. The y-axis in each graph displays the θ from equation (7). The x-axis displays the bandwidth used for the RD regression (6), which we vary from $h^Y \in 2$–10 years relative to the year of the teacher departure. Dashed lines represent 95 percent confidence intervals using standard errors clustered at the school level.

Table A1: RD balance tests

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Male	0.235	0.264	0.004 (0.005)	0.004 (0.006)	0.004 (0.007)	0.993
White	1.000	0.000	0.009 (0.006)			
Black	0.000	0.314	-0.006 (0.004)		-0.010 (0.007)	
Hispanic	0.000	0.686	0.001 (0.005)		0.010 (0.007)	
Mean proportion white at school during K-12	0.741	0.319	0.008* (0.005)	0.003 (0.003)	-0.001 (0.006)	0.485
High school graduation cohort	2004.491	2004.088	0.115** (0.058)	0.035 (0.078)	0.132* (0.075)	0.379
Ever identified as an immigrant during K-12	0.007	0.038	-0.003 (0.002)	-0.002 (0.002)	-0.003 (0.004)	0.744
Economically disadvantaged during K-12	0.077	0.533	0.003 (0.005)	0.007 (0.005)	0.008 (0.009)	0.904
Identified as at risk of dropping out during K-12	0.167	0.357	-0.004 (0.005)	0.002 (0.005)	-0.006 (0.007)	0.352
Gifted education during K-12	0.117	0.134	-0.003 (0.004)	-0.001 (0.005)	-0.005 (0.006)	0.643
In special education during K-12	0.039	0.018	0.002 (0.002)	-0.002 (0.003)	0.004* (0.002)	0.149
Ever in bilingual education during K-12	0.001	0.146	0.004 (0.004)	-0.001 (0.001)	0.010 (0.007)	0.144
Ever in English as a Second Language during K-12	0.003	0.100	-0.003 (0.003)	0.000 (0.001)	-0.005 (0.006)	0.406
Ever had limited English proficiency during K-12	0.004	0.234	0.001 (0.005)	-0.000 (0.001)	0.006 (0.009)	0.501
High school math score	0.356	0.186	0.015 (0.011)	0.016 (0.015)	0.013 (0.015)	0.890
High school ELA score	0.477	0.273	-0.002 (0.010)	-0.003 (0.015)	-0.007 (0.014)	0.812
Had earnings in Texas in prior year	0.797	0.836	-0.001 (0.004)	0.009 (0.007)	-0.006 (0.007)	0.141
Prior year annual earnings ($1000s)	18.082	19.931	-0.322 (0.233)	0.009 (0.318)	-0.251 (0.294)	0.554
N (Male)	16,681	16,894	400,726	227,691	156,455	384,146
P-value: All coefs zero			0.242	0.737	0.524	

Notes: The table presents the RD estimates of the impacts of failing a certification exam on students' pre-determined characteristics. The dependent variables include students' gender, race, participation in K-12 initiatives and support systems, high school achievement, and pre-exam earnings in Texas. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Male)" indicates the sample size for the outcome "Male", serving as a proxy sample size for other outcomes.

Table A2: RD effects of failing a certification exam on exam outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Passed this certification exam by now	1.000	1.000	-0.294***	-0.258***	-0.295***	0.000
			(0.005)	(0.008)	(0.007)	
Passed any certification exam by now	1.000	1.000	-0.191***	-0.152***	-0.198***	0.000
			(0.005)	(0.007)	(0.006)	
# days to pass this exam (if passed)	0.000	0.000	119.648***	111.715***	128.507***	0.000
			(1.304)	(1.967)	(1.948)	
# days to pass any exam (if passed)	0.000	0.000	113.215***	103.849***	124.719***	0.000
			(1.203)	(1.684)	(1.869)	
N (Passed this certification exam by now)	16,683	16,896	506,345	227,740	156,499	384,239
Panel B. 5 years after certification exam						
Passed this certification exam by now	1.000	1.000	-0.260***	-0.232***	-0.251***	0.093
			(0.005)	(0.009)	(0.007)	
Passed any certification exam by now	1.000	1.000	-0.142***	-0.124***	-0.131***	0.421
			(0.004)	(0.007)	(0.006)	
# days to pass this exam (if passed)	0.000	0.000	146.828***	133.259***	161.650***	0.000
			(2.745)	(3.726)	(4.524)	
# days to pass any exam (if passed)	0.000	0.000	152.514***	127.439***	179.677***	0.000
			(2.998)	(3.696)	(5.164)	
N (Passed this certification exam by now)	14,528	14,425	442,510	201,864	133,272	335,136
Panel B. 10 years after certification exam						
Passed this certification exam by now	1.000	1.000	-0.241***	-0.214***	-0.221***	0.582
			(0.006)	(0.010)	(0.008)	
Passed any certification exam by now	1.000	1.000	-0.143***	-0.123***	-0.122***	0.930
			(0.006)	(0.009)	(0.007)	
# days to pass this exam (if passed)	0.000	0.000	160.183***	145.728***	170.863***	0.006
			(4.377)	(6.099)	(6.655)	
# days to pass any exam (if passed)	0.000	0.000	170.482***	150.324***	191.850***	0.000
			(5.461)	(7.231)	(8.244)	
N (Passed this certification exam by now)	9,848	9,734	310,404	145,790	88,718	234,508

Notes: The table presents the RD estimates of the impacts of failing a certification exam on exam outcomes. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether passed the exam by now, an indicator of whether passed any exam by now, the number of days to pass the exam (if passed), and the number of days to pass any exam (if passed). Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the *p*-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Passed this certification exam by now)" indicates the sample size for the outcome "Passed this certification exam by now", serving as a proxy sample size for other outcomes.

Table A3: RD effects of failing a certification exam on earnings outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Annual earnings ($1000s)	38.239	41.432	-2.831***	-1.919***	-3.942***	0.000
			(0.213)	(0.349)	(0.323)	
Annual earnings, including zeroes ($1000s)	34.817	38.053	-2.884***	-1.818***	-3.669***	0.000
			(0.221)	(0.376)	(0.269)	
Earnings from teaching ($1000s)	45.204	46.444	-1.904***	-1.278***	-2.730***	0.001
			(0.174)	(0.287)	(0.292)	
Earnings from teaching, including zeroes ($1000s)	23.153	25.685	-5.229***	-3.640***	-6.589***	0.000
			(0.251)	(0.400)	(0.341)	
N (Annual earnings)	15,190	15,518	450,861	205,999	142,577	348,576
Panel B. 5 years after certification exam						
Annual earnings ($1000s)	48.275	51.687	-0.552*	0.012	-1.110***	0.058
			(0.290)	(0.460)	(0.379)	
Annual earnings, including zeroes ($1000s)	39.323	44.864	-0.898***	-0.248	-1.441***	0.071
			(0.322)	(0.516)	(0.414)	
Earnings from teaching ($1000s)	50.416	52.253	-0.636***	-0.569***	-0.815***	0.410
			(0.142)	(0.201)	(0.213)	
Earnings from teaching, including zeroes ($1000s)	29.750	34.550	-2.452***	-2.143***	-3.149***	0.114
			(0.291)	(0.458)	(0.429)	
N (Annual earnings)	11,834	12,521	342,975	160,902	113,374	274,276
Panel C. 10 years after certification exam						
Annual earnings ($1000s)	54.772	56.658	-1.595***	-2.271**	-1.765**	0.697
			(0.559)	(1.045)	(0.721)	
Annual earnings, including zeroes ($1000s)	37.736	44.679	-0.841**	-0.557	-1.338**	0.410
			(0.416)	(0.653)	(0.677)	
Earnings from teaching ($1000s)	53.565	55.158	-0.625***	-0.589*	-0.549*	0.927
			(0.207)	(0.308)	(0.294)	
Earnings from teaching, including zeroes ($1000s)	22.159	28.270	-1.144***	-1.012*	-1.439***	0.595
			(0.334)	(0.570)	(0.556)	
N (Annual earnings)	6,785	7,676	208,615	99,031	69,173	168,204

Notes: The table presents the RD estimates of the impacts of failing a certification exam on earning outcomes. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: annual earnings, annual earnings (including zeros), earnings from teaching, and earnings from teaching (including zeros). Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Annual earnings)" indicates the sample size for the outcome "Annual earnings ($1000s)", serving as a proxy sample size for other outcomes.

Table A4: RD effects of failing a certification exam on outcomes: Running variable with adjusted scale

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Employed as a teacher	0.497	0.525	-0.116***	-0.084***	-0.149***	0.000
			(0.005)	(0.008)	(0.007)	
Years of teaching experience	0.565	0.597	-0.120***	-0.092***	-0.153***	0.000
			(0.006)	(0.009)	(0.009)	
Annual earnings ($1000s)	38.239	41.432	-2.776***	-1.850***	-3.808***	0.000
			(0.206)	(0.336)	(0.290)	
PV cumulative earnings since exam ($1000s)	49.084	54.467	-4.007***	-2.489***	-5.587***	0.000
			(0.304)	(0.507)	(0.493)	
N (Employed as a teacher)	16,683	16,896	506,345	227,740	156,499	384,239
Panel B. 5 years after certification exam						
Employed as a teacher	0.608	0.677	-0.048***	-0.039***	-0.063***	0.053
			(0.006)	(0.009)	(0.008)	
Years of teaching experience	3.076	3.342	-0.359***	-0.291***	-0.453***	0.002
			(0.026)	(0.038)	(0.035)	
Annual earnings ($1000s)	48.275	51.687	-0.564*	0.037	-1.062***	0.057
			(0.292)	(0.450)	(0.369)	
PV cumulative earnings since exam ($1000s)	173.604	194.717	-8.856***	-5.514***	-12.843***	0.001
			(1.037)	(1.578)	(1.489)	
N (Employed as a teacher)	14,528	14,425	442,510	201,864	133,272	335,136
Panel C. 10 years after certification exam						
Employed as a teacher	0.459	0.553	-0.019***	-0.017*	-0.023**	0.691
			(0.007)	(0.010)	(0.010)	
Years of teaching experience	5.477	6.248	-0.457***	-0.432***	-0.524***	0.450
			(0.060)	(0.089)	(0.080)	
Annual earnings ($1000s)	54.772	56.658	-1.517***	-2.103**	-1.777***	0.791
			(0.512)	(0.992)	(0.677)	
PV cumulative earnings since exam ($1000s)	296.041	338.626	-9.275***	-7.736**	-12.088***	0.377
			(2.351)	(3.784)	(3.039)	
N (Employed as a teacher)	9,848	9,734	310,404	145,790	88,718	234,508

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teaching and earning outcomes, using running variables with adjusted scale. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings since the exam. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A5: RD effects of failing a certification exam on outcomes — Heterogeneity within URM groups

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: B=H
	Black	Hispanic	All URM	Black	Hispanic	
Panel A. 1 year after certification exam						
Employed as a teacher	0.561	0.508	-0.150***	-0.189***	-0.137***	0.002
			(0.007)	(0.014)	(0.009)	
Years of teaching experience	0.654	0.571	-0.152***	-0.183***	-0.144***	0.073
			(0.008)	(0.018)	(0.011)	
Annual earnings ($1000s)	43.243	40.584	-3.942***	-4.561***	-3.593***	0.160
			(0.323)	(0.558)	(0.366)	
PV cumulative earnings since exam ($1000s)	59.873	51.988	-5.672***	-6.634***	-5.093***	0.149
			(0.502)	(0.853)	(0.576)	
N (Employed as a teacher)	5,312	11,584	156,499	47,292	109,207	156,499
Panel B. 5 years after certification exam						
Employed as a teacher	0.650	0.690	-0.062***	-0.085***	-0.055***	0.098
			(0.008)	(0.015)	(0.010)	
Years of teaching experience	3.302	3.361	-0.451***	-0.573***	-0.412***	0.043
			(0.034)	(0.065)	(0.042)	
Annual earnings ($1000s)	52.941	51.110	-1.109***	-2.188***	-0.846*	0.100
			(0.377)	(0.653)	(0.469)	
PV cumulative earnings since exam ($1000s)	202.532	191.085	-12.817***	-16.851***	-10.813***	0.056
			(1.458)	(2.600)	(1.635)	
N (Employed as a teacher)	4,577	9,848	133,272	40,236	93,036	133,272
Panel C. 10 years after certification exam						
Employed as a teacher	0.501	0.577	-0.025***	-0.045**	-0.017	0.199
			(0.009)	(0.018)	(0.012)	
Years of teaching experience	5.911	6.404	-0.536***	-0.729***	-0.483***	0.197
			(0.075)	(0.155)	(0.100)	
Annual earnings ($1000s)	58.556	55.759	-1.764**	-3.501***	-0.724	0.064
			(0.721)	(1.232)	(0.739)	
PV cumulative earnings since exam ($1000s)	347.442	334.551	-12.274***	-17.281***	-10.249***	0.308
			(3.178)	(5.644)	(3.603)	
N (Employed as a teacher)	3,077	6,657	88,718	26,451	62,267	88,718

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teaching and earning outcomes, focusing on heterogeneity within the URM groups. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings since the exam. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between Black and Hispanic groups in Columns (D) and (E), respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A6: RD effects of failing a certification exam on outcomes — Heterogeneity by SES

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: H=L
	High SES	Low SES	All	High SES	Low SES	
Panel A. 1 year after certification exam						
Employed as a teacher	0.522	0.518	-0.114***	-0.099***	-0.135***	0.007
			(0.007)	(0.010)	(0.009)	
Years of teaching experience	0.590	0.574	-0.116***	-0.100***	-0.138***	0.012
			(0.008)	(0.011)	(0.011)	
Annual earnings ($1000s)	39.687	40.208	-3.031***	-2.769***	-3.541***	0.148
			(0.268)	(0.378)	(0.363)	
PV cumulative earnings since exam ($1000s)	51.306	52.551	-4.031***	-3.128***	-5.302***	0.006
			(0.398)	(0.532)	(0.574)	
N (Employed as a teacher)	14,177	10,204	270,856	173,378	97,478	270,856
Panel B. 5 years after certification exam						
Employed as a teacher	0.639	0.705	-0.045***	-0.045***	-0.049***	0.773
			(0.007)	(0.010)	(0.010)	
Years of teaching experience	3.231	3.408	-0.334***	-0.311***	-0.387***	0.235
			(0.031)	(0.044)	(0.045)	
Annual earnings ($1000s)	49.865	51.478	-0.621	-0.015	-1.509***	0.043
			(0.398)	(0.566)	(0.486)	
PV cumulative earnings since exam ($1000s)	183.231	193.917	-8.664***	-7.518***	-11.209***	0.147
			(1.257)	(1.804)	(1.743)	
N (Employed as a teacher)	12,265	8,192	229,901	152,031	77,870	229,901
Panel C. 10 years after certification exam						
Employed as a teacher	0.483	0.571	-0.018*	-0.020*	-0.013	0.716
			(0.010)	(0.012)	(0.016)	
Years of teaching experience	5.780	6.379	-0.407***	-0.402***	-0.462***	0.685
			(0.077)	(0.101)	(0.106)	
Annual earnings ($1000s)	56.897	57.124	-1.949**	-2.509**	-1.587*	0.531
			(0.805)	(1.132)	(0.878)	
PV cumulative earnings since exam ($1000s)	315.340	339.155	-10.119***	-12.429***	-12.213***	0.971
			(3.040)	(4.385)	(3.888)	
N (Employed as a teacher)	8,232	4,725	150,934	106,544	44,390	150,934

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teaching and earning outcomes, focusing on heterogeneity by teacher candidates' socioeconomic status (SES). A teacher candidate is defined as having a low SES if the candidate has ever enrolled in Free or Reduced-Price Lunch during K-12. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings since the exam. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between low and high SES groups in Columns (D) and (E), respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A7: RD effects of failing a certification exam on outcomes — Heterogeneity by Race & SES

	(A)	(B)	(C)	(D)	(E)	(F)
	High SES			Low SES		
	RD coefficients		*P*-value:	RD coefficients		*P*-value:
	White	URM	W=URM	White	URM	W=URM
Panel A. 1 year after certification exam						
Employed as a teacher	-0.075***	-0.181***	0.000	-0.105***	-0.147***	0.009
	(0.011)	(0.018)		(0.013)	(0.009)	
Years of teaching experience	-0.080***	-0.177***	0.000	-0.113***	-0.151***	0.030
	(0.013)	(0.022)		(0.014)	(0.010)	
Annual earnings ($1000s)	-2.242***	-3.930***	0.032	-1.534***	-3.899***	0.000
	(0.443)	(0.616)		(0.534)	(0.354)	
PV cumulative earnings since exam ($1000s)	-2.517***	-5.409***	0.023	-2.934***	-5.805***	0.004
	(0.644)	(1.029)		(0.813)	(0.534)	
N (Employed as a teacher)	135,775	33,987	169,762	91,965	122,512	214,477
Panel B. 5 years after certification exam						
Employed as a teacher	-0.041***	-0.067***	0.263	-0.035**	-0.062***	0.114
	(0.011)	(0.020)		(0.014)	(0.009)	
Years of teaching experience	-0.263***	-0.471***	0.034	-0.340***	-0.453***	0.119
	(0.047)	(0.082)		(0.060)	(0.038)	
Annual earnings ($1000s)	-0.287	0.104	0.729	0.674	-1.462***	0.011
	(0.559)	(1.057)		(0.713)	(0.384)	
PV cumulative earnings since exam ($1000s)	-6.620***	-11.333***	0.234	-4.273	-13.306***	0.005
	(1.990)	(3.301)		(2.692)	(1.534)	
N (Employed as a teacher)	119,628	29,395	149,023	82,236	103,877	186,113
Panel C. 10 years after certification exam						
Employed as a teacher	-0.021	-0.027	0.829	-0.012	-0.022**	0.602
	(0.014)	(0.022)		(0.017)	(0.011)	
Years of teaching experience	-0.424***	-0.480**	0.807	-0.448***	-0.559***	0.525
	(0.110)	(0.190)		(0.143)	(0.092)	
Annual earnings ($1000s)	-2.453*	-1.915	0.800	-0.666	-1.735**	0.451
	(1.353)	(1.477)		(1.152)	(0.727)	
PV cumulative earnings since exam ($1000s)	-12.684**	-12.021*	0.939	-2.793	-13.343***	0.151
	(5.123)	(6.595)		(6.114)	(3.676)	
N (Employed as a teacher)	84,261	20,281	104,542	61,529	68,437	129,966

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teaching and earning outcomes, focusing on heterogeneity by teacher candidates' race and SES. A teacher candidate is defined as having a low SES if the candidate has enrolled in Free or Reduced-Price Lunch during K-12. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings since the exam. Columns (A)–(C) report results for high-SES candidates, while Columns (D)–(F) report results for low-SES candidates. Columns (A) and (B) show RD coefficients using equation (1), and Column (C) indicates the *p*-value testing for differences between White and URM groups. Similarly, Columns (D) and (E) show RD coefficients for low-SES candidates, with Column (F) reporting the corresponding *p*-value. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A8: Extrapolated effects of failing certification exam on outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Exam takers within 10 pts of passing		All exam takers who failed		All exam takers (passed + failed)	
	White	URM	White	URM	White	URM
Panel A. 1 year after certification exam						
Employed as a teacher	-0.107***	-0.168***	-0.134***	-0.203***	-0.021***	-0.082***
	(0.004)	(0.003)	(0.003)	(0.002)	(0.000)	(0.001)
Years of teaching experience	-0.107***	-0.169***	-0.127***	-0.196***	-0.020***	-0.079***
	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)
Annual earnings ($1000s)	-2.585***	-4.301***	-3.013***	-5.646***	-0.463***	-2.268***
	(0.155)	(0.135)	(0.109)	(0.077)	(0.017)	(0.032)
PV cumulative earnings since exam ($1000s)	-3.526***	-6.137***	-4.083***	-8.230***	-0.635***	-3.339***
	(0.236)	(0.204)	(0.166)	(0.115)	(0.026)	(0.048)
N (Employed as a teacher)	15,430	19,688	35,391	63,496	227,740	156,499
Panel B. 5 years after certification exam						
Employed as a teacher	-0.056***	-0.090***	-0.112***	-0.199***	-0.017***	-0.080***
	(0.004)	(0.004)	(0.003)	(0.002)	(0.000)	(0.001)
Years of teaching experience	-0.378***	-0.589***	-0.619***	-1.013***	-0.092***	-0.406***
	(0.018)	(0.016)	(0.012)	(0.009)	(0.002)	(0.004)
Annual earnings ($1000s)	-0.146	-1.800***	-0.564***	-4.130***	-0.084***	-1.638***
	(0.251)	(0.172)	(0.152)	(0.098)	(0.023)	(0.039)
PV cumulative earnings since exam ($1000s)	-7.420***	-14.665***	-11.452***	-26.659***	-1.705***	-10.687***
	(0.749)	(0.621)	(0.506)	(0.372)	(0.076)	(0.153)
N (Employed as a teacher)	13,208	16,484	30,058	53,425	201,864	133,272
Panel C. 10 years after certification exam						
Employed as a teacher	-0.033***	-0.045***	-0.077***	-0.127***	-0.010***	-0.048***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.000)	(0.001)
Years of teaching experience	-0.563***	-0.749***	-0.984***	-1.509***	-0.129***	-0.568***
	(0.044)	(0.039)	(0.029)	(0.022)	(0.004)	(0.009)
Annual earnings ($1000s)	-0.445	-0.621**	-1.133***	-2.330***	-0.148***	-0.857***
	(0.438)	(0.257)	(0.290)	(0.147)	(0.038)	(0.054)
PV cumulative earnings since exam ($1000s)	-10.083***	-15.263***	-15.727***	-31.497***	-2.062***	-11.844***
	(1.785)	(1.362)	(1.178)	(0.802)	(0.155)	(0.306)
N (Employed as a teacher)	8,365	10,185	19,115	33,362	145,790	88,718

Notes: The table presents the extrapolated effects of failing a certification exam on teaching and earning outcomes. The extrapolation uses the observed relationship between the scores and outcomes of samples with an exam score range of 240 and 260, and employs a linear extrapolation to project the outcomes for candidates who are below the passing score. Panels A-C indicate the extrapolated effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether being employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings. Columns (A)-(B), (C)-(D), and (E)-(F) use exam takers with a score within ten points of the passing score, all exam takers with a score below the 240 cutoff, and all exam takers, respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A9: Extrapolated effects of failing certification exam on outcomes
Lower bounds on White/URM gap

	(A)	(B)	(C)	(D)	(E)	(F)
	Exam takers within 10 pts of passing		All exam takers who failed		All exam takers (passed + failed)	
	White	URM	White	URM	White	URM
Panel A. 1 year after certification exam						
Employed as a teacher	-0.107***	-0.168***	-0.134***	-0.203***	-0.021***	-0.082***
	(0.004)	(0.003)	(0.003)	(0.002)	(0.000)	(0.001)
Years of teaching experience	-0.107***	-0.169***	-0.127***	-0.196***	-0.020***	-0.079***
	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)
Annual earnings ($1000s)	-2.824***	-4.395***	-3.013***	-5.646***	-0.426***	-2.152***
	(0.155)	(0.135)	(0.109)	(0.077)	(0.016)	(0.030)
PV cumulative earnings since exam ($1000s)	-3.025***	-5.095***	-3.177***	-6.583***	-0.494***	-2.671***
	(0.236)	(0.204)	(0.166)	(0.115)	(0.026)	(0.047)
N (Employed as a teacher)	15,430	19,688	35,391	63,496	227,740	156,499
Panel B. 5 years after certification exam						
Employed as a teacher	-0.055***	-0.075***	-0.104***	-0.186***	-0.015***	-0.074***
	(0.004)	(0.004)	(0.003)	(0.002)	(0.000)	(0.001)
Years of teaching experience	-0.383***	-0.573***	-0.619***	-1.013***	-0.092***	-0.406***
	(0.018)	(0.016)	(0.012)	(0.009)	(0.002)	(0.004)
Annual earnings ($1000s)	-0.257	-1.466***	-0.899***	-4.420***	-0.013	-1.368***
	(0.251)	(0.172)	(0.152)	(0.098)	(0.019)	(0.036)
PV cumulative earnings since exam ($1000s)	-4.986***	-10.296***	-7.918***	-22.447***	-1.179***	-8.998***
	(0.749)	(0.621)	(0.505)	(0.372)	(0.076)	(0.152)
N (Employed as a teacher)	13,208	16,484	30,058	53,425	201,864	133,272
Panel C. 10 years after certification exam						
Employed as a teacher	-0.043***	-0.046***	-0.098***	-0.132***	-0.013***	-0.050***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.000)	(0.001)
Years of teaching experience	-0.613***	-0.783***	-0.984***	-1.509***	-0.129***	-0.568***
	(0.044)	(0.039)	(0.029)	(0.022)	(0.004)	(0.009)
Annual earnings ($1000s)	-0.445	-0.621**	-1.133***	-2.330***	-0.105***	-0.733***
	(0.438)	(0.257)	(0.290)	(0.147)	(0.027)	(0.046)
PV cumulative earnings since exam ($1000s)	-10.083***	-15.263***	-15.727***	-31.497***	-2.062***	-11.844***
	(1.785)	(1.362)	(1.178)	(0.802)	(0.155)	(0.306)
N (Employed as a teacher)	8,365	10,185	19,115	33,362	145,790	88,718

Notes: The table presents the extrapolated effects of failing a certification exam on teaching and earning outcomes. The extrapolation employs a linear extrapolation to project the outcomes for candidates who are below the passing score, using a sample that generates the lower bounds on the white-URM gap. Panels A-C indicate the extrapolated effects in one, five, and ten years after the exam, respectively. The dependent variables include four outcomes: an indicator of whether being employed as a teacher, years of teaching experience, annual earnings, and the present value of cumulative earnings. Columns (A)-(B), (C)-(D), and (E)-(F) use exam takers with a score within ten points of the passing score, all exam takers with a score below the 240 cutoff, and all exam takers, respectively. "N (Employed as a teacher)" indicates the sample size for the outcome "Employed as a teacher", serving as a proxy sample size for other outcomes.

Table A10: Proportion of first-year teachers who took each TExES certification exam by grade

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)
	First-year teachers in 2005–2010						First-year teachers in 2011–2016					
TExES Exam	3rd grade	4th grade	5th grade	6th grade	7th grade	8th grade	3rd grade	4th grade	5th grade	6th grade	7th grade	8th grade
Panel A. Elementary school exams												
Generalist EC-4	0.753	0.659	0.249	0.107	0.076	0.065	0.184	0.157	0.071	0.045	0.041	0.038
Generalist EC-6	0.004	0.006	0.019	0.015	0.003	0.003	0.630	0.555	0.573	0.309	0.127	0.109
Bilingual Generalist EC-4	0.116	0.099	0.026	0.004	-	0.002	0.024	0.024	0.005	0.003	0.002	-
Bilingual Generalist EC-6	-	-	-	-	-	-	0.054	0.058	0.059	0.011	0.003	-
English as a Second Language (ESL)/Generalist EC-4	0.045	0.042	0.010	0.003	-	0.002	0.008	0.009	0.002	0.002	0.001	-
English as a Second Language (ESL)/Generalist EC-6	-	-	-	-	-	-	0.066	0.058	0.054	0.020	0.004	0.003
Panel B. Middle school exams												
Generalist 4-8	0.040	0.132	0.544	0.403	0.326	0.277	0.052	0.153	0.260	0.473	0.455	0.407
Bilingual Generalist 4-8	0.001	0.008	0.033	0.007	0.004	0.002	0.001	0.004	0.005	0.002	-	-
English as a Second Language (ESL)/Generalist 4-8	0.002	0.004	0.024	0.017	0.010	0.010	0.002	0.004	0.005	0.008	0.005	0.007
Mathematics 4-8	0.005	0.014	0.047	0.182	0.239	0.218	0.010	0.021	0.038	0.156	0.217	0.209
English Language Arts and Reading 4-8	0.009	0.030	0.063	0.191	0.214	0.179	0.012	0.021	0.025	0.114	0.190	0.158
Science 4-8	0.003	0.006	0.044	0.010	0.007	0.005	0.004	0.008	0.023	0.012	0.010	0.006
Social Studies 4-8	0.002	0.006	0.023	0.013	0.009	0.008	0.004	0.005	0.012	0.013	0.018	0.014
Mathematics/Science 4-8	0.001	0.005	0.024	0.032	0.034	0.032	0.005	0.009	0.015	0.031	0.041	0.038
English Language Arts and Reading/Social Studies 4-8	0.004	0.017	0.039	0.059	0.053	0.040	0.007	0.013	0.018	0.042	0.040	0.040
# first-year teachers	7,058	7,593	5,118	3,367	3,999	3,407	6,773	7,148	6,262	3,848	4,036	3,048

Notes: This table shows the proportion of first-year teachers who took each of the TExES certification exams listed in column (A) by grade. Columns (B)–(G) show statistics for first-year teachers in 2005–2010. Columns (H)–(M) show statistics for first-year teachers in 2011–2016. "-" represents values that are censored due to a small number of observations.

Table A11: Teacher non-cognitive value added by certification exam performance and race

	(A)	(B)	(C)	(D)	(E)
	Estimates for all certification exam takers			Estimates for exam takers who passed	
Panel A. Non-cognitive value added (Math teachers)					
URM	0.009***			0.009***	
	(0.002)			(0.002)	
Passed certification exam * White		0.000			
		(0.003)			
Passed certification exam * URM		-0.000			
		(0.003)			
Certification exam score/10 * White			-0.000		-0.000
			(0.000)		(0.001)
Certification exam score/10 * URM			-0.000		0.001
			(0.001)		(0.001)
N (# students with Math scores)	8,904,118	8,904,118	8,904,118	8,904,118	8,904,118
P-value: White=URM	0.000	0.969	0.962	0.000	0.260
Panel B. Non-cognitive value added (ELA teachers)					
URM	0.009***			0.007***	
	(0.001)			(0.002)	
Passed certification exam * White		-0.003			
		(0.003)			
Passed certification exam * URM		-0.004			
		(0.003)			
Certification exam score/10 * White			-0.001*		-0.001*
			(0.001)		(0.001)
Certification exam score/10 * URM			-0.001		-0.000
			(0.001)		(0.001)
N (# students with ELA scores)	9,404,289	9,404,289	9,404,289	9,404,289	9,404,289
P-value: White=URM	0.000	0.698	0.974	0.000	0.413
Fixed effects:					
Exam*Year	Y				
Exam*Year*Race		Y	Y		
Exam*Year*Passed exam				Y	
Exam*Year*Race*Passed exam					Y

Notes: The table presents the correlations between teacher effectiveness by race and certification performance using equations B3. Panels A and B examine the non-cognitive value added for Math and ELA teachers, separately. Estimates in Columns (A)-(C) are obtained using all exam takers, while estimates in Columns (D)-(E) are based on exam takers who passed. "N (#students with Math/ELA scores)" indicates the sample size, while "*P*-value: White = URM" shows *p*-value for estimate differences between white and URM groups.

Table A12: RD effects of failing a certification exam on teacher value added

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Has value added (Math teachers)	0.062	0.085	-0.025***	-0.010	-0.038***	0.009
			(0.005)	(0.007)	(0.008)	
Has value added (ELA teachers)	0.061	0.081	-0.025***	-0.013	-0.037***	0.055
			(0.005)	(0.008)	(0.009)	
Math value added	-0.122	-0.083	0.048**	0.066*	0.048	0.761
			(0.022)	(0.038)	(0.038)	
ELA value added	-0.054	-0.034	0.021	-0.018	0.045	0.132
			(0.018)	(0.023)	(0.032)	
Non-cognitive value added (Math teachers)	-0.014	0.021	-0.013	-0.053	0.004	0.316
			(0.024)	(0.038)	(0.034)	
Non-cognitive value added (ELA teachers)	-0.027	0.022	-0.011	0.024	-0.060	0.168
			(0.026)	(0.041)	(0.041)	
N (Math teachers with value added)	218	275	6,128	2,819	1,747	4,566
N (ELA teachers with value added)	220	272	5,977	2,768	1,653	4,421
Panel B. 2-10 years after certification exam						
Has value added (Math teachers)	0.073	0.090	-0.008**	-0.011**	-0.011**	0.975
			(0.003)	(0.005)	(0.005)	
Has value added (ELA teachers)	0.071	0.084	-0.004	-0.007	-0.003	0.567
			(0.003)	(0.005)	(0.005)	
Math value added	-0.007	0.045	-0.009	-0.018	-0.011	0.721
			(0.010)	(0.015)	(0.015)	
ELA value added	-0.008	0.043	0.003	-0.004	0.005	0.566
			(0.007)	(0.010)	(0.013)	
Non-cognitive value added (Math teachers)	-0.004	0.030	-0.003	-0.000	-0.014	0.306
			(0.005)	(0.010)	(0.009)	
Non-cognitive value added (ELA teachers)	0.006	0.040	-0.002	-0.007	-0.009	0.863
			(0.006)	(0.011)	(0.009)	
N (Math teachers with value added)	2,935	3,960	89,174	41,250	31,320	72,570
N (ELA teachers with value added)	2,832	3,766	84,103	39,765	29,319	69,084

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teacher value added. Panels A-C indicate the RD effects in one, five, and ten years after the exam, respectively. The dependent variables include three outcomes by Math and ELA teachers: an indicator of whether having value added, academic value added, and non-cognitive value added. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the *p*-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Math/ELA teachers with value added)" indicates the sample size for the outcome "Has value added (Math/ELA teachers) ", serving as a proxy sample size for other outcomes.

Table A13: RD effects of failing a certification exam on teacher value added on non-cognitive outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: White=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Attendance value added (Math teachers)	-0.000	0.000	-0.001	-0.002**	-0.001	0.334
			(0.001)	(0.001)	(0.001)	
Attendance value added (ELA teachers)	0.000	0.000	-0.000	0.000	-0.001	0.419
			(0.001)	(0.001)	(0.001)	
Suspension value added (Math teachers)	0.004	-0.009	-0.001	0.006	-0.019	0.245
			(0.008)	(0.013)	(0.014)	
Suspension value added (ELA teachers)	0.018	-0.004	0.000	-0.023*	0.027	0.033
			(0.010)	(0.014)	(0.017)	
Grade retention value added (Math teachers)	0.001	-0.002	0.003	0.001	0.003	0.670
			(0.002)	(0.004)	(0.003)	
Grade retention value added (ELA teachers)	-0.001	-0.002	0.001	0.004	-0.004	0.123
			(0.003)	(0.004)	(0.004)	
N (Math teachers with value added)	218	275	6,128	2,819	1,747	4,566
N (ELA teachers with value added)	220	272	5,977	2,768	1,653	4,421
Panel B. 2-10 years after certification exam						
Attendance value added (Math teachers)	0.000	0.000	-0.000	-0.000	-0.000	0.978
			(0.000)	(0.000)	(0.000)	
Attendance value added (ELA teachers)	0.000	0.001	-0.000	-0.000	0.000	0.389
			(0.000)	(0.000)	(0.000)	
Suspension value added (Math teachers)	0.005	-0.015	0.001	-0.004	0.005	0.074
			(0.002)	(0.003)	(0.003)	
Suspension value added (ELA teachers)	0.001	-0.015	-0.002	0.002	0.001	0.913
			(0.002)	(0.004)	(0.003)	
Grade retention value added (Math teachers)	-0.001	-0.000	0.000	0.001	0.001	0.760
			(0.001)	(0.001)	(0.001)	
Grade retention value added (ELA teachers)	-0.001	-0.001	0.000	-0.001	0.001	0.107
			(0.001)	(0.001)	(0.001)	
N (Math teachers with value added)	2,935	3,960	89,174	41,250	31,320	72,570
N (ELA teachers with value added)	2,832	3,766	84,103	39,765	29,319	69,084

Notes: The table presents the RD estimates of the impacts of failing a certification exam on teacher value added. Panels A and B indicate the RD effects in one, and two to ten years after the exam, respectively. The dependent variables include three non-cognitive value added outcomes by Math and ELA teachers: attendance, suspension, and grade retention value added. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between white and URM groups. "N (Math/ELA teachers with value added)" indicates the sample size by Math and ELA teachers.

Table A14: RD effects of failing a certification exam on outcomes conditional on having value added

	(A)	(B)	(C)	(D)	(E)	(F)
	Mean above threshold		RD coefficients			*P*-value: W=URM
	White	URM	All	White	URM	
Panel A. 1 year after certification exam						
Years of teaching experience (Math teachers)	0.384	0.351	0.014	-0.073	0.245	0.242
			(0.148)	(0.177)	(0.195)	
Years of teaching experience (ELA teachers)	0.321	0.353	-0.015	-0.014	0.008	0.941
			(0.252)	(0.186)	(0.185)	
High school math score (Math teachers)	0.386	0.404	0.006	0.097	-0.017	0.491
			(0.083)	(0.115)	(0.101)	
High school math score (ELA teachers)	0.052	0.092	-0.050	-0.054	-0.157	0.654
			(0.097)	(0.143)	(0.157)	
N (Math teachers with value added)	218	275	6,128	2,819	1,747	4,566
N (ELA teachers with value added)	220	272	5,977	2,768	1,653	4,421
Panel B. 2-10 years after certification exam						
Years of teaching experience (Math teachers)	3.951	4.555	-0.035	-0.104	-0.144	0.817
			(0.096)	(0.123)	(0.117)	
Years of teaching experience (ELA teachers)	3.761	4.499	-0.098	0.032	-0.115	0.442
			(0.139)	(0.129)	(0.134)	
High school math score (Math teachers)	0.381	0.267	0.030	0.022	0.028	0.942
			(0.041)	(0.059)	(0.061)	
High school math score (ELA teachers)	0.061	0.055	-0.031	-0.087	-0.069	0.842
			(0.042)	(0.072)	(0.055)	
N (Math teachers with value added)	2,935	3,960	89,174	41,250	31,320	72,570
N (ELA teachers with value added)	2,832	3,766	84,103	39,765	29,319	69,084

Notes: The table presents the RD estimates of the impacts of failing a certification exam on outcomes conditional on having value added. Panels A and B indicate the RD effects in one, and two to ten years after the exam, respectively. The dependent variables include two outcomes by Math and ELA teachers: years of teaching experience and high school math score. Columns (A) and (B) display the dependent variable means of candidates with exam scores ten points above the threshold. Columns (C)-(E) show the RD coefficients using equation (1). Column (F) indicates the p-value testing the statistical differences of RD effects between white and URM groups in Columns (D) and (E), respectively. "N (Math/ELA teachers with value added)" indicates the sample size by Math and ELA teachers.

Table A15: RD-DiD effects of harder certification exams on other teacher composition outcomes

	(A)	(B)	(C)	(D)	(E)	(F)
	Grade 3–4 departures			Grade 7–8 departures		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Teachers' cert exam outcomes						
Average pass rate of teacher's exam(s)	0.807	-0.020***	-0.006***	-0.002*	-0.005***	-0.018***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Score on first attempt at TExES exam(s)	257.671	-1.266***	0.512	0.009	0.095	-1.692***
		(0.188)	(0.424)	(0.200)	(0.399)	(0.652)
Panel B. Teacher demographics						
White	0.584	-0.003	-0.017***	-0.022***	-0.019***	0.017**
		(0.003)	(0.003)	(0.004)	(0.005)	(0.008)
Hispanic	0.278	0.006**	0.012***	0.012***	0.005	-0.013**
		(0.002)	(0.003)	(0.003)	(0.003)	(0.006)
Black	0.119	-0.004*	0.003	0.007**	0.009***	-0.005
		(0.002)	(0.002)	(0.003)	(0.003)	(0.006)
Male	0.091	-0.003	0.007***	0.018***	0.012***	-0.016**
		(0.002)	(0.003)	(0.004)	(0.004)	(0.006)
High school math score	0.467	-0.034***	-0.013	-0.016	0.089**	0.084
		(0.012)	(0.023)	(0.023)	(0.040)	(0.053)
High school ELA score	0.526	0.004	0.005	0.021	0.023	0.001
		(0.012)	(0.022)	(0.020)	(0.036)	(0.048)
Panel C. Teacher pay						
Earnings growth from last year	3,277	3,213***	4,119***	3,769***	4,517***	-158
		(187)	(281)	(235)	(319)	(520)
Total earnings growth for all teachers	12,976	12,956***	15,635***	7,010***	7,876***	-1,813
		(773)	(1,097)	(454)	(507)	(1,478)
Total earnings growth for URM teachers	5,051	4,448***	6,453***	2,485***	2,092***	-2,397***
		(498)	(647)	(251)	(239)	(868)
Total earnings growth for White teachers	7,577	8,027***	8,308***	4,164***	5,549***	1,105
		(516)	(803)	(321)	(417)	(1,083)
N (# *sty* observations)	4,887	105,829	86,102	123,706	107,812	423,449

Notes: This table displays RD and RD-DiD estimates of the effects of the TExES reform on teacher composition as defined by their certification exam outcomes (Panel A), demographics (Panel B), and pay (Panel C). Column (A) shows the mean of each outcome in the year prior to the teacher departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in the post-reform years (2011–2016). Columns (B)–(E) show RD coefficients β from equation (6) estimated separately for grades 3–4 and 7–8, and for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Column (F) shows the RD-DiD coefficient θ from equation (7). Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: RD-DiD effects of harder certification exams on student achievement with τ_{ty} = -3 to +3 RD window

	(A)	(B)	(C)	(D)	(E)	(F)
	Grade 3–4 departures			Grade 7–8 departures		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Student demographics (balance tests)						
Number of students with test scores	74.197	-0.306	0.972***	-6.529***	-2.280***	2.971**
		(0.263)	(0.326)	(0.778)	(0.836)	(1.214)
Male	0.497	-0.001	0.003***	0.004***	0.006***	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
White	0.270	-0.022***	-0.040***	-0.037***	-0.050***	0.005**
		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Hispanic	0.530	0.018***	0.038***	0.031***	0.043***	-0.008***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Black	0.149	-0.004***	-0.004***	-0.002*	-0.000	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Economically disadvantaged	0.664	0.007***	0.046***	0.026***	0.054***	-0.011***
		(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
In gifted education	0.094	0.001	-0.002**	-0.015***	-0.004***	0.014***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
At risk of dropping out	0.471	0.033***	0.049***	0.087***	0.031***	-0.072***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Demographic index (Math score)	-0.003	0.015***	0.001	0.021***	0.001	-0.006
		(0.001)	(0.002)	(0.002)	(0.002)	(0.004)
Demographic index (ELA score)	-0.007	0.015***	0.002	0.016***	0.007***	0.006
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Panel B. Student achievement						
Math score	-0.028	-0.002	-0.014***	0.007	-0.012**	-0.008
		(0.004)	(0.005)	(0.006)	(0.006)	(0.011)
Math score residuals	-0.007	0.000	-0.003	-0.001	-0.005	-0.000
		(0.003)	(0.003)	(0.004)	(0.003)	(0.007)
ELA score	-0.029	-0.004	-0.012***	0.002	-0.000	0.006
		(0.003)	(0.004)	(0.004)	(0.004)	(0.007)
ELA score residuals	-0.005	-0.002	-0.005**	-0.003*	0.001	0.006
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
N (# *sty* observations)	14,661	34,209	26,551	39,779	33,341	133,880

Notes: This table displays estimates of the effects of the TExES reform on student composition (Panels A) and student achievement (Panel B) from an RD-DiD specification with a larger RD window. Column (A) shows the mean of each outcome in the year prior to the teacher departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in the post-reform years (2011–2016). Columns (B)–(E) show coefficients β from a modified version of equation (6) in which we omit the running variables terms (τ_{ty} and $\mathbf{1}\{\tau_{ty} \geq 0\}\tau_{ty}$) and include only observations from $\tau_{ty} = -3$ to $+3$ in the regression sample. We estimate these β coefficients separately for grades 3–4 and 7–8, and for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Column (F) shows θ coefficients from a version of our RD-DiD specification (8) with the same modifications as the RD regression (see equation B5 in Appendix B.3.2) . Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: RD-DiD effects of harder certification exams on student achievement — Heterogeneity by race

	(A)	(B)	(C)	(D)	(E)	(F)
		Grade 3–4 departures		Grade 7–8 departures		
	Post-reform mean at τ_{ty} = -1	Post-reform RD	Pre-reform RD	Post-reform RD	Pre-reform RD	RD-DiD
Panel A. Math scores						
Math score (All students)	-0.039	-0.004	-0.006	0.014**	-0.002	-0.014
		(0.004)	(0.005)	(0.005)	(0.005)	(0.010)
Math score (White students)	0.292	-0.008	-0.001	0.005	-0.016***	-0.028*
		(0.007)	(0.009)	(0.007)	(0.006)	(0.015)
Math score (URM students)	-0.210	-0.004	0.001	0.015***	0.006	-0.014
		(0.005)	(0.007)	(0.006)	(0.006)	(0.012)
Math score (Hispanic students)	-0.146	-0.005	-0.007	0.015**	0.005	-0.008
		(0.006)	(0.008)	(0.006)	(0.007)	(0.013)
Math score (Black students)	-0.440	0.004	0.016	0.018**	0.007	-0.022
		(0.010)	(0.012)	(0.008)	(0.008)	(0.019)
Panel B. ELA scores						
ELA score (All students)	-0.035	0.000	-0.010**	0.009***	0.006	0.007
		(0.004)	(0.004)	(0.003)	(0.004)	(0.008)
ELA score (White students)	0.377	-0.013*	-0.013*	-0.001	0.007	0.007
		(0.007)	(0.007)	(0.005)	(0.006)	(0.012)
ELA score (URM students)	-0.229	0.003	-0.011*	0.012***	-0.002	-0.000
		(0.004)	(0.006)	(0.004)	(0.007)	(0.011)
ELA score (Hispanic students)	-0.199	0.006	-0.014**	0.013***	-0.008	-0.002
		(0.005)	(0.007)	(0.004)	(0.008)	(0.012)
ELA score (Black students)	-0.336	-0.009	-0.009	0.006	0.009	0.003
		(0.008)	(0.010)	(0.006)	(0.009)	(0.017)
N (# *sty* observations)	4,887	105,829	86,102	123,706	107,812	423,449

Notes: This table displays RD and RD-DiD estimates of the effects of the TExES reform on student math (Panel A) and ELA (Panel B) test scores among different race/ethnicity groups. Column (A) shows the mean of each outcome in the year prior to the teacher departure ($\tau_{ty} = -1$) in school/grades that experienced a departure in the post-reform years (2011–2016). Columns (B)–(E) show RD coefficients β from equation (6) estimated separately for grades 3–4 and 7–8, and for departures in 2011–2016 (post-reform) and 2005–2010 (pre-reform). Column (F) shows the RD-DiD coefficient θ from equation (7). Standard errors in parentheses are clustered at the school level with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Empirical Appendix

B.1 Variable definitions.

- **Annual earnings.** Annual earnings from the TWC data in 2019 dollars. We sum any reported earnings across all four quarters, and convert to 2019 dollars using CPI data from the Bureau of Labor Statistics.
- **At risk of dropping out.** The proportion of students who are reported as being "At Risk of Dropping Out."
- **Certification exam score.** The individual's score on their first attempt at their first TExES content exam. Scores range from 100–300 with 240 representing a passing score.
- **Content exam.** Throughout the paper, we focus on an individual's first TExES content exam. We define content exams to include all TExES exams that lead to classroom teacher certification excluding the the Pedagogy and Professional Responsibilities (PPR) exams. We exclude TExES exams for Principals, School Counselors, School Librarians, Educational Diagnosticians, and the Performance Assessment for School Leaders (PASL). We exclude TExES exams that do not have score ranges in 100–300 (Art, French, German, Latin, Spanish, and AAFCS: Human Development & Family Studies). We also exclude EXCET exams (the precursor to the TExES exams) as well as all other non-TExES exams (ACTFL, PACT, TASC, TASC-ASL, TEXMAT, and edTPA). We identify the earliest date at which an individual took one of the TExES content exams and focus on outcomes related to that exam. If an individual took multiple TExES exams on their earliest date, we include observations for each separate exam.
- **Days between first try and passing.** The number of days between the date at which the individual first took the content exam and the date at which they first passed the exam. This variable is defined only for individuals who ever passed the exam, and is equal to zero for individuals who passed on their first try.
- **Demographic index (math/ELA score).** An index of predicted math/ELA scores (in SD units) based on a large vector of student covariates. The covariates are dummies for the full interaction of sex, race/ethnicity, economic disadvantage, at-risk of dropping out, special education, gifted education, grade, and year. The demographic indices are predicted values from regressions of math/ELA scores (standardized to the exam/year level) on these dummies, estimated separately for each test regime (TAAS, TAKS, and STAAR).
- **Economically disadvantaged.** The proportion of students who are eligible for free or reduced-price meals or who are reported as having "Other economic disadvantages".
- **Employed as a teacher.** The total paid full-time equivalent (FTE) years that the individual worked in any Texas public school in a given year since taking their first TExES content exam.

- **Ever passed any exam.** An indicator equal to one if the individual ever passed any TExES content exam (regardless of whether or not it was the first TExES content exam that they attempted).
- **Ever passed exam.** An indicator equal to one if the individual ever passed the first TExES content exam that they attempted.
- **Experienced URM teacher.** The proportion of teachers who are both URM and received their initial teaching certification more than five years ago relative to the year in which they are teaching
- **Experienced White teacher.** The proportion of teachers who are both White and received their initial teaching certification more than five years ago relative to the year in which they are teaching
- **First-year teacher.** A teacher that meets two criteria: 1) They are in the first academic year in which we observe the teacher in the TEA staff files; and 2) the TEA employment files report that the teacher has 0 years of experience.
- **High school (math/ELA) score.** The math/ELA score we have for a teacher's high school math/ELA exams, standardized at the grade-year level. For the TAAS exam (1994–2002), we use the (mostly 10th grade) math and reading exit scores. For the TAKS exam (2003–2011), we use the grade 10 end-of-grade math and reading/ELA scores. For the STAAR exam (2012–2022), we use the end-of-course Algebra I score and the average of the end-of-course English 1 and English 2 scores. We use the individual's first score from any of these math/ELA exams. See the definition of math/ELA scores below for details on the test scores.
- **In gifted education.** The proportion of students who are reported as being in Gifted and Talented programs.
- **Math/ELA score (SD units).** Scale scores on standardized grade 3–8 math and English Language Arts (ELA) achievement tests, standardized to mean zero and SD one within the population of individuals who took the same exam in the same year. For each student $\times$ year $\times$ subject exam, we use the first non-missing score. Scores are from the three testing regimes that existed in Texas during our time period: TAAS (1994–2002), TAKS (2003–2011), and STAAR (2012–2022). We use Texas Learning Index (TLI) and Normal Curve Equivalent (NCE) scores for the TAAS regime, and scale scores for the TAKS and STAAR regimes. Scores for some exam takers in grades 3–5 are from Spanish language versions of these tests.
- **Math/ELA score residuals (SD units).** Residuals from a regression of math/ELA scores in SD units (defined above) on a large vector of student, school $\times$ grade, and school level controls. The controls variables mirror those in Chetty et al. (2014a)'s teacher value-added specification, except we use school/grade level controls rather than classroom controls because we only observe students' classrooms from 2012 onward. Student-level controls include cubics in lagged math and ELA scores, sex, economic

disadvantage, race/ethnicity dummies, at-risk of dropping out, gifted education, and indicators for missing values of the demographics variables. We also include school × grade and school averages of each of these variables. We interact all covariates with grade dummies, and include grade × year dummies in the regression. For this residual variable we include only students who have lagged test scores in both subjects, with the exception of grade 3 students since there are no grade 2 tests. Thus controls for grade 3 students include only those based on demographic variables, not lagged test scores.

- **New teacher (within 5 years of cert).** The proportion of teachers who received their initial teaching certification within five years relative to the year in which they are teaching.
- **New URM teacher.** The proportion of teachers who are both URM and received their initial teaching certification within five years relative to the year in which they are teaching.
- **New White teacher.** The proportion of teachers who are both White and received their initial teaching certification within five years relative to the year in which they are teaching
- **Number of students with test scores.** The total number of students in the school/-grade with non-missing math scores.
- **Number of times taking this exam.** The number of times the individual attempted the content exam.
- **Passed exam on first try.** An indicator equal to one if the individual passed their first TExES content exam on their first attempt.
- **Pedagogy and Professional Responsibilities (PPR) exam.** In Table 1, we include summary statistics for TExES PPR exams. These include the Pedagogy and Professional Responsibilities EC-4/EC-6/EC-12/4-8/8-12 exams and the Pedagogy & Prof Responsibilities Trade & Industrial Ed 6-12/8-12 exams.
- **Present value (PV) cumulative earnings since exam.** We deflate nominal earnings to the year in which the individual took their first TExES content exam using a 5 percent discount rate, and then sum across all years in the reported time range since taking the exam.
- **Total salary for all teachers.** The sum of total teaching salary from the TEA data for all teachers in a given school × grade × subject × year cell. For this we sum the base pay and other supplementary pay to get the teacher's total annual salary, multiply the annual salary by the proportion of full-time equivalent (FTE) years associated with each teaching assignment, and then sum across all teachers in the school × grade × subject × year.

- **Total salary for URM teachers.** The sum of total teaching salary from the TEA data for all URM teachers in a given school × grade × subject × year cell.
- **Total salary for White teachers.** The sum of total teaching salary from the TEA data for all White teachers in a given school × grade × subject × year cell.
- **Took the Generalist EC-4 exam.** The proportion of teachers who ever took and Generalist EC-4, Bilingual Generalist EC-4, or English as a Second Language (ESL)/Generalist EC-4 exam.
- **Took the Generalist EC-6 exam.** The proportion of teachers who ever took and Generalist EC-6, Bilingual Generalist EC-6, or English as a Second Language (ESL)/Generalist EC-6 exam.
- **Underrepresented minority (URM).** Individuals whose race/ethnicity is reported in the data as Black or Hispanic. Throughout the paper, when we compare outcomes for White and URM individuals, we exclude individuals whose race ethnicity is reported in the data as Native American, Asian, or Two or More Races (although these race/ethnicity groups are included when we examine outcomes for all individuals).
- **Years of teaching experience.** The sum of all paid full-time equivalent (FTE) years that the individual worked in any Texas public school in the reported time range since taking the exam.

B.2 Calculating Value-Added.

B.2.1 Teacher Value-Added Estimation. We obtain the academic and non-cognitive value-added estimate for each teacher j using the following equations (B1) and (B2) separately for each subject sub (math or ELA):

$$A^{sub}_{ijkgst} = \alpha_1 A^{sub}_{it-1} + \alpha_2 A^{-sub}_{it-1} + \gamma X_{it} + \lambda C_{kgst} + \zeta S_{st} + \nu_{gt} + \epsilon_{ikgst} \tag{B1}$$

$$NC^{sub}_{ijkgst} = \alpha_1 NC^{sub}_{it-1} + \alpha_1 A^{sub}_{it-1} + \alpha_2 A^{-sub}_{it-1} + \gamma X_{it} + \lambda C_{kgst} + \zeta S_{st} + \nu_{gt} + \epsilon_{ikgst} \tag{B2}$$

For academic value-added estimation in equation (B1), A^{sub}_{ijkgst} is student i's standardized math or ELA score in year t, grade g, classroom k, and taught by teacher j in school s. We include a set of controls following Chetty et al. (2014b). Student i's A^{sub}_{it-1} and A^{-sub}_{it-1} represent lagged standardized math and ELA scores and their squares and cubes. X_{it} are student characteristics (economic disadvantage, ethnicity/race, sex, whether they are in special education, whether they are at risk, and whether they are gifted). Classroom characteristics, C_{kgst}, and school characteristics, S_{st}, include the mean individual characteristics, mean lagged standardized test scores in math and ELA and their squares and cubes for all students in classroom k and school s, respectively.[33] We also interact all student, class, and

[33]We implement a similar procedure as of Chetty et al. (2014b) to process the missing values when estimating the model: (i) we generate indicators for students with missing values in the other subject and include the interaction terms between the indicators and lagged own-subject test scores; (ii) we do the same thing for class-level lagged test scores across the two subjects; (iii) we include indicators for students with missing values in demographic controls.

school-level controls with grade-level to allow for differences in effect across grades. Lastly, to control for grade-year specific factors affecting all students, we include ν_{gt}.[34]

We implement a similar procedure as of Chetty et al. (2014b) for sampling. We exclude students with missing values in current or lagged test scores in the subject for which we are estimating VA. We restrict students who have only one teacher per subject for the whole year. We restrict to teachers who taught the school-grade-subject cell for the whole year ($\geq$ 275 days), as well as restrict to students who enrolled in the class of a school-grade-subject cell for the whole year ($\geq$ 275 days). We keep classrooms which have 10 and 50 observations with current and lagged scores in the relevant subject. We also keep the observation associated with the most recent classrooms. Finally, we drop teacher-subject cells with fewer than 10 observations.

When estimating non-cognitive value-added, we employ the equation (B2). NC^{sub}_{ijkgst} is an index that measures student i's non-cognitive outcomes in year t, grade g, classroom k, and taught by teacher j in school s. The index is constructed using students' behavioral outcomes: attendance (percentage of days that student was present), grade retention (indicator), and suspension (indicator) (Jackson, 2018).[35] In addition to the same set of controls used in equation (B1), equation (B2) also controls for lagged values of the non-cognitive index, including its linear, squared, and cubic terms, along with their interactions with grade fixed effects.[36]

Lastly, we obtain teacher-year estimates of academic ($\bar{A}_{jt}$) and non-cognitive value-added ($\bar{NC}_{jt}$) by estimating equations (B1) and (B2) and performing the following exercises, separately for each subject:

1. Obtain residuals by taking the difference between the student test score and the estimates from the previous regression with all controls except the teacher dummy:
   $a_{it} = A_{it} - [\hat{\alpha}_1 A^{sub}_{it-1} + \hat{\alpha}_2 A^{-sub}_{it-1} + \hat{\gamma} X_{it} + \hat{\lambda} C_{kgst} + \hat{\zeta} S_{st} + \nu_{gt}]$
   $nc_{it} = NC_{it} - [\hat{\alpha}_1 NC^{sub}_{it-1} + \hat{\alpha}_2 A^{sub}_{it-1} + \hat{\alpha}_3 A^{-sub}_{it-1} + \hat{\gamma} X_{it} + \hat{\lambda} C_{kgst} + \hat{\zeta} S_{st} + \nu_{gt}]$

2. Then average the residuals among teacher j's students at the teacher-year:
   $\bar{A}_{jt} = \text{Mean}[a_{it} | i \in \{i : j(i,t) = j\}]$
   $\bar{NC}_{jt} = \text{Mean}[nc_{it} | i \in \{i : j(i,t) = j\}]$

$\bar{A}_{jt}$ and $\bar{NC}_{jt}$ are our estimates of teacher-year academic and non-cognitive value-added, respectively. The estimated value-added is essentially the classroom-year residual where $\hat{\beta}$ is estimated using within teacher variation. This follows Chetty et al. (2014b) by accounting for non-random sorting of students across teachers and its effect on the estimation of student, class, and school characteristics. We differ from Chetty et al. (2014b) by not completing

[34] We also include fixed-effects for served student population type (e.g., regular, honors, gifted and talented, bilingual, and special education) defined at the classroom-level interacted with grade.

[35] Similar to Jackson (2018), we first estimate an iterated principal factors method on the behavioral outcomes. There is only one principal component (the first eigenvalue is 0.49 and the second and third are 0.0002 and -0.0002). We then computed the unbiased prediction of this sole component using the Bartlett method. The predicted non-cognitive index equals 1.21 (attendance) - 0.53 (grade retention) - 0.56(log of 1 + suspension).

[36] The non-cognitive value-added estimation thus excludes students with missing current or lagged values in grade retention, attendance, or suspension.

their final step which calculates each $\bar{A}_{jt}$ (or $\bar{NC}_{jt}$) as a function of other years' estimates (drift). We deviate because we prefer to have an estimate for teachers with only one year of teaching, which is not possible in their method. In practice, the estimates on value-added between the two approaches are highly correlated.

B.2.2 Differences in Teacher Effectiveness across Groups. Rather than directly estimating teacher value added, we can alternatively assess how teacher effectiveness very across groups following Kane et al. (2008). To do so, we add an additional indicator, URM, into the VA equations (B1) and (B2), which gives the following specification for academic and non-cognitive value added:

$$A^{sub}_{ijkgst} = \alpha_1 A^{sub}_{it-1} + \alpha_2 A^{-sub}_{it-1} + \gamma X_{it} + \lambda C_{kgst} + \zeta S_{st} + \nu_{gt} + URM_{jt} + \epsilon_{ikgst} \tag{B3}$$

$$NC^{sub}_{ijkgst} = \alpha_1 NC^{sub}_{it-1} + \alpha_1 A^{sub}_{it-1} + \alpha_2 A^{-sub}_{it-1} + \gamma X_{it} + \lambda C_{kgst} + \zeta S_{st} + \nu_{gt} + URM_{jt} + \epsilon_{ikgst} \tag{B4}$$

Given that we assess the teacher effectiveness among exam takers in different race groups, we include exam-year FEs as in the baseline RD equation (1). In some models, we further include terms that measure exam performance — an indicator for passing the exam and certification exam scores — to assess how teacher effectiveness vary by performance and race for exam takers who end up teaching. In these models, we include alternative exam-year-race, exam-year-passed exam, or exam-year-race-passed exam FEs. Finally, we cluster standard errors at the school-level.

B.3 RD-DiD specification. This section provides details on our regression discontinuity difference-in-differences (RD-DiD) specification that we use in Section 6

B.3.1 Definition of teacher departures. We define a *teacher departure* as an instance in which a teacher with five or more years of experience leaves a given school. In other words, a departure occurs when a teacher with 5+ years of experience appears in a school one year but does not appear at that same school in the next year. Teachers that change subjects or grades within the same school are not counted as departures. We let t denote calendar years, y denote the year of the teacher departure, and τ_{ty} denote years relative to the departure. We define the departure year y as the first year that the teacher is no longer at the school, or, equivalently, $\tau_{ty} = 0$.

The sample of teacher departures that we include in our RD-DiD analysis includes two main restrictions. First, we consider only the departure of grade 3–4 and grade 7–8 teacher in subjects for which we have test scores. In most of our analyses, we define focus on grade 3–4 *generic* subject teacher, where generic means a teacher that taught any of the following subjects (with associated `subject` codes in the ERC data): General Science (8), Mathematics (10), English (22), Reading (27), Social Studies (38), and Generic (98). We combine all of these subjects into a single generic subject because many elementary teachers teach all of these core subjects, and in many cases the TEA data codes the teaching subject as Generic (98). For Table 8, however, we focus on grade 3–4 teachers who are specifically reported as teaching either Mathematics (`subject` = 10) or English/Reading (`subject` = 22 or 27). For grade 7–8 teachers, we require that the departing teacher taught either Mathematics (`subject` = 10) or English/Reading (`subject` = 22 or 27), and we treat the two subjects separately for our analysis.

Second, we include only teachers who taught one third or more of the students in a given school, grade, and subject in the year prior to their departure year. We sum the teacher's full-time equivalent (FTE) years in the school/grade/subject, divide it by the total FTE in that school/grade/subject, and keep only departures in which the teacher's FTE is one-third or more of the total FTE. This restriction allows us to focus on instances in which the departing teacher causes a significant change in teacher composition at the school/grade/subject level, which is the level at which we can connect teachers to student test scores across all years of our data.

Teacher departures that meet both the subject and the FTE requirements are included in our RD-DiD analysis. We let s denote the school/grade/subject triplets that are associated with each teacher departure, and stack our dataset to include observations associated with each departure event as described in Section 6.1.[37]

B.3.2 Benchmark RD-DiD specifications. This subsection provides details on our RD-DiD regression specifications.

In Tables 6–7 and Appendix Tables A15 and A17), column (F) displays RD-DiD coefficients θ from equation (8) with Treated_g defined as an indicator for grades 3–4 and Post_p defined as an indicator for teacher departures in $y \in$ 2011–2016. As described in Sections 6.1–6.2, the intuition for equation (8) comes from a two-step specification. First, estimate the RD regression (6) separately for each pairwise combination of school exposure group $g \in$ {treated, control} and departure period $p \in$ {2005–2010, 2011–2016}, which gives four RD coefficients β_{gp}. Second, use the resulting RD coefficients β_{gp} as dependent variables in the simple DiD regression (7).

For Appendix Table A16, we estimate a modified version of equation (8) in which we omit the running variables terms (τ_{ty} and $\mathbf{1}\{\tau_{ty} \geq 0\}\tau_{ty}$) and include only observations from $\tau_{ty} = -3$ to $+3$ in the regression sample:

$$Y_{st} = \Big(\phi \text{Treated}_g + \delta \text{Post}_p + \theta \text{Treated}_g \text{Post}_p\Big)\mathbf{1}\{\tau_{ty} \geq 0\} + \gamma_{sy} + \varepsilon_{sty} \quad \text{if } |\tau_{ty}| \leq 3. \tag{B5}$$

In this specification, the intuition for the θ coefficient comes from the two step procedure: 1) Let β_{gp} represent the change in average outcomes in the three years after the teacher departure relative to the three years before the departure (rather than an RD coefficient, as in equation 6) for each pairwise combination of school exposure group $g \in$ {treated, control} and departure period $p \in$ {2005–2010, 2011–2016}; and 2) Use the resulting β_{gp} coefficients as dependent variables in the simple DiD regression (7).

B.3.3 RD-DiD specifications for Table 8. In Table 8, we restrict our sample to students in grades 3–4 who have a departing teacher that is identified in the data as specifically teaching math or ELA (see Appendix B.3.1) and use test scores or students in subjects that are unaffected by the departure as the control group. For this analysis, we also restrict our sample to school/grades that have teachers who are identified as teaching math or ELA specifically in *every* year from 2005–2016.

In Panels A–B of Table 8, we exploit within-school variation by using exam scores in the subject the teacher did *not* teach as a control group. In Panel A, we restrict our regression

[37] Note that a teacher can depart from multiple grades or subjects in the same year if they teach more than one grade or subject.

sample to school/grades with a departing math teacher. In Panel B, we restrict our regression sample to school/grades with a departing ELA teacher. In both cases we reshape our data so that there are two test score observations for each student/year (one in math and the other in ELA), and then estimate the following RD-DiD specification:

$$\begin{aligned} Y_{st} = \Big(\phi \text{SameSubject}_e + \delta \text{Post}_p + \theta \text{SameSubject}_g \text{Post}_p\Big)\mathbf{1}\{\tau_{ty} \geq 0\} + \\ \alpha_{gp}\tau_{ty} + \psi_{gp}\mathbf{1}\{\tau_{ty} \geq 0\}\tau_{ty} + \gamma_{sy} + \varepsilon_{sty} \quad \text{if } |\tau_{ty}| \leq h^Y_{gp}. \end{aligned} \tag{B6}$$

where s now denotes school/grade/*exam-subject* triplets and SameSubject_e is an indicator equal to one if the exam subject is the same as the subject of the departing teacher (e.g., the math score when the math teacher departs, and the ELA score when the ELA teacher departs). All other variables in equation (B6) are defined as in our benchmark RD-DiD specification (8). Thus in Panel A, the θ coefficient shows how the effect of a math teacher departures on math scores changes relative to the effect on ELA scores. In Panel B, the θ coefficient shows how the effect of an ELA teacher departures on ELA scores changes relative to the effect on math scores.

In Panels C–D of Table 8, we exploit across-school variation by using school/grades with a departing teacher who taught the *other* subject as a control group. We again restrict our regression sample to schools with a departing grade 3–4 math or ELA teacher. We estimate the same RD-DiD specification (B6) as above, but s now denotes school/grade/*teaching-subject* triplets. We continue to define SameSubject_e as an indicator equal to one if the exam subject is the same as the subject of the departing teacher. In Panel C, we use math scores as the outcome variable, and thus the θ coefficient shows how the effect of a math teacher departure on math scores changes relative to the effect of an ELA teacher departure on math scores. In Panel D, we use ELA scores as the outcome variable, and thus the θ coefficient shows how the effect of an ELA teacher departure on ELA scores changes relative to the effect of a math teacher departure on ELA scores.