



Screen or Barrier? Science of Reading Certification Exams and Teacher Supply in Texas

Andrew Avitabile
University of Virginia

Twenty states now require prospective teachers to pass a standalone Science of Reading (SOR) exam. These exams restrict teacher supply as a barrier, and whether they are worth that cost depends on whether they also screen on teaching quality. This paper explores this tension in Texas' SOR exam, required since 2021 for new elementary and middle-grades candidates, using administrative data linking exam attempts, certificates, employment, and student achievement. In a regression discontinuity at the passing threshold, narrowly failing lowers the probability of teaching the following year by 6.6 percentage points, driven by alternatively certified candidates. A difference-in-differences design comparing affected to unaffected certification fields within the same preparation programs shows that new certifications in affected fields surged ahead of implementation, then fell, and have not recovered five years later. Against these costs, exam scores predict student reading achievement more strongly than scores on the other certification exams candidates take, and students of teachers who passed on the first attempt score 0.024 standard deviations higher in reading than students of teachers who failed. The exam is thus a rare pre-hire predictor of teacher effectiveness, and lowering the threshold would likely admit less effective teachers. Structured Science of Reading coursework at preparation programs raised first-attempt pass rates about 4 percentage points, suggesting states can lower the barrier without lowering the standard.

VERSION: October 2026

Suggested citation: Avitabile, Andrew. (2026). Screen or Barrier? Science of Reading Certification Exams and Teacher Supply in Texas. (EdWorkingPaper: 26-1615). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/m844-rw67>

Screen or Barrier? Science of Reading Certification Exams and Teacher Supply in Texas *

Andrew Avitabile
University of Virginia
[Click here for the latest version](#)

Last Updated October 5, 2026

Abstract

Twenty states now require prospective teachers to pass a standalone Science of Reading (SOR) exam. These exams restrict teacher supply as a *barrier*, and whether they are worth that cost depends on whether they also *screen* on teaching quality. This paper explores this tension in Texas' SOR exam, required since 2021 for new elementary and middle-grades candidates, using administrative data linking exam attempts, certificates, employment, and student achievement. In a regression discontinuity at the passing threshold, narrowly failing lowers the probability of teaching the following year by 6.6 percentage points, driven by alternatively certified candidates. A difference-in-differences design comparing affected to unaffected certification fields within the same preparation programs shows that new certifications in affected fields surged ahead of implementation, then fell, and have not recovered five years later. Against these costs, exam scores predict student reading achievement more strongly than scores on the other certification exams candidates take, and students of teachers who passed on the first attempt score 0.024 standard deviations higher in reading than students of teachers who failed. The exam is thus a rare pre-hire predictor of teacher effectiveness, and lowering the threshold would likely admit less effective teachers. Structured Science of Reading coursework at preparation programs raised first-attempt pass rates about 4 percentage points, suggesting states can lower the barrier without lowering the standard.

*Andrew Avitabile: yaj3ma@virginia.edu. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200005 to the University of Virginia, and by the Bankard Fund for Political Economy at the University of Virginia. The opinions expressed are those of the author and do not represent views of the Institute, the U.S. Department of Education. The author thanks Brendan Bartanen and Jim Wyckoff for comments on earlier drafts of the paper and seminar participants at the University of Virginia's EdPolicyWorks Policy Lab for helpful suggestions. Any remaining errors are my own.

1 Introduction

Teacher certification exams occupy an uneasy position in education policy. They create a *barrier* to teacher supply (Kleiner, 2006), but they may also be a *screen* if they filter candidates on knowledge relevant to student learning (Stiglitz, 1975). Whether an exam is worth its cost to teacher supply depends on which dominates, and for two decades the empirical evidence has pointed one way. More stringent requirements reduce teacher supply without improving quality (Angrist & Guryan, 2008; Deneault et al., 2026), passing thresholds do a poor job of separating more from less effective candidates (Goldhaber, 2007; Orellana & Winters, 2025), and exam scores are, at best, weakly predictive of teacher effectiveness (Cowan et al., 2023; Goldhaber et al., 2017; Kane et al., 2008; Rockoff et al., 2011). The conclusion drawn from this evidence is that states should lower barriers to entry and instead reward teachers who raise student achievement (Hanushek & Rivkin, 2007).

Following this evidence some states have dropped testing requirements (NJ State Policy Lab, 2025; Swisher, 2022), but two recent developments have pushed others in the opposite direction. First, teachers who entered through relaxed pandemic-era pathways proved somewhat less effective than their traditionally certified peers (Backes et al., 2024; Kirksey, 2024). Second, in 2024, roughly 40 percent of fourth graders scored below basic in reading on the National Assessment of Educational Progress (NAEP), the highest share since 2002 (National Center for Education Statistics, 2025a). States have responded to declining academic achievement in reading by passing legislation that requires literacy instruction that contains explicit, systematic teaching of foundational skills (Castles et al., 2018; Seidenberg, 2017)—called the Science of Reading (SOR). Despite these efforts, many educator preparation programs (EPPs) often do not provide prospective teachers with these skills (Ellis et al., 2023). In response, twenty states now require a standalone SOR exam as a condition of certification (National Council on Teacher Quality, 2024).

Although these exams have become increasingly popular, states lack evidence on how they affect the teacher pipeline. In this paper I ask three questions. To what extent are SOR certification exams a barrier to supply? Do they screen on teacher effectiveness? And what can policymakers and EPPs do to help candidates meet the standard?

I answer these questions in the context of the Texas Examinations of Educator Standards (TExES) SOR exam, announced in June 2019 and required beginning January 1, 2021 of all new candidates in early childhood through grade 6 and grades 4–8 core subjects and English language arts. The data for this study cover the universe of Texas certification exams, certificates, and public school employment, and are linked to student achievement records. To measure the exam as a barrier, I use a regression discontinuity at the passing threshold that identifies the effect of narrowly failing on certification and employment, and a difference-in-differences design that compares the flow of new certifications in fields that began requiring the exam to fields within the same EPPs that did not. To measure the exam as a screen, I explore the extent to which SOR scores predict student achievement and whether the candidates the exam excludes are less effective than the ones it admits. Finally, I ask whether Texas Reading Academies (TRAs), state-developed SOR coursework delivered by EPPs, can raise pass rates on SOR certification exams.

The SOR exam is a significant barrier to teacher supply. Narrowly failing reduces the probability of earning a teaching certificate within six months by 8.0 percentage points and of working as a teacher the following October by 6.6 percentage points. The employment cost depends on where the exam falls in a candidate’s route. Traditional candidates take it at the end of their preparation, and for them a narrow failure delays certification but not subsequent employment. Alternatively certified candidates must pass before their program can place them on the paid internship the route requires, so a narrow failure for those on this route lowers the probability of teaching the next year by 11.4 percentage points. At the market level, new certifications in affected fields spiked ahead of the mandate as candidates rushed to certify under the old rules, then fell, and have not recovered five years later. The size of the decline is imprecisely estimated and depends on how pre-announcement trends are handled, so I treat the pattern as the finding rather than any single number. I also find no evidence of substitution, as candidates who narrowly fail are no more likely to certify in unaffected fields.

On the screening side, I ask whether the exam provides a pre-hire signal of teaching quality and how the candidates the passing threshold excludes compare to the ones it admits. Using student-level value-added models, I find that students assigned to a teacher who scored

one standard deviation higher on the SOR exam score 0.010 standard deviations higher in reading within the same grade and school, about 0.05 standard deviations of teacher effectiveness. Although the size of this correlation is modest, it is notable given how difficult it has proven to predict teacher effectiveness before hire (Clotfelter et al., 2007; Cowan et al., 2023; Kane & Staiger, 2008; Staiger & Rockoff, 2010). This association is also stronger than that between student achievement and other related certification exams these candidates take, and comparable to the strongest licensure-test estimates from other states (Clotfelter et al., 2007; Cowan et al., 2023). I also find evidence that the threshold excludes candidates who are less effective than the ones it admits. Students of teachers who passed the exam on the first attempt score 0.024 standard deviations higher in reading than students of teachers who failed it, which is again larger in magnitude than for other comparable exams. This gap is also likely understated, since the failers who reach a tested classroom are more positively selected on family income and SAT scores than the passers who do. I interpret this as evidence that lowering the threshold would admit teachers who are, on average, less effective at raising reading achievement than those it currently admits.

Taken together, these results put states in a bind. A state deciding whether to adopt or keep an SOR exam is choosing between pre-hire information about teaching quality and a considerable cost to teacher supply. This is occurring at a moment when reading achievement and interest in teaching are at or near their lowest levels in decades (Bartanen & Kwok, 2023; Kraft & Lyon, 2024; National Center for Education Statistics, 2025a). This need not be the only choice if states could keep the standard and help candidates meet it, but prior literature offers almost no evidence on this option. Here I provide preliminary evidence evaluating TRAs, which ten EPPs began delivering to preservice candidates in the 2024–25 school year. First-attempt SOR pass rates at these programs rose about 4 percentage points relative to programs that never adopted. This evidence is early, as the adopting programs are few, pass rates on the pedagogy exam also rose in the adoption year, and a second cohort of adopters fails a pre-trend test. But the results suggest that the supply cost of an exam is not fixed, and that states can lower it without removing minimum competency standards. As more states add SOR requirements, how to help candidates pass these exams deserves as much attention as whether to test.

2 Background

2.1 Conceptual Framework: Licensure as Screen and Barrier

This paper draws on three related economic traditions. The economics of occupational licensing (Kleiner, 2006; Kleiner & Krueger, 2013) treats licensure as a labor market institution performing two functions simultaneously. First, a *screening* function that filters candidates on minimum competency and second a *barrier* function that imposes costs (e.g., fees, preparation time, retake delays, and foregone earnings) on prospective entrants. The welfare consequences of a licensure requirement depend on the balance between these two. If an exam screens effectively, its supply costs may be offset by quality gains. If it imposes barriers without meaningful screening, it restricts supply without benefits.

Whether an exam screens effectively depends on what its scores measure. The classical distinction between *screening* (Stiglitz, 1975) and *signaling* (Spence, 1973) is useful here. A screen identifies differences in candidate knowledge relevant to performance, while a pure signal produces a credential whose value derives from the cost of obtaining it rather than what it certifies. Together, these traditions generate three testable predictions that organize the empirical work. First, if the exam functions as a binding barrier, failing should delay certification and reduce subsequent employment, with larger effects for candidates already at the margin of the labor market. Second, if the exam screens rather than signals, scores should predict student reading achievement and stop less effective candidates from entering the profession. Third, if the exam acts as a barrier due to inadequate preparation rather than fixed differences in candidate ability, then structured preparation that builds this knowledge is an investment in human capital (Becker, 1964) and should improve exam performance. If the exam is instead closer to a pure signal, such preparation may help candidates pass without making them better teachers. The remainder of this section reviews the evidence base motivating SOR exams, the literature on certification exams, teacher supply, and quality, and the Texas institutional context in which these predictions are tested.

2.2 The Science of Reading and Teacher Preparation

The Science of Reading synthesizes decades of empirical research showing that reading acquisition requires explicit, systematic instruction in phonological awareness, phonics, fluency, vocabulary, and comprehension (Castles et al., 2018; Seidenberg, 2017). Meta-analyses consistently demonstrate the effectiveness of systematic phonics instruction for improving literacy outcomes (Ehri et al., 2001; McArthur et al., 2018). Despite this evidence, EPPs have historically emphasized approaches that downplay these methods (e.g., balanced literacy and whole language), leaving classroom literacy instruction without an evidence-based foundation (Moats, 2020). Ellis et al. (2023) analyzed 702 elementary educator preparation programs across all fifty states and found that 30% provide no practice opportunities connected to the core components of the Science of Reading.

States have responded to this evidence aggressively. As of January 2026, 20 states require a standalone certification exam demonstrating minimum competency in the Science of Teaching Reading (National Council on Teacher Quality, 2024). In some cases, broad SOR reforms have coincided with improved student outcomes. For example, quasi-experimental evidence from Mississippi’s comprehensive literacy reforms indicates gains of 0.23 standard deviations in reading (Spencer, 2024), and Novicoff and Dee (2025) find similarly positive effects of California’s Early Literacy Support Block Grant. However, these initiatives bundle curriculum adoption, coaching, professional development, and certification requirements, without isolating the contribution of SOR licensure exams specifically. To date, no studies have directly examined whether Science of Reading certification exams improve student outcomes or how they shape the teacher workforce.

2.3 Teacher Certification Exams, Quality, and Supply

A large literature shows that certification requirements function as binding barriers to entry, shaping both the size and composition of the teacher workforce (Angrist & Guryan, 2008; Boyd et al., 2007; Deneault et al., 2026). Angrist and Guryan (2008) demonstrate that states that added more stringent certification requirements reduced teacher supply and raised teacher wages without measurable improvements in teacher quality. More recently,

Deneault et al. (2026) use Texas certification exam data to document that narrowly failing a certification exam substantially delays entry into teaching and imposes significant earnings losses, with considerably larger effects for underrepresented minority (URM) candidates than for white candidates. These supply costs are especially salient at a moment of declining interest in the teaching profession and persistent gaps between the demographic composition of the teacher workforce and the students it serves (Bartanen & Kwok, 2023; Kraft & Lyon, 2024).

Whether these barriers are justified depends on the exams' screening value, and here the evidence is mixed. Teacher effectiveness has proven difficult to predict from anything observable at hire, but certification exam scores are generally modestly correlated with teacher effectiveness (Clotfelter et al., 2007; Kane et al., 2008; Rockoff et al., 2011). Goldhaber (2007) finds a small positive relationship between North Carolina licensure test scores and student achievement, but estimates that raising the state's cutoff to Connecticut's would exclude fewer than 0.5 percent of the workforce judged very ineffective while excluding 7 percent judged effective. Comparing teachers with the same school and grade in Massachusetts, Cowan et al. (2023) find that a one standard deviation higher licensure score predicts 0.012 standard deviations higher math achievement and 0.005 in English language arts.¹

Other studies have found less positive evidence for certification exams as screens. Buddin and Zamarro (2009) find no relationship between California teachers' Reading Instruction Competence Assessment (RICA) exam scores and reading achievement among Los Angeles elementary teachers. In the same study described above, Deneault et al. (2026) find nearly identical value-added distributions for teachers who passed easier and harder Texas elementary certification exams, suggesting that making exams more difficult does little to improve teacher quality. Orellana and Winters (2025) examine the Praxis II thresholds in Connecticut and find that failing disproportionately deters candidates at least as effective as those who passed. In sum, existing exams pose significant barriers and there is mixed evidence on whether they screen on teacher effectiveness sufficiently enough to justify costs to supply.

Most states have read this evidence and continued to require and expand certification

¹Notably, Cowan et al. (2023) examine the Massachusetts Tests for Educator Licensure (MTEL) as an average across the exams each candidate takes, which for elementary candidates includes the Foundations of Reading test.

testing, likely for two reasons. First, as preparation pathways grow more heterogeneous with the expansion of alternative certification (National Council on Teacher Quality, 2025), exams are one of the few mechanisms states retain to verify minimum competency. This concern is validated by evidence that teachers entering through relaxed pandemic-era emergency pathways were less effective than traditionally prepared peers (Backes et al., 2024; Kirksey & Gottlieb, 2023). Second, as states adopt new instructional approaches to combat declining student performance on national reading assessments (National Center for Education Statistics, 2025b)—such as Science of Reading-based literacy instruction—new exams are created to verify these specific skills.

Other states have considered the evidence on exams as screens and barriers and moved in the opposite direction. For example, New Jersey eliminated its Praxis Core requirement effective January 2025, citing shortages and financial barriers (NJ State Policy Lab, 2025). Similarly, several states have dropped or made optional the edTPA (Amin, 2022)—a performance-based, subject-specific assessment in which teacher candidates assemble a portfolio during their student teaching, including lesson plans, video of their own instruction, and written commentaries, scored externally to gauge classroom readiness.

Much of this policy debate treats testing as binary policy: maintain requirements that restrict supply and diversity or weaken standards and risk admitting underprepared teachers. This study first asks whether Texas’ SOR exam is worth its cost on those terms, but then examines a policy that may help candidates meet the standard rather than choosing whether or not to enforce it.

2.4 Texas’ Science of Reading Exam

In June 2019, Texas passed House Bill 3, which mandated that new teacher candidates demonstrate proficiency in the Science of Reading through a certification exam. Beginning January 1, 2021, candidates pursuing certifications in early childhood through grade 6, core subjects grades 4–8, or English Language Arts grades 4–8 must pass the TExES Science of Reading exam (exam 293). The exam assesses knowledge of phonological awareness, phonics, fluency, vocabulary, and reading comprehension instruction grounded in evidence-based practices. Example items are provided in Appendix B.

The exam consists of 90 multiple-choice questions scored on a scaled system with a passing score of 240. Candidates who fail may retake the exam after a 30-day waiting period, are limited to five attempts absent a waiver, and pay a fee of \$116 per attempt, creating direct financial costs for candidates who must retake. The affected certifications account for 41.7% of newly issued teaching certificates in Texas in the pre-mandate period (Table 2), making the SOR exam a significant checkpoint in the teacher certification pipeline.

The SOR exam is one of three exams required for certification in these fields. Every candidate also takes the Pedagogy and Professional Responsibilities (PPR) exam, a test of general pedagogy required for all Texas teaching certificates, and a content exam specific to the field. For the generalist certificates (Core Subjects EC–6 and Core Subjects 4–8), the content exam is the corresponding Core Subjects exam, a composite of separately scored subject tests. For the English Language Arts and Reading 4–8 certificate, it is an ELA content exam. I use the PPR, ELA, and Core Subjects exams as benchmarks for the SOR exam’s screening value.

2.5 Texas Reading Academies at EPPs

To support preservice teachers’ preparation in evidence-based early literacy instruction, the TEA authorized select EPPs to offer Texas Reading Academy (TRA) coursework beginning in the 2024–25 academic year. These academies extend the TRA, which was initially a statewide professional development program for in-service teachers, to preservice preparation.

TRAs provide a structured, intensive sequence of coursework beyond standard EPP requirements. The curriculum consists of twelve modules covering the full scope of research-based literacy instruction, including oral language and vocabulary development, phonological and phonemic awareness, print concepts and alphabet knowledge, decoding and encoding, reading fluency, reading comprehension, and written composition. As part of TRAs, candidates complete online modules delivered through Canvas, participate in in-person instructional sessions, and demonstrate mastery through graded artifacts, assessments, and coaching, with a minimum mastery threshold of 80 percent across required assignments.

To host TRA, EPPs must apply through a TEA-administered authorization process.

This requires the EPP to name a “Cohort Leader”, who then must pass a required training and onboarding session. Once approved, EPPs retain flexibility in integration: some offer academies as intensive standalone courses, while others embed the content across multiple semesters of coursework.

Importantly, participation is not universal. EPPs self-select into the application process, although essentially all applicants are approved. [Table A16](#) documents the authorized EPPs and each program’s adoption year based on TEA administrative records.² This staggered and incomplete adoption generates variation in preservice teachers’ access to TRA coursework across programs and cohorts, which I exploit in the empirical analysis.

3 Data and Sample

This analysis uses administrative data from the University of Houston Education Research Center (ERC), which links records across multiple state agencies. The ERC integrates data from the TEA, the Texas Higher Education Coordinating Board (THECB), and other state entities through common identifiers, enabling longitudinal tracking of individuals across the K–12, postsecondary, and teacher workforce systems. I draw on several categories of these data: teacher certification exams, teacher certifications, public school employment and student achievement, and postsecondary enrollment records.

3.1 Teacher Certification Exams and Certifications

The State Board for Educator Certification (SBEC) maintains two cumulative administrative files that make up the primary dataset used in this paper. The first is a record of all attempts at SBEC certification exams, covering every exam administered through March 2026. Each record includes the exam type, administration date, scaled score, pass/fail result, certification route (traditional, alternative, or certification by exam), and the EPP that approved the candidate to test. The second file contains all teaching certificates issued in Texas through early 2026. Each certification record includes the certification field, type, and program, the issue and effective dates, the EPP attended, and candidate gender and

²More information on participating EPPs can be found [here](#).

race/ethnicity. Both files contain unique identifiers that allow records to be linked across files and to other ERC data sources.

The exam file lacks demographics, so I construct race and ethnicity measures for exam takers by linking to two external sources: TEA’s Public Education Information Management System (PEIMS) student records, covering individuals who attended Texas public schools as K–12 students, and THECB postsecondary records, covering individuals who applied to or attended Texas public colleges and universities. Together, these linkages yield race and ethnicity for about two-thirds of first-attempt exam takers. The remainder likely attended K–12 school outside Texas, outside the time range covered in my data, and private or out-of-state postsecondary institutions. Analyses requiring demographics restrict to candidates with available covariates, and I note this restriction where it applies.

3.2 Public School Employment and Student Achievement

Employment data come from the PEIMS staff files, which are annual snapshots of all individuals employed in Texas public schools measured as of the last Friday in October. I observe employment status, district and campus assignments, role classifications, and experience at a single point each year. I link these records to the exam and certification files to measure whether exam takers work as teachers in the October following their exam. For analyses linking teachers to students, I use TEA’s class roster data, which connect teachers directly to the students in their classrooms. Student achievement is measured by State of Texas Assessments of Academic Readiness (STAAR) scores in reading and math, and scores are always standardized to mean zero and standard deviation one within grade, year, and subject. Student covariates including race, gender, special education status, gifted status, and economic disadvantage, come from the class rosters.

Value-added models that make use of this dataset use all regular STAAR administrations in grades 4 through 8 in reading and math from 2021-22 through 2023-24 school years, excluding alternate and other specially scored assessments.³

³Grade 3 students are excluded because the models control for prior-year achievement and STAAR testing begins in grade 3, so no lagged score exists for them.

3.3 First-Attempt Certification Exam Takers

The first analytic sample consists of all individuals who sit for the TExES SOR exam for the first time with a valid scaled score, restricted to candidates whose exam date falls at least one year before the end of the employment data.⁴ This ensures a minimum of one year of follow-up for certification and employment outcomes. The resulting sample includes 21,692 first-attempt SOR exam takers from January 1, 2021 through October 23, 2023.

Table 1 presents summary statistics overall and by first-attempt pass status. Approximately 80 percent of candidates pass on their first attempt, with a mean scaled score of 251.6 against a passing score of 240. Candidates who fail are disproportionately on the alternative certification route (52.7 percent, versus 36.3 percent of those who pass), more likely to be non-white (53.9 versus 41.7 percent among the subsample with demographics)⁵, and slightly less likely to be female (88.7 versus 93.0 percent).

A notable feature of the data is that approximately 31 percent of first-attempt SOR exam takers are already employed as teachers in Texas public schools before their exam date, at nearly identical rates among those who pass (31.5 percent) and fail (31.0 percent). This reflects some individuals who have another credential and sit for the examination looking to add an additional teaching area (i.e., what Texas calls “certification by exam”), but also the structure of the Texas teacher labor market, in which individuals can teach on interim credentials, emergency permits, or without formal certification in districts of innovation.⁶ Accordingly, 58.8 percent of candidates earn a *standard certificate* within one year while 70.8 percent are employed as teachers the following October.⁷ This is consistent with a context where many teachers enter classrooms before completing all certification requirements.

⁴The scaled-score restriction also excludes the exam’s initial piloting period, during which SBEC reported only pass/fail results, so the effective sample begins in September 2021 rather than January 2021.

⁵In this sample I compare white candidates with all other candidates. White teachers are overrepresented in the teaching workforce relative to the students they teach, and concern about certification exams has centered on whether they keep candidates of color out of the profession (Angrist & Guryan, 2008; Deneault et al., 2026), so this is the contrast most relevant to that debate. The grouping is a simplification that combines Black, Hispanic, Asian, and other candidates, whose experiences with certification exams may differ. Because race and ethnicity are observed for only about two-thirds of this sample, the number of candidates near the passing threshold is too small to examine these groups separately.

⁶These are districts who are allowed by TEA to opt out of regulations on inputs like teacher certification and class sizes. See Anglin (2024) for more information.

⁷A standard certificate is Texas’ full, five-year, renewable teaching credential.

3.4 Newly Issued Certifications

The second analytic sample consists of all newly issued standalone teaching certificates in Texas from 2016 through 2025, where each observation is the first time an individual receives a particular certification (i.e., individuals who earn multiple standalone certifications may appear multiple times in the dataset).⁸ Here, I observe the certification field, type, issue date, EPP, and candidate demographics for all 342,816 records. I classify each certificate as SOR-required (i.e., in a field that came to require the SOR exam) or comparison following the TEA’s Required Testing Chart. [Table A2](#) lists the full classification.⁹

[Table 2](#) presents summary statistics for this sample, split by whether the certification was issued before or after the SOR mandate’s June 2019 announcement. The two periods are broadly similar in route composition, though out-of-state certifications rise modestly, from 9.9 to 13.6 percent. Descriptively, the share of newly issued certifications in SOR-required fields fell from 41.7 percent pre-announcement to 36.0 percent afterward. The demographic composition also shifted, with the White share declining from 59.2 to 55.2 percent and the Hispanic share rising from 25.1 to 27.6 percent.

To estimate how the SOR requirement affected the flow of new certifications, I collapse this sample to counts by educator preparation program (EPP) $\times$ program-type (e.g., traditional, alternative, etc.) $\times$ certification-field $\times$ half-year (i.e., January to June and July to December).¹⁰ This collapse spans 200 EPP-program-type units, 3 SOR-required and 16 comparison certification fields, and 2016 January to June (H1) through 2025 July to December (H2), yielding 9,585 SOR-required and 51,120 comparison cells. [Table 3](#) describes the resulting panel.¹¹

⁸I begin the sample in 2016 to avoid the January 2015 replacement of the Generalist EC-6 (TExES 191) and Generalist 4–8 (TExES 111) exams with the Core Subjects EC-6 (TExES 291) and 4–8 (TExES 211) exams.

⁹A certification is coded SOR-required if the TExES Science of Teaching Reading exam (293) is among its required content-pedagogy tests. Non-teaching credentials and supplemental certificates that attach to a base certificate rather than certifying a teaching field independently are excluded from the sample. For more information, see the notes to [Table A2](#).

¹⁰Some EPPs offer multiple program types. For example, several universities that offer traditional teacher education programs to undergraduates also operate alternative certification programs.

¹¹I code a cell as zero if the EPP-program-type issued no certifications in that field during that half-year but was otherwise active, defined as having issued a certification in *any* field during that half-year. Periods before an EPP-program-type’s first observed activity, or after its last, are excluded from the panel rather than coded as zero. This distinguishes a program that was operating but chose not to issue a given certification

3.5 Texas Reading Academy Data

For the analysis of TRAs, I obtained the set of EPPs authorized to offer academy coursework, and each program’s adoption year, from TEA administrative records (Table A16). I match this information to EPP identifiers in the exam and certification files to classify candidates by whether their EPP offered a TRA at the time of their first exam attempt.

Table 4 compares pre-period characteristics of the 10 EPPs that adopted academies in 2024–25 and the 94 EPPs that never adopted, measured over the 2021–22 through 2023–24 academic years. EPPs adopting in 2025–26 are excluded from both groups. Most importantly, the two groups perform similarly on both the SOR and PPR exams in the pre-adoption period. Pass rates and scaled scores are statistically indistinguishable for both exams. The groups are also balanced on certification production, demographics, and the share of certifications in SOR-required fields. The most notable compositional difference is that academy EPPs are more likely to be traditional university-based programs (73.5 versus 57.8 percent).

4 Empirical Strategy

4.1 Barrier: The Cost of the Requirement

4.1.1 Individual-Level Cost of Failing

Candidates scoring 240 or above pass the SOR exam and those below fail. Assuming candidates just above and just below this cutoff are comparable in all respects other than pass status, comparing their outcomes identifies the causal effect of passing the exam for candidates at the pass/fail margin.

Using a regression discontinuity, I examine four outcomes that trace the consequences of failing from immediate response to career trajectory: whether candidates retake the SOR exam within one year, whether they earn a standard teaching certificate within six months and within one year of their exam date, and whether they work as a teacher in a Texas from one that had not yet begun, or had ceased, issuing certifications altogether. Under this rule, 51.2 percent of SOR-required cells and 69.9 percent of comparison cells are zero, reflecting the narrowness of many comparison fields (e.g., Foreign Language, Computer Science) relative to the broader SOR-required fields.

public school the following school year.¹²

Formally, I estimate

$$Y_i = \alpha + \beta \cdot \mathbb{I}(\text{Score}_i \geq 240) + f(\text{Score}_i - 240) + \varepsilon_i, \quad (1)$$

where $f(\cdot)$ is implemented as a local linear regression with a triangular kernel and the data-driven bandwidth of Calonico et al. (2014). Since employment is observed only annually while exam dates vary throughout the year, exam timing mechanically determines how long candidates have to secure a position before the annual snapshot. For consistency, all specifications therefore include month-of-exam fixed effects. I also assess robustness to alternative bandwidth selectors. Finally, I estimate the model separately by certification route (e.g., Traditional, Alternative, etc.) to assess for heterogeneity. Because the SOR exam is one of several testing requirements, I also estimate the model separately for candidates who had already passed the PPR and a content exam before their first SOR attempt, for whom the SOR exam was the last remaining requirement, and for all other candidates, following the univariate approach of Wong et al. (2013).

Identification requires that candidates just above and below the cutoff are comparable in all respects other than pass status. The main threat to this is manipulation. If candidates could precisely control their scores, those who landed just above the threshold might differ systematically from those just below. A formal density test (Cattaneo et al., 2018) narrowly fails to reject the null of no manipulation ($p = 0.051$), and the score density rises smoothly through the threshold with no visible bunching on either side (Figure A2). A p -value this close to conventional thresholds warrants scrutiny, so it is worth stating why manipulation is implausible in this setting. Scaled scores are computed from raw scores through a conversion that candidates do not know in advance, candidates cannot observe their standing relative to the threshold while testing, and no mechanism exists to revise a score once submitted. A candidate who wanted to land just above 240 would have no way to do so. The observable evidence is consistent with this. Pre-test covariates are balanced across the threshold

¹²Teachers are observed in Texas public schools once annually. For example, for a candidate who takes the SOR exam between October 1, 2023, and September 30, 2024, employment is measured as of the October 2024 staffing snapshot for the 2024–25 school year.

(Table A1). Candidates just above and just below do not differ in certification route, demographics, or prior teaching status. Manipulation is therefore extremely unlikely in this setting.

4.1.2 Market-Level Effects on the Flow of New Certifications

The RD identifies the cost of failing for candidates near the threshold, but not the margins where the mandate’s barrier may matter most. Candidates who score well below the cutoff face more uncertain paths to passing, and the exam’s direct and indirect costs (e.g., fees, study time, travel to a testing center) may deter prospective candidates from entering the pipeline at all. To examine this margin, I compare the number of newly issued standalone teaching certifications in SOR-required fields to the number in comparison fields, using the sample described in Section 3.4.

Using counts of newly issued certifications at the educator preparation program (EPP) by program type (e.g., Traditional, Alternative) by certification field (e.g., EC–6 Core Subjects) by half-year level, I estimate event study regressions via Poisson pseudo-maximum likelihood:

$$\mathbb{E}[Y_{ekt}] = \exp \left(\sum_{t \neq 2019H1} \beta_t \cdot \mathbb{1}(t) \times \text{SOR}_c + \gamma_{ek} + \delta_c + \lambda_t \right), \quad (2)$$

where Y_{ekt} is the number of new certifications issued by EPP e through program type k in certification field c during half-year t , and SOR_c indicates that field c began requiring the SOR exam in January 2021. The specification absorbs EPP-by-program-type (γ_{ek}), certification field (δ_c), and half-year (λ_t) fixed effects, so identification comes from within-EPP-program-type comparisons of how SOR-required and comparison fields evolve over time. The exponentiated β_t give the rate ratio of certifications in SOR-required relative to comparison fields in each half-year, normalized to 2019H1, the last period before House Bill 3 was enacted in June 2019. PPML accommodates the count nature of the outcome and allows estimation with cells containing zero certifications (Santos Silva & Tenreyro, 2006).¹³ Standard

¹³PPML is the preferred estimator here for two reasons. First, 51.2 percent of SOR-required cells and 69.9 percent of comparison cells are zero, so log-linear models are undefined without an arbitrary transformation of the outcome. Second, exponentiated PPML coefficients are rate ratios and so are invariant to program scale, which matters because EPPs range from those issuing a handful of certifications a year to those issuing several hundred. PPML is also consistent under any conditional mean of the form in Equation 2, this

errors are clustered at the certification field level, the level at which the SOR requirement is assigned.

I summarize the event study with three pooled specifications, each estimated overall and separately by certification route. The first replaces the half-year interactions with a single indicator for all periods after 2019H1. This captures the average post-announcement effect, including any anticipatory response. Once candidates learned of the mandate in June 2019 they had an immediate incentive to certify before the exam became a requirement, and restricting attention to post-2021 changes would understate the requirement’s full effect on the flow of entrants. The second is a donut specification that drops 2019H2 through 2021H2 from the estimation sample and defines the post period as 2022H1 onward, removing the anticipatory window and its immediate aftermath rather than folding it into the post indicator. I interpret this as the effect once the market has settled into a new steady state. The third addresses the possibility that the two groups were already diverging before the announcement. It adds the term $SOR_c \times t$ for $t \leq 2019H1$, allowing the differential between SOR-required and comparison fields to follow its own linear trend through the pre-announcement period. The counterfactual for the post period is then the level that fitted trend reaches in 2019H1 rather than the average pre-announcement differential, so the pooled coefficient is not driven by a gap that was already moving before House Bill 3. I report all three throughout.

With 19 certification field clusters, 3 of which are treated, clustered standard errors at the certification field level may understate the sampling variability of the treatment coefficient (Cameron & Miller, 2015). I therefore assess inference on each pooled specification via randomization inference. Following MacKinnon and Webb (2020), I re-estimate the specification 2,000 times, each time randomly assigning SOR-required status to 3 of the 19 certification fields while holding the number of treated fields fixed, and record the t -statistic on the coefficient of interest. The reported p -value is the share of placebo draws whose t -statistic equals or exceeds in magnitude the t -statistic from the true assignment.¹⁴¹⁵ I report

means the outcome variable need not be Poisson distributed, and its fixed effects are free of the incidental parameters problem (Wooldridge, 1999).

¹⁴The wild cluster bootstrap is the more common alternative in linear settings with few clusters, but has no off-the-shelf implementation for fixed-effects PPML. Randomization inference does not require asymptotics to kick in, so the results remain valid however few certification fields the sample contains.

¹⁵As a further check, I re-estimate the event study and all three pooled specifications by ordinary least squares and wild cluster bootstrap with Webb weights, which are feasible for linear models. Despite the

the randomization inference p -value alongside, rather than in place of, the clustered p -value, and discuss any divergence between them in the results.

Interpreting the post-mandate change as the effect of the requirement rests on the assumption that, absent the mandate, certifications in SOR-required fields would have evolved in parallel to those in comparison fields. This assumption is untestable, but three pieces of evidence bear on it. [Figure 3a](#) plots raw certification counts for the two groups over the full panel. The event study provides pre-announcement coefficients, for which I report a joint test of the null that all are zero. The third pooled specification relaxes parallel trends directly by allowing the two groups to differ in level and in pre-announcement slope. None of these rules out a smooth pre-existing divergence that continues into the post period, so in [Table A6](#) I report how large a violation of parallel trends would have to be to overturn the pooled estimate, following Rambachan and Roth (2023).

Two confounders deserve particular attention. The first is the COVID-19 pandemic, which arrived several months before the mandate took effect. If the pandemic differentially affected demand for or supply of teachers across certification fields, by making SOR-required fields less attractive or shifting hiring toward other fields such as special education, the estimates would partly reflect those changes rather than the SOR requirement. The second is that the mandate coincided with a redesign of the content exams in the affected fields. When the standalone SOR exam was introduced, reading pedagogy content was removed from the EC-6 and ELAR 4-8 content exams and revised versions were launched. The estimates therefore capture the combined effect of the new testing regime rather than the SOR requirement in isolation.

4.2 Screen: Does the Exam Identify Effective Teachers?

The welfare implications of the exam’s supply effects depend on whether it screens for knowledge that matters for teaching effectiveness. If the exam identifies teachers who are more effective in the classroom, the costs documented above may be offset by gains in teacher quality among those who enter the profession. This section evaluates that possibility on two

model parsimony, as described elsewhere, this specification is not preferred because its coefficients are in number of certificates and difficult to interpret across EPPs of different sizes.

margins, whether exam scores predict teacher effectiveness and whether the candidates the passing threshold excludes are less effective than the ones it admits.

4.2.1 Predictive Validity of Certification Exam Scores

One way a certification exam may function as a screen is by providing information about teacher effectiveness before hire. A substantial literature has searched for such signals and largely come up short, though certification exam scores are among the few characteristics that consistently—if modestly—correlate with future effectiveness (Clotfelter et al., 2007; Kane et al., 2008; Rockoff et al., 2011). This section asks whether the SOR exam is such a signal, and whether it is a more informative one than comparable certification exams in Texas.

To assess whether certification exams predict teacher effectiveness, I estimate student-level value-added models of the form

$$A_{ijt} = \gamma \cdot \tilde{S}_j + X'_{ijt}\beta + \bar{X}'_{cjt}\delta + \theta_{e(j,t)} + \psi_{d(j)} + \mu_t + \lambda_{sg} + \varepsilon_{ijt}, \quad (3)$$

where A_{ijt} is the standardized STAAR score of student i taught by teacher j in year t and $\tilde{S}_j$ is teacher j 's first-attempt scaled score on the relevant certification exam, standardized within year and exam. The vector X_{ijt} holds student-level controls: lagged STAAR exam scores in reading and math, entered with quadratic and cubic terms, and indicators for sex, race/ethnicity, limited English proficiency, special education status, gifted status, and economic disadvantage. The vector $\bar{X}_{cjt}$ holds classroom-level averages of these same covariates. Teacher experience ($\theta_{e(j,t)}$; 0 to 9 years, with 10 and above pooled) and degree ($\psi_{d(j)}$; no degree, Bachelors, Master's, or Doctorate) fixed effects absorb returns to experience and education. Year (μ_t) fixed effects absorb statewide shocks common to a testing year and ε_{ijt} is the student-level error, clustered at the teacher level. My preferred specifications also include grade-by-school (λ_{sg}) fixed effects, so that γ is identified from comparisons of teachers within the same grade and school. I also report additional specifications with grade and district, and grade and school fixed effects. I estimate Equation 3 separately for the SOR, Pedagogy and Professional Responsibilities (PPR), and English Language Arts (ELA)

exams and interpret γ as the strength of the exam as a signal of teaching quality.

To test directly whether the SOR exam carries information beyond other exams, I also estimate horse-race specifications that replace $\tilde{S}_j$ with two exam scores (i.e., SOR and PPR scores) entered simultaneously, so that each coefficient reflects the association with student achievement net of the other. A positive and statistically meaningful SOR coefficient conditional on another exam score would indicate the SOR exam captures information not already contained in the other exam.

Finally, I replace $\tilde{S}_j$ with a binary indicator for passing an exam on the first attempt. In these models, γ gives the average difference in student achievement between teachers who passed and failed an exam on their first attempt. As with the scores, I estimate each exam's indicator on its own and then enter the SOR and other exam indicators together. This analysis also lets me include the Core Subjects exams, for which my data record pass status but no single scaled score.

The identifying assumption of this analysis is that, conditional on the controls, students are as good as randomly assigned to teachers with different exam results. Two forms of sorting threaten it. Across schools, teachers with stronger credentials work disproportionately in schools serving more advantaged students (Clotfelter et al., 2005; Lankford et al., 2002). Within schools, principals move teachers across grades in response to testing pressure, placing weaker teachers in untested grades (Fuller & Ladd, 2013; Grissom et al., 2017), and more experienced teachers are often assigned higher-achieving students (Kalogrides et al., 2013). School-by-grade fixed effects address the first two by comparing teachers who work in the same school and grade. Lagged achievement and classroom composition attempt to address the third, although they cannot rule out sorting on student characteristics that neither prior scores nor demographics capture, such as motivation or parental involvement. Experimental and quasi-experimental evidence suggests the remaining bias in models with these controls is small (Chetty et al., 2014; Kane & Staiger, 2008).

The pass/fail comparison faces a further problem. Value-added is observed only for candidates who go on to teach a tested grade, and failing the exam itself may make that less likely, so the failers observed in classrooms may not be representative of failers generally. I examine this directly by comparing the pre-exam characteristics of candidates who do and

do not appear in the value-added sample, separately for those who passed and failed.

4.3 The Effect of Texas Reading Academies

The final analysis tests whether TRA coursework can help candidates pass the SOR exam. To do this, I compare exam outcomes at EPPs that did and did not offer TRAs, before and after implementation, using a difference-in-differences design. Formally, I estimate, via ordinary least squares, two-way fixed effects models of the form:

$$Y_{iet} = \alpha_e + \delta_t + \sum_{k \neq 2023} \beta_k \cdot \text{TRA}_e \times \mathbb{1}(t = k) + \varepsilon_{iet}, \quad (4)$$

where Y_{iet} is an indicator for passing the SOR exam on the first attempt or the first-attempt scaled score, standardized within year, of candidate i at EPP e in academic year t . α_e are EPP fixed effects, δ_t are year fixed effects, and standard errors are clustered at the EPP level, the level at which academies are adopted. The coefficients β_k measure TRA–non-TRA differences in year k relative to the year before implementation. A companion difference-in-differences model replaces the interaction terms with a single $\text{TRA}_e \times \text{Post}_t$ indicator. Since additional EPPs adopted academies in 2025–26, and to avoid issues that arise from comparing already-treated to not-yet-treated programs under staggered adoption (Goodman-Bacon, 2021), I separately analyze the 2024–25 and 2025–26 cohorts against a never treated comparison group of EPPs.

As with the market-level analysis in [Section 4.1.2](#), identification rests on parallel trends: absent the academies, outcomes at adopting and non-adopting EPPs would have evolved similarly. For the 2024–25 cohort the pre-period supports this. The event-study coefficients for the three years before adoption are small and individually and jointly insignificant for both pass rates and scores ([Figure 5](#); joint $p = 0.537$ and 0.524), and adopting and non-adopting programs had similar SOR and PPR pass rates and scores before adoption ([Table 4](#)). I focus the main analysis on this cohort for two reasons. First, the 2025–26 cohort’s pre-period coefficients are large and trend downward, indicating that these programs were already losing ground to never-adopters before they adopted, so the comparison group is a poor counterfactual for them. Second, its post-adoption window currently covers only a partial

school year. I report the 2025–26 cohort’s results in [Table A17](#) for completeness. Since the primary analysis rests on one of the two cohorts, I interpret the results as preliminary evidence on whether structured preparation affects exam performance.

5 Results

The SOR exam first-attempt pass rate hovers around 80 percent over the sample period, which is roughly comparable to the PPR exam and well above the EC-6 Content exam ([Figure A1](#)). The level is high enough that most candidates eventually pass. It is also low enough that one in five fails on the first attempt and faces additional exam fees and preparation time. The analyses below test the extent to which the exam is a barrier at the individual level ([Section 5.1](#)) and the market level ([Section 5.2](#)), whether it screens on teacher effectiveness ([Section 5.3](#)), and whether preparation can mitigate its supply cost ([Section 5.4](#)).

5.1 The Individual Cost of Failing

[Table 5](#) reports estimates of [Equation 1](#) for each outcome. Passing reduces the probability of retaking the exam within a year by 77.9 percentage points, so most candidates who narrowly fail make at least one additional attempt. The more substantive question is whether failure delays or prevents certification and employment. Narrowly passing increases the probability of earning a teaching certificate within six months by 8.0 percentage points. By twelve months the estimate attenuates to 4.0 percentage points and loses statistical significance, which indicates that most candidates who narrowly fail eventually retake, pass, and certify. The friction is not fully absorbed by the certification process, however. Passing increases the probability of working as a teacher the following October by 6.6 percentage points. This effect is not purely mechanical. Individuals in Texas can and do teach without a certificate, and roughly a third of test takers are already in classrooms before their exam ([Table 1](#)), so failing does not by itself bar a candidate from employment. [Figure 1](#) shows these discontinuities graphically.

These results are robust. Point estimates are nearly identical with and without month-of-

exam fixed effects (Table A3) and stable across alternative data-driven bandwidth selectors (Table A4). Some estimates lose statistical significance without controls or under narrower bandwidths, but in each case the point estimates barely move, so these choices affect precision rather than the qualitative result.

The overall results conceal a sharp divide by certification route (Table 6). Among traditional candidates, who are typically finishing university-based programs and not yet employed as teachers, the exam comes at the end of the pipeline. Passing lowers their probability of retaking by 92.8 percentage points and raises six-month certification by 14.8 percentage points, consistent with candidates sitting for the exam on the cusp of completing all other requirements. By twelve months the gap has closed, with a negative and insignificant point estimate (−6.7 percentage points), and the effect on employment the following school year is small and statistically insignificant (1.9 percentage points). For a traditional candidate who just fails, the exam delays time to earning a credential but not entering the classroom as a teacher.

The pattern differs for alternative route candidates. Even among alternative candidates who pass, fewer than a quarter certify within a year (Figure 2, panel (c)), far below every other route. The certification effect of passing is small and insignificant at six months (2.4 percentage points) but grows to 10.5 percentage points at one year, and narrowly failing reduces the probability of teaching the following October by 11.4 percentage points. This pattern is consistent with where the exam comes in the alternative route’s sequence. Many alternative programs require candidates to pass the relevant certification exams before they can be placed on an intern certificate, so passing the exam is one of the first steps toward the classroom rather than one of the last, and failing it blocks entry to the profession.

The out-of-state and certification-by-exam routes resemble the traditional route. Passing raises six-month certification by 16.4 percentage points for out-of-state candidates and 9.9 for by-exam candidates. The gap shrinks by one year, and there is no meaningful employment effect, though both samples are small and only the out-of-state six-month estimate is statistically significant. This pattern fits candidates who typically hold a teaching certificate already and are sitting for the SOR exam to add a field rather than to enter the profession. For them, the exam comes after most of the work of becoming a teacher is done.

A second cut of the data supports this reading. Because the SOR exam is one of several testing requirements, the pooled estimate mixes candidates for whom it is the last one remaining with candidates who have not yet cleared the PPR or a content exam. [Table A8](#) separates them. Among candidates who had already passed both other exams (i.e., the SOR exam was their last testing requirement to earn certificate), passing raises six-month certification by 20.5 percentage points, more than twice the pooled estimate, and has a small, statistically insignificant effect on employment (1.8 percentage points). Among candidates with other requirements still open, the six-month certification effect is 7.9 percentage points, but passing raises employment by 7.7 percentage points. This corresponds to the results by certification route. [Table A9](#) shows that only 12.0 percent of alternative candidates had cleared the other exams before their first SOR attempt, compared with 26.7 percent of traditional and 39.3 percent of by-exam candidates.

These results show that the cost of narrowly failing depends on where the exam falls in a candidate’s route to certification. It comes at the end of a sequence for traditional candidates, at the entry point for alternative candidates, and after certification is already secured for out-of-state and by-exam candidates. But estimates at the passing margin capture only candidates who show up to test and just miss. If the exam also deters people from attempting at all, the fuller effect on supply appears not at the cutoff but in the flow of new certifications, the subject of the next section.

5.2 The Market-Level Cost

Panel (a) of [Figure 3](#) shows the market-level pattern of newly issued standalone certifications in SOR-required and non-SOR fields by half-year. Both series show a pronounced seasonal sawtooth pattern, peaking each January–June as candidates completing traditional programs certify ahead of spring graduation. Both series also drift downward from 2016 through 2019, consistent with the broader national decline in entry into teaching over this period (Bartanen & Kwok, [2023](#); Kraft & Lyon, [2024](#)). The decline is somewhat steeper in fields that came to require the exam, a point I return to below. The COVID-19 pandemic mutes both series’ January–June 2020 peak but disrupts neither series relative to the other. Against this backdrop, the divergence around the mandate is clear. Certifications in fields

that would require the SOR exam surge in the second half of 2020—breaking the usual seasonal trough—as candidates rush to certify before the requirement took effect. The series then drops sharply once the exam became required in January 2021 and settles onto a lower path. The gap between the two series widens at the mandate and has not closed through 2025.

Panel (b) formalizes this pattern by plotting the coefficients estimated from [Equation 2](#). Each point represents the estimated rate ratio of certifications in SOR-required relative to non-SOR fields, normalized to the first half of 2019. Pre-announcement rate ratios are statistically indistinguishable from one, and a joint test fails to reject the null that the pre-period coefficients are jointly zero ($p = 0.700$; [Table 7](#)). They do, however, decline gradually from about 1.09 in 2016 to 1.00 in the reference period, so SOR-required fields were losing ground to comparison fields slowly before the announcement. The rate ratio rises to 1.36 in the second half of 2020 as candidates accelerated certification ahead of the mandate, then falls to 0.67 in the first half of 2021 as that pipeline exhausted. From 2022 onward the rate ratio declines from 0.77 to 0.57, so the gap widens rather than closes over the four years following implementation. Taken at face value, an EPP that would have issued one hundred certifications in SOR-required fields, given its own history and the statewide trend in comparison fields, issues seventy-seven in the first half of 2022 and fifty-seven by the second half of 2025. Part of that decline continues the pre-announcement drift, which the trend-adjusted estimate below nets out.

[Table 8](#) condenses these dynamics into single estimates. Including the anticipatory period, certifications in SOR-required fields fell by 21.5 percent ($0.785 - 1 = -0.215$) relative to non-SOR fields within EPPs. The donut specification, which excludes half-years 2019H2–2021H2 to remove the anticipatory surge and its immediate aftermath, implies a larger decline of 31.4 percent ($0.686 - 1 = -0.314$). The trend-adjusted specification, which measures the post-period level against the endpoint of the pre-announcement trend, implies a decline of 17.7 percent ($0.823 - 1 = -0.177$), smaller than the other two because part of the raw gap reflects the slow pre-announcement drift documented above. I treat this as the preferred estimate of the mandate’s effect, since it is the one that does not attribute the pre-existing divergence to the requirement. All three estimates are significant at the 5 percent level under clustered

standard errors ($p = 0.016, 0.004, \text{ and } 0.003$). Under randomization inference the evidence is weaker, with p -values of 0.128, 0.105, and 0.094 for the full, donut, and trend-adjusted specifications. With only three of nineteen certification fields treated, randomization inference is the more conservative procedure, and I read the market-level decline as a stable point estimate whose statistical significance depends on the inference method. Estimating the same specifications by ordinary least squares yields similar patterns (Table A5).

Sensitivity to violations of parallel trends points the same way. Table A6 shows that, allowing the pre-period trend to continue linearly into the post period, the estimated decline over 2019H2–2025H2 is significant at the 10 percent but not the 5 percent level, while the estimated decline restricted to the post-implementation period, 2022H1–2025H2, is significant at the 5 percent level. Both estimates are sensitive to allowing the pre-period trend to bend beyond a small threshold. The weaker result for the full post-announcement period is consistent with its inclusion of the second-half-2020 anticipatory surge.

Columns (2)–(5) of Table 8 report the same three specifications separately by certification route with Figure 4 showing the corresponding event studies. The contraction is concentrated outside the traditional, university-based route. Under the trend-adjusted specification, certifications fell by 17 percent in the alternative route, 23 percent out of state, and 40 percent in the certification-by-exam route, while the traditional route shows a decline of under 3 percent that is statistically indistinguishable from zero under either inference method. The full specification gives the same ordering; in the donut, the alternative and out-of-state estimates swap places. This is consistent with the RD finding that a narrow failure is only a temporary setback for candidates embedded in university-based preparation, and a larger obstacle for those entering through other routes.

The route-specific estimates warrant caution. The joint pre-period test rejects for the traditional, alternative, and by-exam routes (Table 7), but the pre-trend problems differ by route. For traditional and out-of-state candidates, the pre-period coefficients are tightly estimated and hover near one without a discernible direction, so the traditional rejection reflects precision rather than an economically meaningful trend.¹⁶ For alternative and by-exam can-

¹⁶The joint test does not reject for out-of-state ($p = 0.308$), and its pre-period coefficients are similarly flat.

didates, the pre-period coefficients decline steadily in the run-up to the announcement, running in the same direction as the post-mandate effect. The drift is most severe for by-exam, whose rate ratio falls from 1.31 to 1.09 over the six pre-periods. The trend-adjusted specification is designed to absorb a linear version of this drift. Under randomization inference its by-exam estimate remains significant ($p = 0.043$), its out-of-state estimate is marginal ($p = 0.066$), and its alternative-route estimate is not significant ($p = 0.119$). Because a pre-trend that is not linear would survive this adjustment, I do not treat the alternative and by-exam estimates as cleanly identified, and I read the route-level results as a rough decomposition of where the aggregate decline is concentrated rather than as four separately identified effects. The pooled estimate, whose own pre-trend is modest and is more generally supportive of the parallel trends assumption, remains the primary market-level finding.

One possible explanation for the decline is substitution, with candidates redirecting toward non-SOR fields rather than exiting or delaying. I find no evidence for this. At the passing margin, the regression discontinuity in [Table A7](#) splits the certification outcome by field. Failing the exam sharply reduces the probability of certifying in an SOR-required field, by 11.3 percentage points within six months and 7.9 within a year, while the effect on certifying in a non-SOR field is small and statistically indistinguishable from zero at both horizons. Candidates who narrowly fail do not pivot into unaffected fields. At the market level, certifications in comparison fields would have risen to absorb candidates leaving the affected fields, and the event study in [Figure 3](#) shows no such offsetting increase. This is consistent with the mandate’s effect being a contraction in the affected fields and not a reshuffling across them.

5.3 Does the Exam Screen?

The supply costs documented above are only worth bearing if the exam measures knowledge that matters for student learning. [Table 9](#) reports estimates of [Equation 3](#), relating student reading achievement to their teachers’ certification exam results. Panel A enters each exam’s standardized first-attempt score, first on its own and then alongside the SOR score for the teachers who took both exams. Panel B replaces the score with an indicator for passing the exam on the first attempt.

A one standard deviation increase in a teacher’s SOR score is associated with a 0.010 standard deviation increase in student reading achievement within the same grade and school (column 1). The corresponding estimates for the ELA content exams and the PPR exam are 0.007 and 0.006 (columns 2 and 3). To put the magnitude in context, the standard deviation of teacher value-added in reading in these data is 0.20 student standard deviations, so a one standard deviation increase in SOR score corresponds to about 0.05 standard deviations of teacher effectiveness in reading.

Columns (4) and (5) enter the SOR score together with the PPR or ELA score, on the teachers who have both. In each case the SOR coefficient is unchanged or larger (0.011 and 0.017), while the PPR and ELA coefficients fall to zero (-0.002 and 0.000). Whatever predictive content the other exams carry on their own is absorbed by the SOR score, and the reverse is not true. [Table A11](#) shows that the estimates are robust to alternative fixed effects (e.g., school and grade, rather than school-by-grade), and [Table A12](#) shows that the SOR relationship rises across the score distribution. Restricting to first-year teachers, whose effectiveness is not yet shaped by experience or by who stays in teaching, gives a similar but less precise picture ([Table A13](#)). The point estimates are largely unchanged in this smaller sample, though the SOR coefficient is no longer statistically significant when entered alongside the PPR score.

Whether the SOR score measures knowledge specific to literacy instruction is harder to say. [Table A14](#) shows that SOR scores predict student math achievement more strongly than reading, at 0.019 standard deviations per standard deviation of score and 0.023 for passing on the first attempt, and the same holds among first-year teachers ([Table A15](#)).¹⁷ This might be explained in two ways. First, literacy is foundational to learning in every subject, so a teacher who is better at teaching reading may raise math achievement through students’ ability to read and understand math problems. Second, the exam may capture a general dimension of instructional skill, so that its score predicts effectiveness in any subject rather than in reading specifically. The reading estimates may also understate what the exam measures, since value-added is observed only in grades 4 through 8 and the knowledge

¹⁷Notably, the standard deviation of teacher value-added is 0.26 student standard deviations in mathematics compared with 0.20 in reading, so the math estimate corresponds to about 0.07 standard deviations of teacher effectiveness against 0.05 in reading.

the exam tests is most relevant to teaching emergent readers in earlier grades. In either case, the results suggest that the SOR score is a stronger predictor of teacher effectiveness than related teacher certification exams in Texas.

Panel B of [Table 9](#) examines whether teachers who passed an exam on the first attempt are more effective than those who failed it. Students of teachers who passed the SOR exam on the first attempt score 0.024 standard deviations higher in reading than students of teachers who failed it. This is larger in magnitude than the result for passing the PPR exam (0.008), the ELA content exam (0.007), and the Core Subjects Exams (0.003). Entering the SOR and Core Subjects indicators together leaves the SOR estimate at 0.023 and the Core Subjects estimate at zero (column 5). Among first-year teachers the SOR pass/fail gap is 0.016, but is no longer statistically significant ([Table A13](#)). Among the exams these candidates take, the SOR exam's threshold is the one that most clearly separates more from less effective reading teachers.

The pass/fail comparison is made only among candidates who reached a tested classroom, and the regression discontinuity results show that failing makes that less likely. This could bias these estimates if failing changes who reaches a classroom, so that the failers and passers observed there are not selected in the same way. [Table A10](#) describes this selection by regressing an indicator for being observed with a reading value-added estimate on an indicator for failing the exam on the first attempt. Failing on the first attempt lowers the probability of being observed with a reading value-added estimate by 5.1 percentage points (Panel A, column 1). Among candidates working as a teacher the following October, the difference is 0.8 percentage points and statistically indistinguishable from zero (column 3), so the selection comes from failers not reaching a classroom rather than from failers who teach being placed outside tested grades and subjects.

Panel B of [Table A10](#) compares candidates who are and are not observed with value-added on four pre-exam characteristics, separately for passers and failers. Selection on sex and race is similar in the two groups. Selection on academic preparation and family income is less so. Observed passers score 11.4 points higher on the SAT than unobserved passers, while observed failers score 32.9 points higher than unobserved failers, a difference of 21.5 points ($p = 0.06$). Observed passers are about as likely as unobserved passers to come from

a family earning under \$60,000, while observed failers are 7.1 percentage points less likely than unobserved failers, a difference of 8.2 percentage points ($p < 0.05$). Controlling for the first-attempt score (even-numbered columns) reduces the SAT difference to 6.7 points but leaves the income difference at 10.8 percentage points. The SAT selection therefore runs largely through failers who scored closer to the cutoff being more likely to teach, while the income selection does not. One reading of the income pattern is that the costs of failing, including retake fees and delayed certification, weigh more heavily on candidates with fewer resources.

Failers who reach a tested classroom are thus more positively selected, relative to failers as a whole, than passers who reach one are relative to passers. Insofar as academic preparation is positively, if weakly, correlated with teacher effectiveness (Clotfelter et al., 2007; Rockoff et al., 2011), the pass/fail gap in Panel B of Table 9 is more likely to understate than overstate the difference between all first-attempt failers and passers. The exam therefore screens in two senses. Its score predicts effectiveness better than the other exams these candidates take, and its threshold on average separates candidates who raise reading achievement less from those who raise it more.

5.4 Can Preparation Lower the Barrier?

The results so far establish that the exam is a meaningful barrier to teacher supply and also a real signal of teacher effectiveness. One response may be to lower the threshold. An alternative is to help more candidates clear it. In this section, I test whether TRA coursework helped candidates do this.

For the 2024–25 cohort, the pre-period trends are clean (Figure 5).¹⁸ Academy and non-academy programs moved together on both SOR pass rates and scores before adoption, and the pre-period event-study coefficients are individually and jointly insignificant (Table 10; joint $p = 0.537$ for pass rates, $p = 0.524$ for scores). The two groups then diverge in 2024–25. Academy programs post a 4.0 percentage point gain in first-attempt SOR pass rates (column

¹⁸As discussed above, the 2025–26 cohort’s pre-period coefficients trend downward and reject the joint test of parallel trends, and its post-adoption window covers only a partial school year, so I do not rely on it. Its results are reported in Figure A3 and Table A17.

1) and a 0.15 standard deviation gain in scores (column 2). The gap holds into the partial 2025–26 year, at 5.4 percentage points and 0.19 standard deviations, so the advantage does not fade as the first candidates move through. The pooled difference-in-differences estimates are of similar size for pass rates (3.8 percentage points) and somewhat smaller for scores (0.12 standard deviations), but less precise; the pass-rate estimate is significant only at the 10 percent level and the score estimate is not statistically significant (Panel B).

To check whether the gain is specific to SOR content, I turn to the PPR exam, which includes content the academies do not cover ([Figure 6](#)). Literacy-specific preparation should raise SOR performance more than PPR, and the results are mixed. The PPR pass rate rises 4.2 percentage points in the adoption year (column 3), close to the SOR gain, but falls to a null in the partial 2025–26 year (0.7 percentage points), and the pooled difference-in-differences estimates are smaller than the SOR estimates and insignificant (Panel B, columns 3 and 4). The PPR movement is thus confined to the first year rather than sustained, which cuts against the reading that adopting programs were simply improving across the board. Some spillover is plausible in any case, since coursework on explicit, systematic instruction may sharpen general pedagogical skill.

Taken together, the results suggest the academies help exam performance, but they fall short of proving it. Candidates at EPPs in the first cohort of academy adopters improved on the SOR exam, but a similar first-year gain on the PPR exam means I cannot fully rule out that these programs were improving in more general ways. And the second cohort, which lacks clean pre-trends and a full year of data, cannot corroborate the first. What the design can establish is that adopting programs were on parallel trends beforehand and pulled ahead on the tested content after, which is consistent with the academies working.

6 Discussion

Prior work has largely found that stricter teacher certification requirements restrict supply without improving quality (Angrist & Guryan, [2008](#)), so the natural expectation for adding a new certification exam is that its cost as a barrier will outweigh its value as a screen. The evidence in this study shows that this expectation applies less well to Texas’

Science of Reading exam. This exam is indeed a barrier, but its scores are stronger predictors of teacher effectiveness and candidates who pass the exam on their first attempt are, on average, more effective than those who fail.

The exam’s impact as a barrier depends on where the exam falls in a candidate’s route to the classroom. This source of heterogeneity is distinct from the differences by race that have been emphasized by prior literature (e.g., Deneault et al., 2026). For candidates near the end of the process, those finishing university programs or already certified in another field or state, narrowly failing delays a certificate without affecting employment. For alternative-route candidates, who must pass before their program can place them in a paid internship, failing lowers the probability of teaching the following year by about eleven percentage points. Splitting candidates by whether the SOR exam was their last remaining examination requirement gives the same contrast. Narrowly failing the SOR exam acts as a barrier to entering the teaching profession, even in the Texas context where 15.6 percent of teachers enter the classroom with no certificate at all (Ghazzawi et al., 2025). It also binds hardest on alternative route candidates, who make up an increasingly large share of new teachers (National Council on Teacher Quality, 2025).

The market-level evidence shows the barrier created a behavioral response for candidates beyond those at the margin of passing. Newly issued certificates in areas that would come to require the exam rose sharply relative to comparison areas just before the exam became a requirement. After the requirement, certifications in those fields fell and have not recovered five years later, although the size of the decline is imprecisely estimated and dependent upon modeling decisions. Although teacher supply in affected areas declined, this evidence does not establish that districts struggled to fill positions in affected areas. Fields that started requiring the exam (i.e., elementary and middle school generalist and ELA teachers) are not in chronic shortage in Texas nor are they fields that are generally short on teacher supply elsewhere (Cowan et al., 2016). It is important to caveat that the same requirement in areas that are more commonly short on teachers (e.g., secondary mathematics or special education) could carry a higher cost.

The Science of Reading exam is a better screen on teacher effectiveness in reading than the comparable exams in Texas, and at least as good as any examined in prior literature. A

one standard deviation higher SOR score predicts 0.010 standard deviations higher reading achievement. That is more than the PPR or ELA content scores, which add nothing once the SOR score is entered in the same regression. It is twice what Cowan et al. (2023) find when relating scores on the Massachusetts Tests for Educator Licensure (MTEL) to student achievement in English language arts under a similar specification (a student-level value-added model with school-by-grade fixed effects). It exceeds the reading estimates from North Carolina and California, which range from zero to about 0.005 (Buddin & Zamarro, 2009; Clotfelter et al., 2007; Goldhaber, 2007). Even so, the relationship is modest, about 0.05 standard deviations of teacher effectiveness, so the search for strong pre-hire predictors of teaching quality continues (Rockoff et al., 2011; Staiger & Rockoff, 2010). Moreover, because the SOR score predicts effectiveness in teaching math at least as well as reading, the exam may measure a general instructional skill rather than knowledge specific to teaching reading.

The SOR exam is also a screen in that its passing threshold separates more from less effective teachers. Students of teachers who failed on the first attempt score 0.024 standard deviations lower in reading than students of teachers who passed. Selection is unlikely to explain this gap and, if anything, shrinks it, since the failers who reach a classroom are more positively selected than the passers who do. The regression discontinuity results show that narrowly failing keeps some candidates from teaching the following year. Removing the requirement would let those candidates in, and because they scored below the threshold, they would be, on average, somewhat less effective at teaching reading than the teachers the exam now admits. Two qualifications apply. First, the gap is an average, and the threshold likely excludes some strong teachers while admitting some weak ones, as prior work has shown for other exams in other settings (Goldhaber, 2007; Orellana & Winters, 2025). Second, the net effect on students depends on who fills the position instead. In Texas, the replacement for a candidate who would have taught but scored just short of 240 may be someone with no preparation at all (Anglin, 2024). Evidence from relaxed pandemic-era pathways shows that such teachers are less effective than traditionally prepared peers (Backes et al., 2024; Kirksey, 2024), so turning away a prepared candidate who narrowly failed could lower the quality of instruction. A third consideration runs the other way and suggests these estimates

understate the exam’s benefit. The analysis does not examine what the requirement does to preparation. If programs reorganize coursework around the exam, or candidates learn its content in order to pass, the requirement may improve reading instruction.

These results have four implications for the twenty states that require such an exam and for those considering one. States that start requiring teacher certification exams in the Science of Teaching Reading should expect a supply cost concentrated among alternative-route candidates. Dropping the requirement has a cost as well, since it would likely admit less effective reading teachers on average. The score is among the stronger pre-hire signals of teacher effectiveness in reading available, and using it only for a pass/fail decision discards most of its information. As discussed by Goldhaber (2007) two decades ago, programs and districts should consider ways of using this information to direct coaching and induction support to low performers. Finally, preparation is the one lever that could lower the barrier while keeping the standard, and the licensure literature has paid it little attention. First-attempt pass rates rose about four percentage points at the programs that first adopted Texas Reading Academies. That evidence is early, and it shows that coursework raises exam performance, not yet that it produces better teachers. If the exam measures general instructional skill, coursework aimed at the exam could raise pass rates without raising effectiveness, and the threshold would screen less well. Following academy completers into classrooms will show which it is.

6.1 Limitations

Several limitations bound these conclusions, beyond those mentioned above. The regression discontinuity results are local to the passing margin. The market-level analysis has three treated fields, its statistical significance depends on the inference method, and it cannot separate the SOR requirement from the concurrent redesign of the affected content exams or from pandemic-era disruptions to the teacher labor market. The value-added estimates establish predictive validity, not a causal effect of teacher knowledge on achievement, and they come from grades 4 through 8, whereas the knowledge the exam tests matters most for teaching beginning readers. The pass/fail comparison observes only candidates who reached a tested classroom. The characteristics on which I can assess their selection are weak pre-

dictors of effectiveness, so the direction of the bias is argued rather than established, and the estimates for first-year teachers are imprecise. The Texas Reading Academies evidence comes from just ten EPPs in one adoption year and these programs saw a first-year gain on the pedagogy exam of similar size. Finally, this is one exam in one state, in fields without chronic shortages, and neither the costs nor the screening value necessarily generalize to another exam, subject, or labor market.

7 Conclusion

Texas' Science of Reading exam is both a barrier and a screen. Narrowly failing delays certification and lowers the chance that alternative-route candidates teach the following year. The exam's introduction also prompted a large behavioral response: candidates rushed to certify before the requirement took effect, and new certifications in the affected fields then fell and had not recovered five years later. At the same time, the exam predicts teacher effectiveness better than the other exams these candidates take, and the candidates who fail it are, on average, less effective reading teachers than those who pass. A state deciding whether to adopt or keep such an exam therefore faces a tradeoff between a smaller supply of teachers and a less effective one. Early evidence from Texas Reading Academies suggests the tradeoff can be eased, since structured coursework developed by the state and delivered by EPPs raised pass rates. As more states adopt Science of Reading requirements, research should turn to how candidates can be helped to meet the standard without lowering it.

References

- Amin, R. (2022, April). New york officials vote to scrap edtpa teacher certification exam [Accessed: 2026-01-16].
- Anglin, K. (2024). The Role of State Education Regulation: Evidence From the Texas Districts of Innovation Statute. *Educational Evaluation and Policy Analysis*, 46(3), 534–554.
- Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, 27(5), 483–503.
- Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2024). Four years of pandemic-era emergency licenses: Retention and effectiveness of emergency-licensed Massachusetts teachers over time. *Economics of Education Review*, 101, 102562.
- Bartanen, B., & Kwok, A. (2023). From Interest to Entry: The Teacher Pipeline From College Application to Initial Employment. *American Educational Research Journal*, 60(5), 941–985.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. NBER.
- Boyd, D., Goldhaber, D., Lankford, H., & Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *The Future of Children*, 17(1), 45–68.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs: Robust Nonparametric Confidence Intervals. *Econometrica*, 82(6), 2295–2326.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51.
- Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal*, 18(1), 234–261.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377–392.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673–682.
- Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing Elements in the Discussion of Teacher Shortages. *Educational Researcher*, 45(8), 460–462.
- Cowan, J., Goldhaber, D., Jin, Z., & Theobald, R. (2023). Assessing Licensure Test Performance and Predictive Validity for Different Teacher Subgroups. *American Educational Research Journal*, 60(6), 1095–1138.

- Deneault, C., Riehl, E., & Zou, J. (2026, February). *Disparate Impacts of Teacher Certification Exams* (Working Paper No. 34860). National Bureau of Economic Research.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71(3), 393–447.
- Ellis, C., Holston, S., Drake, G., Putman, H., Swisher, A., & Peske, H. (2023). Teacher prep review: Strengthening elementary reading instruction [Accessed January 2026].
- Fuller, S. C., & Ladd, H. F. (2013). School-based accountability and the distribution of teacher quality across grades in elementary school. *Education Finance and Policy*, 8(4), 528–559.
- Ghazzawi, D., Olofson, M., Landa, J., & Eluru, M. (2025, March). *Uncertified teacher rates 2019–20 through 2024–25* (tech. rep.). Texas Education Agency.
- Goldhaber, D. (2007). Everyone's Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness? *The Journal of Human Resources*, 42(4), 765–794.
- Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education*, 68(4), 377–393.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Grissom, J. A., Kalogrides, D., & Loeb, S. (2017). Strategic staffing? how performance pressures affect the distribution of teachers within schools and resulting student achievement. *American Educational Research Journal*, 54(6), 1079–1116.
- Hanushek, E. A., & Rivkin, S. G. (2007). Pay, Working Conditions, and Teacher Quality. *The Future of Children*, 17(1), 69–86.
- Kalogrides, D., Loeb, S., & Bételle, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, 86(2), 103–123.
- Kane, T. J., Rockoff, J. E., & 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., & Staiger, D. O. (2008, December). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (tech. rep. No. w14607). National Bureau of Economic Research. Cambridge, MA.
- Kirksey, J. J. (2024, July). *Amid Rising Number of Uncertified Teachers, Previous Classroom Experience Proves Vital in Texas. Policy Brief. No. 1* (tech. rep.) (ERIC Number: ED659927). Center for Innovative Research in Change, Leadership, and Education.
- Kirksey, J. J., & Gottlieb, J. J. (2023). Teacher preparation goes virtual in the wild west: The impact of fully-online teacher preparation in Texas.
- Kleiner, M. M. (2006). *Licensing Occupations: Ensuring Quality or Restricting Competition?* Upjohn Press.
- Kleiner, M. M., & Krueger, A. B. (2013). Analyzing the Extent and Influence of Occupational Licensing on the Labor Market. *Journal of Labor Economics*, 31(S1), S173–S202.
- Kraft, M. A., & Lyon, M. A. (2024). The Rise and Fall of the Teaching Profession: Prestige, Interest, Preparation, and Satisfaction over the Last Half Century. *American Educational Research Journal*, 61(6), 1192–1236.

- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62.
- MacKinnon, J. G., & Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2), 435–450.
- McArthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H.-C., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*, (11), CD009115.
- Moats, L. C. (2020). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. In *Reading research anthology: The why? of reading instruction* (pp. 29–73). Arena Press.
- National Center for Education Statistics. (2025a). *2024 NAEP reading assessment: Results at grades 4 and 8 for the nation, states, and districts* (The Nation’s Report Card No. NCES 2024218). U.S. Department of Education, Institute of Education Sciences.
- National Center for Education Statistics. (2025b). The nation’s report card: 2024 naep reading assessment highlights [Accessed May 2026].
- National Council on Teacher Quality. (2024). *Licensure tests - state teacher policy database* [Accessed: 2026-01-16]. National Council on Teacher Quality.
- National Council on Teacher Quality. (2025). *What’s the alternative? how some alternative certification programs are shortchanging teachers and students* (tech. rep.). National Council on Teacher Quality.
- NJ State Policy Lab. (2025, January). Nj ends requirement for praxis exams for teacher certification [Accessed: 2026-01-16].
- Novicoff, S., & Dee, T. S. (2025, January). *The achievement effects of scaling early literacy reforms* (EdWorkingPaper No. 23-887). Annenberg Institute at Brown University.
- Orellana, A., & Winters, M. A. (2025). Licensure Tests and Teacher Supply. *Journal of Human Resources*, 0324–13465R2.
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, 6(1), 43–74.
- Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4), 641–658.
- Seidenberg, M. S. (2017). *Language at the speed of sight: How we read, why so many can’t, and what can be done about it*. Basic Books.
- Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Spencer, N. (2024). Comprehensive early literacy policy and the “Mississippi Miracle”. *Economics of Education Review*, 103, 102595.
- Staiger, D. O., & Rockoff, J. E. (2010). Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives*, 24(3), 97–118.
- Stiglitz, J. (1975). The Theory of “Screening,” Education, and the Distribution of Income. *American Economic Review*, 65(3), 283–300.

- Swisher, A. (2022, July). Setting sights lower: States back away from elementary teacher licensure tests.
- Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2), 107–141.
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1), 77–97.

Tables and Figures

Table 1: Summary Statistics for First-Attempt Science of Reading Exam Takers

	Full Sample			Passed			Failed		
	Mean	SD	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD	<i>N</i>
Passed SOR Exam	0.805	0.396	21692	1.000	0.000	17459	0.000	0.000	4233
Scaled Score on SOR Exam	251.6	17.3	21692	258.1	10.3	17459	224.8	14.3	4233
Certification Route									
Traditional	0.343	0.475	21692	0.354	0.478	17459	0.294	0.456	4233
Alternative	0.395	0.489	21692	0.363	0.481	17459	0.527	0.499	4233
By Exam	0.111	0.314	21692	0.116	0.320	17459	0.089	0.285	4233
Out of State	0.152	0.359	21692	0.167	0.373	17459	0.089	0.285	4233
Teaching before SOR Exam	0.314	0.464	21692	0.315	0.465	17459	0.310	0.463	4233
Demographics									
Has demographic information	0.679	0.467	21692	0.668	0.471	17459	0.723	0.447	4233
Female	0.921	0.270	13448	0.930	0.256	10652	0.887	0.316	2796
Non-white	0.442	0.497	14720	0.417	0.493	11658	0.539	0.499	3062
Outcomes									
Retake Exam within One Year	0.125	0.330	21692	0.000	0.008	17459	0.638	0.481	4233
Earn Standard Cert. in Six Months	0.429	0.495	21692	0.487	0.500	17459	0.189	0.392	4233
Earn Standard Cert. in One Year	0.588	0.492	21692	0.659	0.474	17459	0.296	0.457	4233
Employed as a Teacher Next October	0.708	0.455	21692	0.738	0.439	17459	0.581	0.493	4233

Notes: Sample includes all first-attempt takers of the TExES Science of Reading exam (exam 293) from January 1, 2021 through October 23, 2023, restricted to candidates with valid scaled scores. Passed and Failed columns split the sample by first-attempt pass status (score $\geq$ 240). Demographic information is available for candidates who can be linked to PEIMS K–12 student records or THECB postsecondary enrollment records; the reduced sample sizes for Female and Non-white reflect this linkage. Certification outcomes measure whether the candidate earns any standard teaching certificate within the specified window. Employment is measured as of the last Friday in October following the exam date.

Table 2: Descriptive Statistics for New Certifications

	Pre-Mandate		Post-Mandate	
	Mean	N	Mean	N
SOR-Required Fields	0.417	139,575	0.360	203,241
Non-SOR Fields	0.583	139,575	0.640	203,241
<i>Route</i>				
Traditional	0.240	139,575	0.227	203,241
Alternative	0.338	139,575	0.335	203,241
By Exam	0.323	139,575	0.302	203,241
Out of State	0.099	139,575	0.136	203,241
<i>Demographics</i>				
Female	0.747	139,575	0.757	203,241
White	0.592	139,575	0.552	203,241
Hispanic	0.251	139,575	0.276	203,241
Black	0.108	139,575	0.117	203,241

Notes: Table reports descriptive statistics for the certification-level sample underlying the market-level analysis, prior to collapsing to the educator preparation program (EPP) $\times$ program-type $\times$ certification-field $\times$ half-year panel. Sample includes all newly issued standalone teaching certifications in the SOR-required and comparison fields listed in Table A2, 2016 through 2025. Post-mandate is defined as certifications issued from 2019H2 onward.

Table 3: Structure of the Market-Level Analytic Panel

	SOR-Required	Comparison
Certification fields	3	16
EPP-program-type units	200	200
Panel cells (EPP $\times$ field $\times$ half-year)	9,585	51,120
Share of cells zero-filled	51.2%	69.9%
Total certifications in panel	131,335	211,481

Notes: This table describes the panel formed by collapsing certification records in Table 2 to counts by educator preparation program (EPP) $\times$ program-type $\times$ certification-field $\times$ half-year, spanning 2016H1 through 2025H2. Certification fields are the 19 individual certification categories classified as SOR-required or comparison fields listed in Table A2. EPP-program-type units are the distinct combinations of educator preparation program and program type (e.g., Traditional, Alternative) observed issuing certifications. Panel cells are EPP-program-type $\times$ certification-field $\times$ half-year observations; a cell is coded zero if the EPP-program-type issued no certifications in that field during that half-year but was otherwise active, defined as having issued a certification in any field during that half-year. Periods before an EPP-program-type’s first issued certification, or after its last, are excluded from the panel rather than coded as zero.

Table 4: Pre-Period EPP Characteristics by Texas Reading Academy Status

	Academy EPPs		Non-Academy EPPs		Difference	
	Mean	SD	Mean	SD	Mean	SE
Certification Exams						
N Exam Takers	194.2	184.3	288.7	972.8	94.5	(309.5)
Pass SOR Exam	0.827	0.186	0.843	0.118	0.016	(0.042)
SOR Exam Scaled Score	252.9	10.4	252.9	7.0	0.1	(2.4)
Pass PPR Exam	0.872	0.092	0.899	0.099	0.027	(0.033)
PPR Exam Scaled Score	259.9	8.7	260.5	6.8	0.6	(2.3)
Program Type						
Traditional	0.735	0.408	0.578	0.473	-0.157	(0.156)
Alternative	0.265	0.408	0.411	0.471	0.146	(0.155)
By Exam	0.000	0.000	0.005	0.044	0.005	(0.014)
Out of State	0.000	0.000	0.006	0.059	0.006	(0.019)
Certs Issued						
Certs. Requiring SOR Exam	0.494	0.091	0.501	0.180	0.007	(0.058)
Certs. Not Requiring SOR Exam	0.506	0.091	0.499	0.180	-0.007	(0.058)
Demographics						
Female	0.786	0.073	0.769	0.117	-0.017	(0.038)
Asian	0.016	0.011	0.025	0.032	0.009	(0.010)
Black	0.038	0.040	0.116	0.180	0.079	(0.057)
Other	0.022	0.018	0.023	0.020	0.001	(0.007)
Hispanic	0.449	0.304	0.404	0.254	-0.045	(0.086)
White	0.475	0.273	0.431	0.249	-0.043	(0.083)
N EPPs	10		94		104	

Notes: Table compares pre-period characteristics of Educator Preparation Programs (EPPs) authorized to offer Texas Reading Academy coursework beginning in 2024–25 to EPPs that never adopted Texas Reading Academies during the study period. Means are calculated over academic years 2021–22 through 2023–24 (pre-academy period). The difference column reports the non-academy minus academy difference with standard errors from a two-sample *t*-test in parentheses. EPPs that adopted Texas Reading Academies in 2025–26 are excluded from both groups to avoid contamination from anticipatory changes at programs preparing to adopt. See [Table A16](#) for the full list of EPPs and adoption years.

[†] Demographics presented for individuals who earn Standard teaching certificates associated with an EPP.

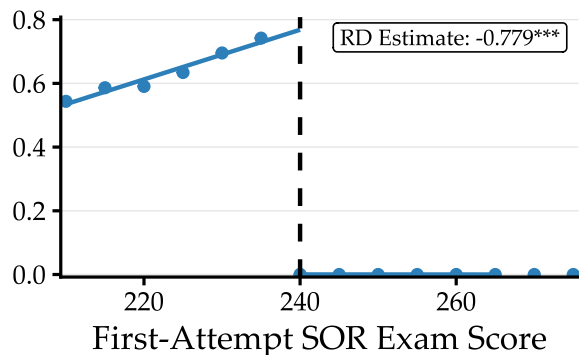
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: RD Estimates of Passing the SOR Exam

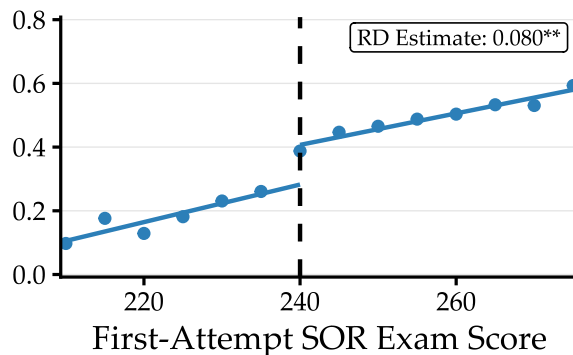
	(1) Retake SOR Exam	(2) Earn Std. Cert. in 6 Months	(3) Earn Std. Cert. in 1 Year	(4) Teach Next October
Conventional	−0.777*** (0.019)	0.085*** (0.025)	0.050 (0.029)	0.065** (0.025)
Bias-corrected	−0.779*** (0.019)	0.080** (0.025)	0.040 (0.029)	0.066** (0.025)
Robust	−0.779*** (0.023)	0.080** (0.031)	0.040 (0.033)	0.066* (0.030)
Month of Exam FE	✓	✓	✓	✓
<i>N</i>	21,692	21,692	21,692	21,692

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam (TExES 293) on teacher outcomes. Each column presents a separate outcome. Columns (1) shows the probability of retaking the SOR exam within one year of the exam date. Columns (2) and (3) show the probability of earning a standard teaching certificate within six months and one year of the exam date, respectively. Column (4) shows the probability of working as a teacher in Texas public schools on the last Friday of October following the exam date. All specifications include month-of-exam fixed effects. All estimates use a local linear regression with triangular kernel weights and data-driven bandwidth selection. Standard errors in parentheses. Conventional and bias-corrected rows report heteroskedasticity-robust standard errors; robust rows report the bias-corrected robust standard errors of Calonico et al. (2014), which account for bias in the local polynomial estimator near the boundary. Sample includes all first-attempt test takers of the SOR exam from January 2021 to October 2023.

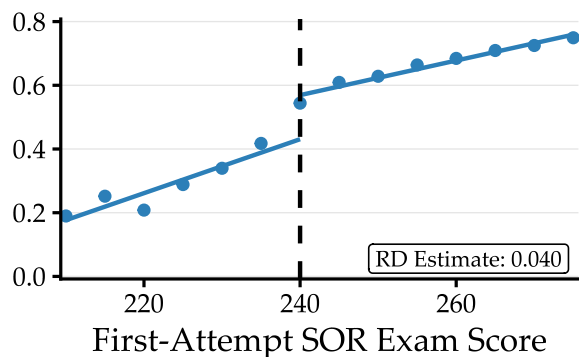
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



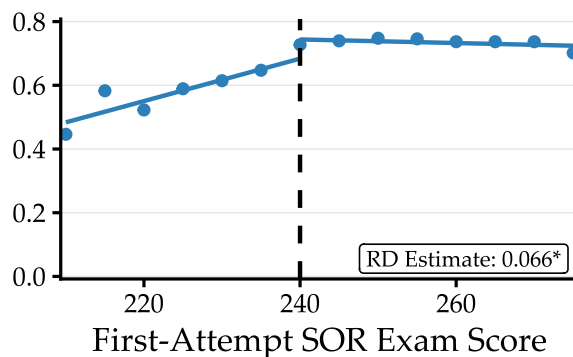
(a) Retake Exam in Twelve Months



(b) Earn Standard Cert. in Six Months



(c) Earn Standard Cert. in Twelve Months



(d) Work as a Teacher Next October

Figure 1: Teacher Outcomes by Science of Reading Exam Score

Notes: Each panel plots outcomes for first-attempt test takers of the TExES Science of Reading (SOR) exam (exam 293) from January 1, 2021 to October 23, 2023. X-axis shows first-attempt exam scores grouped into five-point bins (e.g., scores 220–224 are assigned to bin 220); points represent the mean outcome proportion within each bin. Lines show separate linear fits on each side of the passing threshold of 240 (indicated by the vertical dashed line). Panel (a) shows the proportion who retake the SOR exam within one year. Panel (b) shows the proportion who earn a Standard teaching certificate within six months of the exam date. Panel (c) shows the proportion who earn a Standard teaching certificate within twelve months of the exam date. Panel (d) shows the proportion who work as a teacher in Texas public schools on the last Friday of October following the exam date. To protect individual privacy, cells with proportions below one percent are coded as zero and cells above 95 percent are coded as one. The annotated RD estimate in each panel is the bias-corrected robust regression discontinuity estimate of Calonico et al. (2014), using a local linear regression with triangular kernel weights and data-driven bandwidth selection, corresponding to the robust estimates reported in Table 5.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

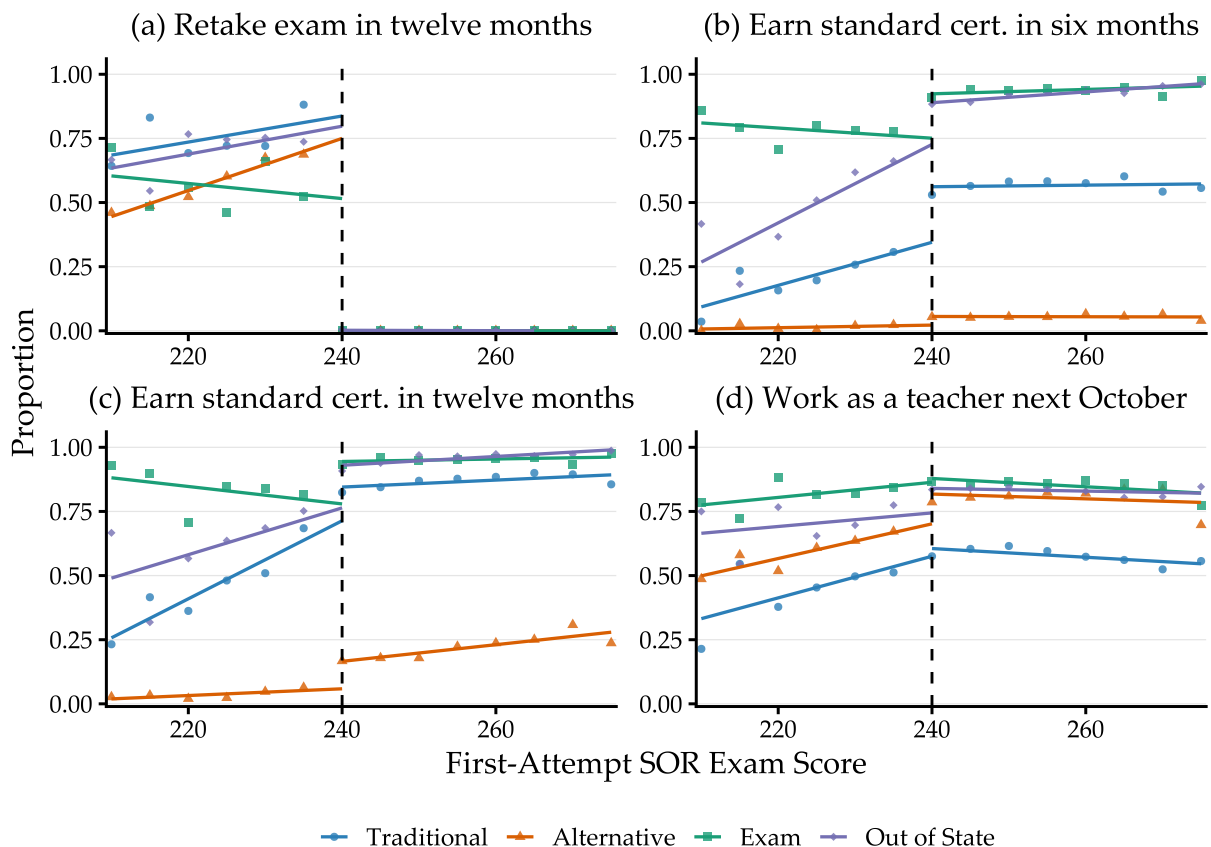


Figure 2: Teacher Outcomes by SOR Exam Score and Certification Route

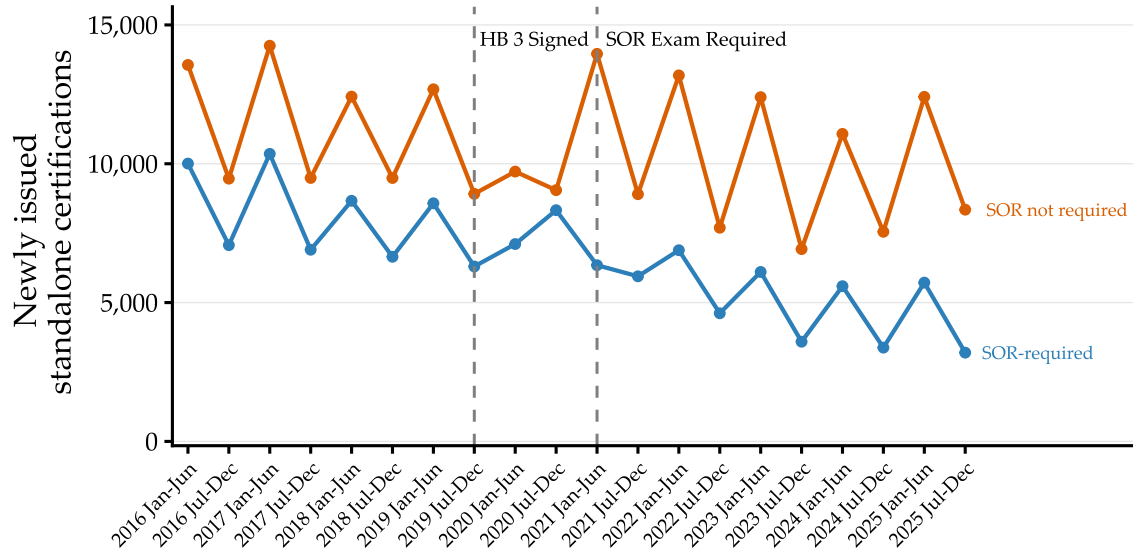
Notes: Each panel plots outcomes by certification route for first-attempt test takers of the TExES Science of Reading (SOR) exam (exam 293) from January 1, 2021 to October 23, 2023. Certification routes are Traditional (university-based educator preparation programs), Alternative (alternative certification providers), By Exam (candidates certified by examination without a preparation program), and Out of State (candidates with out-of-state credentials). X-axis shows first-attempt exam scores grouped into five-point bins (e.g., scores 220–224 are assigned to bin 220); points represent the mean outcome proportion within each bin. Lines show separate linear fits on each side of the passing threshold of 240 (indicated by the vertical dashed line). Panel (a) shows the proportion who retake the SOR exam within one year. Panel (b) shows the proportion who earn a Standard teaching certificate within six months of the exam date. Panel (c) shows the proportion who earn a Standard teaching certificate within twelve months of the exam date. Panel (d) shows the proportion who work as a teacher in Texas public schools on the last Friday of October following the exam date. To protect individual privacy, cells with proportions below one percent are coded as zero and cells above 95 percent are coded as one.

Table 6: RD Estimates of Passing the SOR Exam, by Certification Route

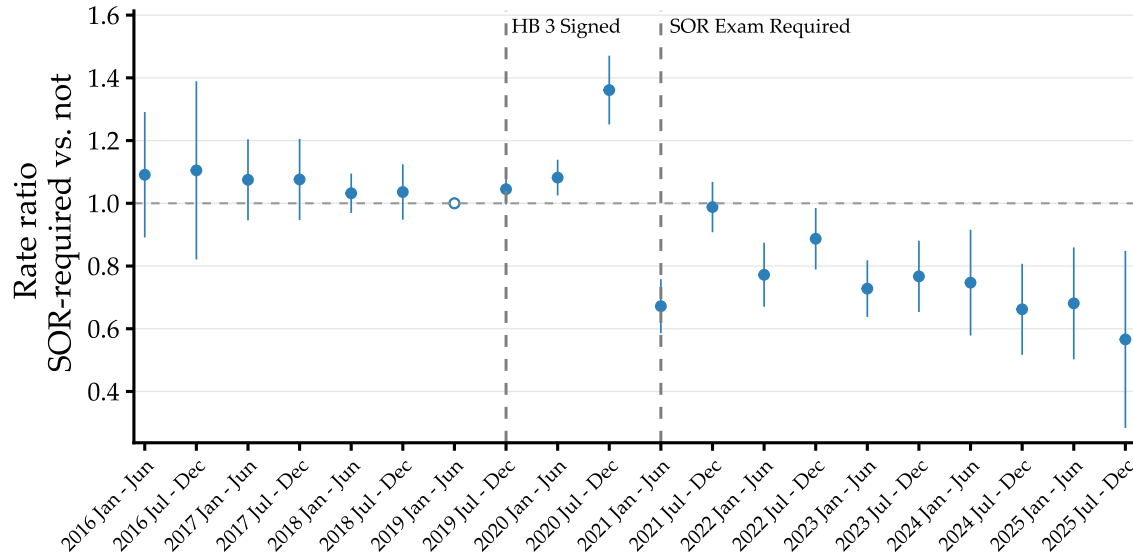
	(1) Retake SOR Exam	(2) Earn Std. Cert. in 6 Months	(3) Earn Std. Cert. in 1 Year	(4) Teach Next October
Traditional	−0.928*** (0.039)	0.148** (0.050)	−0.067 (0.057)	0.019 (0.053)
Alternative	−0.682*** (0.038)	0.024 (0.013)	0.105*** (0.025)	0.114** (0.043)
By Exam	−0.558*** (0.083)	0.099 (0.096)	0.081 (0.091)	0.022 (0.091)
Out of State	−0.802*** (0.059)	0.164* (0.067)	0.073 (0.066)	−0.043 (0.071)

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam (TExES 293) on teacher outcomes, estimated separately for each certification route. Each cell reports the estimate from a separate regression. Column (1) shows the probability of retaking the SOR exam within one year. Columns (2) and (3) show the probability of earning a standard teaching certificate within six months and one year of the exam date, respectively. Column (4) shows the probability of working as a teacher in Texas public schools on the last Friday of October following the exam date. All specifications include month-of-exam fixed effects, which compare candidates who sat for the exam at a similar point relative to the October employment snapshot; they are retained for the certification outcomes for comparability and have little effect on the estimates (Table A3). Estimates use local linear regression with triangular kernel weights and data-driven bandwidth selection. Estimates are bias-corrected and standard errors are the robust standard errors of Calonico et al. (2014), reported in parentheses. Traditional route includes candidates enrolled in university-based educator preparation programs; Alternative includes alternative certification providers; By Exam includes candidates who satisfied certification requirements by examination; Out of State includes candidates with credentials from another state. Sample includes first-attempt SOR exam takers from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



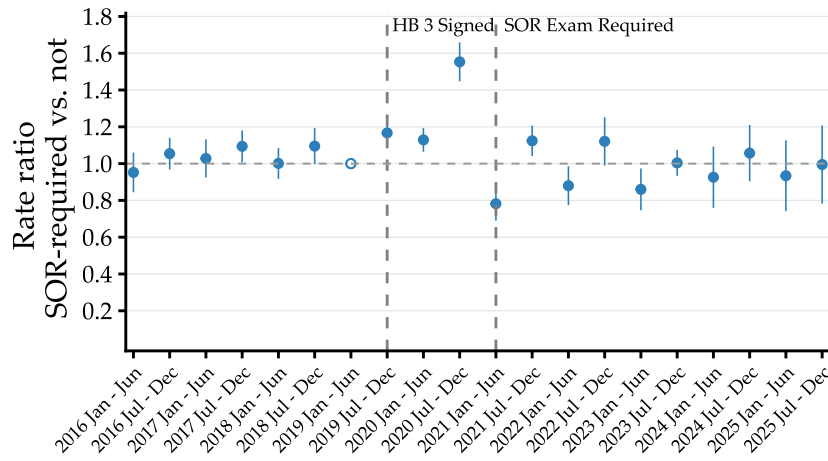
(a) Semi-annual certification counts



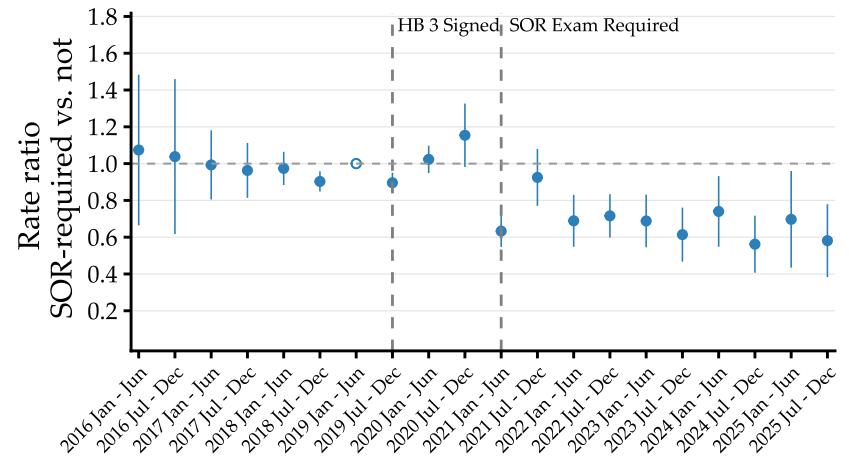
(b) Event study estimates

Figure 3: New Certifications in SOR-Required Fields

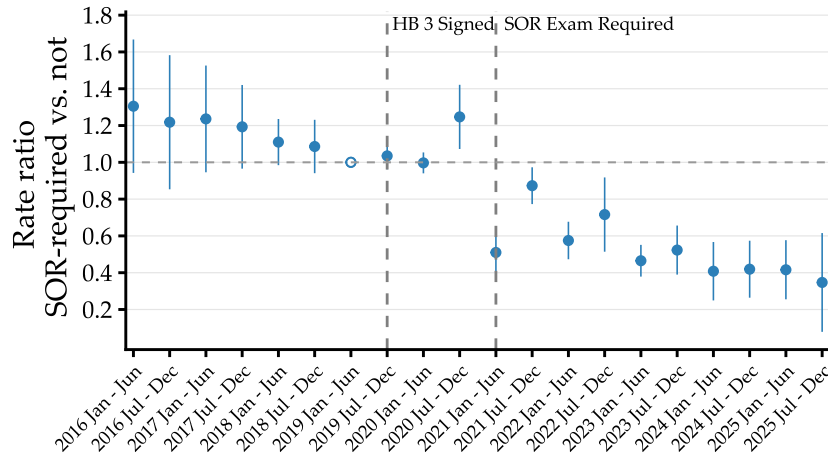
Notes: Panel (a) plots the number of newly issued standalone Texas teaching certifications requiring versus not requiring the TExES Science of Teaching Reading (SOR) exam (TExES 293), by half-year. The dotted line indicates when HB 3 was signed into law (June 2019); the dashed line indicates when the SOR requirement took effect (January 2021). Panel (b) plots coefficients from a difference-in-differences event study estimated by Poisson pseudo-maximum likelihood on data collapsed to the educator preparation program (EPP) $\times$ program type $\times$ half-year $\times$ certification type level, where the outcome is the count of newly issued certifications. Each point gives the rate ratio of certifications in SOR-required relative to non-SOR fields, normalized to the omitted period, 2019 January–June (hollow marker). Specifications include EPP $\times$ program type, certification type, and half-year fixed effects. Vertical bars give 95% confidence intervals; standard errors are clustered at the certification type level. [Table A2](#) lists SOR-required and comparison fields. A newly issued certification refers to the first time an individual receives that certification; non-teaching and supplemental credentials are excluded.



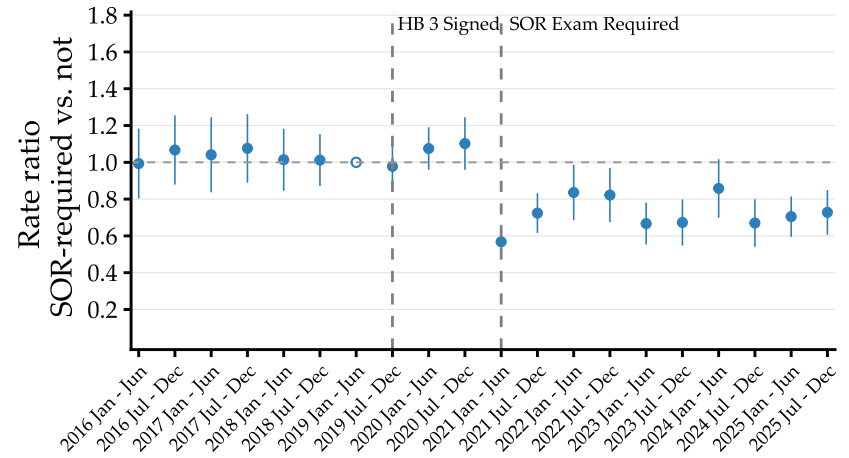
(a) Traditional



(b) Alternative



(c) By Exam



(d) Out of State

Figure 4: New Certifications in SOR-Required Fields Relative to 2019H1, by Certification Route

Notes: Each panel reproduces the market-level difference-in-differences event study of Figure 3 for a single certification route: Traditional (panel a), Alternative (panel b), Certification by Exam (panel c), and Out of State (panel d). Within each route, coefficients are estimated by Poisson pseudo-maximum likelihood on data collapsed to the educator preparation program (EPP) $\times$ program type $\times$ half-year $\times$ certification type level, where the outcome is the count of newly issued certifications. Each point gives the rate ratio of certifications in SOR-required relative to non-SOR fields in that half-year, normalized to the omitted period, 2019 January–June (hollow marker). The dotted vertical line indicates when HB 3 was signed into law (June 2019); the dashed vertical line indicates when the SOR requirement took effect (January 2021). Specifications include EPP $\times$ program type, certification type, and half-year fixed effects. Vertical bars give 95% confidence intervals; standard errors are clustered at the certification type level. Table A2 lists SOR-required and comparison fields. A newly issued certification refers to the first time an individual receives that certification; non-teaching and supplemental credentials are excluded.

Table 7: Effect of the SOR Requirement on New Certifications: Event Study

	All (1)	Traditional (2)	Alternative (3)	By Exam (4)	Out of State (5)
2016 Jan–Jun × SOR Req.	1.091 (0.102)	0.952 (0.055)	1.074 (0.209)	1.305 (0.185)	0.993 (0.097)
2016 Jul–Dec × SOR Req.	1.105 (0.145)	1.054 (0.044)	1.038 (0.215)	1.218 (0.186)	1.067 (0.096)
2017 Jan–Jun × SOR Req.	1.075 (0.066)	1.028 (0.053)	0.993 (0.096)	1.236 (0.148)	1.041 (0.104)
2017 Jul–Dec × SOR Req.	1.076 (0.066)	1.094* (0.044)	0.963 (0.076)	1.193 (0.116)	1.076 (0.095)
2018 Jan–Jun × SOR Req.	1.032 (0.032)	1.001 (0.043)	0.974 (0.046)	1.110 (0.064)	1.014 (0.086)
2018 Jul–Dec × SOR Req.	1.036 (0.045)	1.095* (0.050)	0.903** (0.028)	1.086 (0.074)	1.012 (0.072)
2019 Jan–Jun × SOR Req.	1.000	1.000	1.000	1.000	1.000
2019 Jul–Dec × SOR Req.	1.045 (0.025)	1.167*** (0.041)	0.896*** (0.028)	1.035 (0.037)	0.978 (0.054)
2020 Jan–Jun × SOR Req.	1.082** (0.029)	1.129*** (0.033)	1.023 (0.038)	0.997 (0.029)	1.075 (0.059)
2020 Jul–Dec × SOR Req.	1.361*** (0.056)	1.553*** (0.054)	1.154 (0.088)	1.247** (0.089)	1.102 (0.073)
2021 Jan–Jun × SOR Req.	0.672*** (0.044)	0.782*** (0.047)	0.633*** (0.044)	0.510*** (0.053)	0.568*** (0.047)
2021 Jul–Dec × SOR Req.	0.988 (0.041)	1.124** (0.042)	0.925 (0.079)	0.873* (0.051)	0.724*** (0.055)
2022 Jan–Jun × SOR Req.	0.772*** (0.052)	0.880* (0.054)	0.689*** (0.072)	0.575*** (0.052)	0.836 (0.077)
2022 Jul–Dec × SOR Req.	0.887* (0.050)	1.121 (0.067)	0.716*** (0.060)	0.716* (0.103)	0.822* (0.075)
2023 Jan–Jun × SOR Req.	0.728*** (0.046)	0.860* (0.058)	0.688*** (0.073)	0.465*** (0.044)	0.667*** (0.058)
2023 Jul–Dec × SOR Req.	0.767*** (0.058)	1.004 (0.036)	0.614*** (0.075)	0.523*** (0.068)	0.673*** (0.064)
2024 Jan–Jun × SOR Req.	0.747* (0.086)	0.926 (0.085)	0.740* (0.098)	0.408*** (0.081)	0.858 (0.081)
2024 Jul–Dec × SOR Req.	0.662*** (0.074)	1.057 (0.078)	0.562*** (0.079)	0.419*** (0.079)	0.670*** (0.066)
2025 Jan–Jun × SOR Req.	0.681** (0.091)	0.934 (0.098)	0.697 (0.134)	0.416*** (0.082)	0.705*** (0.056)
2025 Jul–Dec × SOR Req.	0.566* (0.144)	0.995 (0.108)	0.581** (0.101)	0.347** (0.137)	0.728*** (0.062)
<i>N</i>	60,705	28,367	31,426	380	532
Joint <i>F</i> -test of pre-period (<i>p</i> -value)	0.700	< 0.001	< 0.001	0.019	0.308

Notes: Table reports estimates from Poisson pseudo-maximum likelihood regressions of newly issued standalone teaching certifications on an interaction between certification-field SOR-required status and time. All specifications include EPP × program-type, certification-field, and half-year fixed effects. Coefficients are normalized to the omitted reference period, 2019 January–June. Each point-estimate row gives the rate ratio of certifications in SOR-required relative to non-SOR fields in that half-year. Column (1) uses the full sample; columns (2)–(5) restrict to the traditional, alternative, certification-by-examination, and out-of-state routes, respectively. Coefficients are exponentiated and reported as rate ratios. Standard errors clustered at the certification-field level are provided in parenthesis.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Effect of the SOR Requirement on New Certifications: Pooled Estimates

	All (1)	Traditional (2)	Alternative (3)	By Exam (4)	Out of State (5)
Trend-adjusted level shift ^a	0.823	0.973	0.827	0.604	0.766
Clustered SE p -value	[0.003]	[0.636]	[0.004]	[< 0.001]	[< 0.001]
RI p -value ^b	[0.094]	[0.756]	[0.119]	[0.043]	[0.066]
Post $\times$ SOR Req. (2019H2–2025H2)	0.785	0.994	0.779	0.531	0.761
Clustered SE p -value	[0.016]	[0.945]	[0.131]	[< 0.001]	[0.006]
RI p -value ^b	[0.128]	[0.958]	[0.366]	[0.034]	[0.248]
Donut: Post $\times$ SOR Req. (2022H1–2025H2) ^c	0.686	0.931	0.676	0.403	0.727
Clustered SE p -value	[0.004]	[0.466]	[0.053]	[< 0.001]	[0.003]
RI p -value ^b	[0.105]	[0.695]	[0.239]	[0.026]	[0.244]
N	60,705	28,367	31,426	380	532

Notes: Table reports pooled estimates from Poisson pseudo-maximum likelihood regressions of newly issued standalone teaching certifications on the interaction between certification-field SOR-required status and time. All specifications include EPP $\times$ program-type, certification-field, and half-year fixed effects. Coefficients are exponentiated and reported as rate ratios. Column (1) uses the full sample; columns (2)–(5) restrict to the traditional, alternative, certification-by-examination, and out-of-state routes. Each estimate is followed by the p -value from the certification-field clustered standard error and the randomization inference p -value, both reported in brackets.

^aThe trend-adjusted specification includes a differential linear trend for SOR-required fields estimated on pre-announcement half-years only (2016H1–2019H1). The counterfactual for the post period is the level that fitted trend reaches in 2019H1, so the reported level shift measures the average post-announcement gap relative to the endpoint of the pre-mandate trend rather than to the average pre-announcement differential.

^bRandomization inference p -value is calculated by re-estimating the specification 2,000 times, each time randomly assigning SOR-required status to 3 of the 19 certification fields while holding the true number of treated fields fixed, and computing the t -statistic on the coefficient of interest in each placebo estimate. The reported p -value is the share of these placebo t -statistics whose magnitude equals or exceeds that of the t -statistic from the true assignment.

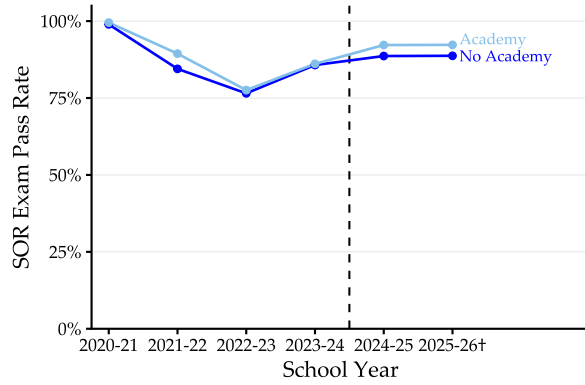
^cThe donut specification excludes half-years 2019H2 through 2021H2, removing the anticipatory certification surge preceding implementation and its immediate aftermath.

Table 9: Certification Exam Results and Student Reading Achievement

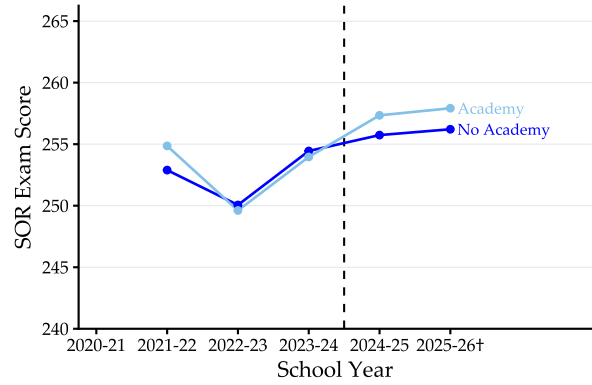
	Student reading achievement (SD)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Standardized first-attempt score					
SOR exam score	0.010*** (0.002)			0.011** (0.003)	0.017*** (0.005)
ELA content exam score		0.007*** (0.001)			0.000 (0.005)
PPR exam score			0.006*** (0.001)	−0.002 (0.003)	
<i>N</i>	399,015	2,026,107	3,824,570	316,163	157,059
Panel B. Passed on the first attempt					
Passed SOR exam	0.024*** (0.006)				0.023** (0.008)
Passed ELA content exam		0.007** (0.003)			
Passed PPR exam			0.008** (0.003)		
Passed Core Subjects exam				0.003 (0.002)	−0.001 (0.004)
<i>N</i>	399,015	2,026,107	3,824,570	1,408,274	506,403

Notes: Table reports estimates of Equation 3, relating student reading achievement on the STAAR to their teachers' certification exam results. The dependent variable is the student's reading score, standardized to mean zero and standard deviation one within grade and year. Panel A enters the teacher's first-attempt scaled score on the Science of Reading (SOR), ELA content, or Pedagogy and Professional Responsibilities (PPR) exam, standardized within exam and administration year. Columns (1)–(3) enter each exam's score on its own; columns (4) and (5) enter the SOR score together with the PPR or ELA score for the teachers who took both exams. The ELA content exams pool the ELA 4–8, ELA 7–12, and ELAR/Social Studies 4–8 exams. Panel B replaces the score with an indicator for passing the exam on the first attempt. Columns (1)–(3) restrict to teachers with valid scaled scores; column (4) uses the Core Subjects (EC–6 and 4–8) exam, for which only pass status is observed; and column (5) enters the SOR and Core Subjects indicators together. All specifications include a cubic in lagged reading and math achievement, student demographics, classroom averages of these covariates including lagged achievement, and school-by-grade, year, teacher experience (0–9 years, with 10 and above pooled), and teacher degree fixed effects. Standard errors clustered at the teacher level are reported in parentheses.

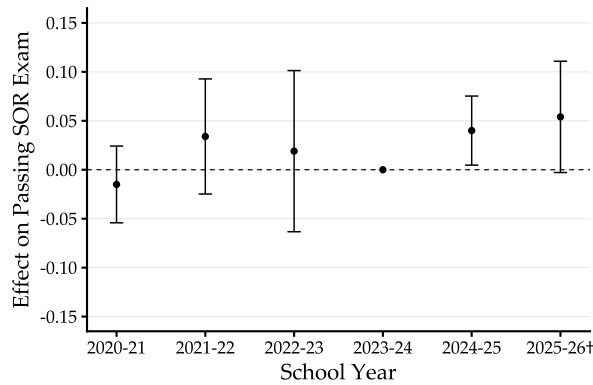
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



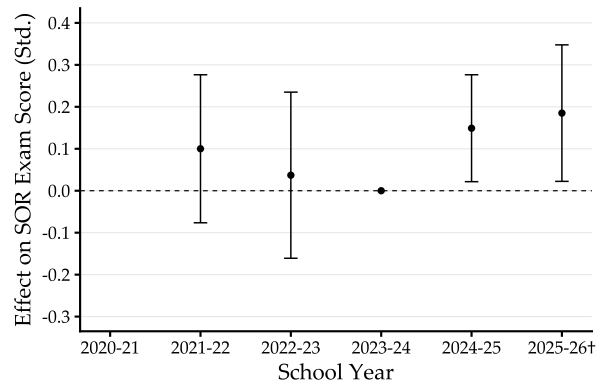
(a) Levels: Pass Rate



(b) Levels: Average Score



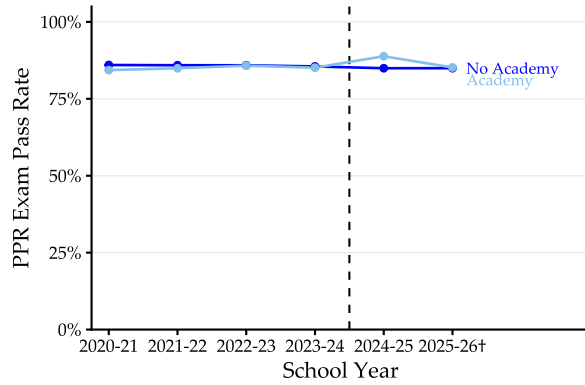
(c) Event Study: Pass Rate



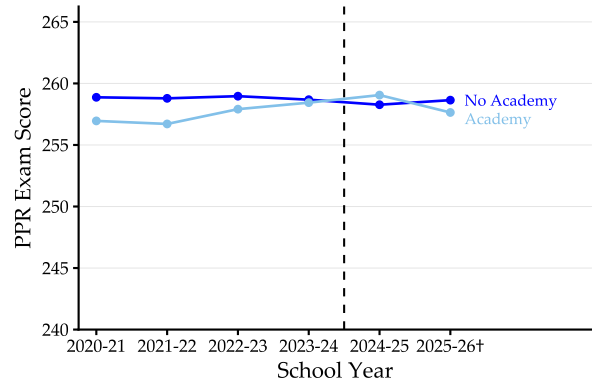
(d) Event Study: Average Score

Figure 5: TExES Science of Reading (SOR) Outcomes by Texas Reading Academy Adoption

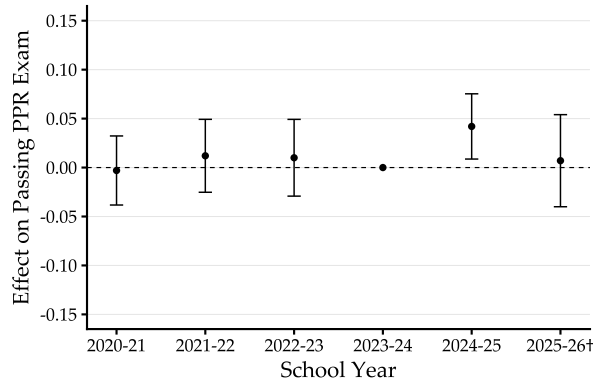
Notes: Panels (a) and (b) show levels of pass rates and average scores, respectively, for the Science of Reading (SOR) exam, comparing EPPs with and without Texas Reading Academies. Panels (c) and (d) show event study estimates for the same outcomes, with the reference year (2023–24) indicated by a hollow marker. Texas Reading Academies were first implemented in the 2024–25 school year. Pass rate is measured as the proportion of first-attempt test takers who pass the exam (score ≥ 240). Average score represents the mean exam score among all first-attempt test takers, standardized to have mean zero and standard deviation of one within each year. [†] Certification exam data through March 2026.



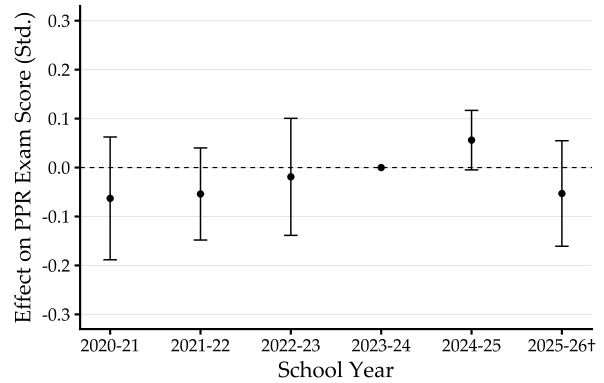
(a) Levels: Pass Rate



(b) Levels: Average Score



(c) Event Study: Pass Rate



(d) Event Study: Average Score

Figure 6: Pedagogy and Professional Responsibilities (PPR) Outcomes by Texas Reading Academy Adoption

Notes: Panels (a) and (b) show levels of pass rates and average scores, respectively, for the Pedagogy and Professional Responsibilities (PPR) exam, comparing EPPs with and without Texas Reading Academies. Panels (c) and (d) show event study estimates for the same outcomes, with the reference year (2023–24) indicated by a hollow marker. Texas Reading Academies were first implemented in the 2024–25 school year. Pass rate is measured as the proportion of first-attempt test takers who pass the PPR exam. Average score represents the mean exam score among all first-attempt test takers, standardized to have mean zero and standard deviation of one within each year. 2020–21 scaled scores are unavailable due to piloting. [†] Certification exam data through March 2026.

Table 10: Effect of Texas Reading Academies on Certification Exam Outcomes

	SOR		PPR	
	Pass Rate (1)	Score (2)	Pass Rate (3)	Score (4)
Panel A. Event Study				
2020–21 × Texas Reading Academy	−0.015 (0.020)		−0.003 (0.018)	−0.063 (0.064)
2021–22 × Texas Reading Academy	0.034 (0.030)	0.100 (0.090)	0.012 (0.019)	−0.054 (0.048)
2022–23 × Texas Reading Academy	0.019 (0.042)	0.037 (0.101)	0.010 (0.020)	−0.019 (0.061)
2023–24 × Texas Reading Academy	0.000 (.)	0.000 (.)	0.000 (.)	0.000 (.)
2024–25 × Texas Reading Academy	0.040* (0.018)	0.149* (0.065)	0.042* (0.017)	0.056+ (0.031)
2025–26 × Texas Reading Academy†	0.054+ (0.029)	0.185* (0.083)	0.007 (0.024)	−0.053 (0.055)
Panel B. Difference-in-Differences				
Post × Texas Reading Academy	0.038+ (0.023)	0.117 (0.099)	0.027 (0.019)	0.056 (0.045)
Joint F-test of pre-period	0.537	0.524	0.634	0.677
<i>N</i>	61,769	40,991	96,254	96,254
Pre-period mean, non-Academy	0.857	0.055	0.855	−0.180
Pre-period mean, Academy	0.861	0.026	0.851	−0.194

Notes: Table reports estimates from event study (Panel A) and difference-in-differences (Panel B) specifications comparing certification exam outcomes at Educator Preparation Programs (EPPs) that adopted Texas Reading Academies in 2024–25 to EPPs that never adopted. Regressions are estimated on first-attempt exam records, so EPPs are weighted by their number of test takers. The outcomes are an indicator for passing on the first attempt and the standardized first-attempt score on the SOR exam (the TExES Science of Teaching Reading exam, exam 293) and the Pedagogy and Professional Responsibilities (PPR) exam. The PPR sample includes all first-attempt PPR takers at each EPP, including candidates in certification fields that do not require the SOR exam. All specifications include EPP fixed effects and year fixed effects. The reference year for the event study is 2023–24, the year immediately preceding academy implementation. The joint F-test reports the p -value from a test that all pre-period interaction coefficients are jointly zero. EPPs that adopted Texas Reading Academies in 2025–26 are excluded from both groups. † Certification exam data through March 1, 2026. Standard errors clustered at the Educator Preparation Program level in parentheses. 2020–21 scaled scores are unavailable due to piloting.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A. Additional Tables and Figures

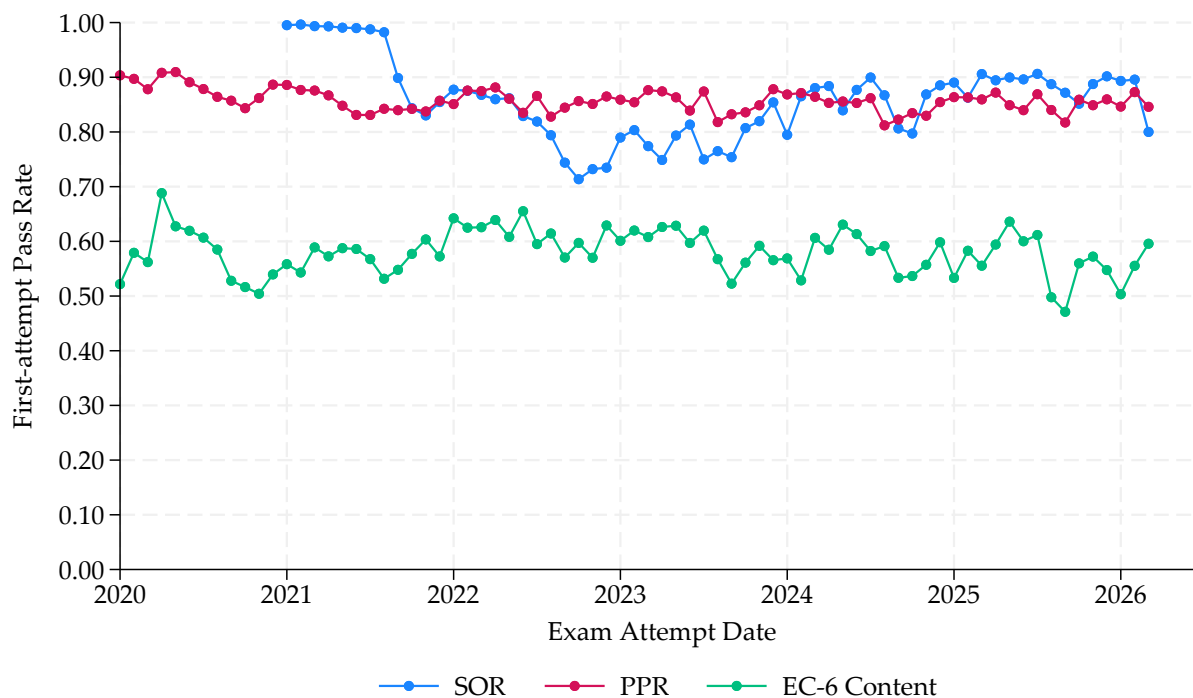


Figure A1: First-Attempt Pass Rates for Texas Teacher EC-6 Certification Exams

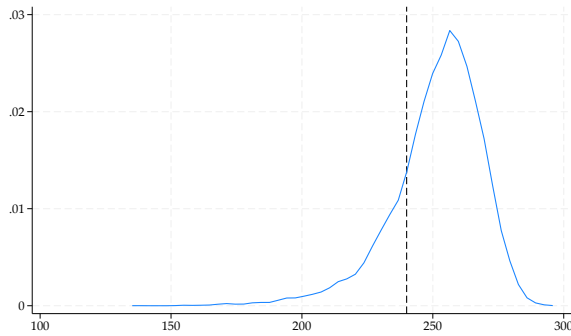
Notes: This figure plots first-attempt pass rates for three Texas teacher certification exams from 2020 to 2025. The blue line shows pass rates for the Science of Reading (SOR) exam (exam 293). The orange line shows pass rates for the Pedagogy and Professional Responsibilities (PPR) exam. The green line shows pass rates for the EC-6 Content exam. The sample includes all first-attempt test takers in Texas. The black vertical line indicates January 1, 2021, when the SOR exam was introduced. The red dashed vertical line indicates when the State Board for Educator Certification (SBEC) began including scaled scores from the SOR exam after initial piloting.

Table A1: Covariates at the SOR Exam Passing Threshold

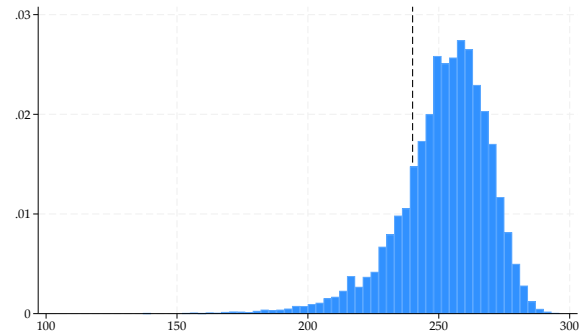
	Conventional	Bias-corrected	Robust
Female	0.010 (0.017)	0.011 (0.017)	0.011 (0.020)
White	0.019 (0.031)	0.022 (0.031)	0.022 (0.038)
Traditional	-0.003 (0.023)	-0.005 (0.023)	-0.005 (0.027)
Alternative	0.012 (0.026)	0.009 (0.026)	0.009 (0.032)
By Exam	0.010 (0.017)	0.016 (0.017)	0.016 (0.020)
Out of State	-0.017 (0.016)	-0.021 (0.016)	-0.021 (0.019)
Teaches Before Exam	-0.008 (0.026)	-0.008 (0.026)	-0.008 (0.032)

Notes: Table reports regression discontinuity estimates at the SOR exam passing threshold (score ≥ 240) with each pre-determined covariate as the outcome. None of the estimates are statistically significant at the 0.05 level, supporting the assumption that candidates just above and below the cutoff are comparable in observable characteristics. All estimates use a local linear regression with triangular kernel weights and data-driven bandwidth selection. Heteroskedasticity-robust standard errors are reported in parentheses. Demographic variables (i.e., Female, White) are available only for candidates linked to PEIMS or THECB records.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



(a) Density Plot



(b) Histogram

Figure A2: Distribution of Science of Reading Exam Scores

Notes: This figure plots the distribution of scores on the Texas Science of Reading (SOR) exam (exam 293) for first-attempt test takers between January 1, 2021 and October 23, 2023. Panel (a) shows a kernel density estimate of the score distribution. Panel (b) shows a histogram with overlaid density estimate. The vertical dotted line indicates the passing threshold of 240. A density discontinuity test using the method of Cattaneo et al. (2018) fails to reject the null hypothesis of no manipulation around the cutoff ($p = 0.051$), suggesting no evidence of sorting around the passing threshold.

Table A2: Classification of Texas Teaching Certificates by SOR-Exam Requirement

Certification Group	SOR (293) Required?
<i>Panel A: SOR-required certifications</i>	
Core Subjects: Early Childhood–Grade 6	Yes
Core Subjects: Grades 4–8	Yes
English Language Arts and Reading: Grades 4–8	Yes
<i>Panel B: Non-SOR certifications (comparison group)</i>	
Computer Science: Grades 8–12	No
English Language Arts and Reading: Grades 7–12	No
Fine Arts: EC–12	No
Foreign Language: EC–12	No
Health & Physical Education	No
Mathematics: Grades 4–8	No
Mathematics: Grades 7–12	No
Science: Grades 4–8	No
Science: Grades 6–12	No
Science: Grades 7–12	No
Social Studies: Grades 4–8	No
Social Studies: Grades 7–12	No
Special Education: EC–12	No
Vocational Education 6–12	No
Vocational Education 8–12	No
Vocational Education EC–12	No

Notes: Table lists the 19 certification groups in the market-level analysis sample, which are also the units at which standard errors are clustered. Classification follows the Texas Education Agency Required Test Chart (May 2024); a certification is coded SOR-required if the TExES Science of Teaching Reading exam (293) appears among its required content-pedagogy tests. Successive versions of a certificate are grouped into a single field, so that each SOR-required group is observed before and after the mandate: Core Subjects and Core Subjects with STR (and the earlier Generalist certificates) are grouped by grade band, Early Childhood: PK–3 is grouped with Core Subjects: EC–6, and English Language Arts and Reading/Social Studies: Grades 4–8 (with or without STR) is grouped with English Language Arts and Reading: Grades 4–8. Non-teaching credentials (Reading Specialist, Counselor, Diagnostician, Librarian, Principal, Superintendent) and supplemental certificates that attach to a base certificate rather than certifying a teaching field independently (Bilingual, ESL, Gifted & Talented, Special Education Supplemental) are excluded from the sample.

Table A3: RD Estimates of Passing the Science of Reading Exam: With and Without Controls

	Retake SOR Exam		Earn Std. Cert. in 6 Months		Earn Std. Cert. in 1 Year		Teach Next October	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	−0.777*** (0.019)	−0.776*** (0.019)	0.085*** (0.025)	0.084** (0.026)	0.050 (0.029)	0.051 (0.029)	0.065** (0.025)	0.059* (0.026)
Bias-corrected	−0.779*** (0.019)	−0.778*** (0.019)	0.080** (0.025)	0.078** (0.026)	0.040 (0.029)	0.040 (0.029)	0.066** (0.025)	0.061* (0.026)
Robust	−0.779*** (0.023)	−0.778*** (0.023)	0.080** (0.031)	0.078* (0.031)	0.040 (0.033)	0.040 (0.034)	0.066* (0.030)	0.061 (0.031)
Month of Exam FE	✓		✓		✓		✓	
<i>N</i>	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam (TExES 293) on teacher outcomes with and without month-of-exam fixed effects. Odd-numbered columns include month-of-exam fixed effects; even-numbered columns do not. Columns (1)–(2) show the probability of retaking the SOR exam within one year of the exam date. Columns (3)–(4) show the probability of earning a Standard teaching certificate within six months of the exam date. Columns (5)–(6) show the probability of earning a Standard teaching certificate within one year of the exam date. Columns (7)–(8) show the probability of working as a teacher in Texas public schools on the last Friday of October following the exam date. Month-of-exam fixed effects account for the mechanical relationship between exam timing and the annual October employment snapshot. All estimates use a local linear regression with triangular kernel weights and data-driven bandwidth selection. Standard errors in parentheses. Conventional and bias-corrected rows report heteroskedasticity-robust standard errors; robust rows report the bias-corrected robust standard errors of Calonico et al. (2014), which account for bias in the local polynomial estimator near the boundary. Sample includes all first-attempt test takers of the SOR exam from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Robustness of RD Estimates to Bandwidth Selection Method

	Retake SOR Exam			Earn Std. Cert. 6 Mo.			Earn Std. Cert. 1 Yr.			Teach Next October		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Conventional	−0.777*** (0.019)	−0.777*** (0.019)	−0.772*** (0.025)	0.085*** (0.025)	0.086*** (0.025)	0.065 (0.034)	0.050 (0.029)	0.054* (0.027)	0.020 (0.038)	0.065** (0.025)	0.065** (0.025)	0.063 (0.032)
Bias-corrected	−0.779*** (0.019)	−0.779*** (0.019)	−0.773*** (0.025)	0.080** (0.025)	0.080** (0.025)	0.063 (0.034)	0.040 (0.029)	0.041 (0.027)	0.016 (0.038)	0.066** (0.025)	0.066** (0.025)	0.064* (0.032)
Robust	−0.779*** (0.023)	−0.779*** (0.023)	−0.773*** (0.027)	0.080** (0.031)	0.080** (0.031)	0.063 (0.036)	0.040 (0.033)	0.041 (0.033)	0.016 (0.040)	0.066* (0.030)	0.066* (0.030)	0.064 (0.035)
BW Method	mserd	msesum	cerrd	mserd	msesum	cerrd	mserd	msesum	cerrd	mserd	msesum	cerrd
Bandwidth	13.4	13.3	8.1	11.7	11.9	7.1	11.2	12.1	6.8	12.4	12.4	7.5
Month of Exam FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>N</i>	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692	21,692

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam under different data-driven bandwidth selection methods from Calonico et al. (2014). Each set of three columns presents a separate outcome estimated using three bandwidth methods: mserd (MSE-optimal, common bandwidth), msesum (MSE-optimal, sum of bandwidths), and cerrd (CER-optimal, common bandwidth). MSE-optimal bandwidths minimize mean squared error and tend to be wider; CER-optimal bandwidths ensure coverage error rate optimality and tend to be narrower. Employment specifications include month-of-exam fixed effects. Standard errors in parentheses. Conventional and bias-corrected rows report heteroskedasticity-robust standard errors; robust rows report the bias-corrected robust standard errors of Calonico et al. (2014), which account for bias in the local polynomial estimator near the boundary. Sample includes all first-attempt test takers of the SOR exam from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Estimating Market-Level Results with OLS

	(1) Event study	(2) Full post	(3) Donut	(4) Pre-trend Adjusted
<i>Panel A. Event study</i>				
2016 Jan–Jun × SOR Req.	4.826 [0.265]			
2016 Jul–Dec × SOR Req.	−0.188 [0.904]			
2017 Jan–Jun × SOR Req.	5.318 [0.261]			
2017 Jul–Dec × SOR Req.	−0.855 [0.615]			
2018 Jan–Jun × SOR Req.	1.087 [0.236]			
2018 Jul–Dec × SOR Req.	−2.153 [0.568]			
2019 Jan–Jun × SOR Req.	0.000			
2019 Jul–Dec × SOR Req.	−2.904 [0.368]			
2020 Jan–Jun × SOR Req.	−1.804 [0.320]			
2020 Jul–Dec × SOR Req.	0.978 [0.087]			
2021 Jan–Jun × SOR Req.	−4.431 [0.061]			
2021 Jul–Dec × SOR Req.	−3.286* [0.041]			
2022 Jan–Jun × SOR Req.	−2.735 [0.056]			
2022 Jul–Dec × SOR Req.	−5.347* [0.021]			
2023 Jan–Jun × SOR Req.	−4.051 [0.055]			
2023 Jul–Dec × SOR Req.	−7.068* [0.013]			
2024 Jan–Jun × SOR Req.	−4.316 [0.062]			
2024 Jul–Dec × SOR Req.	−7.598* [0.011]			
2025 Jan–Jun × SOR Req.	−4.234 [0.052]			
2025 Jul–Dec × SOR Req.	−7.874* [0.038]			
<i>Panel B. Pooled specifications</i>				
Post × SOR Req. (2019H2–2025H2)		−5.148* [0.029]		
Donut: Post × SOR Req. (2022H1–2025H2)			−6.408* [0.036]	
Trend-adjusted level shift				−3.344 [0.132]
<i>N</i>	60,705	60,705	44,517	60,705
Joint <i>F</i> -test of pre-period (<i>p</i> -value)	0.786			

Notes: Table reproduces the market-level analysis by ordinary least squares as a check on the Poisson pseudo-maximum likelihood estimates reported in [Table 7](#) and [Table 8](#). The outcome is the count of newly issued standalone teaching certifications in levels, so coefficients are in certificates per EPP × program-type × certification-field × half-year. All specifications include EPP × program-type, certification-field, and half-year fixed effects. Column (1) reports the event study, normalized to the omitted reference period, 2019 January–June. Column (2) reports the pooled difference-in-differences estimate over the full post-announcement period. Column (3) reports the donut specification, which excludes half-years 2019H2 through 2021H2. Column (4) adds a linear trend in the SOR-required differential through the pre-announcement period and reports the post-announcement level shift relative to that trend. Brackets report wild cluster bootstrap *p*-values with Webb weights and 9,999 replications, clustered at the certification-field level; stars follow the same test. The joint *F*-test reports the *p*-value from a test that the pre-announcement interaction coefficients are jointly zero.

p* < 0.05, *p* < 0.01, ****p* < 0.001

Table A6: Sensitivity of the Market-Level Estimate to Pre-Trend Violations

	2019 Jul–Dec to 2025 Jul–Dec		Donut-weighted: 2022 Jan–Jun to 2025 Jul–Dec	
	90% CI	95% CI	90% CI	95% CI
Original (unrestricted)	[0.730, 0.918]	[0.714, 0.938]	[0.615, 0.844]	[0.597, 0.870]
$M = 0.000$	[0.781, 0.989]	[0.764, 1.011]	[0.681, 0.926]	[0.662, 0.954]
$M = 0.001$	[0.766, 1.024]	[0.746, 1.049]	[0.661, 0.976]	[0.637, 1.008]
$M = 0.002$	[0.738, 1.088]	[0.715, 1.112]	[0.625, 1.063]	[0.599, 1.095]

Notes: Table reports HonestDiD confidence intervals, exponentiated and expressed as rate ratios, under the smoothness restriction Δ^{SD} (Rambachan & Roth, 2023) for the equally-weighted average post-mandate effect over two target windows: the full post-announcement period (2019H2–2025H2, including the anticipatory certification surge preceding the mandate’s January 2021 implementation) and the post-implementation period (2022H1–2025H2). The donut-weighted column places zero weight on half-years 2019H2–2021H2 in the target average, rather than excluding them from the underlying event-study estimation. This allows the counterfactual pre-trend to run smoothly through the anticipatory and initial implementation periods before being evaluated against the post-implementation coefficients, and is more conservative than dropping these periods outright, since it lets any accumulated pre-trend deviation compound over more periods before reaching the target window. M is the maximum permitted period-to-period change in the counterfactual trend’s slope, relative to a linear extrapolation of the pre-period trend; $M = 0$ imposes exact linear continuation.

Table A7: RD Estimates of Passing the SOR Exam on Certification, by Certification Field

	(1) Earn Std. Cert. in 6 Months	(2) Non-SOR Field	(3) Earn Std. Cert. in 1 Year	(4) Non-SOR Field
	SOR Field	Non-SOR Field	SOR Field	Non-SOR Field
Conventional	0.120*** (0.024)	-0.003 (0.014)	0.090*** (0.027)	-0.019 (0.017)
Bias-corrected	0.113*** (0.024)	-0.004 (0.014)	0.079** (0.027)	-0.022 (0.017)
Robust	0.113*** (0.029)	-0.004 (0.018)	0.079* (0.031)	-0.022 (0.021)

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam (TEExES 293) on the probability of earning a Standard teaching certificate, estimated separately by certification field. Each cell reports the estimate from a separate regression. Columns (1) and (3) use as the outcome earning a certificate in a field for which the SOR exam is required; columns (2) and (4) use earning a certificate in a field for which it is not. SOR-required fields comprise early childhood and elementary generalist certifications (Core Subjects: Early Childhood–Grade 6), middle-grades generalist certifications (Core Subjects: Grades 4–8), and middle-grades English language arts and social studies certifications (English Language Arts and Reading: Grades 4–8; and English Language Arts, Reading/Social Studies: Grades 4–8). Because passing the SOR exam is a prerequisite only in the former group, the non-SOR columns serve as a falsification test. Columns (1) and (2) show the probability of earning a certificate within six months of the exam date; columns (3) and (4) within one year. All specifications include month-of-exam fixed effects. Estimates use local linear regression with triangular kernel weights and data-driven bandwidth selection. Conventional, bias-corrected, and robust estimates follow Calonico et al. (2014); standard errors reported in parentheses. Sample includes first-attempt SOR exam takers from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: RD Estimates of Passing the SOR Exam, by Whether It Was the Last Requirement for Certification

	(1) Retake SOR Exam	(2) Earn Std. Cert. in 6 Months	(3) Earn Std. Cert. in 1 Year	(4) Teach Next October
<i>Panel A. SOR exam was the last requirement</i>				
Conventional	−0.810*** (0.060)	0.211** (0.073)	0.109 (0.058)	0.027 (0.066)
Bias-corrected	−0.808*** (0.060)	0.205** (0.073)	0.108 (0.058)	0.018 (0.066)
Robust	−0.808*** (0.074)	0.205* (0.089)	0.108 (0.072)	0.018 (0.078)
<i>N</i>	4,049	4,049	4,049	4,049
<i>Panel B. Other requirements still open</i>				
Conventional	−0.771*** (0.020)	0.083*** (0.025)	0.064* (0.029)	0.074** (0.026)
Bias-corrected	−0.773*** (0.020)	0.079** (0.025)	0.054 (0.029)	0.077** (0.026)
Robust	−0.773*** (0.024)	0.079** (0.030)	0.054 (0.034)	0.077* (0.032)
<i>N</i>	17,643	17,643	17,643	17,643

Notes: Table reports regression discontinuity estimates of the effect of passing the Science of Teaching Reading (SOR) exam (TExES 293), estimated separately for two groups of first-attempt takers, following the univariate approach to frontier-specific effects in Wong et al. (2013). Panel A restricts the sample to candidates who had passed both the Pedagogy and Professional Responsibilities (PPR) exam and a content exam for an SOR-required field before their first SOR attempt, so that the SOR exam was the last exam requirement for a standard certificate. Panel B includes all other first-attempt takers, who had failed or not yet taken at least one of those exams. Outcomes are as in Table 5. All specifications include month-of-exam fixed effects. Estimates use local linear regression with triangular kernel weights and data-driven bandwidth selection. Conventional, bias-corrected, and robust estimates follow Calonico et al. (2014); standard errors reported in parentheses. Table A9 reports the composition of each group by certification route. Sample includes first-attempt SOR exam takers from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: First-Attempt SOR Takers by Certification Route and Whether the SOR Exam Was the Last Requirement

	SOR exam was the last requirement	Other requirements still open	Total
Traditional	3,768 (26.7%)	10,349 (73.3%)	14,117
Alternative	2,484 (12.0%)	18,184 (88.0%)	20,668
By Exam	1,714 (39.3%)	2,652 (60.7%)	4,366
Out of State	885 (20.9%)	3,340 (79.1%)	4,225
Total	8,851 (20.4%)	34,525 (79.6%)	43,376

Notes: Table reports the number of first-attempt Science of Teaching Reading (SOR) exam takers in each certification route, split by whether they had passed both the Pedagogy and Professional Responsibilities (PPR) exam and a content exam for an SOR-required field before their first SOR attempt. Row percentages in parentheses. Counts include all first-attempt takers from January 2021 to October 2023; the regression samples in [Table A8](#) are smaller because they require an observed scaled score.

Table A10: Selection into the Value-Added Sample

<i>Panel A. Probability of being observed with reading value-added</i>								
	All first-attempt takers				Teaching the following October			
	(1)	(2)	(3)	(4)	(3)	(4)	(3)	(4)
Failed on first attempt	−0.051*** (0.006)	−0.054*** (0.006)	−0.008 (0.009)	−0.010 (0.009)				
Certification-route FE		✓		✓				
<i>N</i>	43,375	43,375	30,216	30,216				

<i>Panel B. Pre-exam characteristics of observed and unobserved candidates</i>								
	Female		White		Family income < \$60,000		SAT score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Observed	−0.035*** (0.004)	−0.021*** (0.006)	−0.006 (0.007)	0.006 (0.011)	0.012 (0.010)	0.043** (0.016)	11.369** (3.850)	11.724* (4.837)
Failed	−0.033*** (0.007)	−0.007 (0.010)	−0.083*** (0.010)	0.000 (0.017)	0.071*** (0.015)	0.019 (0.024)	−98.187*** (5.606)	45.101*** (7.971)
Observed × failed	−0.014 (0.017)	−0.025 (0.017)	−0.002 (0.023)	−0.013 (0.025)	−0.082* (0.034)	−0.108** (0.036)	21.506 (11.419)	6.711 (11.870)
First-attempt SOR score		0.001*** (0.000)		0.004*** (0.000)		−0.003*** (0.001)		5.339*** (0.202)
Certification-route FE	✓	✓	✓	✓	✓	✓	✓	✓
Mean of dependent variable	0.918	0.918	0.531	0.531	0.572	0.572	1,037	1,037
<i>N</i>	27,995	13,448	30,365	14,720	13,920	7,010	9,326	4,984

Notes: Table describes selection into the reading value-added sample at the teacher level, with one observation per first-attempt Science of Teaching Reading (SOR) exam taker. Observed means having a reading value-added estimate, which requires teaching reading in a tested grade (4–8). Panel A regresses an indicator for being observed on an indicator for failing on the first attempt (a first-attempt score below 240), among all first-attempt takers (columns 1–2) and among those working as a teacher in a Texas public school the October after the exam (columns 3–4). Panel B regresses the characteristic in the column header on indicators for being observed, failing, and their interaction, with certification-route fixed effects. The coefficient on observed is the difference between observed and unobserved passers, and the coefficient on observed × failed is how much larger that difference is among failers, the differential selection that would bias the pass/fail comparison in Panel B of Table 9. Even-numbered columns add the first-attempt SOR scaled score as a control; the score is observed for about half of takers, so their samples are smaller. The mean of the dependent variable is among all first-attempt takers for whom the characteristic is observed. Family income and SAT score (with ACT scores converted to the SAT scale) come from Texas Higher Education Coordinating Board application records and are observed only for candidates linked to them. Heteroskedasticity-robust standard errors reported in parentheses. Sample includes first-attempt SOR exam takers from January 2021 to October 2023.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: Certification Exam Scores and Student Achievement: Alternative Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Reading									
SOR Exam	0.011*** (0.002)	0.012*** (0.002)	0.009*** (0.002)				0.011*** (0.003)	0.009** (0.003)	0.010*** (0.003)
PPR Exam				−0.000 (0.001)	0.005*** (0.001)	0.004*** (0.001)	−0.003 (0.003)	0.002 (0.003)	−0.003 (0.003)
<i>N</i>	665,412	665,412	665,412	8,146,331	8,146,331	8,146,331	527,910	527,910	527,910
<i>N</i> Teachers	8,002	8,002	8,002	74,057	74,057	74,057	6,312	6,312	6,312
Panel B. Math									
SOR Exam	0.011*** (0.003)	0.013*** (0.003)	0.009*** (0.003)				0.014** (0.005)	0.011** (0.004)	0.007 (0.004)
PPR Exam				−0.000 (0.001)	0.005*** (0.001)	0.004*** (0.001)	−0.004 (0.005)	−0.000 (0.004)	0.004 (0.004)
<i>N</i>	647,378	647,378	647,378	7,665,403	7,665,403	7,665,403	512,947	512,947	512,947
<i>N</i> Teachers	7,993	7,993	7,993	73,409	73,409	73,409	6,303	6,303	6,303
Grade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
District FE		✓			✓			✓	
Campus FE			✓			✓			✓

Notes: Table reports estimates of Equation 3 under alternative fixed-effects specifications to the school-by-grade model in Table 9, for reading (Panel A) and math (Panel B). The dependent variable is the student's STAAR score standardized within grade, year, and subject. Columns (1)–(3) enter the Science of Reading (SOR) exam score, columns (4)–(6) the Pedagogy and Professional Responsibilities (PPR) exam score, and columns (7)–(9) both scores jointly. All specifications control for prior reading and math achievement (quadratic and cubic terms), race, gender, special education status, gifted status, economic disadvantage, and classroom- and school-level averages of these covariates, with grade and year fixed effects. Columns (2), (5), and (8) add district fixed effects and columns (3), (6), and (9) add campus fixed effects; the fixed-effects rows are reported once, in Panel B, and apply to both panels. Standard errors clustered at the teacher level are reported in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A12: Student Achievement by Quintile of Teacher Exam Score

	Reading			Math		
	SOR (1)	ELA content (2)	PPR (3)	SOR (4)	ELA content (5)	PPR (6)
Panel A. All teachers						
Quintile 2	0.016* (0.007)	0.005 (0.003)	0.004 (0.002)	−0.001 (0.012)	−0.017 (0.017)	−0.002 (0.004)
Quintile 3	0.017* (0.007)	0.005 (0.003)	0.007*** (0.002)	0.021 (0.013)	0.001 (0.017)	0.007 (0.004)
Quintile 4	0.023*** (0.007)	0.015*** (0.003)	0.010*** (0.002)	0.043*** (0.013)	0.024 (0.018)	0.009* (0.005)
Quintile 5	0.030*** (0.007)	0.015*** (0.003)	0.015*** (0.002)	0.039** (0.014)	0.015 (0.017)	0.013** (0.005)
<i>N</i>	399,015	2,026,107	3,824,570	215,564	139,919	3,574,150
Panel B. First-year teachers						
Quintile 2	0.023* (0.011)	0.008 (0.009)	0.001 (0.007)	0.026 (0.018)	−0.054 (0.057)	−0.011 (0.012)
Quintile 3	0.007 (0.012)	−0.002 (0.010)	0.015* (0.007)	0.015 (0.018)	0.116* (0.058)	−0.020 (0.012)
Quintile 4	0.013 (0.011)	0.005 (0.010)	0.006 (0.007)	0.075*** (0.019)	−0.036 (0.059)	−0.008 (0.012)
Quintile 5	0.034** (0.012)	0.011 (0.011)	0.011 (0.007)	0.046* (0.020)	0.160** (0.050)	0.002 (0.011)
<i>N</i>	142,841	199,403	334,332	82,030	10,151	291,207

Notes: Table reports estimates of Equation 3 with the teacher’s standardized first-attempt score replaced by indicators for the quintile of that score among test takers; the bottom quintile is the omitted category. Each column within a panel is a separate regression with the same controls and fixed effects as Table 9. The dependent variable is the student’s STAAR reading score (columns 1–3) or math score (columns 4–6), standardized within grade and year. Panel B restricts the sample to teachers in their first year (zero years of experience). Standard errors clustered at the teacher level are reported in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A13: Certification Exam Results and Student Reading Achievement: First-Year Teachers

	Student reading achievement (SD)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Standardized first-attempt score					
SOR exam score	0.010** (0.004)			0.008 (0.006)	0.023** (0.008)
ELA content exam score		0.004 (0.003)			-0.005 (0.008)
PPR exam score			0.006* (0.002)	-0.004 (0.005)	
<i>N</i>	142,841	199,403	334,332	115,233	49,467
Panel B. Passed on the first attempt					
Passed SOR exam	0.016 (0.010)				0.020 (0.012)
Passed ELA content exam		0.001 (0.007)			
Passed PPR exam			0.013 (0.007)		
Passed Core Subjects exam				-0.013* (0.006)	-0.008 (0.008)
<i>N</i>	142,841	199,403	334,332	187,106	140,626

Notes: Table repeats [Table 9](#) for teachers in their first year (zero years of experience). The dependent variable is the student's STAAR reading score, standardized within grade and year. Because every teacher in the sample has zero years of experience, the experience fixed effects drop out; all other controls and fixed effects are as in [Table 9](#). Standard errors clustered at the teacher level are reported in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A14: Certification Exam Results and Student Math Achievement

	Student math achievement (SD)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Standardized first-attempt score					
SOR exam score	0.019*** (0.005)			0.021*** (0.006)	−0.003 (0.029)
ELA content exam score		0.006 (0.006)			0.015 (0.026)
PPR exam score			0.006*** (0.002)	−0.008 (0.006)	
<i>N</i>	215,564	139,919	3,574,150	170,524	13,023
Panel B. Passed on the first attempt					
Passed SOR exam	0.023* (0.011)				0.016 (0.010)
Passed ELA content exam		0.023 (0.014)			
Passed PPR exam			0.008 (0.005)		
Passed Core Subjects exam				0.001 (0.004)	0.007 (0.006)
<i>N</i>	215,564	139,919	3,574,150	1,282,957	414,924

Notes: Table repeats [Table 9](#) with the student's STAAR math score, standardized within grade and year, as the dependent variable. Controls and fixed effects are as in [Table 9](#). Standard errors clustered at the teacher level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A15: Certification Exam Results and Student Math Achievement: First-Year Teachers

	Student math achievement (SD)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Standardized first-attempt score					
SOR exam score	0.021** (0.007)			0.031*** (0.009)	0.004 (0.034)
ELA content exam score		0.035 (0.020)			0.224*** (0.039)
PPR exam score			0.003 (0.004)	−0.012 (0.009)	
<i>N</i>	82,030	10,151	291,207	65,818	3,347
Panel B. Passed on the first attempt					
Passed SOR exam	0.039* (0.016)				0.058*** (0.016)
Passed ELA content exam		0.063 (0.049)			
Passed PPR exam			−0.002 (0.011)		
Passed Core Subjects exam				−0.004 (0.009)	−0.011 (0.010)
<i>N</i>	82,030	10,151	291,207	167,941	119,140

Notes: Table repeats [Table 9](#) with the student’s STAAR math score, standardized within grade and year, as the dependent variable, for teachers in their first year (zero years of experience). Because every teacher in the sample has zero years of experience, the experience fixed effects drop out; all other controls and fixed effects are as in [Table 9](#). Standard errors clustered at the teacher level are reported in parentheses.

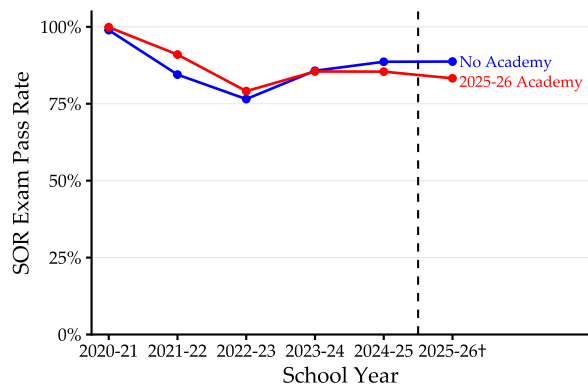
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A16: EPPs Offering Texas Reading Academies by Year of Adoption

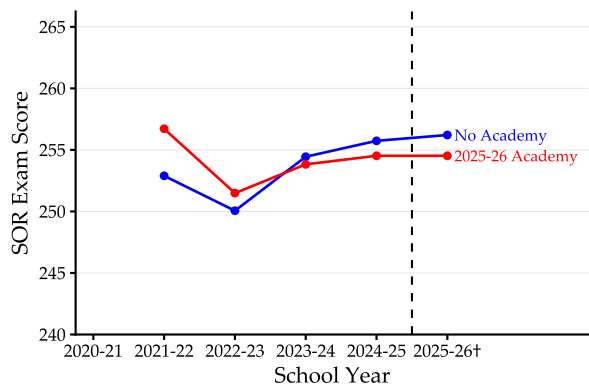
Educator Preparation Program
<i>2024–25 Academic Year</i>
ESC19 Teacher Preparation & Certification Program
Lamar University
Teach Us Texas
Texas A&M University Commerce
Texas Christian University
Texas Tech University
Trinity University
University of Texas at El Paso (UTEP)
University of Texas at Tyler
West Texas A&M University
<i>2025–26 Academic Year</i>
240 Certification
ACTRGV/Pharr San Juan Alamo
Baylor University [†]
Dallas College
East Texas Baptist University
Houston Community College
LeTourneau University [†]
McMurry University
Region 4 Inspire Texas
Responsive Ed 180 Educator Preparation Program
Stephen F. Austin State University Department of Education Studies
Texas Woman’s University
The University of Texas at Dallas - Teacher Development Center

Notes: Table lists Educator Preparation Programs (EPPs) authorized to offer Texas Reading Academies (TRA) coursework to preservice elementary teachers by year of adoption. TRA is a statewide professional development program that engages elementary educators with high-quality instructional strategies rooted in the Science of Reading (SOR). EPPs that adopted TRA in 2024–25 and continued in 2025–26 are listed only under 2024–25. Texas A&M University-Commerce was renamed East Texas A&M University in 2024, and the ESC19 Teacher Preparation & Certification Program appears on later TEA rosters as “Region 19”; both are 2024–25 adopters and are listed only under that year. Sources: [2024–25 cohort](#); [2025–26 cohort](#); [current roster](#).

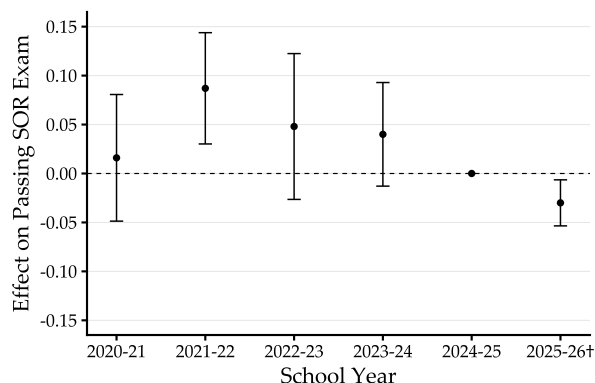
[†] Listed on TEA’s 2025–26 roster as of September 2025 but since removed, indicating the program dropped TRA.



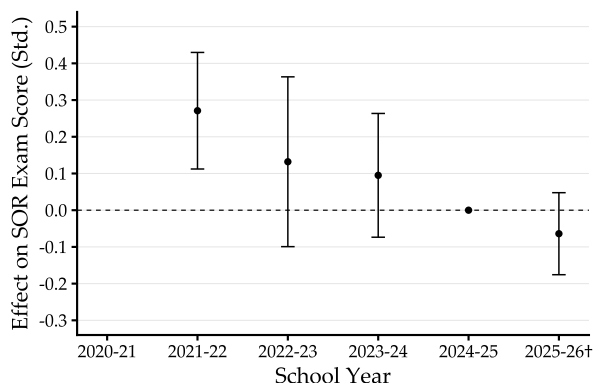
(a) Levels: Pass Rate



(b) Levels: Average Score



(c) Event Study: Pass Rate



(d) Event Study: Average Score

Figure A3: TExES Science of Reading (SOR) Outcomes by Texas Reading Academy Adoption, 2025–26 Cohort

Notes: Panels (a) and (b) show levels of pass rates and average scores, respectively, for the Science of Reading (SOR) exam, comparing EPPs with and without Texas Reading Academies. Panels (c) and (d) show event study estimates for the same outcomes, with the reference year (2024–25) indicated by a hollow marker. Texas Reading Academies were first implemented in the 2025–26 school year. Pass rate is measured as the proportion of first-attempt test takers who pass the exam (score ≥ 240). Average score represents the mean exam score among all first-attempt test takers, standardized to have mean zero and standard deviation of one within each year. [†] Certification exam data through March 2026.

Table A17: Effect of Texas Reading Academies on Certification Exam Outcomes, 2025-26 Cohort

	SOR		PPR	
	Pass Rate (1)	Score (2)	Pass Rate (3)	Score (4)
Panel A. Event Study				
2020–21 × Texas Reading Academy	0.016 (0.033)		0.011 (0.026)	0.075 (0.092)
2021–22 × Texas Reading Academy	0.087** (0.029)	0.271** (0.081)	−0.006 (0.018)	0.103 (0.067)
2022–23 × Texas Reading Academy	0.048 (0.038)	0.132 (0.118)	0.026+ (0.015)	0.092* (0.038)
2023–24 × Texas Reading Academy	0.040 (0.027)	0.095 (0.086)	0.005 (0.015)	0.073 (0.048)
2024–25 × Texas Reading Academy	0.000 (.)	0.000 (.)	0.000 (.)	0.000 (.)
2025–26 × Texas Reading Academy†	−0.030* (0.012)	−0.064 (0.057)	0.028+ (0.015)	−0.002 (0.050)
Panel B. Difference-in-Differences				
Post × Texas Reading Academy	−0.057** (0.021)	−0.151* (0.064)	0.021 (0.013)	−0.060 (0.065)
Joint F-test of pre-period	<0.001	<0.001	0.057	0.178
<i>N</i>	61,423	40,842	93,857	93,857
Pre-period mean, non-Academy	0.887	0.131	0.849	−0.203
Pre-period mean, Academy	0.854	0.059	0.876	−0.099

Notes: Table reports estimates from event study (Panel A) and difference-in-differences (Panel B) specifications comparing certification exam outcomes at Educator Preparation Programs (EPPs) that adopted Texas Reading Academies in 2025–26 to EPPs that never adopted. Regressions are estimated on first-attempt exam records, so EPPs are weighted by their number of test takers. The outcomes are an indicator for passing on the first attempt and the standardized first-attempt score on the SOR exam (the TExES Science of Teaching Reading exam, exam 293) and the Pedagogy and Professional Responsibilities (PPR) exam. The PPR sample includes all first-attempt PPR takers at each EPP, including candidates in certification fields that do not require the SOR exam. All specifications include EPP fixed effects and year fixed effects. The reference year for the event study is 2024–25, the year immediately preceding academy implementation for the second cohort. The joint F-test reports the *p*-value from a test that all pre-period interaction coefficients are jointly zero. EPPs that adopted Texas Reading Academies in 2024–25 are excluded from both groups. † Certification exam data through March 1, 2026. Standard errors clustered at the Educator Preparation Program level in parentheses.

+*p* < 0.10, **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Appendix B. Texas Teacher Certification Exam Descriptions

This appendix provides an overview of the Pedagogy and Professional Responsibilities (PPR) and Science of Reading (SOR) Texas teacher certification exams referenced in this study, including their domain structures and representative sample items.

Pedagogy and Professional Responsibilities EC–12 (Test 160)

The Pedagogy and Professional Responsibilities (PPR) EC–12 exam assesses candidates' knowledge of instructional design, classroom management, and professional conduct. The exam is organized into four domains:

1. **Domain I: Designing Instruction and Assessment to Promote Student Learning.** This domain addresses human developmental processes, learner diversity, instructional planning, and factors that impact student learning. Topics include developmentally appropriate practice, differentiated instruction, curriculum implementation, and critical thinking.
2. **Domain II: Creating a Positive, Productive Classroom Environment.** This domain covers establishing a safe and productive classroom climate, organizing the learning environment, and managing student behavior. Topics include physical classroom arrangement, fostering a passion for learning, formative monitoring, and cooperative group work.
3. **Domain III: Implementing Effective, Responsive Instruction and Assessment.** This domain focuses on communication strategies, active student engagement, technology integration, and monitoring student performance. Topics include questioning techniques, English-language learner supports, BYOD policies, and providing quality feedback.
4. **Domain IV: Fulfilling Professional Roles and Responsibilities.** This domain addresses family involvement, professional development, and legal and ethical requirements for Texas educators. Topics include parent partnerships, mentoring, vertical and horizontal teaming, mandatory reporting of suspected abuse, and IEP development.

Example items:

Which of the following activities would be the most developmentally appropriate for a first-grade teacher to include in a unit on the life cycle of plants?

- (A) Planting a bean and observing the stages of growth
- (B) Reading nonfiction books about plants and creating posters
- (C) Visiting a science center and touring the plant exhibits
- (D) Watching a video about plants and drawing the bean life cycle

Which of the following instructional practices best ensures that a teacher structures the learning environment to address the needs of all students?

- (A) Differentiation (B) Remediation (C) Enrichment (D) Modeling

Which of the following is the most appropriate first step to take if a teacher is concerned that a student is suffering from abuse?

- (A) Calling the parents for a conference to discuss the suspicions
(B) Reporting the suspicions to the child protective services agency
(C) Sharing the concerns with the designated school personnel
(D) Asking the student pointed questions to ensure that all is well at home

Science of Teaching Reading (Test 293)

The Science of Teaching Reading (STR) exam assesses candidates' knowledge of evidence-based reading instruction for early childhood through grade six. The exam is organized into three domains:

1. **Domain I: Reading Pedagogy.** This domain covers foundational concepts in the science of reading, including assets-based approaches, dyslexia intervention, and reading assessment. Topics include systematic explicit instruction, formative and diagnostic assessment, and prevention of reading difficulties.
2. **Domain II: Reading Development—Foundational Skills.** This domain addresses phonological and phonemic awareness, print concepts and alphabet knowledge, phonics and word identification, syllabication and morphemic analysis, and reading fluency. Topics include phoneme-grapheme mapping, the alphabetic principle, decodable texts, inflectional endings, prosody, and automaticity.
3. **Domain III: Reading Development—Comprehension.** This domain covers vocabulary development (including Tier Two and Tier Three words), reading comprehension strategies, comprehension of literary texts, and comprehension of informational texts. Topics include background knowledge activation, text structure, metacognitive strategies, and scaffolding for English learners.

Example items:

A first-grade student has been identified as having dyslexia and has begun intervention. Which of the following approaches to instruction would be most effective to enhance the student's reading development?

- (A) Allowing the student to use colored overlays on all classroom texts
(B) Using reading materials with specialized fonts designed for people with dyslexia
(C) Arranging for the student to use a working-memory training program daily
(D) Providing systematic, explicit multimodal instruction in all essential, evidence-based components of reading

Which of the following sets of words would be most appropriate to categorize as Tier Two words?

- (A) fossil, air, soil (B) clock, book, floor
(C) arrange, observe, predict (D) custom, shelter, community

A third-grade English learner has grade-level decoding skills and scores around the grade-level benchmark for oral reading fluency. However, the student's text comprehension is mixed. When the student is having difficulty with a text, the teacher's best initial response should be to:

- (A) Provide the student with a list of probing questions to answer after reading
(B) Talk with the student to informally assess background knowledge about the topic
(C) Advise the student to read more slowly and focus on comprehension
(D) Encourage the student to develop a written summary while reading

Score Reporting to Candidates

Candidates receive TExES score reports through their Pearson account beginning at 10:00pm Central time on the score report date, typically within 7–10 business days for selected-response exams; exams with constructed-response components take longer because trained scorers grade those responses. Figure B1 presents an example report for the English Language Arts and Reading 7–12 exam (Test 331). The report displays the candidate's total scaled score, pass/not-pass status, the scaled score range (100–300), and the minimum passing score (240). It also provides a performance breakdown by domain and competency, reporting points earned out of points possible for each, with constructed-response items reported separately. Under Texas Education Code, candidates are limited to five attempts per exam, but may appeal this with the Texas Education Agency.



[This barcode contains unique candidate information for security purposes.]



Examinee Score Report

Test: 331 ENG LANGUAGE ARTS AND READING 7-12

Total Scaled Score: 262

Status: Passed

Scaled Score Range: 100-300

Passing Score: 240

Test Date: MM/DD/YYYY

FIRSTNAME M LASTNAME

123 SAMPLE STREET

CITY, TX 99999

TEA ID: 9999999

Performance by Domain and Competency	Points Possible	Points Earned
I. Reading Instruction and Assessment	23	14
001 Foundations of Reading Instruction and Assessment	8	4
002 Vocabulary Development	8	5
003 Reading Comprehension	7	5
II. Text Comprehension and Analysis	14	8
004 Reading Literary Texts	7	3
005 Reading Informational and Argumentative Texts	7	5
III. Oral and Written Communication	23	12
006 Composition	8	3
007 Inquiry and Research	8	4
008 Listening and Speaking	7	5
IV. Educating All Learners and Professional Practice	12	10
009 Differentiation Strategies in Planning and Practice	4	3
010 Culturally Responsive Practices	4	3
011 Data-Driven Practice and Formal/Informal Assessment	4	4
V. Constructed Response	8	8

Your results for the constructed-response item are reported above. For more information about the constructed-response item, see the preparation manual for this exam available on the testing program website at www.tx.nesinc.com.

See next page for additional information.

Figure B1: Example TExES Score Report

Notes: This figure shows an example score report for the TExES English Language Arts and Reading 7–12 (Test 331) exam. The report displays the candidate’s Total Scaled Score, pass/not-pass status, the scaled score range (100–300), and the minimum passing score (240), followed by a Performance by Domain and Competency breakdown reporting points earned out of points possible for each domain and competency. Source: Texas Educator Certification Examination Program (www.tx.nesinc.com).