The Effects of Early Literacy Policies on Student Achievement[1]

John Westall[2]

Amy Cummings[3]

September 30, 2022

Disclaimer: This is a preliminary draft. Please do not cite or circulate without the authors' permission.

Abstract:

Given the importance of early literacy to long-term student success, by 2021, 41 states and the District of Columbia had adopted an early literacy policy to improve student literacy by the end of third grade. Using an event study approach, we examine the impacts of these policies on high- and low-stakes test scores. We find that adopting an early literacy policy improves elementary students' reading achievement on high-stakes assessments. The largest gains occur in third grade and states with comprehensive early literacy policies and those with third-grade retention requirements. Furthermore, we find suggestive evidence that early literacy policies reduce socioeconomic and racial high-stakes achievement gaps in reading. However, we find little evidence of significant gains in low-stakes test scores except in states with comprehensive policies. Our results provide compelling evidence that early literacy policies improve achievement on high-stakes assessments, and these gains are equitably distributed. However, except under comprehensive policies, these high-stakes test-score gains may not necessarily lead to increased literacy learning, as evidenced by limited effects on low-stakes tests.

Keywords: accountability, early literacy, reading, retention, student achievement
JEL Codes: I20, I21, I24

[1] Acknowledgements: We thank Scott Imberman, Melinda Morrill, Katharine Strunk, and participants at the AEFP 2022 Annual Conference and workshops at Michigan State University for their helpful feedback, discussion, and comments. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200009 to Michigan State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

[2] Corresponding author, Postdoctoral Research Fellow, Education Policy Innovation Collaborative, Michigan State University, Contact: westall2@msu.edu

[3] Ph.D. Student, Education Policy, Michigan State University, Contact: cummi264@msu.edu

1. Introduction

Over the past two decades, policymakers have paid increasing attention to early literacy, with a specific focus on reading proficiency by the end of third grade—seen as a critical benchmark for further learning and later outcomes (Cunningham & Stanovich, 1997; Fiester & Smith, 2010; Hernandez, 2011; Sparks et al., 2014). By 2021, 41 states and the District of Columbia had an early literacy policy (ExcelinEd, 2021). These policies vary in content and intensity but share many core components, including diagnostic early literacy assessments beginning in kindergarten, evidence-based literacy instruction and interventions, parental involvement, and professional development for educators—all to improve students' literacy in grades K-3.

Despite the ubiquity of early literacy policies, there is limited research regarding their effects on student achievement or literacy learning. Notable exceptions include longitudinal studies in Florida and Michigan. In Florida, prior research found that third-grade retention under the state's early literacy policy led to short-term positive effects on reading achievement, but these effects dissipate as students progress through school (Greene & Winters, 2004, 2006, 2007, 2009; Schwerdt et al., 2017). Ongoing work in Michigan finds improved achievement under the state's early literacy policy, even before the state implemented the retention component of its law (Strunk et al., 2021). However, to our knowledge, there are no national assessments of early literacy policies across states. We explore the following questions: (1) What are the effects of early literacy policies on reading and math achievement on both low- and high-stakes assessments? (2) Do effects depend on whether the policy includes mandatory retention or is comprehensive (i.e., includes all possible policy components)? (3) Do early literacy policies impact economic or racial test-score gaps?

This paper contributes to existing research in several ways. First, prior work has focused only on the impacts of early literacy policies on high-stakes assessments (Greene & Winters, 2004, 2006, 2007, 2009; Schwerdt et al., 2017; Strunk et al., 2021). While increases in high-stakes test scores could indicate improvements in student literacy skills, they might also reflect "teaching to the test" or other policy-induced behavioral changes (Popham, 2001; Jacob, 2005). To understand the impact of early literacy policies on student learning more thoroughly, we use both high- and low-stakes assessments as outcomes. Second, prior research has evaluated early literacy policies in single states, examining both net impacts (Strunk et al., 2021) and the effects of specific elements, like retention (Greene & Winters, 2004, 2006, 2007, 2009; Schwerdt et al., 2017). Our national setting and detailed policy data allow us to estimate the net impact of early literacy policies across the nation, providing external validity to our findings. Moreover, our national setting provides variation in the content of early literacy policies across states, allowing us to examine whether specific components improve student achievement.

Our analysis uses three publicly available datasets. First, we use a state-level early literacy policy database published by ExcelinEd in 2021 that contains information on all states' early literacy policies, including core components and dates of passage (ExcelinEd, 2021). To measure high-stakes assessment outcomes, we rely on the Stanford Education Data Archive (SEDA), which provides yearly data on aggregated performance on high-stakes reading/language arts and math assessments across states from 2009 to 2018 (S. Reardon et al., 2021). Third, we measure low-stakes assessment outcomes using the National Assessment of Educational Progress (NAEP), which provides average performance on low-stakes reading and math assessments in fourth and eighth grades across states every two years from 1992 to 2019 (The Nation's Report Card, n.d.).

Due to the staggered adoption of early literacy policies across states, we draw from growing literature on robust difference-in-differences estimators (Callaway & Sant'Anna, 2021; Roth et al., 2022; Sun & Abraham, 2021). We use event-study models that leverage variation in the adoption and content of early literacy policies across states and over time to identify the impact of these policies on student achievement. First, we examine the overall effects of early literacy policies on low- and high-stakes reading and math achievement by grade level. Second, we explore whether the content of early literacy policies, particularly the inclusion of a mandatory third-grade retention component and having a so-called "comprehensive" policy, drives any of the effects on student achievement. Third-grade retention is frequently considered the most controversial element of early literacy policies (Burns, 2016; Fiester, 2013; Starr, 2019). Having a comprehensive early literacy policy means that the state mandates third-grade retention and includes provisions for literacy professional development for teachers, literacy screening and parental notification for K-3 students, and a host of interventions for students who need additional support (ExcelinEd, 2021). Third, we examine socioeconomic and racial achievement gaps on high-stakes assessments.

We find that having an early literacy policy improves students' performance on third- through fifth-grade high-stakes reading assessments, with the largest effects appearing in third grade and diminishing after that, even after students exposed to the policy age into later grades. We also provide evidence of positive spillovers of early literacy policies to high-stakes math achievement. However, we find limited evidence of improvements in low-stakes reading or math test scores—except in states with comprehensive policies. The positive effects of early literacy policies are substantially larger in states with comprehensive policies or third-grade retention requirements than in states without. Lastly, we find suggestive evidence that early literacy policies reduce socioeconomic and racial high-stakes reading test-score gaps, though these estimates are noisy and small in magnitude. Our results are robust to different estimators and pass a series of falsification tests. Altogether, these findings suggest that early literacy policies successfully

improve student reading scores on high-stakes stakes assessments, but these improvements may not extend to overall literacy learning, as measured by low-stakes test scores.

Our paper proceeds as follows. Section 2 overviews early literacy policies and the literature surrounding them. We then describe our data and methods in Section 3. In Section 4, we describe our results. Section 5 presents additional robustness checks. We discuss the policy implications of our findings and conclude in Section 6.

2. Background and Relevant Literature

While there is well-established research confirming the importance of early literacy skills for students' cognitive and noncognitive outcomes (e.g., Cunningham & Stanovich, 1997; Fiester & Smith, 2010; Hernandez, 2011; Sparks et al., 2014), there have been few studies assessing the net impacts of early literacy policies on student achievement—despite the ubiquity of these policies. National attention to early literacy in the late 1990s and early 2000s (Pearson et al., 2020; U.S. Department of Education, 2015) spurred states to implement their own early literacy policies, and by 2021, 41 states and the District of Columbia had enacted one of these policies (ExcelinEd, 2021). Figure 1 shows the number of states that passed early literacy policies by year and indicates that most states passed their policies in the 2010s. Early literacy policies vary by state in their content and intensity, though they share many components. ExcelinEd (2021) defines 16 possible early literacy policy components under four broad categories: (1) supports for teachers and policy implementation, (2) assessment and parent notification, (3) instruction and intervention, and (4) retention and intensive intervention. Table 1 shows the number of states that have each of the 16 components.

We focus on three key distinctions between states. The first is between states that have any early literacy policy (i.e., a policy requiring at least one of the 16 components listed in Table 1) and those that do not. Forty-one states and D.C. have an early literacy policy. The second distinction is between states that have third-grade retention mandates and those that do not. Twenty-two states require retention. The third distinction is between states that have "comprehensive" early literacy policies and states that do not. We follow ExcelinEd (2021) in defining a comprehensive policy as having all 16 policy components listed in Table 1. Twelve states have a comprehensive early literacy policy. Because comprehensive policies include all policy components, they are a subset of states with retention components.

While most states have early literacy policies, there is limited research assessing them. Greene and Winters (2004, 2006, 2007, 2009) have conducted multiple quasi-experimental analyses of Florida's early literacy policy, primarily relying on discontinuities created by the retention component of the policy. They find positive short-term effects on students' reading achievement on Florida's state assessment (and positive spillover effects on third-grade math achievement) but that these effects dissipate as students progress through school (Greene & Winters, 2004, 2006, 2007, 2009; Schwerdt et al., 2017). Moreover, studies using

regression discontinuity designs provide only local treatment effect estimates that are valid near the discontinuity. Florida's fourth-grade reading scores on NAEP also improved under its early literacy policy, though it is unclear whether retention or other policy components drove this improvement (Duke et al., 2014). There is also an ongoing effort to evaluate Michigan's early literacy policy (Strunk et al., 2021, 2022). Early results from this evaluation indicate that third-grade student reading achievement on the state assessment has improved and that educators attribute achievement gains to the literacy supports outlined in the policy (Strunk et al., 2021). Notably, researchers rely on high-stakes assessments to evaluate Michigan's and Florida's policies.

We contribute to the existing literature by providing the first known national assessment of early literacy policies and examining their effect on student achievement on high- and low-stakes assessments. Evaluating high- and low-stakes assessment outcomes is important. While increases in high-stakes test scores could indicate improvements in student literacy skills, they could also reflect other phenomena, such as "teaching to the test" (Popham, 2001) or changes in tested content. Because states' high-stakes summative assessments often relate to states' standards and curriculum, the material by states' high-stakes tests may dictate the early literacy skills taught to students. Further, states with third-grade retention requirements identify retention-eligible students using their state's high-stakes assessment, meaning these states have an extra incentive to focus on tested skills. Thus, if early literacy policies also positively affect low-stakes test scores, this could provide further evidence of improved literacy skills—not just better performance on a high-stakes assessment.

We also examine the effects of early literacy policies across grade levels. How early literacy policies affect student achievement across grade levels will be determined by whether the effects are lasting or transient and whether the policies increase or decrease resources or interventions in other grades. If early literacy policies improve students' literacy skills in grades K-3 as intended, we would expect to see improved reading scores in subsequent grade levels as students who are affected by the policy progress through school. On the other hand, if early literacy policies improve test performance through an emphasis on teaching to the test or test-specific skills, or if students' learning improvements do not persist, then there may be little impact in subsequent grades. Additionally, early literacy policies may affect resources and teacher quality in grades outside of K-3. For instance, districts might adopt new literacy curricula for all elementary schools, or literacy professional development might be available to all teachers. In these cases, literacy achievement in higher grades could improve immediately following the adoption of an early literacy policy. Alternatively, districts may focus resources on grades K-3, diverting resources away from later elementary or middle school grade levels. In these cases, literacy achievement in higher grades may decrease following the passage of an early literacy policy.

Early literacy policies may also improve student achievement outside of reading. Improved literacy skills might improve students' test-taking ability since they can better comprehend test questions. However, the increased funding for literacy interventions induced by early literacy policies might also reduce funds available for other subjects. We focus on math achievement because previous research has found that Florida's early literacy policy positively affected students' test performance in this area (Greene & Winters, 2004; Winters & Greene, 2012).

In addition to examining the overall effects of having any early literacy policy on students' reading and math achievement, we explore the differential effects of having two early literacy policy compositions: (1) one that includes third-grade retention and (2) comprehensive policies. Third-grade retention is frequently the most controversial component of early literacy policies, and there is mixed evidence on the effectiveness of grade retention for short- and longer-term cognitive and noncognitive outcomes (e.g., Eren et al., 2022; Greene & Winters, 2004, 2006, 2007; Holmes & Matthews, 1984; Hong & Yu, 2007; Jacob, 2005; Jacob & Lefgren, 2004; Jimerson, 2001; Lorence, 2014; McCombs et al., 2009; Nagaoka & Roderick, 2004; Roderick & Nagaoka, Jenny, 2005; Schwerdt et al., 2017; Weiss et al., 2018; Winters & Greene, 2012; Wu et al., 2010). Comprehensive early literacy policies include other components beyond retention (e.g., professional development for teachers, interventions for students who need additional literacy support), many of which have been found to have positive effects (see Strunk et al., [2021] for a thorough review). Prior research has found positive effects of early literacy policies in states with comprehensive policies. Thus, exploring whether policy composition matters will have important implications for policymakers.

Lastly, we examine the effect of early literacy policies on socioeconomic and racial achievement gaps. These gaps are well established in the literature (Fryer & Levitt, 2004; S. F. Reardon et al., 2019), but whether they increase, decrease, or remain the same under early literacy policies depends both on the allocation of early literacy policies' interventions and resources and the heterogeneity of their efficacy. Even if policymakers target interventions toward historically underserved students, if the interventions are less effective at improving these students' outcomes, achievement gaps might not improve. Further, early literacy policies with third-grade retention requirements have been criticized for their potential for discrimination and disparate outcomes (Greene & Winters, 2009; Licalsi et al., 2019; Livingston & Livingston, 2002; Valencia & Villarreal, 2004), leading to potential increases in test-score gaps.

This paper provides the first known national assessment of the effects of early literacy policies on student reading achievement and gaps in both high- and low-stakes assessments and examines whether policy composition drives any of these effects. Given that most states have early literacy policies, it is vital to provide policymakers with evidence of these policies' effects on student outcomes and socioeconomic and racial achievement. This information will be necessary to policymakers in states currently implementing or considering adopting such policies.

3. Data and Methods

3.1. *Data*

There is no single dataset linking early literacy policies to high- or low-stakes assessment data. We, therefore, combine data from three publicly available sources to assess whether early literacy policies improve student reading achievement: (1) ExcelinEd's (2021) early literacy policy database, (2) the Stanford Education Data Archive (SEDA), and (3) the National Assessment of Educational Progress (NAEP). The ExcelinEd data allow us to determine which states have an early literacy policy, when they first implemented it, and its content. The SEDA enables us to evaluate early literacy policies' effect on high-stakes assessments. The NAEP allows us to assess the effect of early literacy policies on a low-stakes assessment. By comparing the impact of these policies on high- and low-stakes assessments, we can discern whether early literacy policies only improve students' assessment performance or human capital more generally.

3.1.1. *Early Literacy Policy Data*

ExcelinEd's (2021) early literacy policy database documents each state's early literacy policy (if they had one), including the composition of the policy (i.e., the different components included) and the year implemented. Table 1 details the various policy components. States sometimes amended or otherwise altered their early literacy policy, leading to multiple passage dates. We say a state is "treated" if they ever implemented any early literacy policy before the last year of test-score availability (2018 for the SEDA and 2019 for the NAEP). States that never implemented an early literacy policy before the last year of test-score availability form the "never-treated" comparison group.

Variation in policy composition and changes in early literacy policies over time introduces complexity to our analyses. ExcelinEd's (2021) data do not indicate the years in which states adopted specific policy components. We, therefore, always use the earliest adoption date to define treatment status. Moreover, the order in which states adopted policy components is unclear, as ExcelinEd's (2021) dataset does not indicate the year in which policy components were added. For this reason, we focus on the three discernable distinctions between states' policies we describe in Section 2: (1) any early literacy policy, (2) comprehensive early literacy policies, and (3) early literacy policies with third-grade retention requirements.

Figure 2 illustrates the geographic variation in states with different types of early literacy policies. States without shading do not have an early literacy policy by 2021. States shaded without diagonal or hash marks have an early literacy policy that does not require retention and is not comprehensive. Shaded states diagonal marks require third-grade retention for underperforming students but lack some other policy components such that the policy is not comprehensive. Shaded states with hash marks have comprehensive

early literacy policies. We see that states with comprehensive policies and retention mandates are dispersed across the United States.

3.1.2. *SEDA Data*

From the SEDA, we obtain state-level average third- through eighth-grade reading and math achievement scores as well as non-economically disadvantaged-economically disadvantaged and White-Black test-score gaps from 2009 to 2018 (S. Reardon et al., 2021). SEDA test scores are derived from state-level high-stakes testing data from assessments that states use for accountability purposes. Because each state can use a different assessment, SEDA norms test scores to the NAEP to allow cross-state comparisons. The assessments underlying the SEDA data also generally determine third-grade retention eligibility under early literacy policies. Thus, we may see effects on high-stakes tests if educators respond to the threat of retention and work to prepare students specifically for these exams or if the exams themselves change, but we might not see these effects translate to low-stakes tests like the NAEP.

We also use state-by-year-level data from the SEDA regarding average demographics and macroeconomic conditions. These data include average grade-level enrollment and the proportion of students by urbanicity, race and ethnicity, economic disadvantage status, English learner status, and special education status. Additionally, the data includes the proportion of households with a single mother and a parent with at least a bachelor's degree. The data further includes macroeconomic conditions, including log median income, household poverty rates, and unemployment rates. We use these data to compare the characteristics of states with and without early literacy policies and to test whether differential trends in these characteristics explain the impacts of early literacy policies on student achievement.

3.1.3. *NAEP Data*

The NAEP is a low-stakes assessment not tied to any accountability requirements or early literacy policy mandates (e.g., third-grade retention) (The Nation's Report Card, n.d.). It is designed to be comparable over time and across states. Thus, avoiding issues of accountability-related high-stakes test-score inflation (Jacob, 2005). We use state-level average fourth- and eighth-grade reading and math scale scores from the 2003 to 2019 NAEP assessments to assess early literacy policies.[4] Although these policies aim to improve literacy in kindergarten through third grade, third-grade students do not take the NAEP. Further, if states implement policies to improve early literacy in kindergarten through third grade, these efforts should be reflected in fourth-grade reading achievement, although direct effects may lag by one year. We also examine eighth-grade NAEP reading scores to determine whether any effects of early literacy policies are sustained as students progress through school and fourth- and eighth-grade math scores to assess potential spillover effects.

[4] The NAEP is administered approximately every two years. 2019 is the most recent year available.

3.1.4. *Summary Statistics*

Table 2 shows the number of states and state-by-year observations available for our analysis by grade, subject, outcome, and data source. Our analysis sample excludes any states that adopted an early literacy policy before our analysis period because these states cannot provide any identifying variation (i.e., we don't see a *change* in policy that might inform estimates of the impact of policy adoption).[5] For the SEDA, overall average reading scores include 45 states because the remaining states adopted an early literacy policy before 2009, and we exclude them from the sample. The state-by-year observations range from 391 to 400 state-year observations. For SEDA overall math scores, the sample again includes 45 states in third- through sixth grade but only 44 states in seventh and eighth grades. The sample sizes range from 352 to 400 state-year observations. While the same number of states have test-score gap information available, states less frequently reported this data. Thus, these outcomes have consistently fewer state-year observations than the overall average scores.[6] The NAEP data are complete for 47 states adopting early literacy policies after 2003 in each available year, forming a balanced panel of 423 state-year observations.

Table 3 shows summary statistics for the outcome variables of interest in fourth and eighth grade—the two grade levels for which we have data from both SEDA and NAEP.[7] The first six columns show state average reading and math achievement scores from the SEDA data, including overall and by treatment status. SEDA test scores are measured in standard deviations from the national mean. The following six columns present state average NAEP reading and math scale scores, again overall and by treatment status. We additionally include White-Black and non-economically disadvantaged-economically disadvantaged student test-score gaps, which are available in the SEDA but not the NAEP.

Overall, these summary statistics suggest that states with early literacy policies had substantially higher average reading and math scores on high-stakes state assessments than states that never passed early literacy policies from 2009 to 2018 (as evidenced by the SEDA data). However, treated states also had larger socioeconomic and racial test-score gaps. Additionally, low-stakes NAEP scores were nearly identical in treated and never-treated states during this period (as evidenced by the NAEP data).

Table 4 shows summary statistics for student demographic characteristics. These data are only available in the SEDA from 2009 to 2019. Treated states have higher proportions of urban, suburban, and Black students, while never-treated states have higher proportions of town, rural, White, and Asian students. Regarding macroeconomic conditions, treated states have higher unemployment rates and higher proportions of single-mother households, but never-treated states have higher proportions of economically disadvantaged students. Overall, this suggests that states with early literacy policies may serve somewhat

[5] Before 2009 for the SEDA data and before 2003 for the NAEP data.
[6] We examine the possibility that missing test-score data correlates with early literacy policies in Section 5.4.
[7] The full table, including all grade levels, is available in Appendix Table A1.

higher proportions of historically underserved students, but the two groups are not substantially different. Moreover, our empirical strategy controls for time-invariant differences across states using state fixed-effects. We also test for and find no evidence of differential trends in state characteristics in Section 5.1.

3.2. *Methods*

We leverage the differential timing of states' initial early literacy policy passage as well as differences in policy content as sources of quasi-experimental variation. We employ an event-study identification strategy to estimate the effects of early literacy policies on student achievement. We use data on states' average test scores before and after the passage of an early literacy policy, as well as a comparison group of states that do not (or did not yet) have early literacy policies.

Importantly, states passed their policies at different times (see Figure 1). A growing literature has identified issues with using two-way fixed-effects and OLS to estimate the average effect of the treatment on the treated (ATT) in difference-in-differences and event-study research designs in settings with staggered adoption and treatment effect heterogeneity (e.g., Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Roth et al., 2022; Sun & Abraham, 2021). The former is inevitable, and the latter is likely in our setting. Accordingly, we estimate our models using the method proposed by Callaway & Sant'Anna (2021), wherein we estimate group-time average treatment effects for each group of states g treated in a given year at time t. The group-time average treatment effect $ATT(g,t)$ is identified nonparametrically by computing

$$\mathrm{ATT}(\mathrm{g},\mathrm{t}) = (\mathrm{E}[\mathrm{Y}(\mathrm{g})_{\mathrm{t}}] - E[Y(NT)_t]) - \left(E\left[Y(g)_{g-\delta-1}\right] - E\left[Y(NT)_{g-\delta-1}\right]\right) \qquad (1)$$

Where $\delta \in \mathbb{N}_0$. Here $E[\mathrm{Y}(\mathrm{g})_{\mathrm{t}}]$ represents average test scores in year t for states adopting an early literacy policy in year g. $E[Y(NT)_t]$ represents average test scores in year t for never-treated states (states that had not passed an early literacy policy by the end of the analysis period). $E\left[Y(g)_{g-\delta-1}\right]$ represents average test scores in states treated in year g in year $g-\delta-1$, which is some reference year before g. $E\left[Y(NT)_{g-\delta-1}\right]$ is the average test scores in the same reference year, $g-\delta-1$, for never-treated states. If $t \geq g$, meaning after the adoption of an early literacy policy, then $\delta = 0$, and the reference year is the year just before policy adoption, $g-1$. If $t < g$, meaning pre-early literacy policy, then $\delta = g-t$ and the reference period is the prior year, $t-1$.

We estimate $ATT(\mathrm{g},\mathrm{t})$ for each combination of treatment year g and calendar year t using the Sant'Anna & Zhao (2020) doubly robust difference-in-differences estimator based on stabilized inverse probability weighting and ordinary least squares. This estimator uses state fixed-effects, and we include no additional covariates. These $ATT(\mathrm{g},\mathrm{t})$ estimates represent all the possible 2x2 difference-in-differences where the comparison groups are never-treated states and the reference year is the prior year ($t-1$) pre-treatment and the year just before treatment ($g-1$) post-treatment. We then aggregate the $ATT(\mathrm{g},\mathrm{t})$

estimates to form event-study estimates weighted by the number of observations in a year relative to treatment. The aggregation is done as follows:

$$\theta_e^{event} = \sum_{g=1}^{\mathrm{T}} \mathbf{1}(g + e \leq \mathrm{T})\, ATT(g, g+e) P(G = g | G + e \leq \mathrm{T}, NT = 0) \qquad (2)$$

Where e represents the years relative to treatment, g represents the year of treatment (and $g = 1$ indicates the first treatment year), and T is the last year of data availability. θ_e^{event} is the event-study coefficient estimate e years relative to treatment, and it equals a weighted average of the $ATT(g,t)$ estimates such that $t = g + e$ (the ATT estimate for e years relative to treatment in year g). The weights, $P(G = g | G + e \leq \mathrm{T}, NT = 0)$, equal the probability that a state first adopted a policy in year g, given a state was ever treated and that data are available. We perform this aggregation for five years before and after treatment in the SEDA, and eight years before treatment and nine years after treatment in the NAEP.[8]

Because of our reference-group selection, we can estimate event-study estimates for all years relative to treatment, unlike traditional event-study models, which must omit at least one year. Furthermore, pre-policy event-study coefficients are interpreted differently than a standard two-way fixed effects event study where g-1 is the common reference year. Instead of positive (negative) pre-policy coefficients representing negative (positive) pre-policy trends, the Callaway & Sant'Anna (2021) pre-policy event-study coefficients have the opposite interpretation. This difference is because the reference year for the pre-policy Callaway & Sant'Anna (2021) event-study estimates is t-1, the year before the estimated effect, rather than g-1, the year before treatment onset. Thus, we interpret positive pre-treatment coefficient estimates as increases in the outcome relative to never treated states before adopting an early literacy policy and negative pre-treatment coefficients as relative decreases. We cluster standard errors at the state level.

3.2.1 *Identifying Assumptions*

For our estimates to have a causal interpretation, our data must satisfy two key identifying assumptions. The first is the parallel trends assumption, which states that the average test scores of the states first treated in year g and the never-treated states would have followed parallel paths in the absence of the treatment. We test the plausibility of the parallel trends assumption in two ways. First, our event study allows us to test for differential test-score trends before early literacy policy implementation. Significant pre-policy event-study coefficients could indicate a violation of this assumption. Second, we test for differential trends in state-level average demographic characteristics and macroeconomic conditions. In this exercise, we estimate our event-study models with various demographic characteristics and macroeconomic conditions as the outcome. Importantly, we would not expect early literacy laws to affect these variables. Significant estimates in these falsification tests would suggest a violation of parallel trends. Overall, we find no

[8] Because the NAEP is administered every two years, we bin the event-study estimates into two-year bins such that we have five bins before treatment (starting 7 to 8 years pre-policy) and five bins after treatment (ending 8 to 9 years post-policy).

substantial evidence of violations of the parallel trends assumption. We present the detailed results of these falsification tests in Section 5.1.

The second assumption is that states do not anticipate their early literacy policy in ways that would alter their average test scores. This assumption might be violated if, for example, educators know that an early literacy policy – perhaps with a retention component – will be passed in the next year, so they change their behavior to improve students' literacy performance in anticipation. Anticipation of the policy can be detected in the pre-policy event-study estimates by looking at the event study estimate of the impact of early literacy laws in the year before adoption.

4. Results

4.1. *Impacts on High-and Low-Stakes Reading and Math Achievement, by Grade Level*

We begin by examining whether early literacy policies impact high- and low-stakes reading and math test scores by grade level. The SEDA data provide high-stakes test scores, and low-stakes test scores come from the NAEP data.

4.1.1. *High-Stakes Reading Test Scores*

Figure 3 shows event-study estimates of the ATT of early literacy policies on high-stakes reading scores relative to a state's policy enactment year for each grade level three through eight. The x-axis represents years relative to treatment, where zero is the year in which the state passed its early literacy policy, negative values represent years until treatment (i.e., the number of years before the state passed its policy), and positive values represent years since treatment (i.e., the number of years since the state passed its policy).[9] Panels A through F generally show no evidence of differential pre-policy reading-score trends. The one exception is in Panel F, which shows that reading scores in treated states increased slightly relative to never-treated states four- and five-years pre-policy. However, there are no significant differential pre-policy trends in the years closer to the adoption of early literacy policies (t=-1 through t=-3). These systematically insignificant pre-policy event-study coefficient estimates support the parallel trends assumption.

Figure 3, Panel A shows that third-grade high-stakes reading scores begin to increase immediately following the passage of an early literacy policy. The effect's magnitude increases during the first three years, from roughly a 0.02 to 0.06 standard deviation increase in high-stakes reading test scores in states that passed early literacy policies relative to those that did not. The third-grade reading effect appears to diminish marginally four and five years post-policy, but the differences in effect magnitude are not statistically significant. Together, these results indicate that early literacy policies persistently improve third-grade reading achievement up to five years after implementation.

[9] Appendix Table A2 presents detailed coefficient estimates.

Panels B through F show the estimates for grades four through eight. Panels B and C suggest that fourth- and fifth-grade high-stakes reading performance increased immediately following the introduction of an early literacy policy, suggesting that these policies significantly improve high-stakes reading scores in the elementary years. The impacts follow similar dynamics as in third grade, increasing in magnitude in the first three years post-policy then stabilizing (or marginally diminishing) in years four and five. However, the effect's magnitude diminishes from third to fifth grade. The policies' effects of a 0.06 standard deviation increase in third-grade reading scores three years post-policy decreases to a 0.04 standard deviation increase in fifth grade three years post-policy.

Further, there is little evidence of delayed impacts expected as cohorts move through grades over time. For instance, we might expect the 0.06 standard deviation increase we see in third grade three years post-policy to appear in fifth grade five years post-policy as students age into those grades. However, we only see slight and insignificant increases in fifth grade at that time. These results support the hypothesis that early literacy policies improve reading achievement by increasing educational resources or resource quality—particularly in third grade but also throughout elementary school. However, these test-performance gains might not translate into sustained achievement growth.

4.1.2. *High-Stakes Math Test Scores*

Figure 4 shows event-study estimates of the ATT of early literacy policies on high-stakes math scores and parallels Figure 3.[10] Panel A shows that, like reading scores, third-grade math scores begin to increase immediately following the passage of an early literacy policy. The magnitude of this effect increases during the first three years from roughly a 0.02 to 0.1 standard deviation increase in states that passed early literacy policies relative to those that did not. Also, like reading scores, the third-grade math effect diminishes marginally four- and five-years post-policy and as students progress through school (the latter shown in Panels B through F).

4.1.3. *Low-Stakes Reading Test Scores*

Figure 5, Panels A and B present event-study estimates of early literacy policies' effects on state-level average fourth- and eighth-grade reading NAEP scale scores. We use NAEP scores from 2001 to 2019. During this period, the NAEP was administered every two years. For this reason, we create two-year bins of years relative to treatment in the NAEP event-study analysis, so our results cover seven to eight years pre-policy to eight to nine years post-policy.

As with the high-stakes test outcomes in SEDA, we find no evidence of statistically significant pre-policy trends in either fourth- or eighth-grade reading. However, unlike the high-stakes test outcomes, we find little evidence of changes in reading scores *after* the passage of an early literacy policy. Our estimates

[10] Appendix Table A3 presents detailed coefficient estimates.

are insignificant in every grade, subject, and relative year. The maximum of the 95% confidence intervals would imply a four-scale point increase in fourth- and eighth-grade reading. The national standard deviation is roughly 36 scale score points. Thus, these estimates indicate a 0.11 standard deviation increase in NAEP scores. While this is an economically meaningful effect size and around the upper bound of the 95% confidence interval of the effect on high-stakes SEDA reading scores (Figure 3), the estimates are not statistically significant at any conventional level.

4.1.4. *Low-Stakes Math Test Scores*

Figure 5, Panels C and D present event-study estimates of early literacy policies' effects on state-level average fourth- and eighth-grade math NAEP scale scores. Like NAEP reading scores, we find little evidence of changes in math scores after the passage of an early literacy policy. The maximum 95% confidence intervals would imply a five scale-score point increase in eighth-grade math performance and a four-point increase in fourth grade, representing a 0.13 and 0.11 standard deviation increase. Again, the estimates are not statistically significant.[11]

4.2. *The Role of Early Literacy Policy Components*

This section explores the role of different early literacy policy components in improving students' high-stakes reading scores. We focus our analysis on two discernable distinctions between states' early literacy policy components in our dataset: (1) policies with a third-grade retention requirement compared to those without one, and (2) states with comprehensive early literacy policies compared to those without comprehensive policies. The impact of having a retention component measures both the effect of retention and the threat of retention in the future. The same is true with respect to comprehensive early literacy policies. Our estimates here thus reflect an intent-to-treat effect.

4.2.1. *Third-Grade Retention*

We begin by examining states with policies that include retention. Figure 6 compares estimates of the impacts of early literacy policies in states with retention components to those without them.[12] Since we only detect significant overall impacts in third through fifth grades above, we focus on these grades in these analyses.[13] In the first column, we restrict the sample of treated states to those that have or will have retention components, and in the second column, we restrict the treated states to those that never have a retention component. In both columns, never-treated states form the comparison group. States with a retention component see significant and persistent increases in high-stakes reading scores in all grades, third through fifth. The magnitude is quite similar across grades, with the impact increasing through roughly

[11] We present detailed coefficient estimates from Figure 5 in Appendix Table A4

[12] Appendix Table A5 contains coefficient estimates.

[13] In results available on request, we examine grades six through eight and continue to find no evidence of significant effects.

three years post-policy to around a 0.06 standard deviation increase. This estimate is in line with the overall impact estimated above, suggesting that states with retention components explain essentially all of the average effects of early literacy policies on high-stakes reading scores. In contrast, there is no consistent evidence that third- through fifth-grade high-stakes reading scores increased in states without a retention component.

Figure 7 examines the impact of retention components on high-stakes math test scores.[14] Again, we see significant test score gains in states with retention requirements. High-stakes math test scores increased significantly and persistently in third through fifth grades. On the other hand, we find no statistically significant impacts in states without a retention component.

Figures 8 and 9 compare the effects of early literacy policies with retention requirements (Column (1)) to those without (Column (2)) on low-stakes NAEP reading and math assessments in fourth and eighth grades, respectively.[15] Across fourth- and eighth-grade math and reading NAEP scores, we find no consistent evidence of statistically significant increases in low-stakes tests in states with or without early literacy policy retention components. Overall, our results provide compelling evidence that early literacy policies with retention requirements improve high-stakes reading scores more than those without, but these gains may not translate into increases in low-stakes test scores.

4.2.2. *Comprehensive Early Literacy Policies*

Figure 10 shows the effects of comprehensive and non-comprehensive policies on high-stakes reading scores.[16] States with comprehensive early literacy policies experience substantial and sustained increases in high-stakes reading scores following the adoption of their policies. Third and fourth grades see a roughly 0.1 standard deviation increase four- and five-years post-policy, while fifth-grade scores increase by around 0.08 standard deviations. In all cases, this is substantially larger than the overall effect, suggesting that comprehensive policies play an important role in explaining reading-score increases from early literacy policies. Further, unlike the overall results, in states with comprehensive policies, the effects continue to increase several years post-policy passage. We find limited evidence of increased high-stakes reading achievement in states without comprehensive policies.

We compare the impacts of comprehensive and non-comprehensive early literacy policies on high-stakes test scores in Figure 11.[17] We see substantial but noisy increases in high-stakes math scores in third through fifth grade following the introduction of a comprehensive early literacy policy. In contrast, we find

[14] Coefficient estimates are in Appendix Table A6.
[15] Detailed coefficient estimates are in Appendix Tables 7 and 8, respectively.
[16] Appendix Table A9 contains detailed coefficient estimates.
[17] Coefficient estimates are in Appendix Table A10.

no evidence of statistically significant high-stakes math gains in states with non-comprehensive early literacy policies.

Figure 12 examines the effects of early literacy policies on fourth-grade reading and math NAEP scale scores in states with comprehensive early literacy policies (Column (1)) and non-comprehensive policies (Column (2)).[18] Here, we see significant increases in low-stakes fourth-grade reading scores in states with comprehensive policies. Similar to the high-stakes reading results, the impacts of comprehensive early literacy policies on fourth-grade NAEP scores phase in over time until we see a statistically significant, roughly five scale-score point increase six to seven years after the passage of the policy. We also see suggestive evidence of an increase in fourth-grade math scores associated with a comprehensive policy but no statistically significant changes in low-stakes fourth-grade scores in states with non-comprehensive policies. However, we do note some evidence of marginally significant pre-policy decreases in 4th-grade NAEP reading and math in states with comprehensive early literacy policies relative to never-treated states. While these pre-trend coefficients are only marginally significant and several years before a policy's passage, we interpret these results cautiously.

Figure 13 shows parallel results for eighth grade, where we again find evidence of statistically significant increases in low-stakes reading scores associated with a comprehensive policy.[19] This time, introducing a comprehensive early literacy policy is associated with a roughly four scale-score point increase eight to nine years post-adoption. We also see a significant increase in math scores of approximately the same magnitude at around the same time. The timing of these eighth-grade impacts is roughly consistent with when fourth-grade students positively affected by comprehensive early literacy policies would age into eighth grade. Overall, our estimates suggest that states with comprehensive early literacy policies experience larger high- and low-stakes test scores gains than states with non-comprehensive policies.

4.3. *Impacts of High-Stakes Socioeconomic Achievement Gaps*

This section examines the impacts of early literacy policies on socioeconomic and racial test-score gaps. The results so far show a substantial increase in high-stakes reading scores in third through fifth grades following the adoption of an early literacy policy, as well as increases in low-stakes scores in states with third-grade retention requirements and with comprehensive policies. However, these estimates consider only effects on average test scores and ignore treatment-effect heterogeneity. We examine high-stakes reading test score gaps between non-economically disadvantaged and economically disadvantaged students and between White and Black students.

[18] See Appendix Table A11 for detailed coefficient estimates.

[19] Appendix Table A12 has detailed coefficient estimates.

Figure 14 shows non-economically disadvantaged-economically disadvantaged test-score gaps on high-stakes reading assessments in Column (1).[20] We again focus on third through fifth grade. Across these grades, we find no systematic evidence of statistically significant increases in the non-economically disadvantaged-economically disadvantaged test score gap following the introduction of an early literacy policy.

When we look at White-Black achievement gaps in Column (2), we again find little evidence of systematic changes. The coefficient estimates are generally negative, though not statistically significant in most cases. Altogether, we find no systematic evidence that early literacy policies increase achievement gaps and some suggestive evidence that achievement gaps might shrink following the introduction of any early literacy policy.

We also examine how early literacy policies requiring retention impact socioeconomic and racial test-score gaps, but we find no substantial evidence that early literacy policies requiring third-grade retention increase the non-economically disadvantaged-economically disadvantaged or White-Black gaps in high-stakes reading scores. In fact, the post-policy event-study estimates are generally negative and at times indicate a statistically significant decrease in test-score gaps, implying that these gaps are diminishing in some cases.

Altogether, we find no compelling evidence that early literacy policies increase socioeconomic or racial test-score gaps, even when third-grade retention is required. Our estimates are generally noisy but suggest that gaps significantly diminish in some cases following the introduction of an early literacy policy.[21]

5. Additional Robustness Checks

5.1. *Falsification Tests*

Our event-study estimates provide no evidence of systematic differential pre-policy trends in high- or low-stakes reading or math scores across treated and never-treated states, supporting the parallel trends assumption. In this section, we provide further evidence that early literacy policies drive our findings by reporting the results of a series of falsification tests wherein we estimate the impact of early literacy policies on state average demographic characteristics and macroeconomic conditions. We examine the proportion of a state's population that is non-White, the proportion living in urban areas, the proportion with a bachelor's degree or higher, and log median total household income. We use these variables as the outcomes in our event-study model. While early literacy policies could impact earnings and educational attainment in the long run, we do not expect early literacy policies to have any meaningful impact on these state

[20] Appendix Table A13 contains coefficient estimates.

[21] We also examine how early literacy policies requiring retention impact socioeconomic and racial test-score gaps in Appendix Figure A1. We find no substantial evidence that early literacy policies requiring third-grade retention increase the non-economically disadvantaged-economically disadvantaged or White-Black gaps in high-stakes reading scores.

characteristics during our analysis period. Any significant estimates would suggest differential trends in these characteristics across treated and never-treated states might explain changes in test scores we attribute to early literacy policies.

Appendix Figure A2, Panels A through D, show the results of the falsification tests. We do not find evidence of significant differential changes in the proportion of non-White students, the proportion of a state's population in urban areas, the proportion of a state's population who have a Bachelor's degree or higher, or a state's log median household income before or after the introduction of an early literacy policy, providing additional support for the interpretation of our results as the causal effects of early literacy policies.

5.2. *Contemporaneous Policy Shocks*

We also test the robustness of our findings to contemporaneous education policies that may also have affected reading and math achievement. In 2010, around the same time that many states began adopting early literacy policies (see Figure 1), the U.S. Department of Education's RTTT program began providing billions of dollars in education grants to states with winning applications (The White House, 2017). The federal government awarded 19 states a total of over $4 billion in 2010 and 2011 over three phases to implement innovative reform plans for assessments, data systems, educator support, and interventions to turn around low-performing schools (The White House, 2017). RTTT incentivized states to adopt standards such as the Common Core by giving priority in the grant competition. The Common Core included standards in both ELA and math (Common Core State Standards Initiative, n.d.), so we may expect receiving RTTT funds to affect achievement in these subject areas.

Appendix Figure A3 compares the effect of early literacy policies on reading achievement separately for states that did not receive RTTT funding (Column (1)) to those that did (Column (2)). In third grade, states that did not receive RTTT funding and those that did both experience an increase in reading scores following the passage of an early literacy policy. However, the increases in states without RTTT funding were more often statistically significant. In fourth through fifth grade, RTTT states experience slightly larger gains in reading scores following the passage of an early literacy policy than their non-RTTT counterparts. Still, these are only statistically significant three years after policy passage. Overall, these results suggest that RTTT funding is not driving the effect of early literacy policies on reading achievement but is an important complement given the larger magnitude of impacts found in RTTT states.

Appendix Figure A4 similarly compares the effect of early literacy policies on math achievement separately for states that did not receive RTTT funding (Column (1)) to those that did (Column (2)). These results suggest a more substantial RTTT effect. In all grades, states with early literacy policies that did not receive RTTT funding experienced no significant gains in math scores, but those that did receive RTTT funding experienced significant math-score improvements. This suggests that the positive spillover effects

of early literacy policies on math scores may be driven by the receipt of RTTT funding, suggesting these math achievement gains may not be directly attributable to early literacy policies. Therefore, we interpret the math results with caution.

5.3. *Permutation Tests*

Next, we examine whether the documented impacts of early literacy policies on student achievement could result from random chance. We examine the likelihood that we would detect treatment effect estimates of a similar magnitude if we randomly assigned treatment status to states by randomly assigning treatment status 500 times, keeping the time distribution of adoption constant (i.e., the same number of states adopt an early literacy policy each year as were adopted in reality). Then, we estimate the event-study model and collect the coefficient estimates. We construct empirical p-values by computing the proportion of placebo treatment effects greater than the actual treatment effect and empirical confidence intervals by calculating the 5^{th} and 95^{th} percentiles of the placebo treatment effect distribution.

Appendix Figure A5 presents the results of this exercise for high-stakes third- through eighth-grade reading SEDA scores. The solid lines represent our event-study estimates with states' actual treatment status, as shown in Figure 3, Panel A. The shaded areas represent the empirical 95% confidence intervals. Event-study estimates outside of those confidence intervals are larger in magnitude than the 95^{th} or 5^{th} percentile of placebo event study estimates, suggesting a statistically significant estimate. Our results are robust to these empirical confidence intervals. In the years leading up to the adoption of an early literacy policy, we see no evidence of significant differential changes in high-stakes reading scores, except in eighth grade. Immediately following the introduction of an early literacy policy, we see that our estimated increases in reading scores are significantly beyond what one might expect if treatment were randomly assigned in third through fifth grade, but still inside the empirical confidence intervals in sixth through eighth grade, consistent with our prior findings.

5.4. *Missing Data Analysis*

Another concern regarding our analyses' validity is differential attrition from the data. We have complete NAEP data for overall average test scores throughout the analysis period. However, some states did not report their high-stakes standardized test scores in specific years, and these outcomes are missing in the SEDA data. If the probability that a state is missing test score data in the SEDA correlates with having an early literacy policy, our estimates may be biased. This bias would be particularly concerning if states were less likely to report their scores in years with poor performance. This attrition would bias our estimates upward, potentially leading us to conclude that early literacy policies improve achievement when no such effect exists.

We test for differential attrition by estimating our preferred event-study model with an indicator for missing test score data in a given year. Appendix Figure A6 shows the results of this exercise for missing

SEDA reading scores for third through eighth grades. We find no evidence that states adopting early literacy laws were more or less likely to have missing test score information than never-treated states, suggesting that differential attrition is not driving our findings.

5.5 *Alternative Estimators*

Finally, we test the robustness of our main estimates using the Callaway and Sant'Anna (2021) estimator to two alternative difference-in-differences estimators: (1) two-way fixed-effects using ordinary least squares (OLS) and (2) the interaction-weighted event-study estimator proposed by Sun and Abraham (2021). Both estimation methods model the event study as:

$$Y_{st} = \alpha + \sum_{j=[-5,5]} \delta_j \, \mathbf{1}(t - E_s = j) + \theta_s + \tau_t + \epsilon_{st} \quad (3)$$

Again, the outcome, Y_{st}, is the average test score in state *s* in year *t*. E_s is the year in which state *s* first implemented an early literacy policy. $\mathbf{1}(t - E_s = j)$ are indicators that equal one when the time relative to treatment $(t - E_s)$ equals *j*. We omit *t-1* as the reference year. θ_s are state fixed effects and τ_t are year fixed effects. ϵ_{st} are state-year specific idiosyncratic errors. The primary coefficients of interest are δ_j, which measure how different average test scores are *j* years after (or before for negative values of *j*) the introduction of an early literacy policy relative to states without policies in the same period.

First, we estimate the model using OLS. Recent econometrics research has identified issues with this method in the presence of staggered adoption and treatment effect heterogeneity (e.g., Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Roth et al., 2022; Sun & Abraham, 2021). In particular, the event-study parameters represent weighted averages of all possible 2x2 difference-in-differences where the weights may be negative. However, the extent to which this is an issue is not clear a priori. We present the OLS estimates of Equation (3) for SEDA reading scores in Appendix Figure A7. These results are entirely consistent with our preferred estimation strategy. We find no evidence of statistically significant pre-policy trends and evidence of significant increases in elementary reading scores of roughly 0.05 SDs.

Next, we estimate Equation (3) using the interaction-weighted event-study estimator from Sun and Abraham (2021). This method begins by estimating Equation (3) using OLS, then reweights the event-study parameters such that the weights are non-negative. We present these estimates in Appendix Figure A8. Again, we find broadly similar results to our preferred estimation method. This makes sense given the Sun and Abraham (2021) estimator is simply a special case and aggregation of the Callaway and Sant'Anna (2021) estimator in models without covariates. Altogether, these exercises provide that our choice of an estimator is not driving the findings.

6. Discussion and Conclusion

This paper leverages differences in the adoption and content of early literacy policies across states and over time in an event-study research design to identify the causal effects of early literacy policies on high-

and low-stakes reading and math test scores. Our findings add to the extant literature on early literacy policies in several ways. First, we examine the impact on both high- and low-stakes assessments, allowing us to differentiate between test-taking and human capital improvements. Second, we leverage cross-state variation in the content of early literacy policies to examine the effects of specific components. Finally, our national setting provides external validity to our findings.

Our results provide compelling evidence that early literacy policies improve high-stakes achievement. Having any early literacy policy improves the high-stakes reading scores of elementary school students, with the largest impact in third grade. We also find substantial increases in high-stakes math performance following the introduction of an early literacy policy, but further analysis suggests that these math gains are due to other education policy changes, including RTTT and Common Core. However, we find little evidence of significant increases in low-stakes reading test scores, except for moderate gains in states with the most comprehensive policies. We similarly find that the largest impacts on high-stakes test scores are in states with comprehensive early literacy policies and those with third-grade retention requirements. Finally, the high-stakes test-score gains appear to be distributed equitably, with potential decreases in socioeconomic and racial test-score gaps.

Altogether, our results suggest that early literacy policies provide potentially superficial gains in reading achievement and inform policy by demonstrating the importance of using low-stakes test scores to evaluate the impacts of education policies. A focus on high-stakes test scores can mislead policymakers if there are policy-induced changes in high-stakes test-taking or tested materials that do not reflect changes in human capital. Examining low-stakes tests can provide a better measure of changes in actual learning. Our results also highlight the importance of the content and incentives of early literacy policies. The best evidence for significant increases in both low- and high-stakes test scores comes from states with the most comprehensive early literacy policies, including those with third-grade retention requirements. Altogether, these results suggest that the full set of interventions available under early literacy policies is important in improving literacy achievement and skills.

Our study faces several limitations that point to avenues for future research. First, while we provide evidence that comprehensive early literacy policies and retention mandates play an important role in improving state summative assessment scores, we cannot examine the mechanisms through which these policy components improve outcomes. Further research into the implementation of these policy components is vital to understanding how early literacy policies operate. A second limitation is that we focus only on short-run test-score outcomes. However, prior work has established the importance of early literacy skills in determining noncognitive outcomes and long-run student success (Cunningham & Stanovich, 1997; Fiester & Smith, 2010; Hernandez, 2011; Sparks et al., 2014). To fully understand the benefits of early literacy policies, it is important to enumerate their noncognitive and long-term impacts. Finally, this study

does not examine the costs associated with early literacy policies. While we show substantial short-run high-stakes test score gains, policymakers must weigh all benefits, in the short and long run, against the costs of early literacy policies.

References

Burns, J. (2016, March 10). Should struggling readers be retained in 3rd grade? [Michigan State University College of Education]. *Green & Write*. https://education.msu.edu/green-and-write/2016/should-struggling-readers-be-retained-in-3rd-grade/

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*, 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Common Core State Standards Initiative. (n.d.). *About the standards*. Common Core State Standards Initiative. Retrieved July 11, 2022, from http://www.corestandards.org/about-the-standards/

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945.

Duke, N. K., Moje, E. B., & Palincsar, A. S. (2014). *Three IRA Literacy Research Panel Members Comment on Third Grade Retention Laws*. *47*(1), 6.

Eren, O., Lovenheim, M. F., & Mocan, H. N. (2022). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. *Journal of Labor Economics*, *40*(2).

ExcelinEd. (2021). *Comprehensive early literacy policy: State-by-state analysis of fundamental principles*. ExcelinEd. https://excelined.org/wp-content/uploads/2021/10/ExcelinEd_PolicyToolkit_EarlyLiteracy_StatebyStateAnalysis_2021.pdf

Fiester, L. (2013). *Early warning confirmed: A research update on third-grade reading*. The Annie E. Casey Foundation.

Fiester, L., & Smith, R. (2010). *Early warning! Why reading by the end of third grade matters*. The Annie E. Casey Foundation.

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, *86*(2), 447–464. https://doi.org/10.1162/003465304323031049

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2021.03.014

Greene, J. P., & Winters, M. A. (2004). *An evaluation of Florida's program to end social promotion* (No. 7; Education Working Paper). Center for Civic Innovation at the Manhattan Institute.

Greene, J. P., & Winters, M. A. (2006). *Getting farther ahead by staying behind: A second-year evaluation of Florida's policy to end social promotion* (No. 49; Civic Report). Center for Civic Innovation at the Manhattan Institute.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, *2*(4), 319–342.

Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, *28*, 135–142.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. The Annie E. Casey Foundation.

Holmes, C. T., & Matthews, K. M. (1984). The effects of non promotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research*, *52*(4), 225–236. https://doi.org/10.3102%2F00346543054002225

Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, *29*(4), 239–261. https://doi.org/10.3102%2F0162373707309073

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5), 761–796. https://doi.org/10.1016/j.jpubeco.2004.08.004

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, *86*(1), 226–244.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437.

Licalsi, C., Ozek, U., & Figlio, D. (2019). The uneven implementation of universal school policies: Maternal education and Florida's mandatory grade retention policy. *Education Finance and Policy*, *14*(3), 383–413. https://doi.org/10.1162/edfp_a_00252

Livingston, D. R., & Livingston, S. M. (2002). Failing Georgia: The case against the ban on social promotion. *Education Policy Analysis Archives*, *10*(49).

Lorence, J. (2014). Third-grade retention and reading achievement in Texas: A nine year panel study. *Social Science Research*, *48*, 1–19.

McCombs, J. S., Kirby, S. N., & Mariano, L. T. (2009). *Ending social promotion without leaving children behind: The case of New York City* (1st ed.). RAND Corporation. http://www.jstor.org/stable/10.7249/mg894nycdoe

Nagaoka, J., & Roderick, M. (2004). *Ending social promotion: The effects of retention* (Charting Reform in Chicago Series). Consortium on Chicago School Research.

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (2020). *Reaping the Rewards of the Reading for Understanding Initiative*. National Academy of Education. https://naeducation.org/wp-content/uploads/2020/06/Reaping-the-Rewards-of-the-Reading-for-Understanding-Initiative.pdf

Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, *58*(6), 16–20.

Reardon, S. F., Weathers, E., Fahle, E., Jang, H., & Kalogrides, D. (2019). *Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps*. https://vtechworks.lib.vt.edu/handle/10919/97804

Reardon, S., Kalogrides, D., Ho, A., Shear, B., Fahle, E., Jang, H., & Chavez, B. (2021). *Stanford Education Data Archive (SEDA)* [English]. Stanford University. https://exhibits.stanford.edu/data/catalog/db586ns4974

Roderick, M., & Nagaoka, Jenny. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, *27*(4), 309–340.

Roth, J., Sant'Anna, P. H. C., Bikinski, A., & Poe, J. (2022). *What's trending in difference-in-differences? A synthesis of the recent econometrics literature*. https://jonathandroth.github.io/assets/files/DiD_Review_Paper.pdf

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/j.jeconom.2020.06.003

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, *152*, 154–169.

Sparks, R. L., Patton, J., & Murdoch, A. (2014). Early reading success and its relationship to reading achievement and reading volume: Replication of "10 years later." *Read Writ*, *27*, 189–211.

Starr, A. (2019, July 13). States are ratcheting up reading expectations for 3rd-graders. *National Public Radio*. https://www.npr.org/2019/07/13/741156019/states-are-ratcheting-up-reading-expectations-for-3rd-graders

Strunk, K. O., Wright, T. S., Kilbride, T., Zhu, Q., Cummings, A., West, J., Turner, M., & De Voto, C. (2021). *Michigan's Read by Grade Three Law: Year one report*. Education Policy Innovation Collaborative. https://epicedpolicy.org/wp-content/uploads/2021/03/Year_One_RBG3_Report.pdf

Strunk, K. O., Wright, T. S., Westall, J., Zhu, Q., Kilbride, T., Cummings, A., Utter, A., & Mavrogordato, M. (2022). *Michigan's Read by Grade Three Law: Year two report*. Education Policy Innovation Collaborative. https://epicedpolicy.org/wp-content/uploads/2022/02/RBG3_Rpt_Yr2_Feb2022.pdf

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*, 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

The Nation's Report Card. (n.d.). *NAEP data explorer*. Retrieved January 26, 2022, from https://www.nationsreportcard.gov/ndecore/xplore/NDE

The White House. (2017). *Race to the Top*. The White House, President Barack Obama. https://obamawhitehouse.archives.gov/issues/education/k-12/race-to-the-top

U.S. Department of Education. (2015, December 7). *Reading First* [Program Home Page]. U.S. Department of Education. https://www2.ed.gov/programs/readingfirst/index.html

Valencia, R. R., & Villarreal, B. J. (2004). Texas' second wave of high-stakes testing: Anti-social promotion legislation, grade retention, and adverse impact on minorities. In *Leaving children behind: How "Texas-style" accountability fails Latino youth*. SUNY Press.

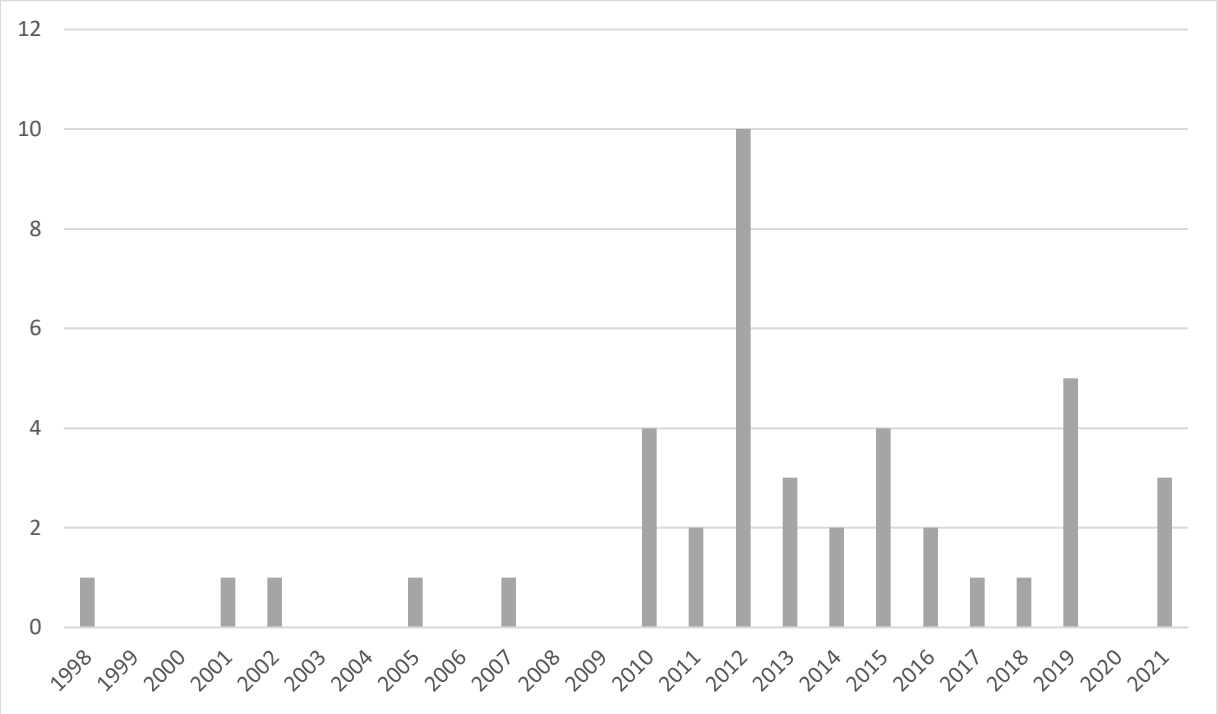
Weiss, S., Stallings, D. T., & Porter, S. (2018). *Is Read to Achieve making the grade? An assessment of North Carolina's elementary reading proficiency initiative*. North Carolina State University College of Education.

Winters, M. A., & Greene, J. P. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy*, *7*(3), 305–330.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology*, *102*(1), 135–152. https://psycnet.apa.org/doi/10.1037/a0016664

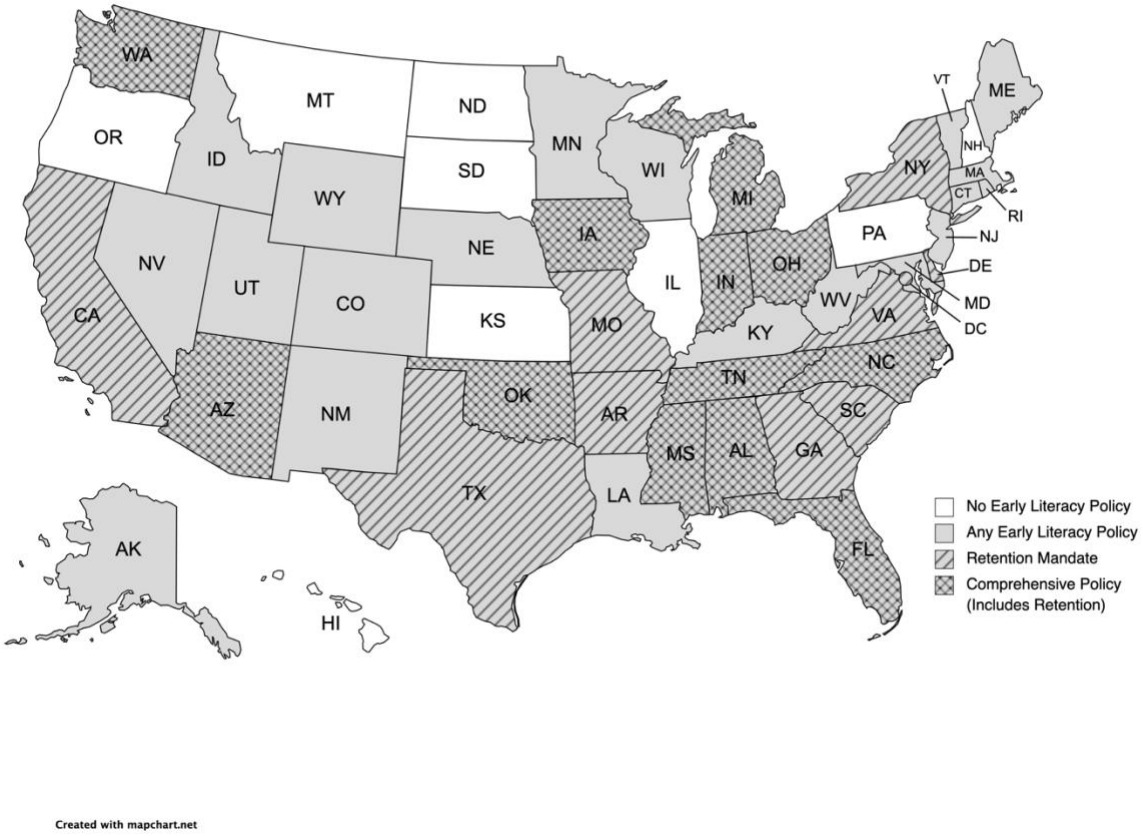
Figures and Tables

Figure 1. Number of States Passing Early Literacy Policies, by Year



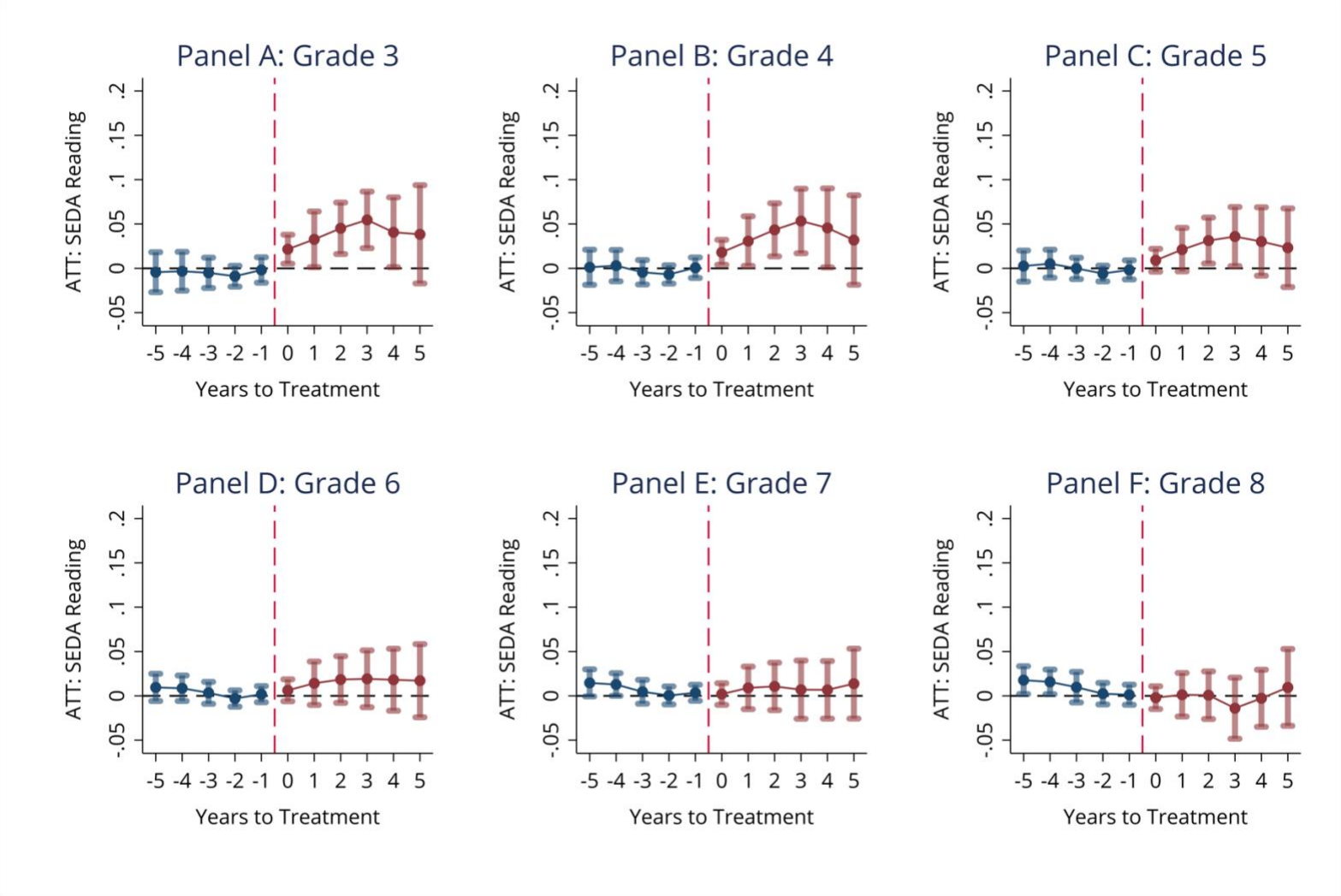
Note. Data derived from ExcelinEd (2021). The figure shows the number of states passing their first early literacy policy in the year given on the x-axis.

Figure 2. Map of States' Early Literacy Policy Components



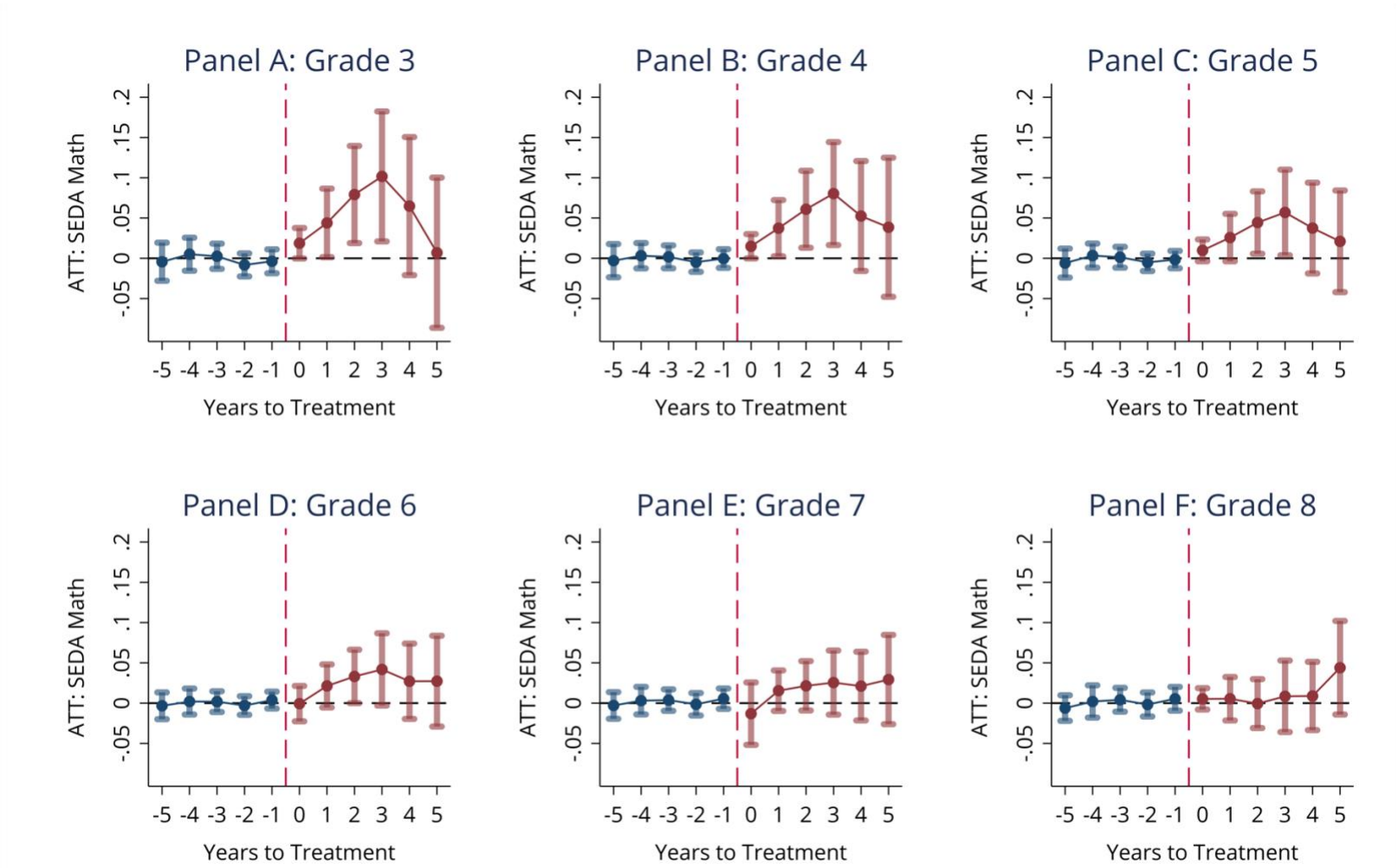
Note: Data are derived from ExcelInEd (2021).

Figure 3. Early Literacy Policies and High-Stakes Reading Scores



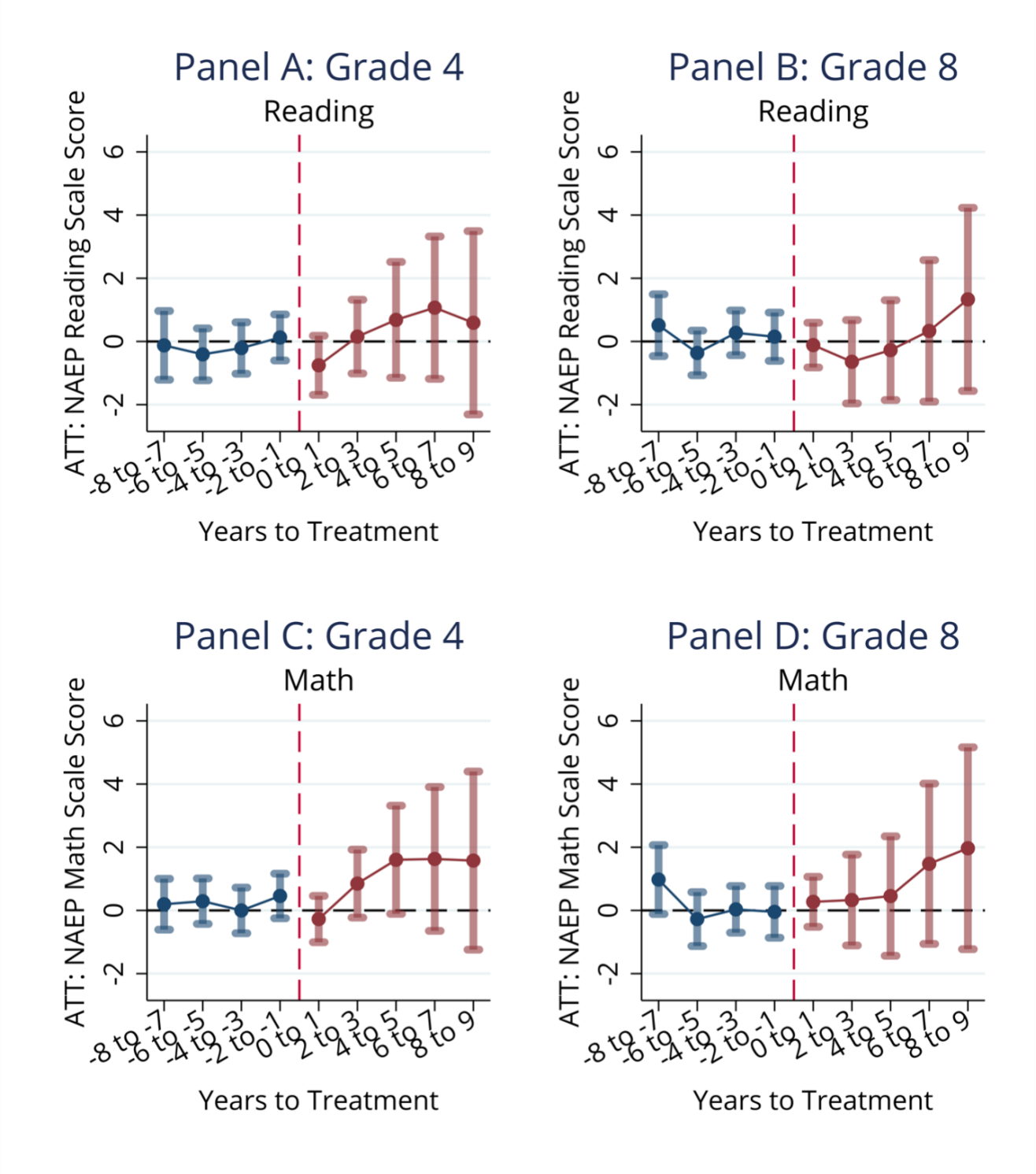
Note: Data are from overall average SEDA reading scores, 2009-2018. The sample sizes range from 391 to 400 state-year observations. Detailed sample sizes can be found in Table 2. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A2.

Figure 4. Early Literacy Laws and High-Stakes Math Achievement



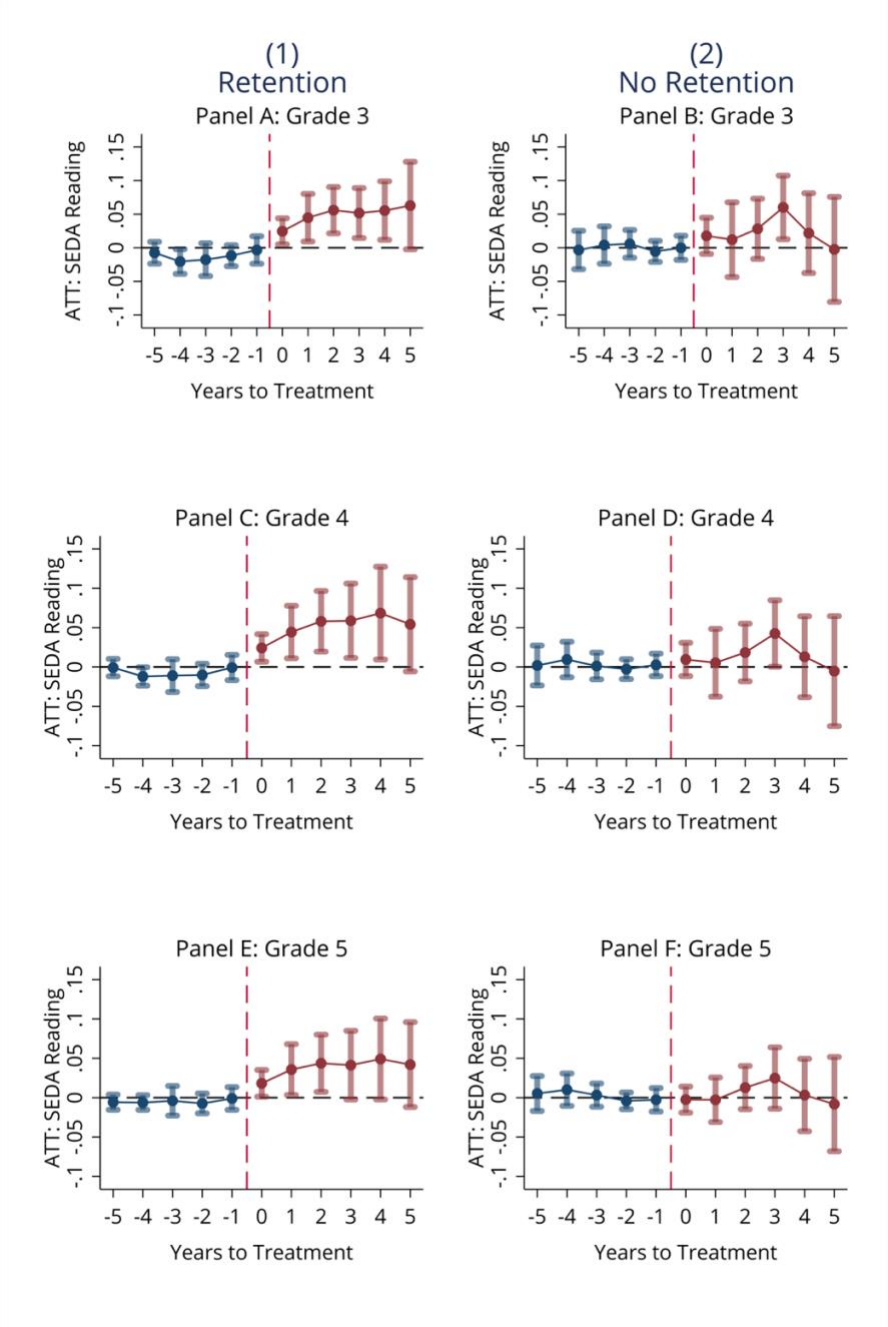
Note: Data are from overall average SEDA math scores, 2009-2018. The sample sizes range from 391 to 400 state-year observations. Detailed sample sizes can be found in Table 2. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A3.

Figure 5. Early Literacy Policies and Low-Stakes Achievement



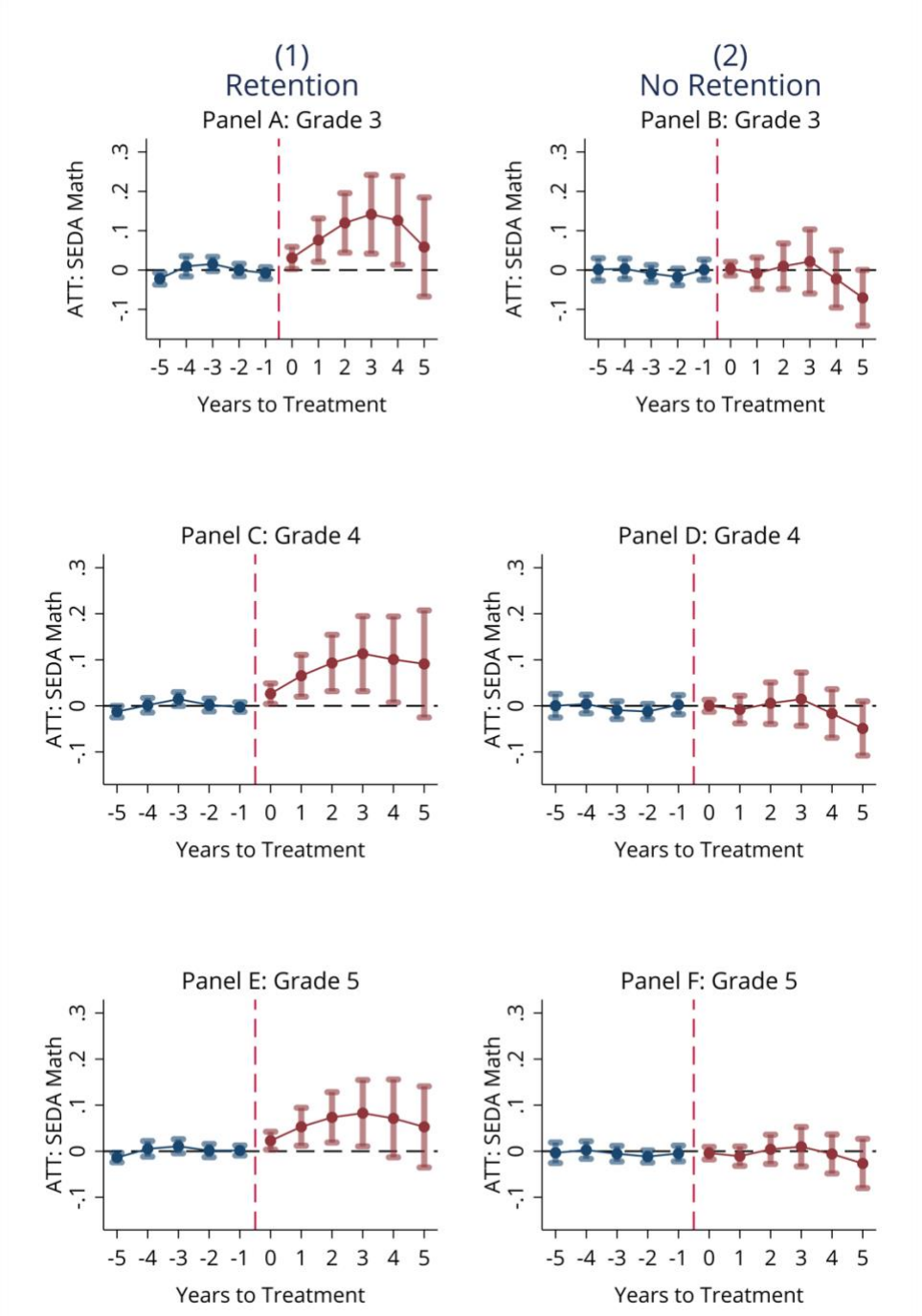
Note: Data are from overall average NAEP fourth- and eighth-grade reading and math scale scores, 2003-2019. The sample sizes are 423 state-year observations. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A4.

Figure 6. Impacts of Early Literacy Policies with Retention Requirements on High-Stakes Reading Scores



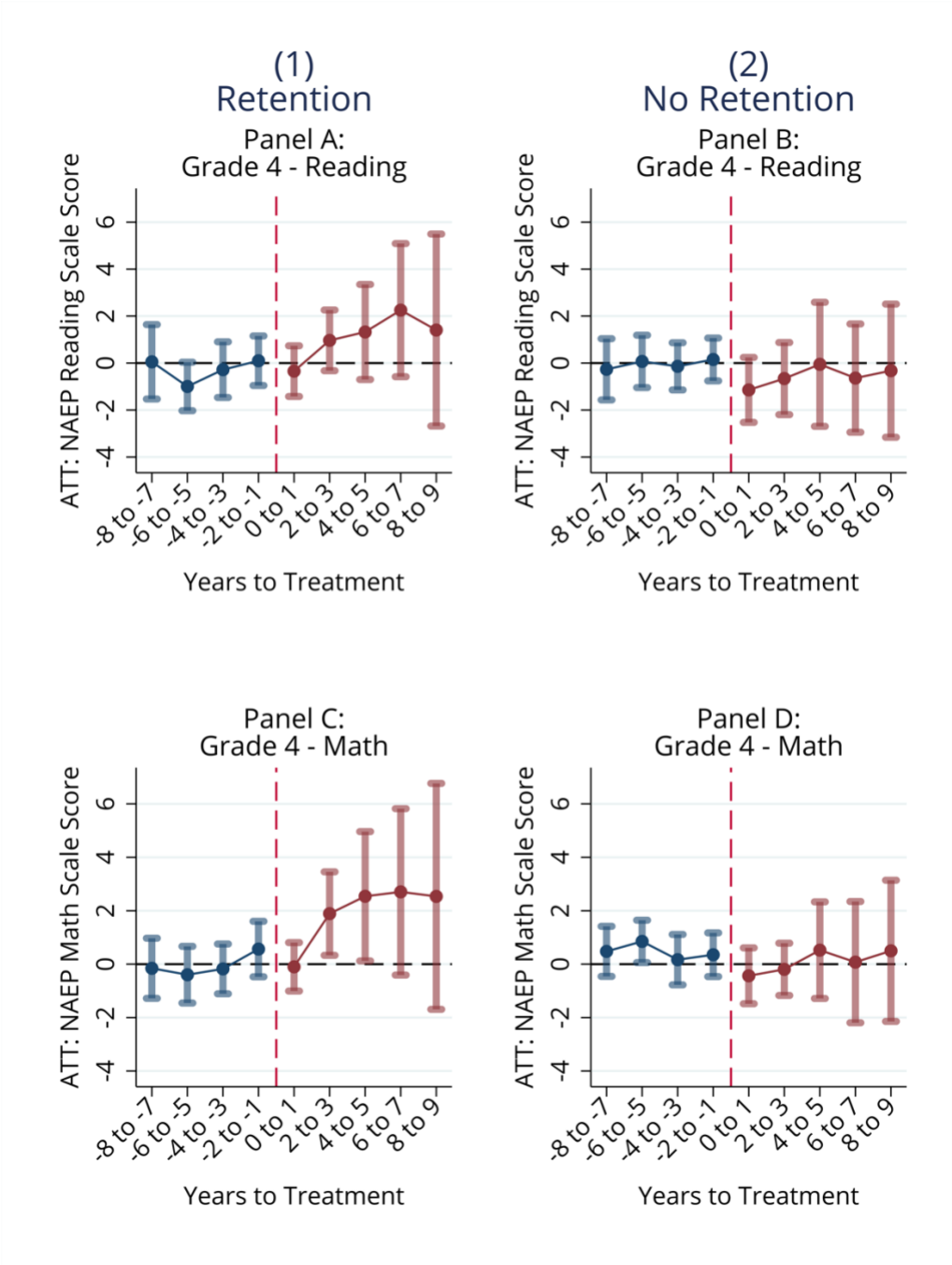
Note: Data are from overall average SEDA reading scores, 2009-2018. Column (1) includes states with retention requirements and never-treated states. Column (2) includes states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A5.

Figure 7. Impacts of Early Literacy Policies with Retention Requirements on High-Stakes Math Scores



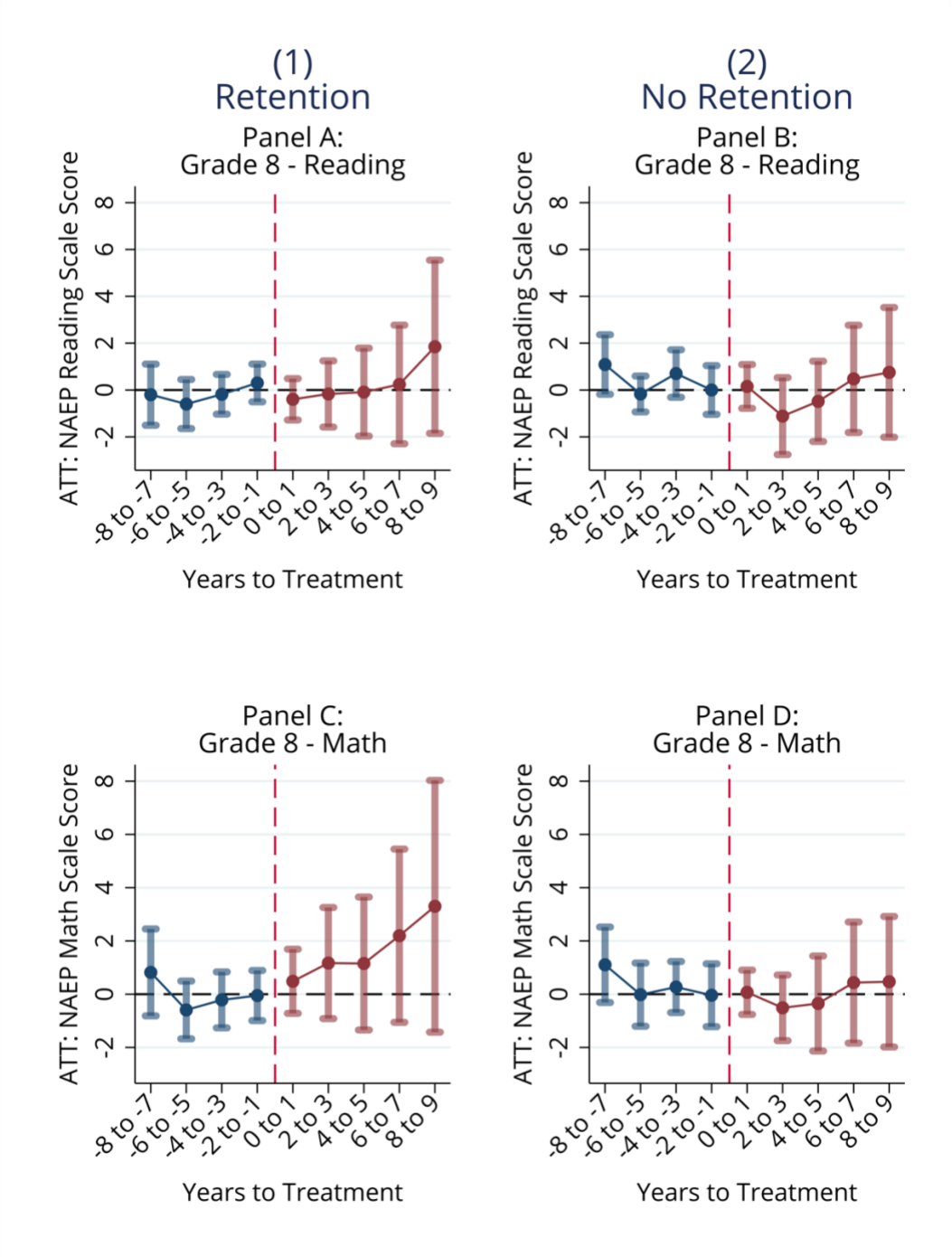
Note: Data are from overall average SEDA math scores, 2009-2018. Column (1) includes states with retention requirements and never-treated states. Column (2) includes states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A6.

Figure 8. Impacts of Early Literacy Policies with Retention Requirements on Low-Stakes Test Score (4th Grade)



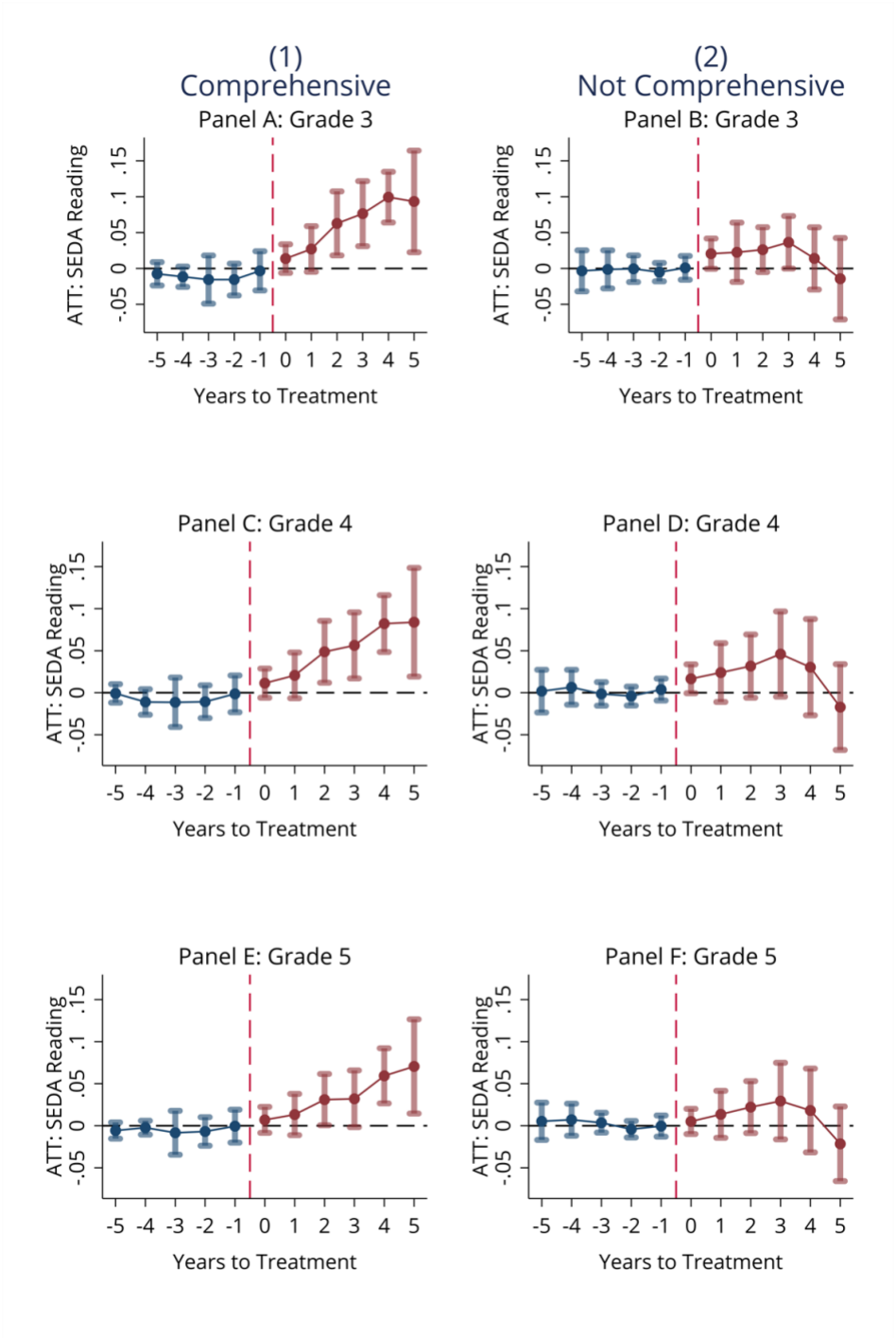
Note: Data are from overall average NAEP fourth-grade reading and math scale scores, 2003-2019. Column (1) includes states with retention requirements and never-treated states. Column (2) includes states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A7.

Figure 9. Impacts of Early Literacy Policies with Retention Requirements on Low-Stakes Test Score (8th Grade)



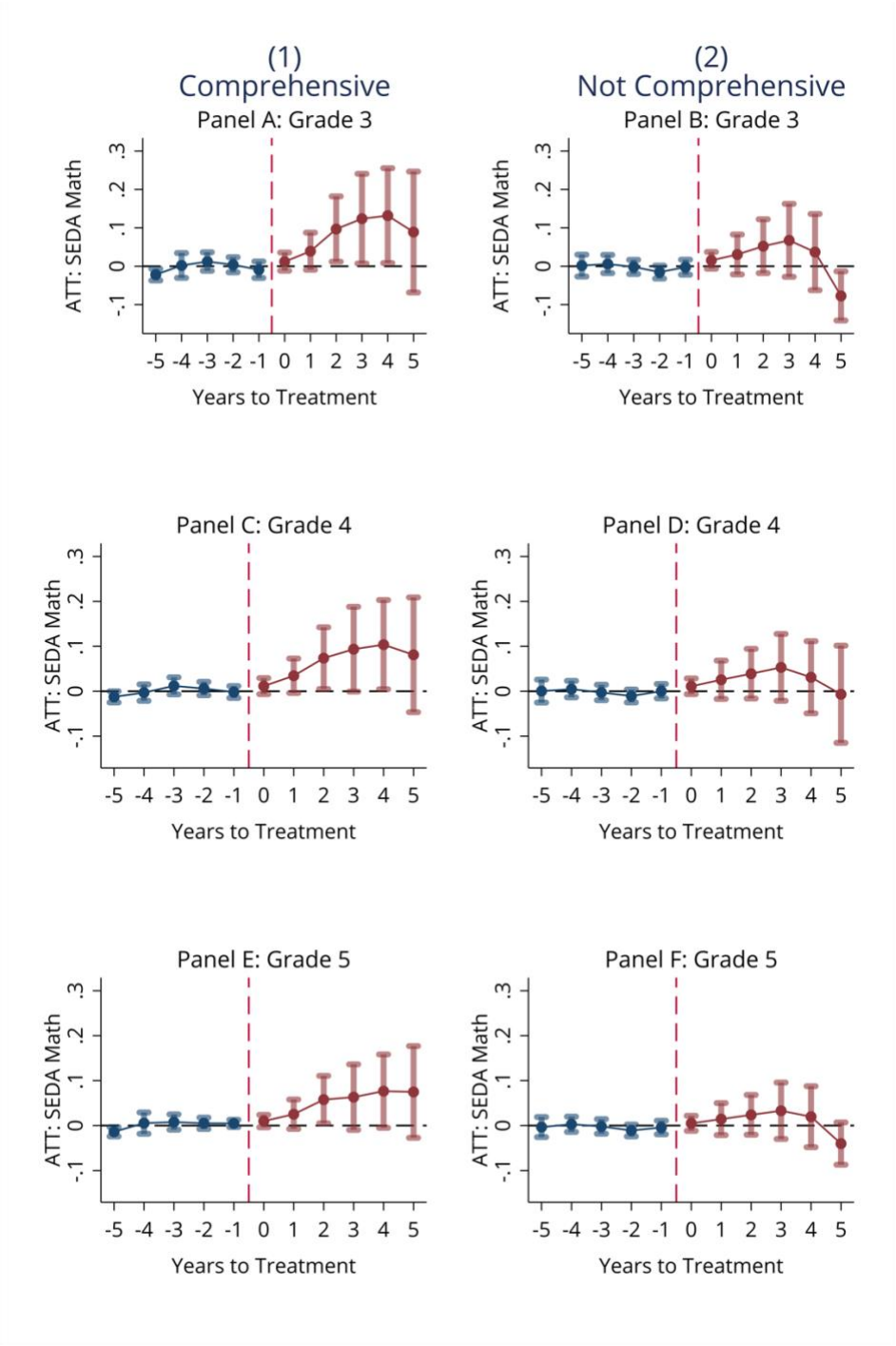
Note: Data are from overall average NAEP eighth-grade reading and math scale scores, 2003-2019. Column (1) includes states with retention requirements and never-treated states. Column (2) includes states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A8.

Figure 10. Impacts of Comprehensive Early Literacy Policies on High-Stakes Reading Scores



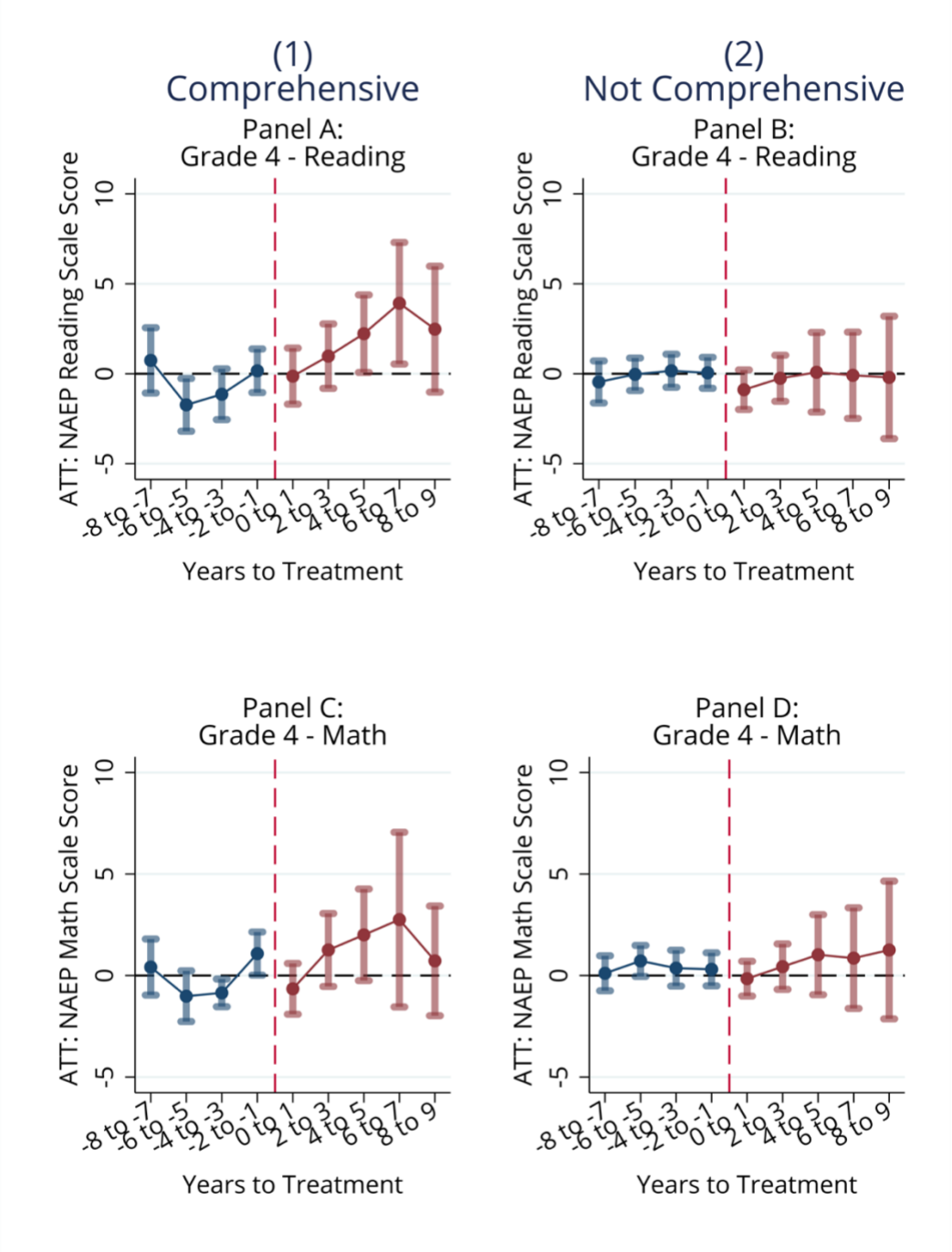
Note: Data are from overall average SEDA reading scores, 2009-2018. Column (1) includes states with comprehensive policies and never-treated states. Column (2) includes states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A9.

Figure 11. Impacts of Comprehensive Early Literacy Policies on High-Stakes Math Scores



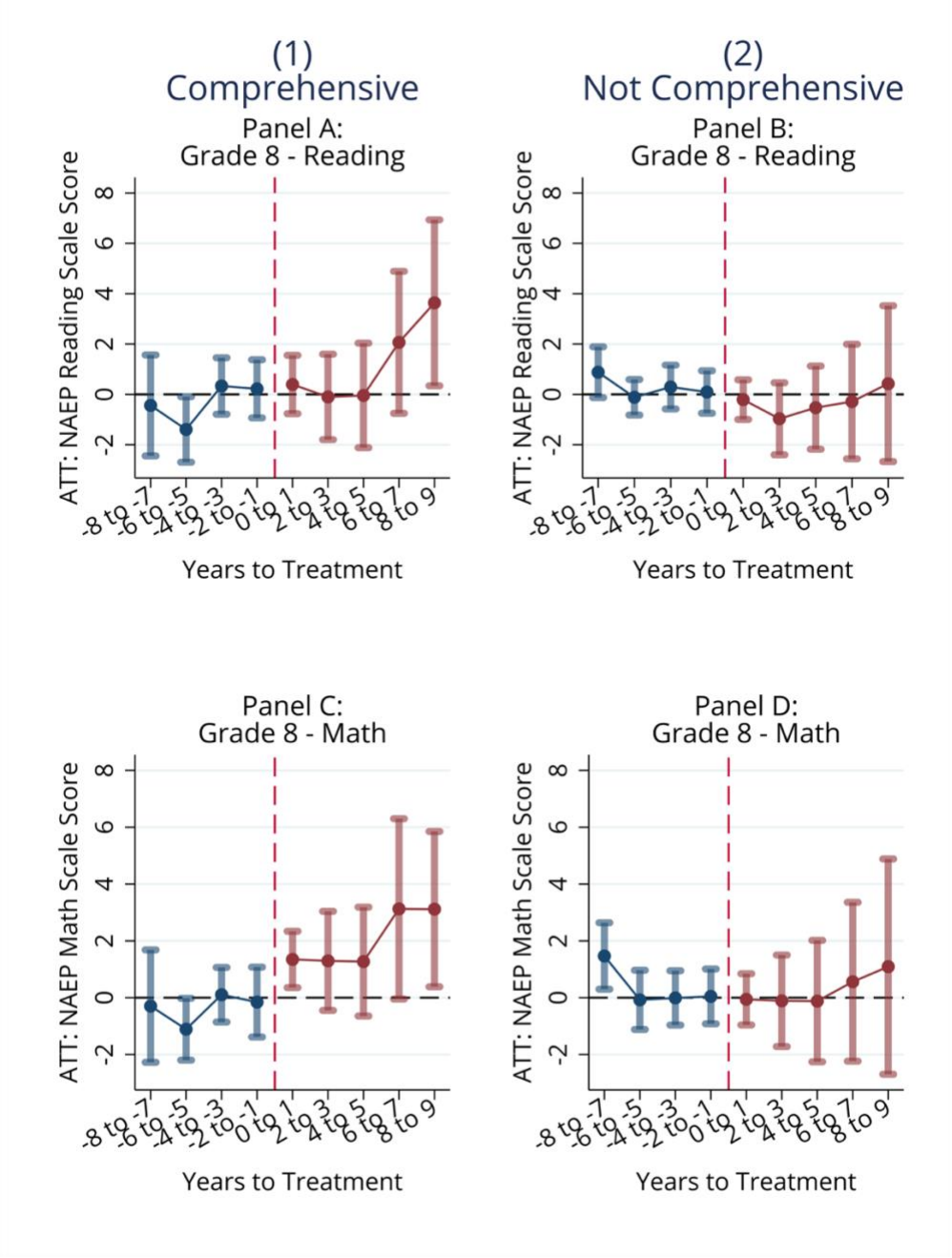
Note: Data are from overall average SEDA math scores, 2009-2018. Column (1) includes states with comprehensive policies and never-treated states. Column (2) includes states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A10.

Figure 12. Comprehensive Early Literacy Policies and Low-Stakes Achievement (4th Grade)



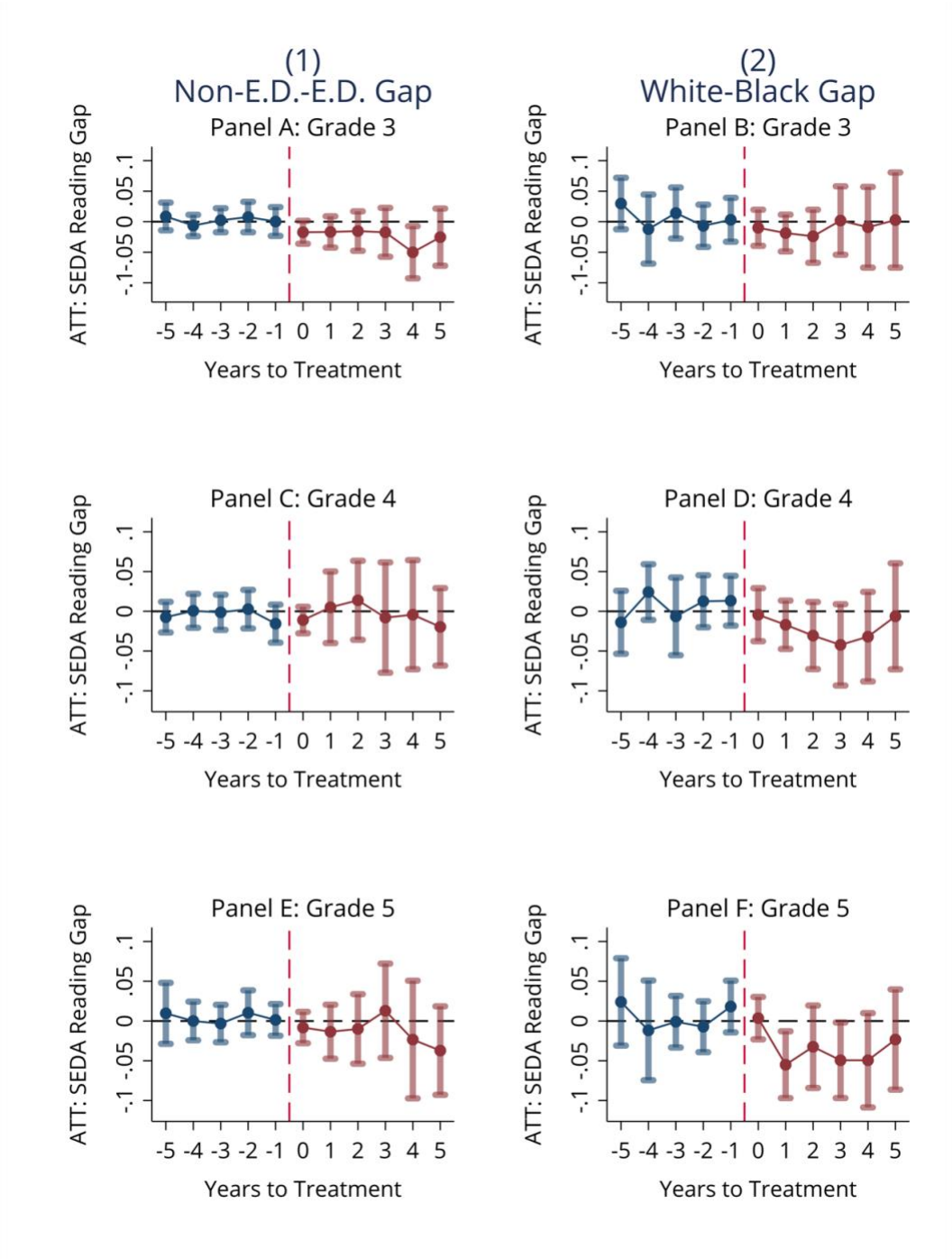
Note: Data are from overall average NAEP fourth-grade reading and math scale scores, 2003-2019. Column (1) includes states with comprehensive policies and never-treated states. Column (2) includes states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A11.

Figure 13. Comprehensive Early Literacy Policies and Low-Stakes Achievement (8th Grade)



Note: Data are from overall average NAEP eighth-grade reading and math scale scores, 2003-2019. Column (1) includes states with comprehensive policies and never-treated states. Column (2) includes states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A12.

Figure 14. Early Literacy Policies and High-Stakes Reading Test Score Gaps



Note: Data are from average SEDA reading test-scores gaps, 2009-2018. Column (1) is the non-economically disadvantaged-economically disadvantaged reading test score gap. Column (2) is the White-Black reading test score gap. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A13.

Table 1. Number of States Including Early Literacy Policy Components

Policy Component	Number of States
Supports for Teachers & Policy	
Science of Reading Training	30
Literacy/Reading Coaches	23
Teacher Prep Program Alignment to SOR and/or SOR Assessment	38
Funding for Literacy Efforts	37
Assessment & Parent Notification	
Universal Screener Identify Students with Reading Deficiency (K-3)	39
Dyslexia Screener for At-Risk Students	22
Notify Parents of Students Identified with Reading Deficiency	32
Instruction & Intervention	
District Adoption of High-Quality Instructional Materials	24
Individual Reading Plan and/or Intervention for Students w/ a Reading Deficiency	38
Monitor Progress Students with Reading Deficiency (K-3)	36
Intervention During Summer/Before, During, and/or After School Hours	33
Summer Reading Camps/Innovative Summer Reading Programs	30
Parent Engagement At-Home Reading Strategies	29
Retention & Intensive Intervention	
Statewide: Initial Determinant Retention at 3rd Grade Based on State Assessment (Cut Score)	22
Multiple Options for Promotion	20
Good Cause Exemptions (GCEs) for Some Students	16

Note. Source: (ExcelinEd, 2021)

Table 2. SEDA and NAEP Sample Sizes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	SEDA (2009-2018)						NAEP (2003-2019)	
	Overall		Non-ED-ED Gap		White-Black Gap		Overall	
	States	State-Years	States	State-Years	States	State-Years	States	State-Years
Panel A: Reading								
Grade 3	45	391	45	373	45	362	-	-
Grade 4	45	398	45	380	45	377	47	423
Grade 5	45	400	45	381	45	387	-	-
Grade 6	45	397	45	383	45	378	-	-
Grade 7	45	394	45	379	45	375	-	-
Grade 8	45	391	45	378	45	371	47	423
Panel B: Math								
Grade 3	45	398	45	380	45	380	-	-
Grade 4	45	400	45	382	45	391	47	423
Grade 5	45	393	45	374	45	384	-	-
Grade 6	45	388	45	374	45	376	-	-
Grade 7	44	371	44	355	44	356	-	-
Grade 8	44	352	44	339	44	335	47	423

Note: These are the effective number of state-by-year observations used in our preferred event-study estimator (Callaway & Sant'Anna, 2021). These sample sizes exclude any states adopting early literacy policies before the analysis period. The SEDA data are yearly from 2009 to 2018. Some states are missing SEDA data in any given year. The NAEP data are every two years from 2003 to 2019. The NAEP data form a balanced panel.

Table 3. Outcome Data Summary Statistics

	SEDA						NAEP					
	Reading			Math			Reading			Math		
	Overall	Treated	Never Treated	Overall	Treated	Never Treated	Overall	Treated	Never Treated	Overall	Treated	Never Treated
4th Grade												
Overall	0.012	0.021	-0.003	0.005	0.017	-0.016	220.935	221.175	220.179	240.343	240.433	240.059
	(0.164)	(0.178)	(0.139)	(0.193)	(0.201)	(0.179)	(6.216)	(6.328)	(5.837)	(5.607)	(5.620)	(5.606)
White-Black Gap	0.717	0.770	0.628	0.772	0.824	0.684						
	(0.209)	(0.206)	(0.184)	(0.214)	(0.216)	(0.181)						
Non-Econ. Dis.-Econ. Dis. Gap	0.729	0.737	0.718	0.711	0.725	0.687						
	(0.122)	(0.121)	(0.122)	(0.134)	(0.134)	(0.131)						
8th Grade												
Overall	-0.003	0.010	-0.023	0.010	0.020	-0.007	264.932	264.857	265.170	282.660	282.422	283.410
	(0.166)	(0.179)	(0.142)	(0.202)	(0.208)	(0.191)	(6.090)	(6.208)	(5.746)	(7.472)	(7.573)	(7.158)
White-Black Gap	0.684	0.744	0.587	0.752	0.815	0.647						
	(0.220)	(0.224)	(0.174)	(0.207)	(0.203)	(0.168)						
Non-Econ. Dis.-Econ. Dis. Gap	0.685	0.695	0.670	0.685	0.697	0.666						
	(0.108)	(0.110)	(0.104)	(0.114)	(0.117)	(0.106)						

Note. These statistics are derived from NAEP and SEDA data from 2009 to 2018. The SEDA data are measured in standard deviations. The NAEP data are measured in scale score points. Standard deviations are in parentheses.

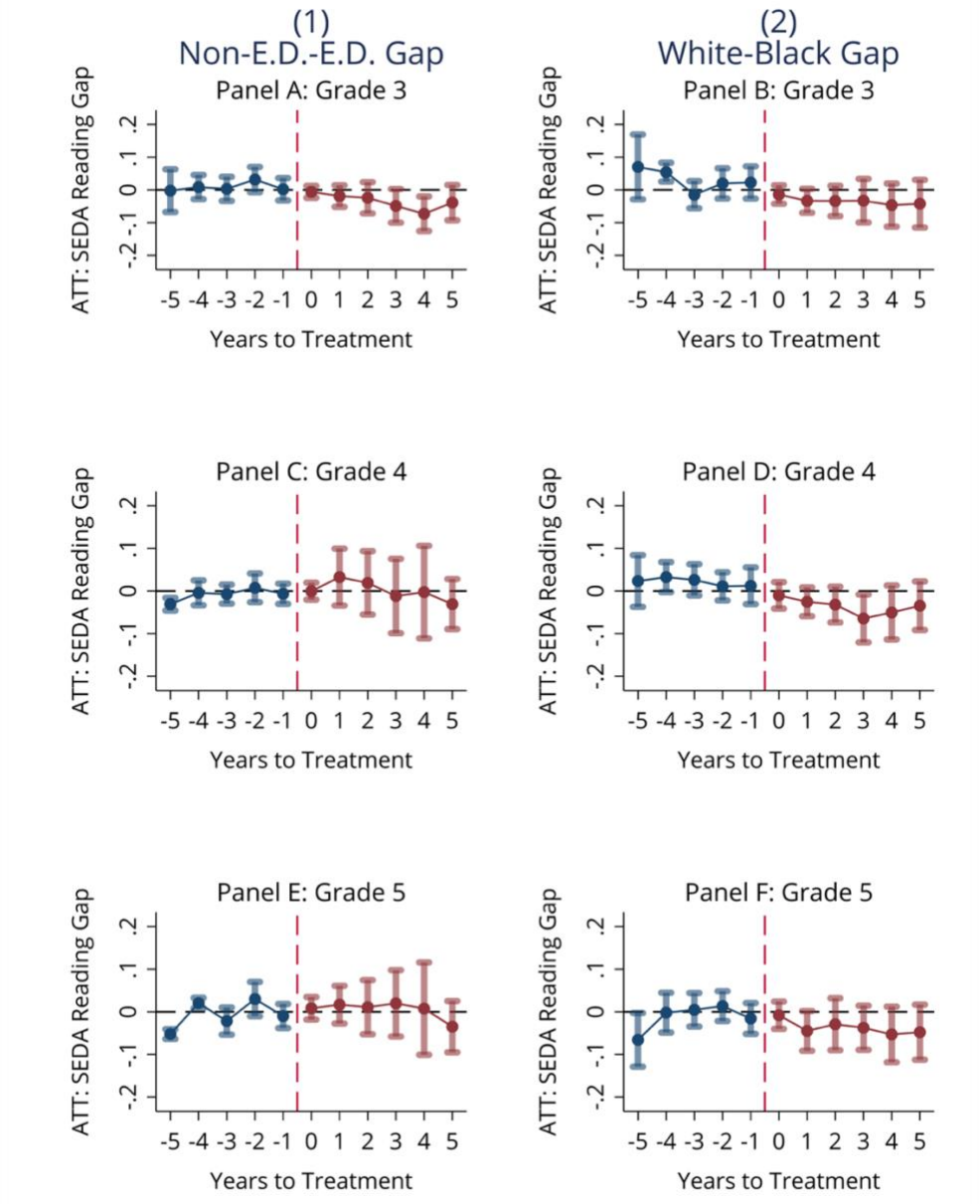
Table 4. Student Demographic Summary Statistics

	Overall	Treated	Never Treated
% Urban	0.272	0.280	0.259
	(0.141)	(0.168)	(0.080)
% Suburb	0.305	0.317	0.287
	(0.190)	(0.199)	(0.175)
% Town	0.159	0.145	0.181
	(0.091)	(0.081)	(0.102)
% Rural	0.264	0.258	0.273
	(0.129)	(0.130)	(0.128)
Average Per-Grade Enrollment	69,637.98	56,856.07	89,926.52
	(84,759.47)	(34,484.73)	(126,626.82)
% Black	0.160	0.179	0.129
	(0.158)	(0.166)	(0.138)
% Asian	0.160	0.036	0.066
	(0.158)	(0.024)	(0.136)
% Hispanic	0.161	0.162	0.161
	(0.139)	(0.130)	(0.152)
% Native American	0.021	0.019	0.025
	(0.042)	(0.044)	(0.040)
% White	0.611	0.605	0.620
	(0.194)	(0.180)	(0.215)
% English Learner	0.064	0.065	0.063
	(0.043)	(0.034)	(0.054)
% Special Education	0.134	0.135	0.133
	(0.027)	(0.027)	(0.027)
% Economically Disadvantaged	0.502	0.497	0.509
	(0.112)	(0.108)	(0.118)
Unemployment Rate	0.079	0.082	0.075
	(0.027)	(0.028)	(0.027)
Poverty Rate	0.143	0.142	0.143
	(0.033)	(0.034)	(0.031)
BA+ Rate	0.295	0.301	0.286
	(0.061)	(0.069)	(0.042)
Single-Mother Household Rate	0.188	0.192	0.182
	(0.043)	(0.047)	(0.037)

Note: Data are from the SEDA. State-level averages from 2009 to 2018. Standard deviations are in parentheses.

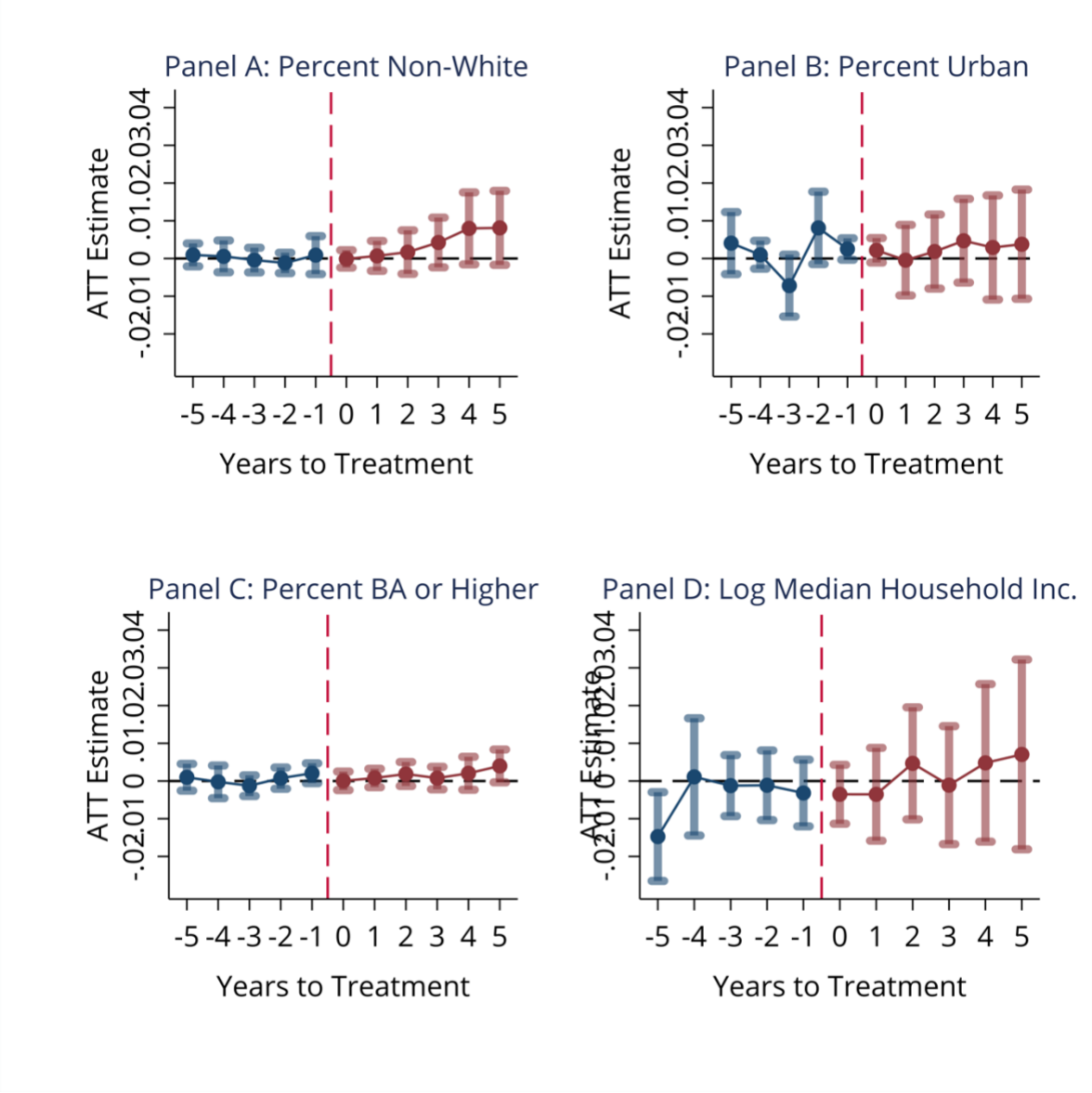
Online Appendix

Appendix Figure A1. Retention Mandate Early Literacy Policies and High-Stakes Reading Test-Score Gaps



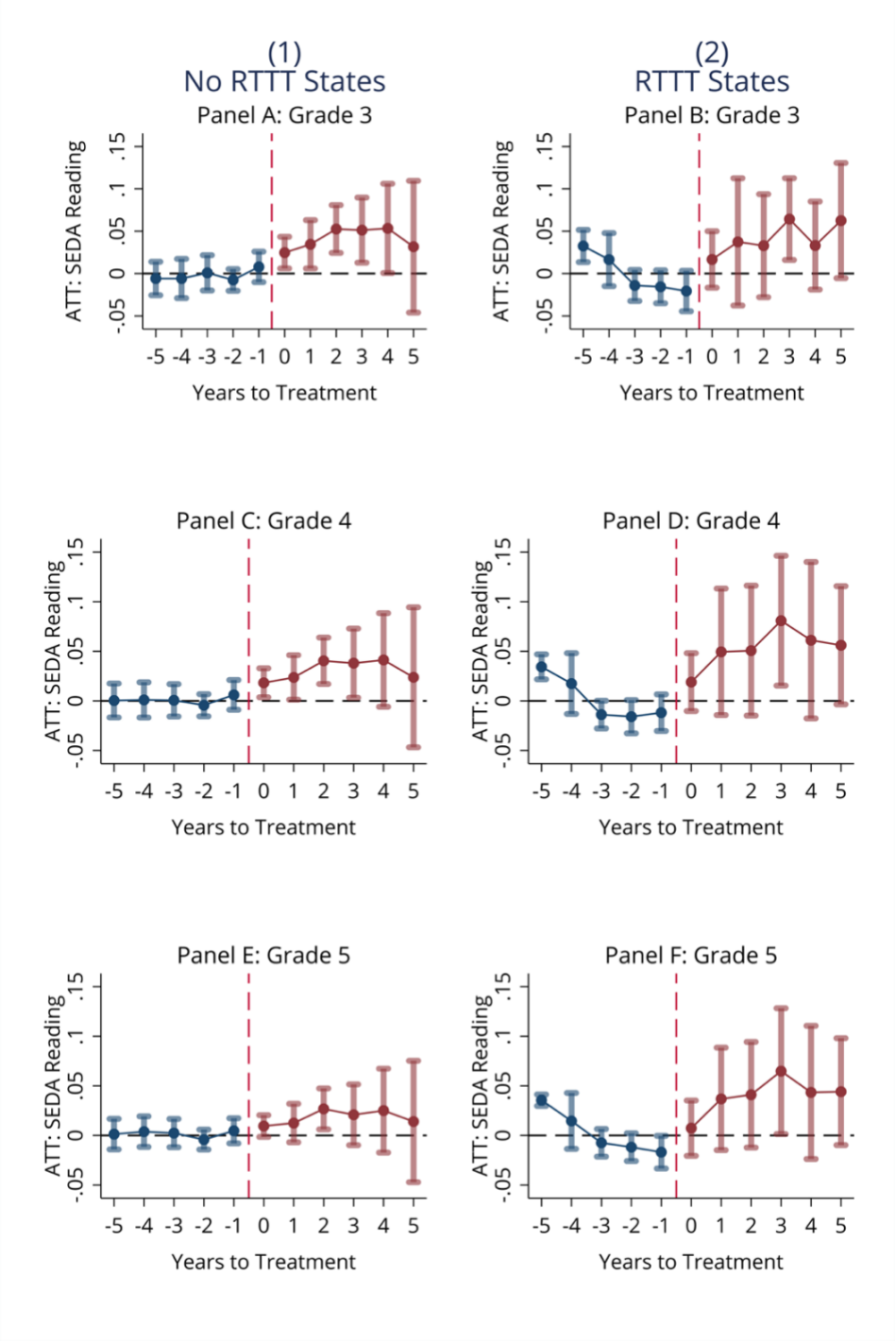
Note: Data are from average SEDA reading test-scores gaps, 2009-2018. The sample includes states with retention requirements and never-treated states. Column (1) is the non-economically disadvantaged-economically disadvantaged reading test score gap. Column (2) is the White-Black reading test score gap. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level. Detailed coefficient estimates are in Appendix Table A6.

Appendix Figure A2. Falsification Tests – State Demographic Characteristics and Macroeconomic Conditions



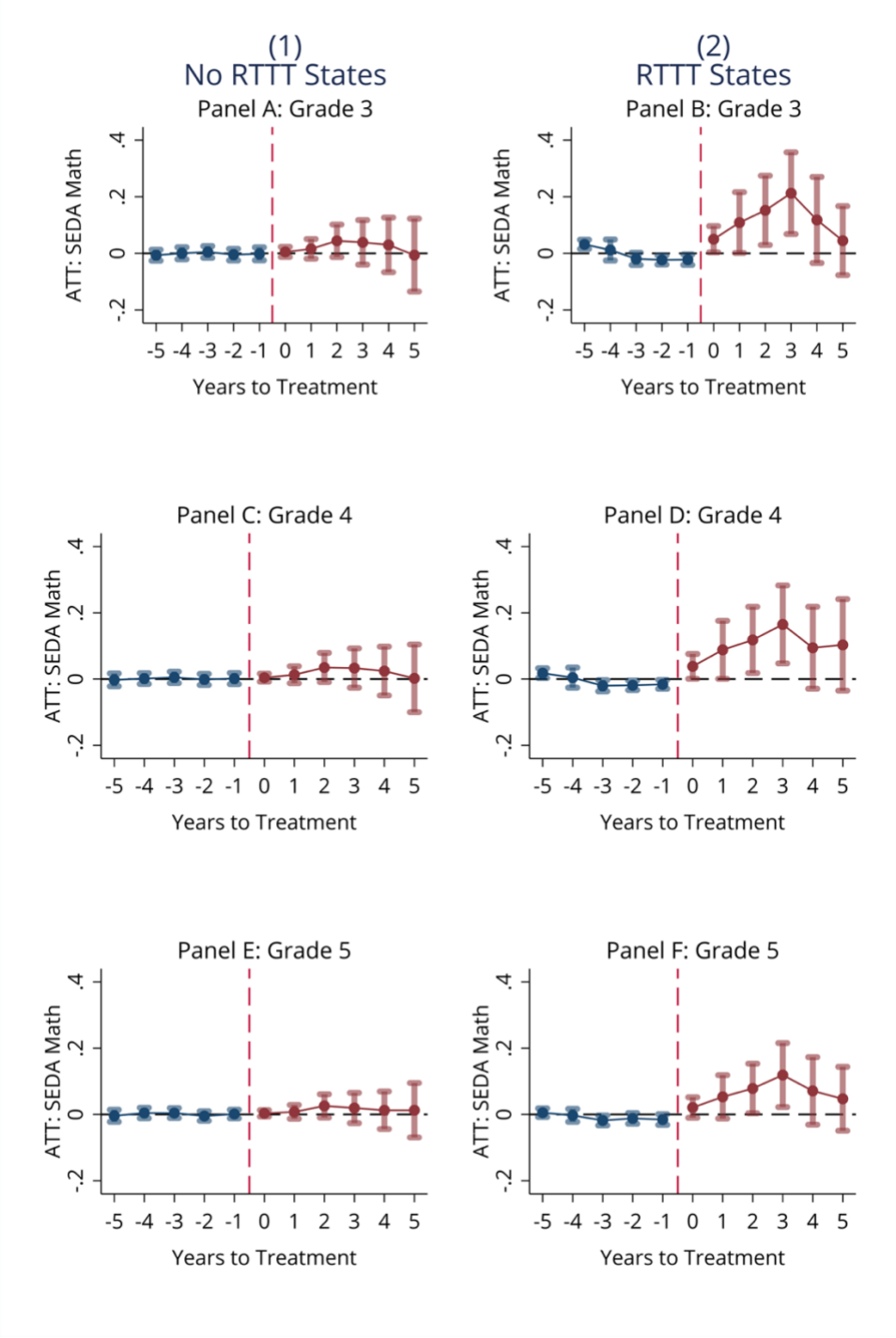
Note: Data are from the SEDA, 2009-2018. In Panel A, the outcome is the percent non-white in a state-year. In Panel B, the outcome is percent urban in a state-year. In Panel C, the outcome is percent with a bachelor's degree or higher in a state-year. In Panel D, the outcome is log median household income in a state-year. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. 95% confidence intervals from standard errors clustered at the state level.

Appendix Figure A3. Early Literacy Policies, High-Stakes Reading Achievement, and Race to the Top



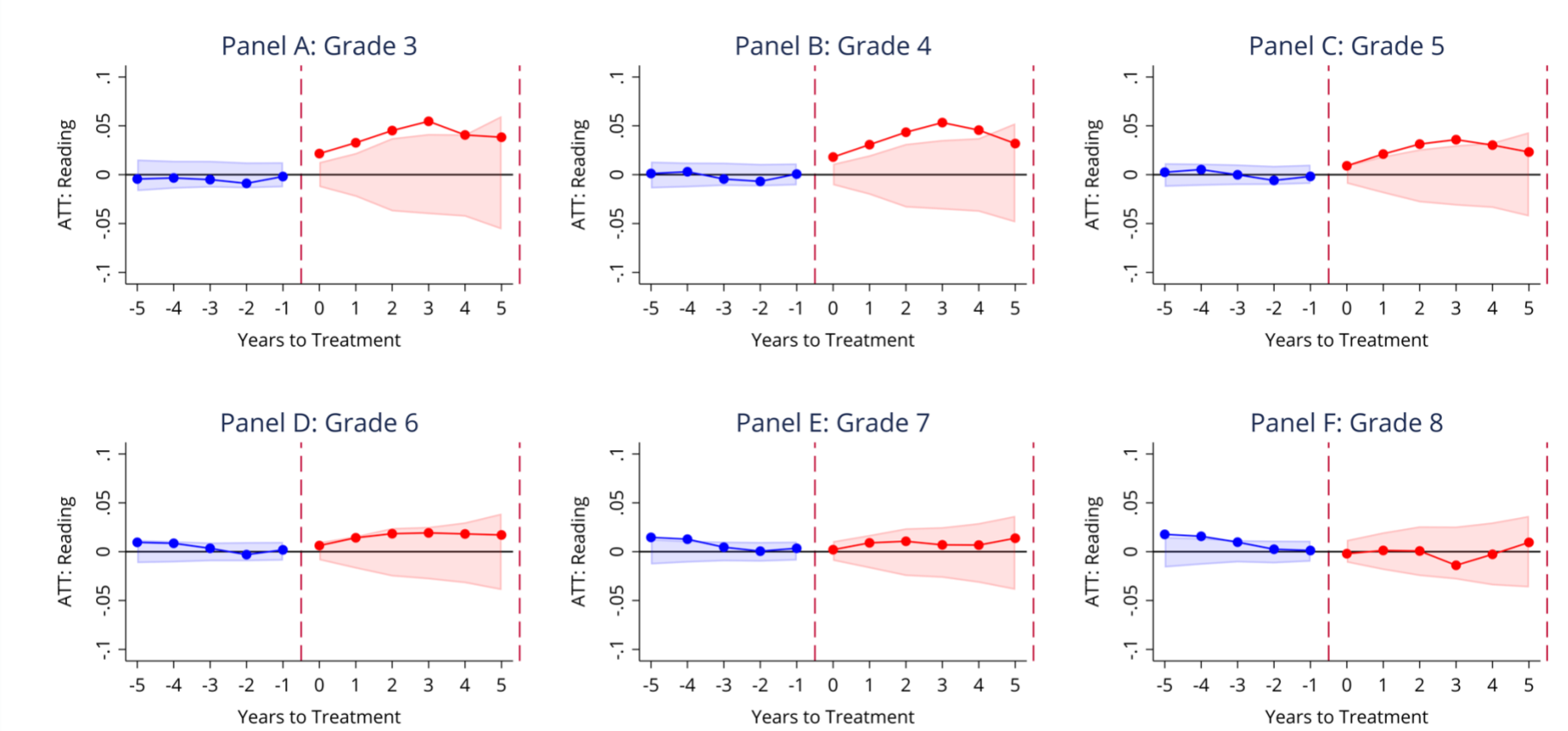
Note: Data are from overall average SEDA reading scores, 2009-2018. Column (1) includes states that did not receive Race to the Top (RTTT) funds and never-treated states. Column (2) includes states that did receive RTTT funds and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level.

Appendix Figure A4. Early Literacy Policies, Math Achievement, and Race to the Top



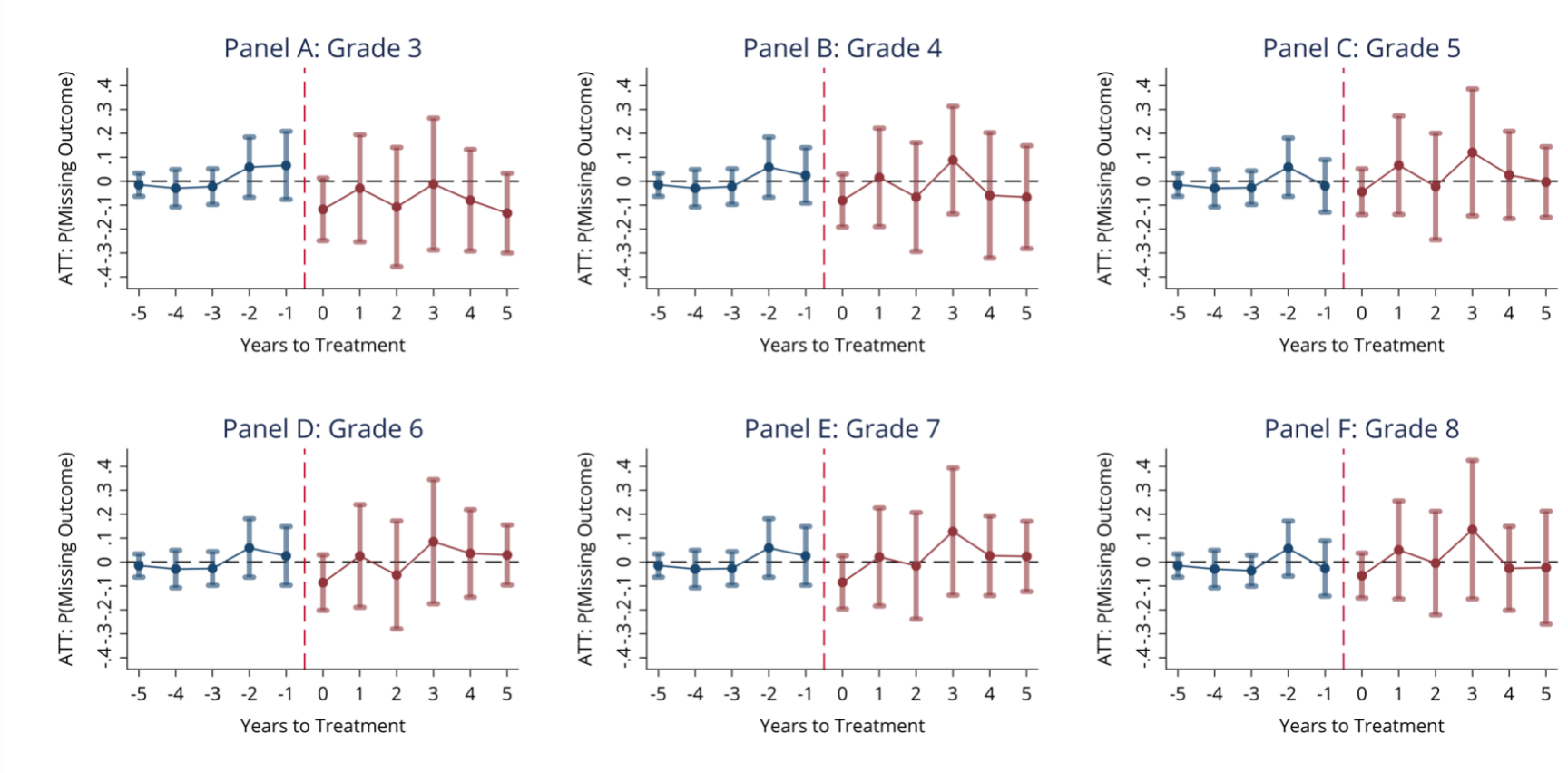
Note: Data are from overall average SEDA math scores, 2009-2018. Column (1) includes states that did not receive Race to the Top (RTTT) funds and never-treated states. Column (2) includes states that did receive RTTT funds and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level.

Appendix Figure A5. Placebo Test – Random Treatment Assignment (High-Stakes SEDA Reading)



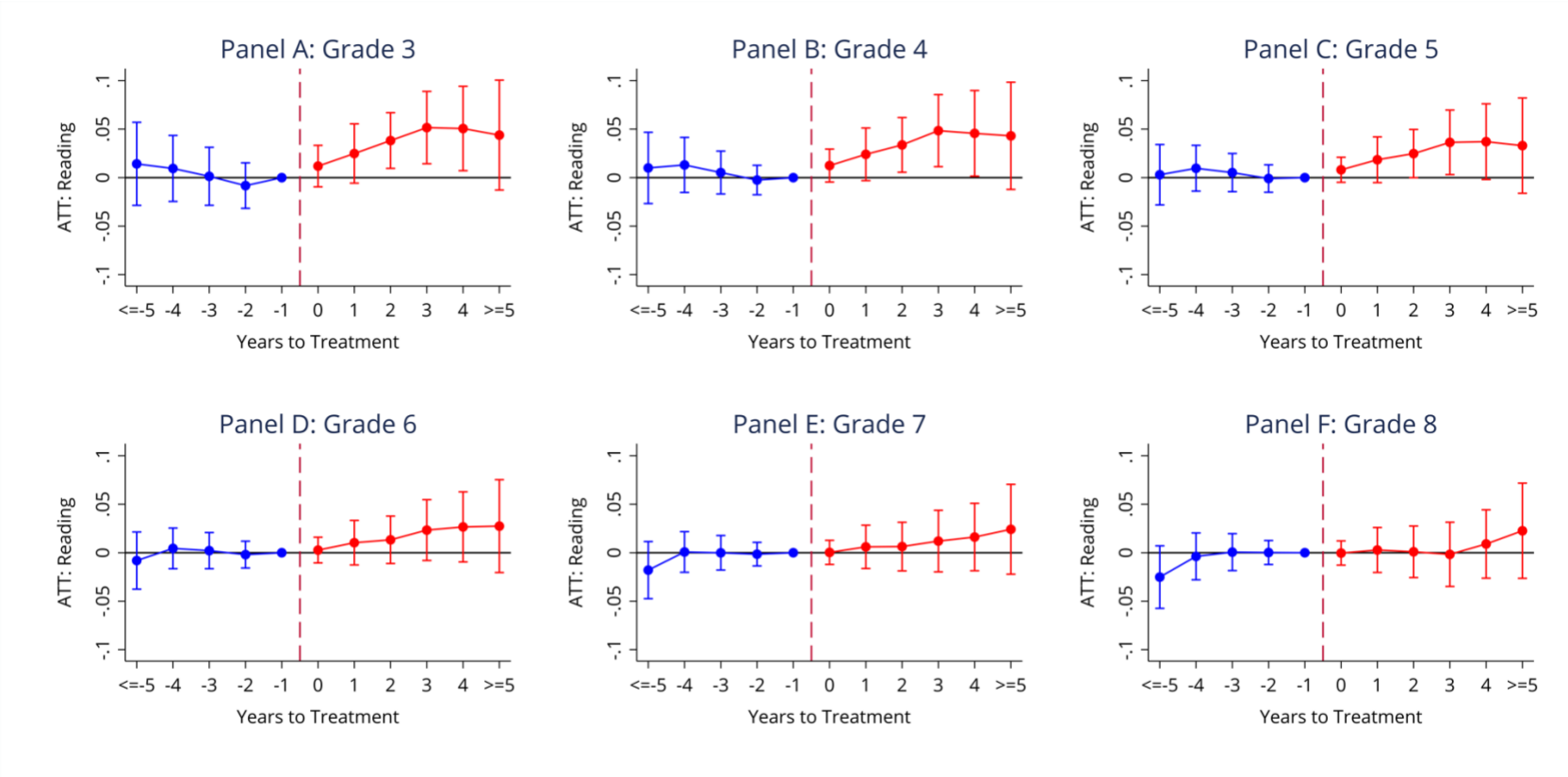
Note: Data are from the third- through eighth-grade overall average SEDA reading scores, 2009-2018. The shaded region represents the empirical 95% confidence interval constructed by randomly assigning treatment status to states 500 times (keeping the time distribution of adoption constant). Line represents the actual event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator shown in Figure 3, Panel A.

Appendix Figure A6. Missing Data Analysis



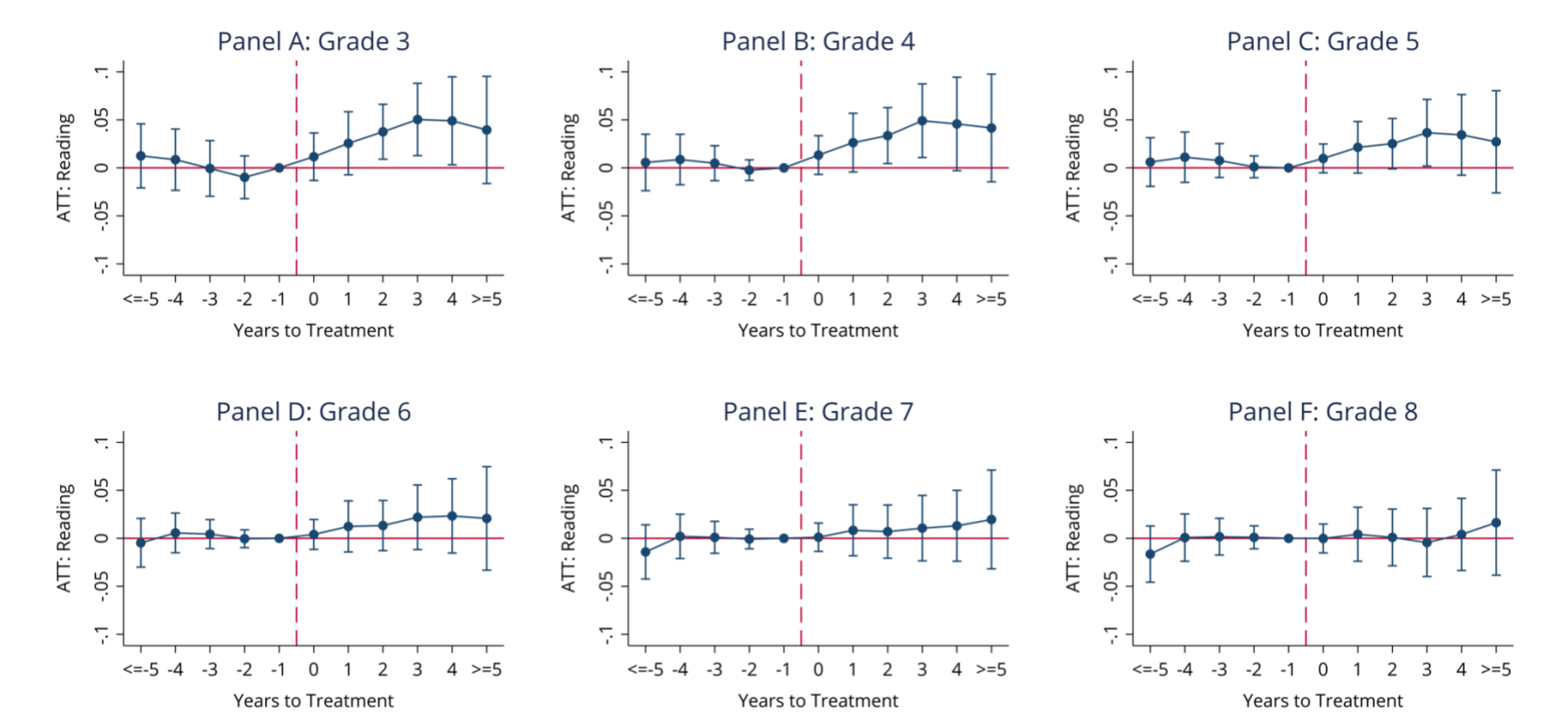
Note: Data are derived from overall average SEDA reading scores, 2009-2018. The outcome in each panel is an indicator for missing data in a given state-year. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. 95% confidence intervals from standard errors clustered at the state level.

Appendix Figure A7. Two-Way Fixed Effects Event Study (OLS)



Note: Data are from overall average SEDA reading scores, 2009-2018. The sample sizes range from 391 to 400 state-year observations. Detailed sample sizes can be found in Table 2. Event study coefficient estimates from a two-way fixed effects event study model estimated using OLS. 95% confidence intervals from standard errors clustered at the state level.

Appendix Figure A8. Interaction-Weighted Event-Study Estimator (Sun & Abraham, 2021)



Note: Data are from overall average SEDA reading scores, 2009-2018. The sample sizes range from 391 to 400 state-year observations. Detailed sample sizes can be found in Table 2. Event study coefficient estimates from the interaction-weighted event-study estimator (Sun & Abraham, 2021). 95% confidence intervals from standard errors clustered at the state level.

Appendix Table A1. Outcome Data Summary Statistics

	SEDA						NAEP					
	Reading			Math			Reading			Math		
	Overall	Treated	Never Treated	Overall	Treated	Never Treated	Overall	Treated	Never Treated	Overall	Treated	Never Treated
3rd Grade												
Overall	0.023	0.033	0.007	0.008	0.023	-0.017						
	(0.166)	(0.180)	(0.141)	(0.196)	(0.203)	(0.182)						
White-Black Gap	0.693	0.740	0.617	0.765	0.809	0.692						
	(0.185)	(0.177)	(0.174)	(0.203)	(0.206)	(0.175)						
Non-Econ. Dis.-Econ. Dis. Gap	0.719	0.723	0.712	0.707	0.717	0.691						
	(0.119)	(0.119)	(0.120)	(0.129)	(0.129)	(0.128)						
4th Grade												
Overall	0.012	0.021	-0.003	0.005	0.017	-0.016	220.935	221.175	220.179	240.343	240.433	240.059
	(0.164)	(0.178)	(0.139)	(0.193)	(0.201)	(0.179)	(6.216)	(6.328)	(5.837)	(5.607)	(5.620)	(5.606)
White-Black Gap	0.717	0.770	0.628	0.772	0.824	0.684						
	(0.209)	(0.206)	(0.184)	(0.214)	(0.216)	(0.181)						
Non-Econ. Dis.-Econ. Dis. Gap	0.729	0.737	0.718	0.711	0.725	0.687						
	(0.122)	(0.121)	(0.122)	(0.134)	(0.134)	(0.131)						
5th Grade												
Overall	0.002	0.012	-0.014	0.000	0.011	-0.017						
	(0.163)	(0.176)	(0.139)	(0.193)	(0.201)	(0.180)						
White-Black Gap	0.706	0.765	0.612	0.755	0.811	0.665						
	(0.209)	(0.204)	(0.182)	(0.207)	(0.205)	(0.178)						
Non-Econ. Dis.-Econ. Dis. Gap	0.725	0.736	0.708	0.702	0.716	0.681						
	(0.119)	(0.118)	(0.119)	(0.121)	(0.121)	(0.119)						
6th Grade												

Overall	-0.008	0.002	-0.023	-0.003	0.007	-0.017						
	(0.163)	(0.176)	(0.140)	(0.197)	(0.206)	(0.183)						
White-Black Gap	0.709	0.767	0.616	0.771	0.832	0.674						
	(0.213)	(0.207)	(0.188)	(0.217)	(0.217)	(0.179)						
Non-Econ. Dis.-Econ. Dis. Gap	0.725	0.731	0.717	0.713	0.726	0.694						
	(0.106)	(0.107)	(0.105)	(0.114)	(0.115)	(0.111)						
7th Grade												
Overall	-0.006	0.003	-0.021	0.001	0.007	-0.010						
	(0.165)	(0.178)	(0.143)	(0.203)	(0.211)	(0.189)						
White-Black Gap	0.704	0.767	0.601	0.770	0.838	0.649						
	(0.232)	(0.238)	(0.180)	(0.228)	(0.227)	(0.173)						
Non-Econ. Dis.-Econ. Dis. Gap	0.713	0.720	0.702	0.708	0.719	0.690						
	(0.105)	(0.106)	(0.103)	(0.115)	(0.118)	(0.108)						
8th Grade												
Overall	-0.003	0.010	-0.023	0.010	0.020	-0.007	264.932	264.857	265.170	282.660	282.422	283.410
	(0.166)	(0.179)	(0.142)	(0.202)	(0.208)	(0.191)	(6.090)	(6.208)	(5.746)	(7.472)	(7.573)	(7.158)
White-Black Gap	0.684	0.744	0.587	0.752	0.815	0.647						
	(0.220)	(0.224)	(0.174)	(0.207)	(0.203)	(0.168)						
Non-Econ. Dis.-Econ. Dis. Gap	0.685	0.695	0.670	0.685	0.697	0.666						
	(0.108)	(0.110)	(0.104)	(0.114)	(0.117)	(0.106)						

Note. These statistics are derived from NAEP and SEDA data from 2009 to 2018. The SEDA data are measured in standard deviations. The NAEP data are measured in scale score points. Standard deviations are in parentheses.

Appendix Table A2. Coefficient Estimates: Early Literacy Policies and High-Stakes Reading Scores

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
t-5	-0.004	0.001	0.003	0.010	0.015*	0.018**
	(0.012)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)
t-4	-0.003	0.003	0.005	0.009	0.013**	0.016**
	(0.011)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
t-3	-0.005	-0.004	-0.000	0.003	0.005	0.010
	(0.009)	(0.007)	(0.006)	(0.006)	(0.007)	(0.009)
t-2	-0.009	-0.007	-0.006	-0.003	0.001	0.002
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
t-1	-0.002	0.001	-0.002	0.002	0.003	0.001
	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
t+0	0.022***	0.018**	0.009	0.006	0.002	-0.002
	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)
t+1	0.033**	0.031**	0.021*	0.014	0.009	0.001
	(0.016)	(0.014)	(0.013)	(0.012)	(0.012)	(0.013)
t+2	0.045***	0.043***	0.031**	0.018	0.011	0.001
	(0.015)	(0.015)	(0.013)	(0.013)	(0.014)	(0.014)
t+3	0.055***	0.053***	0.036**	0.019	0.007	-0.014
	(0.016)	(0.019)	(0.017)	(0.016)	(0.017)	(0.018)
t+4	0.041**	0.046**	0.030	0.018	0.007	-0.003
	(0.020)	(0.023)	(0.020)	(0.018)	(0.017)	(0.016)
t+5	0.038	0.032	0.023	0.017	0.014	0.010
	(0.028)	(0.026)	(0.023)	(0.021)	(0.020)	(0.022)
N	391	398	400	397	394	391

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes are overall average SEDA reading scores. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A3. Coefficient Estimates: Early Literacy Laws and High-Stakes Math Achievement

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
t-5	-0.004	-0.003	-0.006	-0.003	-0.003	-0.006
	(0.012)	(0.011)	(0.009)	(0.008)	(0.008)	(0.008)
t-4	0.005	0.003	0.003	0.002	0.003	0.002
	(0.010)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)
t-3	0.003	0.002	0.001	0.002	0.004	0.004
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
t-2	-0.008	-0.005	-0.005	-0.003	-0.001	-0.002
	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)
t-1	-0.004	-0.000	-0.002	0.004	0.006	0.005
	(0.008)	(0.006)	(0.006)	(0.005)	(0.006)	(0.008)
t+0	0.019*	0.015*	0.010	-0.001	-0.013	0.005
	(0.010)	(0.008)	(0.007)	(0.011)	(0.020)	(0.007)
t+1	0.044**	0.037**	0.026*	0.021	0.015	0.005
	(0.022)	(0.018)	(0.015)	(0.014)	(0.013)	(0.014)
t+2	0.079**	0.061**	0.044**	0.033*	0.021	-0.001
	(0.031)	(0.024)	(0.020)	(0.017)	(0.016)	(0.015)
t+3	0.102**	0.080**	0.057**	0.042*	0.025	0.008
	(0.041)	(0.033)	(0.027)	(0.023)	(0.020)	(0.023)
t+4	0.065	0.052	0.038	0.027	0.021	0.009
	(0.044)	(0.035)	(0.029)	(0.024)	(0.022)	(0.022)
t+5	0.007	0.039	0.021	0.027	0.029	0.044
	(0.048)	(0.044)	(0.032)	(0.029)	(0.028)	(0.030)
N	398	400	393	388	371	352

Note: Data are from overall average SEDA math scores, 2009-2018. The outcomes are overall average SEDA reading scores. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A4. Coefficient Estimates: Early Literacy Policies and Low-Stakes Achievement

	(1) Grade 4 - Reading	(2) Grade 8 - Reading	(3) Grade 4 - Math	(4) Grade 8 - Math
t-7 & 8	-0.123	0.515	0.197	0.977*
	(0.555)	(0.499)	(0.411)	(0.556)
t- 5 & 6	-0.407	-0.363	0.293	-0.275
	(0.419)	(0.360)	(0.367)	(0.436)
t- 3 & 4	-0.208	0.273	0.000	0.034
	(0.415)	(0.361)	(0.368)	(0.377)
t- 1 & 2	0.126	0.150	0.458	-0.045
	(0.372)	(0.390)	(0.360)	(0.418)
t+ 0 & 1	-0.755	-0.115	-0.273	0.272
	(0.477)	(0.361)	(0.374)	(0.404)
t+ 2 & 3	0.153	-0.642	0.850	0.330
	(0.596)	(0.673)	(0.549)	(0.733)
t+ 4 & 5	0.685	-0.274	1.604*	0.455
	(0.935)	(0.806)	(0.874)	(0.964)
t+ 6 & 7	1.070	0.335	1.630	1.476
	(1.150)	(1.143)	(1.162)	(1.294)
t+ 8 & 9	0.592	1.332	1.578	1.968
	(1.480)	(1.478)	(1.438)	(1.630)
N	423	423	423	423

Note: Data are from overall average NAEP reading and math scale scores, 2003-2019. The outcomes are overall average NAEP reading and math scale scores. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A5. Coefficient Estimates: Impacts of Early Literacy Policies with Retention Requirements on High-Stakes Reading Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Retention Required			Retention Not Required		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
t-5	-0.007	-0.001	-0.006	-0.003	0.002	0.005
	(0.008)	(0.006)	(0.005)	(0.015)	(0.013)	(0.011)
t-4	-0.020**	-0.012**	-0.006	0.004	0.010	0.010
	(0.009)	(0.006)	(0.005)	(0.014)	(0.012)	(0.011)
t-3	-0.018	-0.011	-0.004	0.006	0.001	0.003
	(0.012)	(0.011)	(0.010)	(0.011)	(0.009)	(0.008)
t-2	-0.012	-0.010	-0.007	-0.005	-0.003	-0.004
	(0.008)	(0.007)	(0.006)	(0.008)	(0.006)	(0.005)
t-1	-0.003	-0.001	-0.001	-0.000	0.003	-0.003
	(0.010)	(0.008)	(0.007)	(0.009)	(0.007)	(0.008)
t+0	0.025**	0.024***	0.018**	0.018	0.010	-0.002
	(0.010)	(0.009)	(0.009)	(0.014)	(0.011)	(0.008)
t+1	0.045**	0.044***	0.036**	0.012	0.005	-0.003
	(0.018)	(0.017)	(0.016)	(0.028)	(0.022)	(0.014)
t+2	0.056***	0.058***	0.044**	0.028	0.018	0.013
	(0.018)	(0.020)	(0.019)	(0.023)	(0.019)	(0.014)
t+3	0.052***	0.059**	0.041*	0.060**	0.043**	0.025
	(0.019)	(0.024)	(0.022)	(0.024)	(0.022)	(0.020)
t+4	0.055**	0.068**	0.049*	0.022	0.013	0.003
	(0.022)	(0.030)	(0.026)	(0.030)	(0.026)	(0.023)
t+5	0.063*	0.054*	0.042	-0.002	-0.005	-0.008
	(0.033)	(0.031)	(0.028)	(0.040)	(0.036)	(0.031)
N	283	290	290	266	266	268

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes are overall average SEDA reading scores. Columns (1) to (3) include states with retention requirements and never-treated states. Columns (4) to (6) include states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A6. Coefficient Estimates: Impacts of Early Literacy Policies with Retention Requirements on High-Stakes Math Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Retention Required			Retention Not Required		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
t-5	-0.022***	-0.012**	-0.014***	0.002	0.000	-0.003
	(0.008)	(0.006)	(0.005)	(0.015)	(0.013)	(0.011)
t-4	0.010	0.001	0.005	0.003	0.004	0.003
	(0.013)	(0.008)	(0.009)	(0.013)	(0.010)	(0.010)
t-3	0.015*	0.015*	0.011	-0.009	-0.009	-0.005
	(0.009)	(0.008)	(0.008)	(0.011)	(0.010)	(0.009)
t-2	0.000	0.002	0.001	-0.017	-0.012	-0.011*
	(0.008)	(0.007)	(0.007)	(0.011)	(0.008)	(0.007)
t-1	-0.008	-0.002	0.002	0.001	0.002	-0.005
	(0.008)	(0.005)	(0.006)	(0.013)	(0.011)	(0.009)
t+0	0.031**	0.027**	0.023**	0.003	0.000	-0.004
	(0.014)	(0.011)	(0.010)	(0.009)	(0.007)	(0.007)
t+1	0.076***	0.065***	0.053**	-0.008	-0.008	-0.011
	(0.028)	(0.023)	(0.021)	(0.021)	(0.015)	(0.011)
t+2	0.120***	0.093***	0.074***	0.010	0.006	0.004
	(0.039)	(0.031)	(0.028)	(0.030)	(0.023)	(0.016)
t+3	0.142***	0.113***	0.083**	0.022	0.015	0.010
	(0.051)	(0.042)	(0.037)	(0.042)	(0.030)	(0.022)
t+4	0.126**	0.101**	0.071*	-0.023	-0.016	-0.006
	(0.058)	(0.048)	(0.043)	(0.037)	(0.027)	(0.022)
t+5	0.059	0.091	0.053	-0.071*	-0.049	-0.027
	(0.064)	(0.059)	(0.045)	(0.036)	(0.030)	(0.027)
N	287	289	280	267	268	271

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes are overall average SEDA reading scores. Columns (1) to (3) include states with retention requirements and never-treated states. Columns (4) to (6) include states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A7. Coefficient Estimates: Early Literacy Policies with Retention Requirements and Low-Stakes Achievement (4th Grade)

	(1)	(2)	(3)	(4)
	Retention		No Retention	
	Grade 4 - Reading	Grade 4 - Math	Grade 4 - Reading	Grade 4 - Math
t-7 & 8	0.054	-0.155	-0.265	0.478
	(0.808)	(0.574)	(0.665)	(0.478)
t- 5 & 6	-0.996*	-0.395	0.071	0.852**
	(0.526)	(0.541)	(0.569)	(0.404)
t- 3 & 4	-0.281	-0.177	-0.138	0.168
	(0.605)	(0.475)	(0.512)	(0.481)
t- 1 & 2	0.099	0.564	0.153	0.352
	(0.538)	(0.530)	(0.465)	(0.418)
t+ 0 & 1	-0.344	-0.100	-1.143	-0.436
	(0.551)	(0.463)	(0.705)	(0.534)
t+ 2 & 3	0.967	1.893**	-0.660	-0.193
	(0.659)	(0.798)	(0.784)	(0.499)
t+ 4 & 5	1.324	2.541**	-0.053	0.523
	(1.033)	(1.235)	(1.347)	(0.922)
t+ 6 & 7	2.254	2.706*	-0.640	0.075
	(1.447)	(1.589)	(1.176)	(1.158)
t+ 8 & 9	1.406	2.537	-0.324	0.499
	(2.086)	(2.158)	(1.446)	(1.348)
N	261	261	270	270

Note: Data are from overall average NAEP reading and math scale scores, 2003-2019. The outcomes are overall average NAEP reading and math scale scores. Columns (1) and (2) include states with retention requirements and never-treated states. Columns (3) and (4) include states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. Standard errors in parentheses clustered at the state level.

* p<0.1 **p<0.05 ***p<0.01

Appendix Table A8. Coefficient Estimates: Early Literacy Policies with Retention Requirements and Low-Stakes Achievement (8th Grade)

	(1)	(2)	(3)	(4)
	Retention		No Retention	
	Grade 8 - Reading	Grade 8 - Math	Grade 8 - Reading	Grade 8 - Math
t-7 & 8	-0.201	0.818	1.087*	1.105
	(0.666)	(0.833)	(0.649)	(0.723)
t- 5 & 6	-0.600	-0.591	-0.170	-0.017
	(0.535)	(0.553)	(0.391)	(0.606)
t- 3 & 4	-0.183	-0.214	0.702	0.268
	(0.433)	(0.536)	(0.517)	(0.490)
t- 1 & 2	0.301	-0.051	-0.000	-0.039
	(0.412)	(0.480)	(0.532)	(0.602)
t+ 0 & 1	-0.398	0.485	0.152	0.071
	(0.451)	(0.615)	(0.477)	(0.425)
t+ 2 & 3	-0.169	1.170	-1.115	-0.511
	(0.721)	(1.065)	(0.838)	(0.629)
t+ 4 & 5	-0.091	1.153	-0.486	-0.349
	(0.958)	(1.273)	(0.875)	(0.910)
t+ 6 & 7	0.238	2.196	0.475	0.437
	(1.292)	(1.660)	(1.168)	(1.160)
t+ 8 & 9	1.847	3.301	0.753	0.469
	(1.887)	(2.412)	(1.413)	(1.250)
N	261	261	270	270

Note: Data are from overall average NAEP reading and math scale scores, 2003-2019. The outcomes are overall average NAEP reading and math scale scores. Columns (1) and (2) include states with retention requirements and never-treated states. Columns (3) and (4) include states without retention requirements and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. Standard errors in parentheses clustered at the state level.

* p<0.1 **p<0.05 ***p<0.01

Appendix Table A9. Coefficient Estimates: Impacts of Comprehensive Early Literacy Policies on High-Stakes Reading Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Comprehensive			Not Comprehensive		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
t-5	-0.007	-0.001	-0.006	-0.003	0.002	0.005
	(0.008)	(0.006)	(0.005)	(0.015)	(0.013)	(0.011)
t-4	-0.012	-0.011	-0.002	-0.001	0.007	0.007
	(0.007)	(0.008)	(0.004)	(0.014)	(0.011)	(0.010)
t-3	-0.015	-0.011	-0.008	-0.000	-0.001	0.004
	(0.017)	(0.015)	(0.013)	(0.009)	(0.007)	(0.006)
t-2	-0.015	-0.011	-0.007	-0.005	-0.004	-0.004
	(0.011)	(0.010)	(0.009)	(0.007)	(0.006)	(0.005)
t-1	-0.003	-0.001	-0.001	0.001	0.004	-0.000
	(0.014)	(0.011)	(0.010)	(0.008)	(0.007)	(0.006)
t+0	0.014	0.011	0.007	0.021*	0.017*	0.005
	(0.010)	(0.009)	(0.008)	(0.011)	(0.009)	(0.008)
t+1	0.027*	0.021	0.013	0.023	0.024	0.014
	(0.016)	(0.014)	(0.013)	(0.021)	(0.018)	(0.014)
t+2	0.063***	0.049***	0.031**	0.026	0.032*	0.022
	(0.023)	(0.019)	(0.016)	(0.016)	(0.019)	(0.016)
t+3	0.076***	0.056***	0.032*	0.036*	0.046*	0.029
	(0.023)	(0.020)	(0.017)	(0.019)	(0.026)	(0.023)
t+4	0.099***	0.082***	0.059***	0.014	0.030	0.018
	(0.018)	(0.017)	(0.017)	(0.022)	(0.029)	(0.025)
t+5	0.093***	0.084**	0.071**	-0.014	-0.017	-0.022
	(0.036)	(0.033)	(0.029)	(0.029)	(0.026)	(0.023)
N	232	232	232	308	315	317

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes are overall average SEDA reading scores. Columns (1) to (3) include states with comprehensive policies and never-treated states. Columns (4) to (6) include states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A10. Coefficient Estimates: Impacts of Comprehensive Early Literacy Policies on High-Stakes Math Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Comprehensive			Not Comprehensive		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
t-5	-0.022***	-0.012**	-0.014***	0.002	0.000	-0.003
	(0.008)	(0.006)	(0.005)	(0.015)	(0.013)	(0.011)
t-4	0.002	-0.003	0.005	0.006	0.005	0.003
	(0.016)	(0.009)	(0.012)	(0.012)	(0.009)	(0.009)
t-3	0.012	0.012	0.008	-0.002	-0.003	-0.002
	(0.012)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)
t-2	0.004	0.006	0.005	-0.015*	-0.010	-0.011
	(0.010)	(0.008)	(0.006)	(0.009)	(0.007)	(0.007)
t-1	-0.009	-0.002	0.005	-0.002	0.000	-0.004
	(0.011)	(0.007)	(0.004)	(0.010)	(0.008)	(0.008)
t+0	0.012	0.011	0.010	0.015	0.011	0.005
	(0.012)	(0.009)	(0.007)	(0.011)	(0.009)	(0.009)
t+1	0.039	0.034*	0.025	0.031	0.026	0.014
	(0.025)	(0.020)	(0.017)	(0.027)	(0.022)	(0.018)
t+2	0.097**	0.074**	0.058**	0.052	0.039	0.024
	(0.043)	(0.035)	(0.027)	(0.036)	(0.028)	(0.023)
t+3	0.124**	0.094*	0.063*	0.067	0.053	0.033
	(0.060)	(0.048)	(0.037)	(0.049)	(0.038)	(0.032)
t+4	0.132**	0.104**	0.077*	0.037	0.031	0.020
	(0.063)	(0.051)	(0.042)	(0.051)	(0.041)	(0.035)
t+5	0.089	0.081	0.075	-0.077**	-0.007	-0.040*
	(0.080)	(0.065)	(0.052)	(0.033)	(0.055)	(0.024)
N	230	231	232	315	317	310

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes are overall average SEDA reading scores. Columns (1) to (3) include states with comprehensive policies and never-treated states. Columns (4) to (6) include states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01

Appendix Table A11. Coefficient Estimates: Comprehensive Early Literacy Policies and Low-Stakes Achievement (4th Grade)

	(1)	(2)	(3)	(4)
	Comprehensive		Not Comprehensive	
	Grade 4 - Reading	Grade 4 - Math	Grade 4 - Reading	Grade 4 - Math
t-7 & 8	0.743	0.418	-0.456	0.112
	(0.926)	(0.704)	(0.597)	(0.440)
t- 5 & 6	-1.728**	-1.016	-0.031	0.720*
	(0.748)	(0.637)	(0.461)	(0.390)
t- 3 & 4	-1.139	-0.853**	0.167	0.364
	(0.721)	(0.347)	(0.470)	(0.450)
t- 1 & 2	0.165	1.082**	0.053	0.307
	(0.619)	(0.542)	(0.440)	(0.415)
t+ 0 & 1	-0.133	-0.654	-0.887	-0.154
	(0.794)	(0.639)	(0.560)	(0.437)
t+ 2 & 3	0.981	1.261	-0.251	0.438
	(0.915)	(0.917)	(0.654)	(0.572)
t+ 4 & 5	2.228**	2.005*	0.083	1.027
	(1.099)	(1.151)	(1.129)	(1.006)
t+ 6 & 7	3.920**	2.756	-0.084	0.857
	(1.725)	(2.195)	(1.224)	(1.266)
t+ 8 & 9	2.478	0.722	-0.204	1.258
	(1.786)	(1.378)	(1.735)	(1.731)
N	189	189	333	333

Note: Data are from overall average NAEP reading and math scale scores, 2003-2019. The outcomes are overall average NAEP reading and math scale scores. Columns (1) and (2) include states with comprehensive policies and never-treated states. Columns (3) and (4) include states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. Standard errors in parentheses clustered at the state level.

* p<0.1 **p<0.05 ***p<0.01

Appendix Table A12. Coefficient Estimates: Comprehensive Early Literacy Policies and Low-Stakes Achievement (8th Grade)

	(1)	(2)	(3)	(4)
	Comprehensive		Not Comprehensive	
	Grade 8 - Reading	Grade 8 - Math	Grade 8 - Reading	Grade 8 - Math
t-7 & 8	-0.438	-0.296	0.881*	1.467**
	(1.023)	(1.009)	(0.515)	(0.597)
t- 5 & 6	-1.392**	-1.110**	-0.114	-0.077
	(0.664)	(0.555)	(0.360)	(0.532)
t- 3 & 4	0.335	0.104	0.293	-0.008
	(0.572)	(0.490)	(0.445)	(0.487)
t- 1 & 2	0.222	-0.154	0.097	0.046
	(0.589)	(0.627)	(0.431)	(0.493)
t+ 0 & 1	0.395	1.347***	-0.207	-0.057
	(0.591)	(0.506)	(0.402)	(0.460)
t+ 2 & 3	-0.100	1.296	-0.966	-0.110
	(0.866)	(0.889)	(0.730)	(0.822)
t+ 4 & 5	-0.040	1.271	-0.526	-0.120
	(1.060)	(0.978)	(0.842)	(1.090)
t+ 6 & 7	2.068	3.127*	-0.283	0.561
	(1.439)	(1.620)	(1.162)	(1.426)
t+ 8 & 9	3.639**	3.118**	0.427	1.093
	(1.682)	(1.394)	(1.579)	(1.933)
N	189	189	333	333

Note: Data are from overall average NAEP reading and math scale scores, 2003-2019. The outcomes are overall average NAEP reading and math scale scores. Columns (1) and (2) include states with comprehensive policies and never-treated states. Columns (3) and (4) include states with non-comprehensive policies and never-treated states. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Relative years binned into two-year bins due to the biennial nature of NAEP administration. Standard errors in parentheses clustered at the state level.

* p<0.1 **p<0.05 ***p<0.01

Appendix Table A13. Coefficient Estimates: Early Literacy Policies and High-Stakes Reading Test Score Gaps

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-Econ. Dis.-Econ. Dis Reading Gap			White-Black Reading Gap		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
t-5	0.012	-0.006	0.011	0.031	-0.010	0.016
	(0.011)	(0.010)	(0.019)	(0.020)	(0.022)	(0.030)
t-4	0.005	0.005	0.007	-0.007	0.037**	-0.032
	(0.009)	(0.011)	(0.010)	(0.029)	(0.014)	(0.027)
t-3	-0.001	-0.004	-0.011	0.010	-0.013	-0.008
	(0.013)	(0.012)	(0.013)	(0.020)	(0.026)	(0.017)
t-2	0.017	0.003	0.016	0.007	0.013	-0.012
	(0.013)	(0.012)	(0.014)	(0.017)	(0.015)	(0.019)
t-1	0.000	-0.009	-0.007	0.001	0.018	-0.000
	(0.012)	(0.012)	(0.010)	(0.019)	(0.017)	(0.015)
t+0	-0.016*	-0.010	-0.002	-0.009	-0.008	0.003
	(0.009)	(0.009)	(0.010)	(0.015)	(0.017)	(0.015)
t+1	-0.021	0.004	-0.011	-0.026	-0.027*	-0.052***
	(0.015)	(0.024)	(0.018)	(0.016)	(0.014)	(0.020)
t+2	-0.024	0.006	-0.008	-0.040**	-0.025	-0.024
	(0.018)	(0.027)	(0.024)	(0.020)	(0.019)	(0.026)
t+3	-0.034	-0.023	0.010	-0.017	-0.051**	-0.036
	(0.022)	(0.036)	(0.032)	(0.029)	(0.025)	(0.024)
t+4	-0.067***	-0.014	-0.020	-0.035	-0.032	-0.041
	(0.022)	(0.037)	(0.039)	(0.032)	(0.031)	(0.032)
t+5	-0.031	-0.028	-0.034	-0.007	-0.019	-0.017
	(0.024)	(0.025)	(0.030)	(0.041)	(0.033)	(0.034)
N	373	380	381	362	377	387

Note: Data are from overall average SEDA reading scores, 2009-2018. The outcomes in Columns (1) to (3) are the non-economically disadvantaged-economically disadvantaged reading test score gap. The outcomes in Columns (4) to (6) are is the White-Black reading test score gap. Event study coefficient estimates from the Callaway & Sant'Anna (2021) estimator. Standard errors in parentheses clustered at the state level.
* p<0.1 **p<0.05 ***p<0.01