ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-887

The Achievement Effects of Scaling Early Literacy Reforms

Sarah Novicoff
Stanford University

Thomas S. Dee
Stanford University

While policymakers have demonstrated considerable enthusiasm for "science of reading" initiatives, the evidence on the impact of related reforms when implemented at scale is limited. In this pre-registered, quasi-experimental study, we examine California's recent initiative to improve early literacy across the state's lowest-performing elementary schools. The Early Literacy Support Block Grant (ELSBG) provided teacher professional development grounded in the science of reading as well as aligned supports (e.g., assessments and interventions), new funding (about $1000 per student), spending flexibility within specified guidelines, and expert facilitation and oversight of school-based planning. We find that ELSBG generated significant (and cost-effective) improvements in ELA achievement in its first two years of implementation (0.14 SD) as well as smaller, spillover improvements in math achievement.

Keywords: Achievement, early literacy, targeting, support, implementation, oversight, flexibility, funding, science of reading, curriculum, pedagogy

VERSION: December 2023

Suggested citation: Novicoff, Sarah, and Thomas S. Dee. (2023). The Achievement Effects of Scaling Early Literacy Reforms. (EdWorkingPaper: 23-887). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/jnmt-2093

A broad consensus views early literacy as a critically important foundational skill for longer-term academic success. However, persistently low-levels of reading achievement suggest a large-scale and long-standing failure to provide students in the U.S. with the early-literacy skills relevant to realizing their academic potential. For example, in the most recent National Assessment of Education Progress (NAEP), only 1 in 3 of U.S. fourth graders performed at or above proficiency in reading (*NAEP Reading*, 2022)—a proficiency rate that varies little from its level three decades prior (i.e., 29% in 1992). Civil-rights groups (e.g., Carr, 2022) have also recognized the racialized gaps in early literacy as a key dimension of the inequality in educational opportunity with important implications for subsequent outcomes such as exclusionary discipline, special-education referrals, and high-school graduation. Furthermore, the considerable challenges of academic recovery from the COVID-19 pandemic have heightened policy concerns about student achievement generally.

The opportunities and challenges implied by such factors have seeded a long-term and contentious debate (Preston, 2022; Schwartz, 2023b) over the best pedagogical approach to early literacy (i.e., the "Reading Wars"). On one side, the "science of reading" describes a process by which the sequential emphasis of five pillars–phonemic awareness, phonics, fluency, vocabulary, and comprehension–can improve overall literacy by building understanding of sounds, letters, words, sentences, and eventually paragraphs step-by-step (National Reading Panel, 2000; Adams, 1990; Anderson et al., 1985; Chall, 1967; Flesch, 1955). The science of reading contrasts with a "whole language" approach. Whole language relies on the idea that students will learn to read if reading happens around them frequently and if reading is made to be engaging. It also emphasizes identifying words based on contextual cues rather than through their sounds and composite parts (Goodman, 1989; Watson, 1989). In the words of its proponents, "balanced literacy" blends these two contrasting approaches with students receiving both skills-based instruction with phonics and holistic word-based lessons (Pressley et al., 2023). A recent survey of early-elementary teachers found that roughly three out of four stated they used balanced-literacy techniques, guiding students to identify unfamiliar words both by sounding them out (i.e., a phonics-based approach) and through "three-cueing" methods grounded in whole-language concepts (Kurtz et al., 2020).

A growing and high-profile enthusiasm for the science of reading (and a corresponding concern about the persistent prevalence of three-cueing instruction) has recently motivated a substantial number of multi-faceted state-level policy efforts to transform early-literacy pedagogy (i.e., 32 states and the District of Columbia passing laws since 2013; Schwartz, 2023a). However, the limited evidence currently available on these state initiatives is largely descriptive and complicated by the presence of confounding factors or fails to validate the current enthusiasm for the science of reading. For example, prominent claims of a "Mississippi Miracle" in reading achievement (e.g., Kristof, 2023) appeal informally to the state's distinctive test-score trends following a 2013 state initiative that featured the science of reading. Similarly, trend data from Michigan indicate that the state's 2016 "Ready by Grade 3" law coincided with an arrested decline in grade-3 reading achievement (Strunk et al., 2021, Figure 6.7.1). However, the evidence based on quasi-experimental designs is less encouraging. For example, a recent study based on comprehensive state-by-year panel data (Westall and Cummings, 2023) presents event-study evidence based on "difference in differences" (DID) designs. They find "little evidence of significant increases" in NAEP reading scores following these state initiatives with the exception of those states adopting fully comprehensive literacy reforms (e.g., teacher supports, assessments, interventions, and grade retention).[1] Similarly, the earlier federally funded evaluation of Reading First, based on a regression-discontinuity (RD) design, found that this initiative significantly increased instructional time and practices aligned with the science of reading but

[1] We note that the frequent bundling of new grade-retention policies with the literacy reforms creates evaluation challenges (e.g., compositional change and age-at-test confounds) that make it difficult to isolate the impact of policy efforts to promote science-of-reading pedagogy. However, the state literacy reform studied in this paper did not include grade-retention changes. Moreover, our robustness checks assess and dismiss enrollment changes as an internal-validity threat.

did not significantly improve students' reading comprehension in grades one, two, or three (Gamse et al., 2008).

This study provides quasi-experimental evidence on the early achievement impact of California's recent large-scale initiative to promote early literacy among K-3 students served by the state's most chronically underperforming elementary schools. Specifically, California's Early Literacy Support Block Grant (ELSBG) combined new state funding (i.e., over $50 million) with a focused effort to promote pedagogy aligned with the science of reading in grades K-3 in identified schools. ELSBG featured several notable design details relevant to the character and fidelity of its implementation. Specifically, this targeted state funding supported school-specific needs assessments and "Literacy Action Plans," supports aligned to the science of reading for teachers, students, parents, and communities, spending flexibility within specified parameters, and support and oversight managed by a competitively selected "Expert Lead in Literacy." In brief, this study finds robust evidence, based on synthetic "difference-in-differences" and "difference-in-difference-in-differences" designs, that ELSBG significantly increased grade-3 ELA achievement by 0.14 SD (i.e., roughly 25 percent of a year of learning at this age) with smaller spillover benefits for grade-3 math achievement.

These results provide encouraging evidence on the promise of promoting pedagogy linked to the science of reading at some scale. This study can also be situated within several other important and related literatures. For example, the challenges of changing how educators use research-based insights in their daily practice (e.g., Joyce & Cartwright, 2020) are widely recognized as a central impediment to evidence-based reforms. ELBSG's design (i.e., funding, broad supports, a blend of local flexibility and oversight) and the apparent quality of its early implementation suggest it provides a compelling case of how to bridge the "research to practice gap" effectively. Second, this evidence is also illustrative of what may be required to realize, at scale in real-world settings, the much-discussed promise of curricula-based reform (e.g., Whitehurst, 2009). Third, because ELSBG also encouraged planning (and provided local flexibility) within a subset of targeted schools, it also has strong parallels with the school-level reforms currently required under the federal Every Student Succeeds Act (ESSA). Fourth, this study provides evidence of an education reform that was effective within the uniquely strained context of academic recovery from the COVID-19 pandemic. The complementary quasi-experimental research designs employed here also illustrate strategies for assessing the empirical relevance of the potential confounds unique to the pandemic context. Finally, because we pre-registered an analysis plan, this study offers a novel example of transparency with regard to researcher discretion in a quasi-experimental study (Nosek et al., 2018).

The Early Literacy Support Block Grant (ELSBG)

In 2017, a California lawsuit (i.e., *Ella T. vs. the State of California*) alleged that, by sending plaintiffs to schools that did not teach them to read, the state violated the right to an education articulated in the state constitution. The original complaint argued "An education that does not provide access to literacy cannot be called an education at all" (Public Counsel & Morrison & Foerster LLP, 2017). As part of a 2020 settlement to this case, the state agreed to allocate $50 million in support of the Early Literacy Support Block Grant (ELSBG)—a targeted and multi-faceted initiative to improve reading outcomes at the lowest-performing 75 elementary schools in the state (*Ella T. v. State of California: Settlement Implementation Agreement*, 2020).[2]

The state identified ELSBG-eligible schools by averaging the percent of grade-3 students across the 2017-18 and 2018-19 school years who scored at the lowest (i.e., "Standard Not Met") of the four levels on the state's English Language Arts (ELA) assessment and weighting the average by the number of test-takers in each of the years (Authorization of the ESLB Grant, 2020). Because one

[2] The California Department of Education (CDE) retained $3 million of this appropriation to support its administration and oversight of the ELSBG program.

of the 75 eligible schools closed before the case was settled, CDE expanded eligibility to the 76th lowest-performing school based on this baseline score.

In August 2020, the California Department of Education notified the relevant school districts and charter management organizations (CMOs) of their eligible schools. Districts and CMOs interested in the ELSBG program received $40,000 plus $10,000 per eligible school to conduct a root-causes analysis and needs assessment that would inform a required three-year "Literacy Action Plan" proposing how they would improve (St. Andre, 2020). The state disbursed these initial planning funds to eligible and interested school districts in the middle of the 2020-21 school year.. The take-up of ELSBG eligibility was nearly universal. Thirty-five out of the eligible 37 school districts and CMOs ultimately submitted applications for their eligible schools, representing 73 of the 75 eligible and open schools. The state approved all of the resulting plans with budgets totaling $46.86 million (i.e., 3-year budgets that averaged roughly $642,000 per school). Implementation of those plans began in July 2021 with schools receiving their first-year allocations whenever their plans were approved. The state map in Figure A1 shows that the ELBSG-eligible schools are located throughout the state, including urban, suburban, and rural settings.

In addition to this targeted state funding, the ELSBG program has three other broad but notable design features relevant to its effort to improve early literacy at scale across all of these schools. First, the authorizing legislation for the ELSBG program required the selection of a County Office of Education as the statewide "Expert Lead in Literacy" that would support grantees with professional learning networks and technical assistance focused on effective literacy instruction in their early grades (i.e., transitional kindergarten through third grade).[3] Through a competitive selection process, the state

[3] In California, County Offices of Education (COEs) are administrative units that provide various services and specific educational programs in support of area school districts.

selected the Sacramento County Office of Education (SCOE) as the Expert Lead and provided it with $3 million in support of this effort, allotted separately from the $50 million lawsuit settlement.

Second, the ELSBG program identified four specific categories of allowable grant expenditures but allowed schools the flexibility to design their Literacy Action Plans within these requirements. These expenditure categories included (1) high-quality literacy teaching (e.g., new instructional coaches, increased professional development), (2) support for literacy learning (e.g., diagnostic assessment tools, instructional materials), (3) pupil supports (e.g., tutoring, after-school programming) and (4) family and community supports (e.g., mental-health resources, parental outreach and training). The statutory language also required schools to "consult with stakeholders, including school staff, school leaders, parents, and community members" when creating their Literacy Action Plans and for plans to be approved by the school district or CMO governing board during a public meeting to ensure that the plans were informed by the needs of the specific school site (Authorization of the ESLB Grant, 2020).

Third, ELSBG articulated specific restrictions in support of its policy goal to improve literacy. In particular, the program required that schools use these resources to supplement, "not supplant," existing activities and to focus these new resources at the targeted early grades. In support of oversight on these requirements, grantees also had to submit quarterly reports showing expenditures consistent with the approved budget and an annual report examining progress towards the activities and explicit goals articulated in the Literacy Action Plan. Funding in the second and third years is contingent on the submission of such quarterly and annual reports.

These design features of the ELSBG program—targeted state funding, external support from a competitively selected county office, spending flexibility within specified guidelines, and oversight—share the common motivation of supporting a high-fidelity implementation of effective literacy practices at scale across the state's lowest-performing schools. To examine how these funds were spent and the character of ELSBG-funded activities, we relied on several different sources of information including documentation from the Expert Lead in Literacy, school budgets and expenditure forms, and news accounts.

During the drafting period for school-level Literacy Action Plans (i.e., December 2020 to June 2021), SCOE hosted 36 sessions for 3,300 participants including staff in eligible counties, districts, and schools covering nine different topics related to specific literacy-improvement strategies and emphasizing the "science of reading" (Sullivan, 2020). For example, Session 1 introduced the concepts of phonological awareness, phonemic awareness, letter knowledge, decoding, and word recognition and provided links to free assessment tools that schools could use to evaluate the current state of these skills in their students. Sessions 2 focused on vocabulary and comprehension, while Session 3 focused on how best to select texts for read-a-louds and how to monitor and assess reading skills in students. In Session 4, schools brought their monitoring data to discuss what components of reading instruction seemed to present the most challenges for their students and what strategies could be employed as part of a Literacy Action Plan to address this.

As the implementation of the Literacy Action Plans began in July 2021, SCOE sponsored the participation of 336 coaches, teachers, and administrators in the "Online Elementary Reading Academy." A non-profit group, CORE Learning, hosted this asynchronous virtual course focused on effective instructional practices linked to the science of reading. SCOE also contracted with Pivot Learning, of which CORE is a subsidiary, to facilitate a series of Plan-Do-Study-Act sessions supporting school-site teams in identifying and implementing changes in their literacy-related practices. Across these sessions, an average of 58 ELSBG-eligible schools participated. SCOE itself facilitated monthly sessions for literacy coaches and provided ongoing assistance in office hours totaling 748 hours of "direct school support" with an additional 948 hours spent planning, hosting, or attending professional-development offerings as well as sending weekly emails with resources and programming reminders to ELSBG principals and district leads (Sullivan, 2022).

To determine which actions were taken by districts themselves, we collected budgets from all 35 ELSBG-funded school districts or CMOs, either by locating them on their websites or by contacting district or school-level staff directly. We found that staff compensation (i.e., salaries and benefits) represented 69 percent of the budgeted expenditures for the ELSBG funds in the first year. About 80% of this staffing budget paid certificated staff (i.e., school employees with a license for their position such as teachers, instructional coaches, or administrators) while the remainder supported "classified" staff (i.e., non-licensed school staff such as parent engagement coordinators and paraprofessionals).

Expenditure forms and recent news accounts provide more granular detail on how ELSBG funds were spent. For example, at one ELSBG school on the Central Coast, the school successfully hired a new Curriculum Coach and a new Parent Liaison. This literacy curriculum coach then trained staff on phonemic awareness while the Parent Engagement Specialist organized a Family Literacy Night (Klappenback & Marsh, 2022). At a different ELSBG school in Southern California near the Mexican border, the school administered the Basic Phonics Skills Test to all K-3 students at the beginning and end of the year but could not conduct the planned data discussions with teachers due to the limited availability of substitutes during the COVID-19 pandemic (Huerta-Price & Sanchez, 2022). A third ELSBG school in San Jose used the grant to hire a part-time literacy coach who met with teachers weekly to "support developing word recognition scope and sequence and instructional guidelines" and led professional development; the same school also purchased a new assessment and data system to monitor student progress (Black & Corrie, 2022). At an ELSBG school in Sacramento, the principal hired a literacy coach and two instructional aides. The school also spent money on purchasing new books for the school library with more culturally relevant material. Another ELSBG school in Los Angeles purchased and implemented a new curriculum that includes dedicated time for phonemic awareness, phonics, and reading comprehension (Lambert et al., 2022).

Public comments by the Expert Lead in Literacy provide a summative characterization that stressed what is observed in the news accounts: the flexibility to tailor ELSBG programming to local contexts. At a roundtable hosted by the education journalism outlet *EdSource*, Becky Sullivan–Project Lead for SCOE–explained the grant in her own words. She said, the goal at the beginning was to "get common language out there among all the participants in the grant… It was all based on the site and district data and their needs and their context. We did not tell them what to do, what to buy, who [sic] to hire. We introduced them to a process, and we are training them, giving them information" about the science of reading (D'Souza & Vasquez, 2022). Other roundtable participants underscored how ELSBG increased practitioners' understanding of and appreciation for the science of reading. One principal noted "One of the things I think this grant brought to us was the shared common understanding of what the science of reading is and that we do have the ability to teach our students in a way that is research-based with best practices… We had been looking for how do we meet our students' needs." (D'Souza & Vasquez, 2022).

However, whether these efforts were actually successful in improving early literacy outcomes for targeted students is an open empirical question. As noted earlier, the limited evidence available on other initiatives grounded in the science of reading (e.g., Gamse et al., 2008; Westall and Cummings, 2023) is not encouraging. Relatedly, the ELSBG initiative also has close parallels to the targeted and differentiated school-accountability policies that characterized the waiver era under No Child Left Behind (NCLB) and current policy under the Every Student Succeeds Act (ESSA). The evidence on the implementation quality and impact of those reforms is at best mixed (e.g., Bonilla & Dee, 2020; Dee & Dizon-Ross, 2019; Hemelt & Jacob, 2017). In the next sections, we turn to the data and quasi-experimental research designs that will allow us to provide evidence on how the ELSBG initiative influenced student achievement during its first two years of implementation.

Data

Our study relies on the publicly available data from the state of California's assessment system for public schools: the California Assessment of Student Performance and Progress (CAASPP). Specifically, we constructed panel data at the school, subject, and year levels using scores on the Smarter Balanced Summative Assessments in English Language Arts/Literacy (ELA) and mathematics among both third graders (i.e., the only tested grade that is ELSBG-eligible) and fifth graders as a comparison group.[4] These annual data span the period from the beginning of CAASPP in the 2014-15 school year to the 2022-23 school year. This implies seven years of available data given the necessary exclusion of the spring 2020 and 2021 assessments, which were either not given or taken by very few students due to the disruptions of the COVID-19 pandemic. Our two years of post-treatment test scores (i.e., those taken in spring 2022 and spring 2023) correspond to the first two years of ELSBG implementation.

Using data from the California Department of Education's "Public Schools and Districts Directory" file, we began by identifying all the conventional elementary-grade public schools, both traditional and charter, operational between 2015 and 2023 (i.e., 6,717 schools). We then excluded 400 schools with unconventional school structures (e.g., juvenile-justice halls, home and hospital programs, and dedicated special-education schools). We also excluded 139 schools identified as offering "Primarily or Exclusively Virtual Instruction" because only conventional in-person schools were eligible for ELSBG. Finally, we dropped schools who were not eligible for the grant because

[4] Each ITT school also served grade-5 students. However, almost half of the ITT schools do not serve students in grade 6 or above. This leads us to use grade-5 achievement as a comparison group within each school and year for some of our results (i.e., those based on a DDD design).

they did not report test scores in 2018 and 2019 when the assignment variable was calculated. Specifically, because California does not report test scores for any group with 11 or fewer students, 671 small schools reported missing test scores for third grade in both 2018 and 2019, were ineligible for ELSBG support, and are excluded from the sample. This leaves us with an unbalanced panel of 5,507 unique schools. In most of our analyses, though, we also exclude unbalanced panel observations (i.e., schools without reading-achievement data in each of the seven school years) as our preferred research design (i.e., synthetic difference in differences) requires a balanced panel. The modest degree of missingness associated with unbalanced panel observations reflects a variety of factors such as small schools with suppressed test-score data and some school closures or openings from 2015 to 2023. However, we find in auxiliary regressions (see Table A1) that this missingness of school-year observations is unrelated to ELSBG eligibility.

Our main analytical sample therefore consists of a balanced panel of 5,256 unique elementary schools with reading-achievement data for grade-3 in each of the seven school years (i.e., n = 36,792). This sample includes 66 intent-to-treat (ITT) schools (i.e., schools eligible for ELSBG), all but two of whom participated in the state initiative.[5] We note that the number of balanced school-year observations for other grades and subjects (i.e., grade-3 math, grade-5 math and ELA) varies slightly due to the censoring of those test outcomes when there were few test takers. Similarly, in specifications that condition on school-year covariates (i.e., percent White, percent eligible for free/reduced-price lunch, and the natural log of enrollment), sample sizes are somewhat smaller due to missingness. We

[5] As noted earlier, one school closed after the 2018-19 school year but before the grant application had opened. One school opened in August 2016, leading to missingness in 2015 test scores. Another school closed in June 2021, after the planning year of the grant, while two additional schools closed after the 2021-22 school year. Five schools remained open from 2015 to 2023 but had their test score data censored in one of the seven years of our study because they had fewer than 11 third graders take the ELA test that year. These ten schools reduce the ITT sample for which we can observe outcomes from 76 to 66. Two schools declined to apply for the grant but remain in our ITT sample as their decision to not apply may be endogenous to their outcomes.

also note that one of the ITT schools in our main analytical sample is a charter school and that our results are similar when excluding all charter schools from our analysis.[6]

We present the school-by-year academic achievement of California elementary schools in Table 1. As expected, we observe that ITT schools (i.e., those offered the opportunity to apply for the ELSBG based on their low performance on 2017-18 and 2018-19 standardized ELA tests) have much lower test scores in ELA than comparison schools. Specifically, in ITT schools, only 31.15 percent of students score at a Percent Level 2 or higher (Standard Nearly Met, Standard Met, or Standard Exceeded). In other words, more than two-thirds of students in these schools are scoring at the lowest level (Level 1, or Standard Not Met) on their standardized tests in ELA. In schools that were ELSBG-ineligible, the average of this reading proficiency rate was over twice as large (i.e., 67.87 percent). We also constructed parallel test-score measures for grade-3 mathematics and for grade-5 mathematics and ELA. These measures allow us to assess the potential spillover effects of the ELSBG initiative. Additionally, under the assumption of no spillover effects, they also make it possible to estimate the effect of the ELSBG initiative in difference-in-difference-in-differences (i.e., "triple diff") specifications that control for confounds that are both school-specific and time-varying.

Our data also include school-year measures of student demographic and socioeconomic traits as well as school enrollment based on the National Center for Education Statistics' Common Core of Data. In Table A2, we present the baseline (i.e., 2014-15 to 2018-19) averages of these variables by ITT status. These data indicate that ITT schools were, on average, smaller and served substantially higher concentrations of economically disadvantaged students as well as Black and Hispanic students. For example, roughly 90 percent of students in ELSBG-eligible schools were also eligible for free or

[6] In the fuller unbalanced panel that includes 76 ITT schools, four are charter schools. Two were open throughout the time period but had their data censored due to their small size, while one opened in August 2016 and is thus missing 2015 test scores.

reduced-price lunches and nearly 20 percent were Black. The corresponding averages in ELSBG-ineligible schools were 61 percent and 5 percent, respectively.

Methods

Pre-Registration

The growing concern over the credibility of scientific conclusions that rely on multiple forms of researcher discretion (e.g., the choice of outcome variables and research designs) motivated our approach to examining the achievement impact of the ELSBG initiative. In particular, evidence for the prevalence of publication biases and/or specification searching (i.e., "p-hacking") exists across multiple disciplines. Moreover, it appears to be a particular concern in quasi-experimental settings like ours (Brodeur et al., 2020). We pre-registered our preferred analysis plan to address this fundamentally important concern and to provide a transparent "decision tree" for our subsequent design choices (Nosek et al., 2018).[7]

We initially proposed a regression-discontinuity (RD) design that leveraged the cross-sectional variation in a baseline school-level test score (i.e., the percent of grade-3 students scoring at Level 2 or higher on the ELA exams during the 2018 and 2019 assessments) used to identify the ELSBG-eligible schools. We correspondingly designated the Percent Level 2 or higher on the post-treatment grade-3 ELA exam as the single confirmatory outcome measure. The key results of this RD approach are presented visually in Figure A2 and parametrically in Table A3.

These results show that schools with baseline test scores below the eligibility threshold were 96 percentage points more likely to participate in the ELSBG initiative (Table A3)—a virtually "sharp" assignment to treatment that is represented visually in panel A of Figure A2. Furthermore, the full-

[7] Our pre-registration is available at osf.io/5jgwu and dated April 26, 2022. Our outcome data (i.e., spring 2022 assessments) were not publicly available until October 24, 2022.

sample results based on this RD design indicate that this ITT (i.e., ELSBG eligibility) increased post-treatment grade-3 ELA scores (i.e., the percent Level 2 or higher) by nearly 8 percentage points in the first year of implementation (p-value < 0.01; column (2) in Table A3). However, we also find that this statistically significant finding is not robust to alternative functional forms (i.e., local linear regressions and quadratic splines of the assignment variable as seen in Table A3) nor does it persist into the second year of treatment (i.e., columns 4 and 5). Furthermore, a visualization of the reduced-form relationship between the post-treatment test measure and the assignment variable (i.e., panel B of Figure A2) does not provide clear evidence of an impact, possibly due to the lack of power from the small sample size.

Difference-in-Differences (DID) Design

In light of the ambiguity in these findings, we turned to a "difference-in-differences" (DID) design based on school-year panel data. While the internal validity of a DID design is generally more difficult to establish relative to an RD design, it also has two distinct advantages because it no longer relies on observations close to the eligibility threshold. One is a likely increase in statistical power. Second, the causal estimand from a DID design more reliably identifies the average impact of ELSBG eligibility rather than an effect that is potentially distinctive to observations local to the eligibility threshold.

However, the shift to a DID approach introduces an important issue of construct validity with respect to our pre-registered outcome measure. Specifically, once we turn from an RD design to an DID design and thus enlarge our comparison group beyond the threshold, there are potentially serious difficulties in interpreting comparative changes in a proficiency-rate measure (i.e., the pre-post change in ELSBG-eligible schools relative to the contemporaneous changes in comparison schools). Prior studies have carefully explicated this issue (Ho, 2008; Holland, 2002). In our DID context, the specific issue with proficiency-rate outcomes reflects the fact that the ELSBG-eligible schools are, by construction, drawn from the left tail of the test-score distribution of schools (e.g., 31 percent Level

2 or higher, Table 1) while the comparison schools have a right-shifted test-score distribution (i.e., 68 percent Level 2 or higher). This implies that, if test-score distributions in both treatment and comparison schools changed *by the same amount* before and after treatment occurred (i.e., no treatment effect), the proficiency-rate changes across treatment and comparison schools could differ. That is, DID-based treatment estimates using a proficiency-rate outcome—including the one we pre-registered—are now subject to potential biases. Given this important issue (and, also, the possibility that the ELSBG initiative has heterogeneous effects across the test-score distribution), our analysis not only focuses on all proficiency-rate measures (i.e., Level 2 or higher, Level 3 or higher, Level 4) but, also includes scale scores as a key outcome measure.

Our initial DID analysis focused on a conventional two-way fixed effects (TWFE) specification:

$$Y_{st} = \alpha_s + \beta_t + \tau D_{st} + \varepsilon_{st} \qquad (1)$$

in which outcome Y_{st} in school s and year t is a function of school and year fixed effects (i.e., α_s, δ_t), a binary indicator for ELSBG eligibility (i.e., D_{st}), and a mean-zero error term. We note that TWFE-based estimates of the parameter of interest, τ, have a DID interpretation because there is no variation in treatment timing with all ITT schools offered the opportunity to apply to the grant at the same time. However, the internal validity of this approach relies critically on a parallel-trends assumption that states the outcome changes in comparison schools over time provide a valid measure for how the untreated potential outcomes of the ITT schools (which are unobservable) would have changed over time. Event-study estimates (see Figure A3) provide evidence inconsistent with this assumption. Specifically, across all 4 test measures, we see that the ELA scores of ITT schools were trending significantly downward relative to comparison schools *before* the ELSBG initiative began.

Synthetic Difference-in-Differences Design

We address the implied internal-validity threat to this DID-based approach in several ways. Our main approach is to rely on a synthetic difference-in-differences (SDID) estimator (Arkhangelsky et al., 2021). The SDID approach combines attractive features of both DID and synthetic-control procedures. Like DID, it is invariant to additive unit fixed effects (i.e., different outcome levels) and allows for valid large-panel inference. Critically, like synthetic control, it also weakens the reliance on a parallel-trend assumption by constructing unit-specific weights, $\hat{\omega}_s^{sdid}$, that optimally align pre-treatment trends across treated and comparison units. The SDID procedure also introduces time-specific weights, $\hat{\lambda}_t^{sdid}$, that place more emphasis on pre-treatment periods that are similar to the post-treatment period.[8] Given these weights, the SDID procedure forms an estimate of the effect of interest (i.e., $\hat{\tau}^{sdid}$) through this least-squares minimization:

$$\left(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\operatorname{argmin}} \left\{ \sum_{s=1}^{N} \sum_{t=1}^{T} (Y_{st} - \mu - \alpha_s - \beta_t - D_{st}\tau)^2 * \hat{\omega}_s^{sdid} * \hat{\lambda}_t^{sdid} \right\} \quad (2)$$

To conduct statistical inference for estimates based on equation (2), we rely on a block-bootstrap procedure (Arkhangelsky et al., 2021; Algorithm 2), which, though computationally intensive, performs well particularly in settings like ours where the number of treated units is large (Clarke et al., 2023). The central appeal of the SDID procedure is that it addresses internal-validity concerns by focusing on comparisons between treated units and similar comparison units (i.e., a type of "localness" noted by Arkhangelsky et al., 2021). The simultaneous use of both unit and time weights also enhances this localness by giving this procedure "a type of double robustness property" that reduces the influence of potential biases related to any one weight that may be misspecified (Arkhangelsky et al., 2021; Liu et al., 2022). Arkhangelsky et al. (2021) also note that an important but less intuitive benefit of SDID's localness is that it is likely to improve statistical precision through weighting that systematically removes the predictable components of the outcome measures.

[8] See Algorithm 1 in Arkhangelsky et al. (2021) and Clarke et al. (2023) for details on the construction of these weights.

Robustness Checks

We explore the credibility and robustness of SDID-based results in several ways. First, we present results with and without covariate adjustments.[9] Second, we present event-study estimates based on equation (2), which provides visual evidence on whether the SDID procedure effectively addressed the parallel-trend violations observed in the conventional TWFE approach (Figure A3). Third, we also present results that crudely enforce a type of localness. The 76 ITT schools in our study are, by construction, the lowest-performing schools in the state on a baseline ELA test measure while the nearly 5,300 comparison schools represent the remainder of the entire state. In some reported results, we limit the analytical sample to schools that are more similar to the ITT schools on the baseline test measure (e.g., the bottom 4,000 schools, 3,000 schools, etc.).[10]

Fourth, we present results based on three different "difference-in-difference-in-differences" (DDD) designs that provide alternative approaches to addressing parallel-trend violations. The DDD approach tacitly assumes that non-focal school and year-specific test scores (e.g., math in grade-3, math and ELA in grade-5) do not reflect spillover effects of the ELSBG initiative but provide a potential control for unobserved confounds specific to each school-year observation. Specifically, our DDD design conditions on an unrestricted set of two-way fixed effects (i.e., school-by-year, school-by-subject, subject-by-year) and estimates the effects of the three-way interaction of interest (i.e., the treated subject in treated schools observed in the post-treatment period).

We also note that our DDD results provide important evidence on the empirical relevance of another potential internal-validity threat that may not be well-addressed by the SDID procedure.

[9] Covariates included are the percent of a school that is White, percent of a school that receives free or reduced-price lunch, and the natural log of a school's total enrollment. Percent White tends to be more commonly missing, as some schools without any White students leave the item blank rather than entering a zero. For the 184 school-year observations that are missing a percent White but do have other enrollment data, if a school's enrollment in other racial/ethnic groups equals at least 97 percent of their total enrollment, we impute that their percent White is zero.

[10] We note that these ad-hoc sample restrictions are inconsistent with the data-driven choices made by the SDID procedure. Specifically, the SDID unit weights are positive for over half the schools in the data and draw heavily from schools throughout the distribution of the baseline ELA measure used to determine ELSBG eligibility.

Specifically, evidence clearly suggests that the COVID-19 pandemic had negative effects on measures of learning and that these effects were larger among more disadvantaged students (Kuhfeld et al., 2022). Given that the ELSBG-eligible schools are drawn from the bottom of the ELA-score distribution, a possible concern is that they have a post-pandemic shock to test scores that is distinct from their comparison units (e.g., a unique negative bias that imparts a downward bias). In other words, SDID may fail to achieve "localness" because its weighting largely relies on pre-treatment (i.e., pre-pandemic) data—an approach that may confound the effects of the ELSBG initiative with the pandemic's effects that are unique to ELSBG-eligible schools. The DDD results we present provide direct evidence on this issue because they condition on school-by-year fixed effects and effectively rely on within-school comparisons across grades and subjects that are made entirely in the post-pandemic period. We note that these DDD estimates would represent a lower bound on the true impact of ELSBG eligibility on grade-3 ELA achievement if the reading reforms had spillover benefits for grade-3 math achievement or for grade-5 math and ELA outcomes.[11] We present direct evidence on this question by presenting SDID estimates of the effect of ELSBG eligibility on these other achievement outcomes.

Two other robustness checks are of note. First, for 4.6 percent of the unique schools in the sample, the school-year panel data used are unbalanced because test-score outcomes are missing due to closures and the censoring of data from schools with fewer than 11 test-takers in a grade and subject. If ELSBG eligibility influenced this missingness (e.g., through effects on closures or the number of test takers), it could introduce a form of selection bias. We present auxiliary regression SDID estimates in Table A1, which indicate that this missingness is unrelated to ELSBG eligibility. Second, because our results rely on school-level data, it is possible that treatment-endogenous sorting

[11] ELA-focused gains in grade-3 teacher performance and student outcomes could conceivably improve grade-3 math outcomes. Similarly, reform-related changes in the ELA pedagogy of teachers in grade K-3 could influence teachers and students in other grades as well. We also note that 2022-23 fifth graders were in grade 3 during the ELSBG planning year.

(e.g., choosing to enroll or remain in a school because of its ELSBG eligibility) is an internal-validity threat. We present auxiliary regressions in Table A4, which demonstrate that the numbers of test-takers across the four subject-grade combinations are generally unrelated to ELSBG eligibility across a wide variety of specifications.

Results

We present our main results (i.e., the estimated effects of ELSBG eligibility on the four different measures of ELA achievement) in Table 2. These SDID-based estimates consistently indicate that ELSBG eligibility had positive and statistically significant effects on grade-3 ELA test scores. The results based on proficiency-rate outcomes indicate that the program was not only successful at improving achievement in the left tail of the ELA-score distribution (i.e., the group targeted by the initiative) but also raised achievement elsewhere in the distribution. Specifically, the estimated increase in the percent of students scoring at Level 2 or higher on the ELA assessment due to ELSBG eligibility was 6.00 percentage points (p-value < 0.01). To put this estimated effect into perspective, we note that it constitutes a 20 percent increase relative to the baseline level of students at or above Level 2 in ELSBG-eligible schools.[12] The estimated effects on the share of students scoring at or above Levels 3 and 4—4.98 percentage points and 1.82 percentage points, respectively—are also statistically significant (p-value < 0.01). Because so few students scored at the upper end of the distribution at ELSBG-eligible schools in pre-ELSBG years, these represent large percent changes; ELSBG led to a 42 percent change in the percent of students at or above Level 3 and a 59 percent change in the percent of students at Level 4. Notably, the results based on scale scores indicate that ELSBG eligibility increased ELA achievement by 14 percent of a student-level standard deviation (i.e., 0.14

[12] We also note that this impact estimate is similar to the full-sample result based on our pre-registered regression-discontinuity (RD) design (i.e., column 2 in Table A2).

SD, $p < 0.01$). Given that the annual reading-achievement gains of children between grades 2 and 3 are, on average, 0.60 standard deviations (Hill et al., 2008), this effect size associated with ELSBG eligibility implies a gain of nearly a quarter of a year of learning.

The SDID estimates in Table 3 assess whether ELSBG eligibility influenced grade-3 math achievement or grade-5 achievement in ELA or math. In theory, such spillover effects could have been negative if initiatives focused on literacy in early grades detracted effort and attention from learning opportunities in other grades and subjects. Alternatively, the impact of the ELSBG initiative on non-focal grades and subjects could have been positive by building teacher capacity in eligible schools and improving student literacy in ways that supported learning in math. The results in Table 3 suggest that ELSBG eligibility had positive spillover effects for math performance among the focal grade-3 students. Specifically, this estimated effect size—a gain of 0.11 standard deviations—is equivalent to 12 percent of a year of learning in mathematics at this age (Hill et al., 2008). In contrast, we do not see consistent evidence of effects on ELA or math test scores among 5th graders, who were outside ELSBG's focal grades.

We examine the robustness of the results in Tables 2 and 3 in several ways. First, we note that the estimates in Tables 2 and 3 are similar in specifications that condition on outcome-relevant covariates that vary within schools over time.[13] Second, for each of the four testing outcomes across each subject-grade combination, we constructed event-study estimates that identify how the ELSBG-eligible schools and their weighted comparisons trended in each period before treatment. For example, the results for the focal grade-3 ELA measures, presented in Figure 1, suggest that the SDID procedure was effective in eliminating the parallel-trend violations that were apparent in conventional

[13] We also underscore relevant robustness checks noted earlier. In auxiliary regressions, we do not find any consistent evidence that either missingness in the school-year panel data or the number of test-takers is treatment-related (Tables A1 and A4).

DID estimates based on TWFE specifications (Figure A3).[14] The event-study results for the test measures associated with grade-3 math and grade-5 ELA and math similarly suggest the absence of parallel-trend violations (Figures A4, A5, A6). Third, we also note that, for all subject-grade combinations, we find similar results when we increasingly limit the set of comparison schools to those that, like the ELSBG-eligible schools, were in the bottom of the distribution of baseline ELA schools (Tables A6, A7, A8, and A9). We find that this is true even as the set of unique schools available to the SDID procedure shrinks from over 5,000 to only 500.

Fourth, we estimated the effects of ELSBG eligibility on the grade-3 ELA score measures across three types of DDD specifications that relied on different grade and subject groups (i.e., grade-3 math, grade-5 ELA and math) as a placebo. Those results consistently indicate, across all test-score measures and comparison groups, that ELSBG eligibility increased grade-3 ELA achievement (Table A10). The DDD estimates are smallest (i.e., effect size = 0.05; p-value < 0.01) when grade-3 math is treated as the comparison condition, which is to be expected given the evidence that ELSBG implied positive spillover effects on math achievement. We note that these DDD results provide an important complement to our main findings because they condition on fixed effects unique to each school-year combination in the data. This may be particularly important given the concern that the COVID-19 pandemic uniquely harmed the learning opportunities in ELSBG-eligible schools which, by construction, were at the bottom of the state test-score distribution. The fact that DDD estimates with different grade-subject comparison groups consistently indicate that ELSBG eligibility increased grade-3 ELA achievement suggests that this is not an empirically salient confound.

Fifth, we present our results excluding charter schools. Though we only have one charter school in the balanced sample of ITT schools, the traditional SDID methodology used to create Table

[14] Interestingly, despite the evidence of pre-trends in TWFE-based estimates, the impact estimates based on that approach (Table A5) are similar to those reported in Table 2.

2 and Table 3 allows for charter schools to be used as comparison schools for all ITT schools. When we exclude charter schools from sample entirely, our results remain the same (Table A11).

Discussion

As the result of a legal challenge, the state of California recently undertook a focused effort to improve early literacy among K-3 students at more than 70 of the state's lowest-performing public schools. This initiative focused on promoting literacy practices grounded in the "science of reading" and featured several other distinctive design features relevant to its implementation. These included external support and oversight from a competitively selected county office, the development of school-specific and community-informed Literacy Action Plans, additional state resources, and flexibility in the use of those resources subject to state guidelines.

This study provides quasi-experimental evidence on the early impact of this state-level effort on ELA achievement. Specifically, we find that that ELSBG eligibility increased ELA test scores by 0.14 standard deviations among the more than 7,000 third graders served by the targeted schools over the first two years of the grant. This is a larger effect size than almost 90 percent of educational interventions serving more than 2,000 students (Kraft, 2023). Similarly, it increased the share of students performing at Level 2 or higher by 20 percent (i.e., a 6.00 percentage-point gain relative to a pre-treatment baseline of 30.56 percent). We also find that this initiative also led to smaller gains in grade-3 math achievement (i.e., 0.11 standard deviations) and, as intended, had no effects among grade-5 students outside the program's focus. These results are also particularly notable because this effort to close the research-to-practice gap in early literacy occurred in the lowest-performing schools in the state during an unprecedented global pandemic. As Becky Sullivan, Project Lead for the Sacramento County Office of Education and an architect of the ELSBG roll-out, said: "If the lowest schools in the state can show gains under the conditions we've had the last two years, it's definitely a win" (D'Souza, 2022).

Three caveats to these encouraging findings also merit attention. First, we are only able to track the direct outcomes of this new initiative over its first two years. Whether schools—and participating students—are able to sustain these gains is an open question, especially given the evidence that the benefits of reading interventions sometimes phase out over time (May et al., 2022). We also note that teacher turnover (e.g., the loss of newly trained teachers) may mediate the capacity of these targeted schools to sustain these improvements. Second, while it is possible that ELSBG's impact will strengthen as both students and teachers extend their program participation, we do not see evidence of this across the first two program years. Third graders in the first year of implementation (i.e., the 2021-22 school year) scored similarly to 2022-23 third graders though most of the latter had also been exposed to the ELSBG initiative in the second grade and were being taught by teachers in their second ELSBG year. Third, our study focused almost exclusively on grade-3 ELA achievement, the key intended outcome of the ELSBG initiative. Whether the gains on California's high-stakes assessment map onto broader skill gains (Volante, 2004; Westall & Cummings, 2023) and other educational outcomes is also an open question. We do note, though, that our evidence suggests the ELSBG initiative generated spillover benefits for math achievement among grade-3 students, which indicates the broader relevance of this study's main findings.

We also note that the test-score gains attributable to the ELSBG initiative should be evaluated with regard to their costs. The first-year ELSBG implementation budget was $17.8 million (i.e., $15.8 million allocated to schools, $1 million spent by the Sacramento County Office of Education as the Expert Lead in Literacy, and $1 million spent by the California Department of Education). These resources supported the program among 15,541 K-3 students in 75 schools, implying an average one-year cost of $1,144 per pupil. The ELA learning gains per dollar spent on the ELSBG initiative in the first year—0.13 SD per $1,000 (2021 dollars)—compare favorably to other notable interventions focused on children at these grade levels. For example, the learning return on this investment far exceeds that associated with the class-size reductions in Project STAR. Specifically, Krueger (1999) argues that the one-year cost of Project STAR's learning gains (0.22 SD) is equivalent to a third of total expenditures per pupil (i.e., roughly $5,500 per pupil in California). This implies a return of 0.04 SD per $1,000 spent on early-grade class-size reductions, which is less than a third of the return on ELSBG spending indicated by this study's results. Another highly policy-relevant point of comparison for this highly targeted initiative is the return on unrestricted increases in school spending. Results from Jackson & Mackevicius (2021) suggest that a one-year spending increase of $1,000 in 2021 dollars increases test scores by 0.0097 SD.[15] This implies that ELSBG's targeted spending (i.e., focusing on early literacy in the lowest-performing schools) is 13 times more cost effective than a generalized increase in school spending.

These findings suggest that programmatic efforts similar to the ESLBG initiative merit continued interest from policymakers and practitioners. However, replicating (or taking to greater scale) ELSBG's encouraging early results is unlikely to be straightforward. In particular, its distinctive design and implementation details (e.g., resources linked to evidence-based practices and flexible implementation within considered guidelines and oversight) are likely to be critical contributors to the findings reported here. Nonetheless, these results provide a proof point for how such focused efforts can help to realize, in a cost-effective manner, the educational potential of students served by our lowest-performing schools.

[15] Jackson & Mackevicius (2021) show that the average $1000 (in 2018 dollars) increase in per-pupil public school spending over four years (i.e., roughly a $4000 increase per student) increases test scores by 0.0352 SD. To make this comparable with ELSBG cost estimates for the first year of programming in SY 2021-22, we use the Consumer Price Index to adjust for inflation and divide by four to obtain an estimate for a single year.

References

Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. MIT Press.

Anderson, R. C., Hiebert, E. H., Scott, J. A., Wilkinson, I. A. G., & Reading, T. C. on. (1985). *Becoming a Nation of Readers: The Report of the Commission on Reading* (no edition stated). National Academy of Education.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, *111*(12), 4088–4118. https://doi.org/10.1257/aer.20190159

Black, A., & Corrie, E. (2022). *McKinley Elementary: Early Literacy Support Block Grant Annual Report*. McKinley Elementary. https://drive.google.com/file/d/1AyFewzkTWyLFdk_U1xHZY_DIgfNdkljd/view

Bonilla, S., & Dee, T. S. (2020). The Effects of School Reform under NCLB Waivers: Evidence from Focus Schools in Kentucky. *Education Finance and Policy*, *15*(1), 75–103. https://doi.org/10.1162/edfp_a_00275

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review*, *110*(11), 3634–3660. https://doi.org/10.1257/aer.20190687

Carr, S. (2022, December 27). NAACP targets a new civil rights issue—Reading. *The Hechinger Report*. http://hechingerreport.org/naacp-targets-a-new-civil-rights-issue-reading/

Chall, J. (1967). *Learning to Read: The Great Debate*. McGraw Hill.

Clarke, D., Pailañir, D., Athey, S., & Imbens, G. (2023). *Synthetic Difference In Differences Estimation* (arXiv:2301.11859). arXiv. https://doi.org/10.48550/arXiv.2301.11859

Dee, T. S., & Dizon-Ross, E. (2019). School Performance, Accountability, and Waiver Reforms: Evidence From Louisiana. *Educational Evaluation and Policy Analysis*, *41*(3), 316–349. https://doi.org/10.3102/0162373719849944

D'Souza, K. (2022, November 2). How badly did the pandemic deepen California's early reading crisis? *EdSource*. https://edsource.org/2022/how-badly-did-the-pandemic-deepen-californias-early-reading-crisis/680490

D'Souza, K., & Vasquez, A. (2022, April 28). *An unmet promise: Early literacy for all in California* [Roundtable]. https://edsource.org/broadcasts/early-literacy-for-all-in-california

Ella T. v. State of California: Settlement Implementation Agreement. (2020, February 20). Superior Court of the State of California, County of Los Angeles. https://media2.mofo.com/documents/171205-ellla-t-v-california-complaint.pdf

Flesch, R. (1955). *Why Johnny Can't Read and What You Can Do About It*. Harper & Row, Publishers.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., Unlu, F., Bozzi, L., Caswell, L., Rodger, C., Smith, W. C., Brigham, N., & Rosenblum, S. (2008). *Reading First Impact Study*. Institute of Education Sciences. https://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20094038

Goodman, K. S. (1989). Whole-Language Research: Foundations and Development. *The Elementary School Journal*, *90*(2), 207–221. https://doi.org/10.1086/461613

Hemelt, S. W., & Jacob, B. (2017). *Differentiated Accountability and Education Production: Evidence from NCLB Waivers* (Working Paper 23461). National Bureau of Economic Research. https://doi.org/10.3386/w23461

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Ho, A. D. (2008). The Problem with "Proficiency": Limitations of Statistics and Policy Under No Child Left Behind. *Educational Researcher*, *37*(6), 351–360.

Holland, P. W. (2002). Two Measures of Change in the Gaps Between the CDFs of Test-Score Distributions. *Journal of Educational and Behavioral Statistics*, *27*(1), 3–17. https://doi.org/10.3102/10769986027001003

Huerta-Price, B., & Sanchez, C. (2022). *CUSD ELSB Grant Annual Report Y1*. Calexico Unified School District. https://docs.google.com/document/d/1WI6EZrqyk3CaEoG3VczcIQ_lTYsBLGrJ/edit

Jackson, C. K., & Mackevicius, C. (2021). The Distribution of School Spending Impacts. In *National Bureau of Economic Research*. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28517/w28517.pdf

Joyce, K. E., & Cartwright, N. (2020). Bridging the Gap Between Research and Practice: Predicting What Will Work Locally. *American Educational Research Journal*, *57*(3), 1045–1082. https://doi.org/10.3102/0002831219866687

Klappenback, K., & Marsh, N. (2022). *Calabasas Elementary: Early Literacy Support Block Grant Annual Report*. Calabasas Elementary. https://www.pvusd.net/documents/Calabasas-ELSB-Grant-Annual-Report-Y1-21-22.pdf

Kraft, M. A. (2023). The Effect-Size Benchmark That Matters Most: Education Interventions Often Fail. *Educational Researcher*, 0013189X231155154. https://doi.org/10.3102/0013189X231155154

Kristof, N. (2023, May 31). Opinion | Mississippi Is Offering Lessons for America on Education. *The New York Times*. https://www.nytimes.com/2023/05/31/opinion/mississippi-education-poverty.html

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, *114*(2), 497–532.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test Score Patterns Across Three COVID-19-Impacted School Years. *Educational Researcher*, *51*(7), 500–506. https://doi.org/10.3102/0013189X221109178

Kurtz, H., Lloyd, S., Harwin, A., Chen, V., & Furuya, Y. (2020). *Early Reading Instruction: Results from a National Survey*. EdWeek Research Center.

Lambert, D., Tadayon, A., Sequeira, K., & D'Souza, K. (2022, October 6). The youngest readers tackle sounds, words and the science of reading. *EdSource*. https://edsource.org/2022/the-youngest-readers-tackle-sounds-words-and-the-science-of-reading/679248

Liu, L., Wang, Y., & Xu, Y. (2022). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*. https://doi.org/10.1111/ajps.12723

May, H., Blakeney, A., Shrestha, P., Mazal, M., & Kennedy, N. (2022, April 23). *Long-Term Impacts of Reading Recovery through Third and Fourth Grade: A Regression Discontinuity Study from 2011-12 Through 2016-17*. American Education Research Association, San Diego, CA.

NAEP Reading: National Achievement-Level Results. (2022). National Assessment of Education Progress. https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4

National Reading Panel. (2000). *Teaching Children To Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. https://eric.ed.gov/?id=ED444126

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Pressley, T., Allington, R. L., & Pressley, M. (2023). *Reading Instruction That Works*. Guilford Publications.

Preston, T. (2022, May). A chronicle of Kappan's coverage of the reading wars. *Kappan*, *103*(8), 5–7.

Public Counsel, & Morrison & Foerster LLP. (2017). *Ella T. v. State of California: Complaint*. Superior Court of the State of California, County of Los Angeles. https://media2.mofo.com/documents/171205-ellla-t-v-california-complaint.pdf

SB 98 Authorization of the ELSB Grant, (2020). https://www.cde.ca.gov/pd/ps/elsbgrantsb98.asp

Schwartz, S. (2023a, July 21). Which States Have Passed "Science of Reading" Laws? What's in Them? *Education Week*. https://www.edweek.org/teaching-learning/which-states-have-passed-science-of-reading-laws-whats-in-them/2022/07

Schwartz, S. (2023b, August 31). To Move Past the Reading Wars, We Must Understand Where They Started. *Education Week*. https://www.edweek.org/teaching-learning/to-move-past-the-reading-wars-we-must-understand-where-they-started/2023/08

St. Andre, E. (2020). *Early Literacy Support Block Grant Local Educational Agency Funding and Next Steps*. California Department of Education, Educator Excellence and Equity Division. https://www.cde.ca.gov/pd/ps/documents/elsbleawebinar.pdf

Sullivan, B. (2020). *ELSBG Year 1 PL Sessions*. Sacramento County Office of Education.

Sullivan, B. (2022). *Early Literacy Support Block Grant, Expert Lead in Literacy Annual Report Year 2: 2021-22*. https://drive.google.com/file/d/1cu3zRCNsNHSAMWMG24gdOVlHyVet7NRx/view

Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*. https://eric.ed.gov/?id=EJ848235

Watson, D. J. (1989). Defining and Describing Whole Language. *The Elementary School Journal*, *90*(2), 129–141.

Westall, J., & Cummings, A. (2023). *The Effects of Early Literacy Policies on Student Achievement* (23–788). EdWorking Papers. https://edworkingpapers.com/ai23-788

Whitehurst, G. J. (2009, October 14). Don't Forget Curriculum. *Brookings Institution*. https://www.brookings.edu/articles/dont-forget-curriculum/

Table 1—Descriptive Statistics for Test Scores by Grade, Subject, and Intent to Treat (ITT)

	Grade 3		Grade 5	
	Intent to Treat (ITT) Status		Intent to Treat (ITT) Status	
Test Outcome	ITT = 1	ITT = 0	ITT = 1	ITT = 0
English Language Arts (ELA)				
Pct Level 2 or Higher	31.15	67.87	36.87	68.29
	(11.27)	(17.76)	(12.53)	(17.25)
Pct Level 3 or Higher	11.69	43.66	17.54	47.62
	(7.66)	(20.74)	(8.86)	(20.35)
Pct Level 4	3.32	22.76	3.96	20.50
	(3.69)	(16.76)	(3.79)	(15.94)
Standardized Scale Score	-0.84	0.00	-0.76	0.00
	(0.25)	(0.47)	(0.26)	(0.47)
Sample Size	462	36,330	448	35,329
Mathematics				
Pct Level 2 or Higher	35.48	70.09	27.98	60.72
	(12.91)	(17.53)	(12.67)	(20.43)
Pct Level 3 or Higher	13.68	45.81	8.43	33.46
	(8.96)	(21.43)	(6.92)	(21.55)
Pct Level 4	2.68	18.86	2.70	17.58
	(3.78)	(16.30)	(3.35)	(16.65)
Standardized Scale Score	-0.82	0.00	-0.75	0.00
	(0.28)	(0.49)	(0.27)	(0.51)
Sample Size	469	36,330	448	35,322

Note: School-year test data are based on the California Assessment of Student Performance and Progress (CAASPP). Level 2 indicates Standard Nearly Met or higher, Level 3 Indicates Standard Met or higher, and Level 4 indicates Standard Exceeded. The standard deviation is indicated in parentheses below the mean. These are based on a balanced panel of all California elementary schools who report test scores in all 7 school years from 2014-15 to 2022-23, excluding SY 2019-20 when tests were not administered and SY 2020-21 when test participation was highly inconsistent due to the COVID-19 pandemic. The balanced panel of schools with grade 3 ELA test scores includes 5,256 unique schools, of which 66 are ITT.

Table 2—Estimated Effect of ELSBG on 3rd Grade ELA Test Scores

Dependent variable	(1)	(2)
Pct Level 2 or Higher	6.00***	5.74***
	(1.25)	(1.16)
Pct Level 3 or Higher	4.98***	4.61***
	(0.86)	(0.89)
Pct Level 4	1.98***	1.82***
	(0.51)	(0.49)
Standardized Scale Score	0.14***	0.14***
	(0.02)	(0.02)
Covariates?		X
N	36,792	34,384

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023) in grade-3 ELA. Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The school-year covariates are percent White, percent FRPL, and ln(enrollment). Because demographic data from the National Center for Education Statistics are not available for 2023 yet, covariates from 2022 are carried forward into that year. *p<.1. **p < .05. ***p < .01

Table 3—Estimated Effect of ELSBG on Other Grade-Subject Test Scores

	Grade 3		Grade 5			
Dependent variable	Math		ELA		Math	
Pct Level 2 or Higher	3.56***	3.45**	-0.56	-0.64	-1.15	-1.53
	(1.35)	(1.34)	(1.38)	(1.31)	(0.91)	(1.14)
Pct Level 3 or Higher	3.98***	3.65***	0.28	0.01	0.84	0.32
	(1.00)	(1.13)	(0.92)	(0.95)	(0.56)	(0.62)
Pct Level 4	0.63***	0.57	-0.54	-0.77*	0.37	0.16
	(0.50)	(0.53)	(0.46)	(0.43)	(0.30)	(0.33)
Standardized Scale Score	0.11***	0.11***	0.00	-0.01	-0.03	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Covariates?		X		X		X
N	36,799	34,384	35,777	33,537	35,770	33,530

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported the grade-subject test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The school-year covariates are percent White, percent FRPL, and ln(enrollment). Because demographic data from the National Center for Education Statistics are not available for 2023 yet, covariates from 2022 are carried forward into that year. *p<.1. **p < .05. ***p < .01

Figure 1—ELSBG Event-Study Estimates for Grade-3 ELA Test Scores

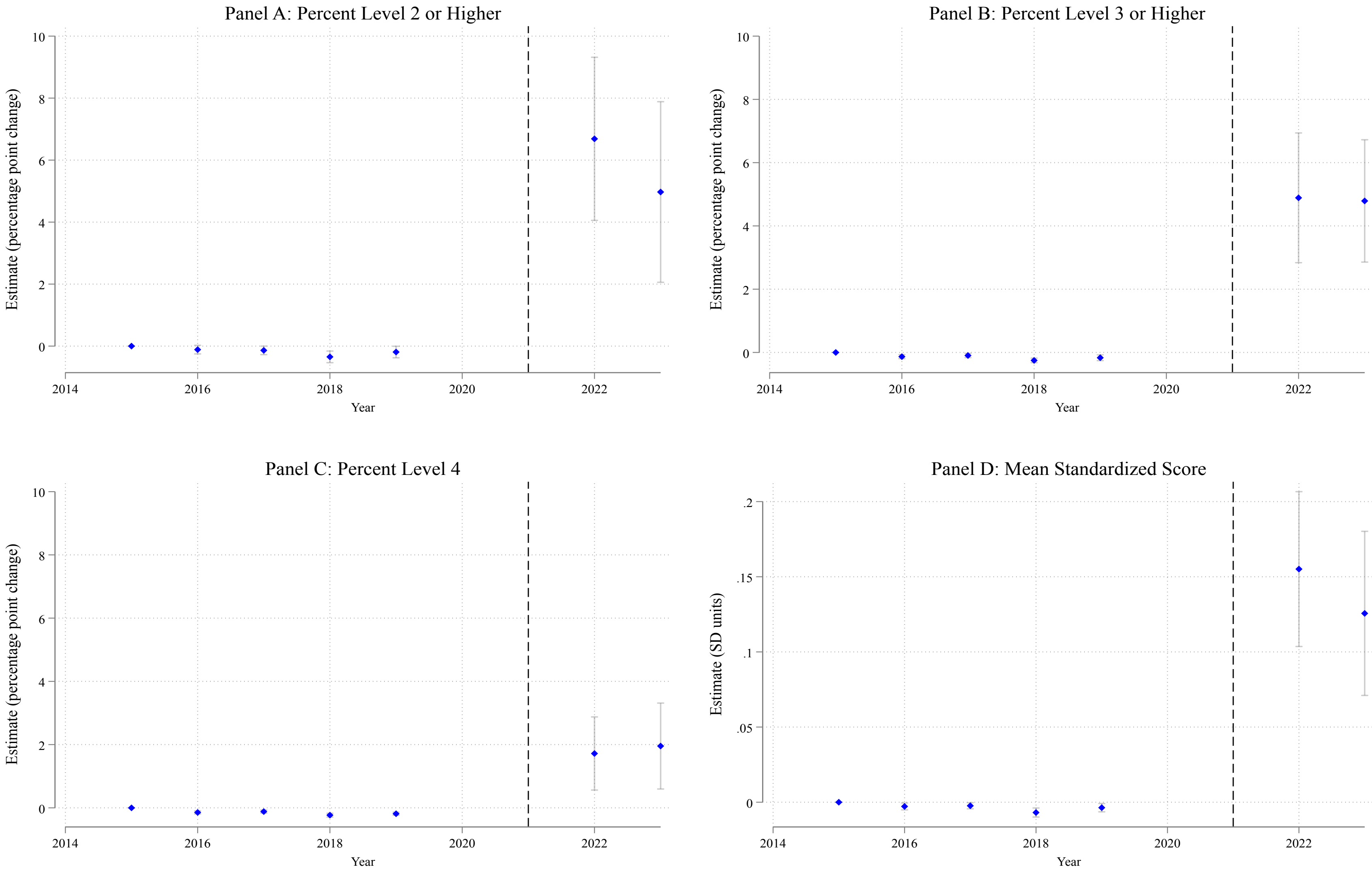


Table A1—Estimated Effects of ELSBG on Missingness

Sample Construction	Grade 3 ELA	Grade 3 Math	Grade 5 ELA	Grade 5 Math
Full sample	0.04	0.03	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
N	38,549	38,549	37,940	37,947
Bottom 4000 Schools	0.01	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
N	28,000	28,000	27,622	27,629
Bottom 3000 Schools	0.01	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
N	21,000	21,000	20,699	20,706
Bottom 2000 Schools	0.01	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
N	14,000	14,000	13,839	13,839
Bottom 1000 Schools	0.01	-0.00	0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)
N	7,000	7,000	6,944	6,944
Bottom 500 Schools	0.01	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
N	3,500	3,500	3,479	3,479

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of 5,507 California elementary schools that were in the risk set (i.e. reported test scores in 2017-18 and 2018-19 and thus were eligible for assignment to ELSBG). The dependent variable is missingness of grade-subject test scores for any reason (i.e., school closed, school opened, or school too small for data reporting). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A2—Descriptive Statistics for Baseline School Traits by ITT Status

	Intent to Treat (ITT) Status of School	
Variable	ITT = 1	ITT = 0
Percent Asian	6.02	11.13
	(7.24)	(15.53)
Percent Black	19.19	5.39
	(17.29)	(8.52)
Percent Hispanic	63.18	53.91
	(24.48)	(29.35)
Percent White	7.28	24.53
	(12.90)	(23.49)
Percent FRPL	88.93	60.92
	(8.29)	(29.13)
Enrolled Students	467.46	565.23
	(188.20)	(233.75)
Sample Size	329	25,665

Note: Cells indicate the conditional mean with the standard deviation in parentheses and are based only on school-year observations from the 5 pre-treatment years (2015-2019). The included schools are those in a balanced panel of California elementary schools that reported grade-3 ELA test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). The source is the National Center for Education Statistics Common Core of Data (CCD).

Table A3—First-Stage and Reduced-form Regression Discontinuity (RD) Estimates

	First Stage	Reduced Form—SY 2021-22		Reduced Form—SY 2022-23	
Sample Construction	(1)	(2)	(3)	(4)	(5)
Full Sample	0.96***	7.72***	-0.58	4.27*	-7.14**
	(0.03)	(2.23)	(3.05)	(2.46)	(3.36)
N	5507	5427	5427	5407	5407
+/- 2.0 SDs	0.96***	0.59	0.45	-3.26	-5.14
	(0.03)	(2.32)	(3.28)	(2.55)	(3.63)
N	2098	2052	2052	2038	2038
+/- 1.5 SDs	0.96***	0.55	-1.08	-2.97	-5.83
	(0.03)	(2.41)	(3.46)	(2.66)	(3.82)
N	1249	1222	1222	1210	1210
+/- 1.0 SDs	0.98***	-0.04	-5.01	-4.45	-8.88*
	(0.03)	(2.82)	(4.08)	(3.26)	(4.60)
N	593	575	575	570	570
+/- 0.5 SDs	1.02***	-3.92	-6.82	-7.08*	-11.58*
	(0.03)	(3.73)	(6.04)	(4.12)	(6.22)
N	201	193	193	190	190
Weighted (triangular kernel)	0.99***	-1.85	-5.13	-6.00*	-8.52*
	(0.02)	(3.02)	(4.35)	(3.34)	(4.68)
N	593	575	575	570	570
Optimal Bandwidth	1.00	-3.85	-6.57	-7.79	-13.72**
	(0.00)	(4.75)	(6.49)	(4.92)	(6.87)
N	104	119	170	135	125
Quadratic			X		X

Note: The first-stage dependent variable is ELSBG participation. The reduced-form dependent variable is the share of students scoring Level 2 or higher on the Grade-3 ELA exam. These estimates condition on linear splines of the assignment variable. Robust standard errors are reported in parentheses. The optimal bandwidth is based on Calonico et al (2014). *p<.1. **p < .05. ***p < .01

Table A4—Estimated Effect of ELSBG on the Number of Test Takers by Grade and Subject

Sample Construction	Grade-3 ELA	Grade-3 Math	Grade-5 ELA	Grade-5 Math
Full Sample	-0.04	-0.05	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
N	36,792	36,799	35,777	35,770
Bottom 4000 Schools	-0.03	-0.03	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
N	27,090	27,090	26,334	26,327
Bottom 3000 Schools	-0.01	-0.02	0.04	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
N	20,328	20,335	19,754	19,747
Bottom 2000 Schools	-0.01	-0.01	0.04	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
N	13,545	13,552	13,160	13,153
Bottom 1000 Schools	0.01	-0.00	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)
N	6,769	6,783	6,622	6,622
Bottom 500 Schools	-0.01	-0.01	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
N	3,388	3,388	3,283	3,283

Note: The dependent variables are the natural log of the number of test takers in the given subject and grade. These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported the grade-subject test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A5—Estimated Effect of ELSBG on Test Scores, TWFE-DID Specifications

Dependent Variable	Grade-3 ELA	Grade-3 Math	Grade-5 ELA	Grade-5 Math
Pct Level 2 or Higher	5.69***	1.82	-1.50	-1.70
	(1.26)	(1.38)	(1.65)	(1.38)
Pct Level 3 or Higher	3.79***	2.32**	-1.14	0.72
	(0.96)	(0.98)	(1.15)	(0.70)
Pct Level 4	1.11**	-0.04	-0.67	0.69*
	(0.46)	(0.41)	(0.46)	(0.38)
Standardized Scale Score	0.13***	0.07***	-0.03	-0.05*
	(0.03)	(0.03)	(0.03)	(0.03)
N	36,792	36,799	35,777	35,770

Note: These TWFE-DID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023) in grade-3 ELA. Robust standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. *p<.1. **p < .05. ***p < .01

Table A6—Estimated Effect of ELSBG on Grade-3 ELA Test Scores by Sample

Sample Construction	Pct Level 2 or Higher	Pct Level 3 or Higher	Pct Level 4	Std. Scale Score
Full Sample	6.00***	4.98***	1.98***	0.14***
	(1.25)	(0.86)	(0.51)	(0.02)
Bottom 4,000 Schools	6.71***	5.08***	1.85***	0.14***
	(1.36)	(0.91)	(0.54)	(0.03)
Bottom 3,000 Schools	6.96***	4.90***	1.77***	0.14***
	(1.12)	(0.69)	(0.42)	(0.02)
Bottom 2,000 Schools	6.62***	4.25***	1.52***	0.13***
	(1.30)	(0.89)	(0.48)	(0.03)
Bottom 1,000 Schools	5.46***	3.09***	0.89*	0.10***
	(1.09)	(0.68)	(0.45)	(0.02)
Bottom 500 Schools	4.94***	2.74**	0.74	0.09***
	(1.41)	(1.10)	(0.67)	(0.03)

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023) in grade-3 ELA. Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A7—Estimated Effect of ELSBG on Grade-3 Math Test Scores by Sample

Sample Construction	Pct Level 2 or Higher	Pct Level 3 or Higher	Pct Level 4	Std. Scale Score
Full Sample	3.58***	3.98***	0.63	0.11***
	(1.35)	(1.00)	(0.50)	(0.03)
Bottom 4000 Schools	4.41***	4.26***	0.74	0.12***
	(1.40)	(1.21)	(0.51)	(0.03)
Bottom 3000 Schools	4.94***	4.36***	0.70	0.12***
	(1.53)	(1.01)	(0.52)	(0.03)
Bottom 2000 Schools	4.71***	3.77***	0.53	0.11***
	(1.23)	(1.06)	(0.50)	(0.03)
Bottom 1000 Schools	3.68***	2.64**	0.13	0.08***
	(1.36)	(1.05)	(0.48)	(0.03)
Bottom 500 Schools	3.64**	2.49**	0.33	0.08**
	(1.47)	(1.25)	(0.66)	(0.03)

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A8—Estimated Effect of ELSBG on Grade-5 ELA Test Scores by Sample

Sample Construction	Pct Level 2 or Higher	Pct Level 3 or Higher	Pct Level 4	Std. Scale Score
Full Sample	-0.56	0.28	-0.54	0.00
	(1.38)	(0.92)	(0.46)	(0.03)
Bottom 4000 Schools	-0.21	0.42	-0.55	0.00
	(1.28)	(0.92)	(0.37)	(0.02)
Bottom 3000 Schools	0.02	0.43	-0.53	-0.01
	(1.36)	(0.85)	(0.40)	(0.03)
Bottom 2000 Schools	-0.05	0.29	-0.55	-0.01
	(1.44)	(1.07)	(0.42)	(0.03)
Bottom 1000 Schools	-0.36	-0.24	-0.74*	-0.02
	(1.79)	(1.15)	(0.43)	(0.03)
Bottom 500 Schools	-0.22	0.22	-0.66*	-0.02
	(1.40)	(0.90)	(0.34)	(0.03)

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A9—Estimated Effect of ELSBG on Grade-5 Math Test Scores by Sample

Sample Construction	Pct Level 2 or Higher	Pct Level 3 or Higher	Pct Level 4	Std. Scale Score
Full Sample	-1.15	0.84	0.37	-0.03
	(0.91)	(0.56)	(0.30)	(0.02)
Bottom 4000 Schools	-0.53	0.97*	0.34	-0.02
	(0.98)	(0.56)	(0.31)	(0.02)
Bottom 3000 Schools	-0.06	0.98	0.31	-0.02
	(1.18)	(0.73)	(0.32)	(0.03)
Bottom 2000 Schools	-0.08	0.82	0.29	-0.02
	(1.20)	(0.79)	(0.38)	(0.03)
Bottom 1000 Schools	-0.56	0.23	0.12	-0.03
	(1.31)	(0.76)	(0.36)	(0.03)
Bottom 500 Schools	-0.75	0.02	-0.04	-0.03
	(1.26)	(0.73)	(0.39)	(0.03)

Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The restricted sample constructions are based on the baseline assignment variable. *p<.1. **p < .05. ***p < .01

Table A10—Estimated Effect of ELSBG on Grade-3 ELA Test Outcomes, DDD Specifications

	Comparison Grade-Subject		
Dependent variable	Grade-3 Math	Grade-5 ELA	Grade-5 Math
Pct Level 2 or Higher	3.52***	6.89***	6.94***
	(0.97)	(1.93)	(1.70)
Pct Level 3 or Higher	1.32	4.80***	2.75**
	(0.91)	(1.51)	(1.23)
Pct Level 4	1.10**	1.66***	0.35
	(0.48)	(0.63)	(0.57)
Standardized Scale Score	0.05***	0.14***	0.17***
	(0.02)	(0.04)	(0.04)
N	73,542	71,120	71,120

Note: These DDD intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023). Robust standard errors are in parentheses and clustered at the school level. All specifications condition on fixed effects unique to each school-year, school-grade-subject, and year-grade-subject interaction. The treatment indicator of interest is a binary indicator for the three-way interaction identifying Grade-3 ELA observations in ITT schools in the post-treament period. *p<.1. **p < .05. ***p < .01

Table A11—Estimated Effect of ELSBG on Grade-Subject Test Scores excluding Charter Schools

Dependent variable	Grade 3 ELA	Grade 3 Math	Grade 5 ELA	Grade 5 Math
Pct Level 2 or Higher	6.01***	3.59**	-0.71	-1.36
	(1.24)	(1.43)	(1.71)	(1.19)
Pct Level 3 or Higher	4.92***	3.82***	0.05	0.59
	(0.79)	(1.11)	(0.95)	(0.74)
Pct Level 4	1.91***	0.49	-0.64	0.22
	(0.44)	(0.46)	(0.43)	(0.43)
Standardized Scale Score	0.14***	0.11***	-0.01	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
N	33,880	33,887	33,040	33,033

Note: These SDID intent-to-treat (ITT) estimates are based on balanced panels of California elementary schools that reported the grade-subject test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023), excluding charter schools. Bootstrapped standard errors are in parentheses and clustered at the school level. All specifications condition on school and year fixed effects. The school-year covariates are percent White, percent FRPL, and ln(enrollment). Because demographic data from the National Center for Education Statistics are not available for 2023 yet, covariates from 2022 are carried forward into that year. *p<.1. **p < .05. ***p < .01

Figure A1—Map of ELSBG ITT Schools

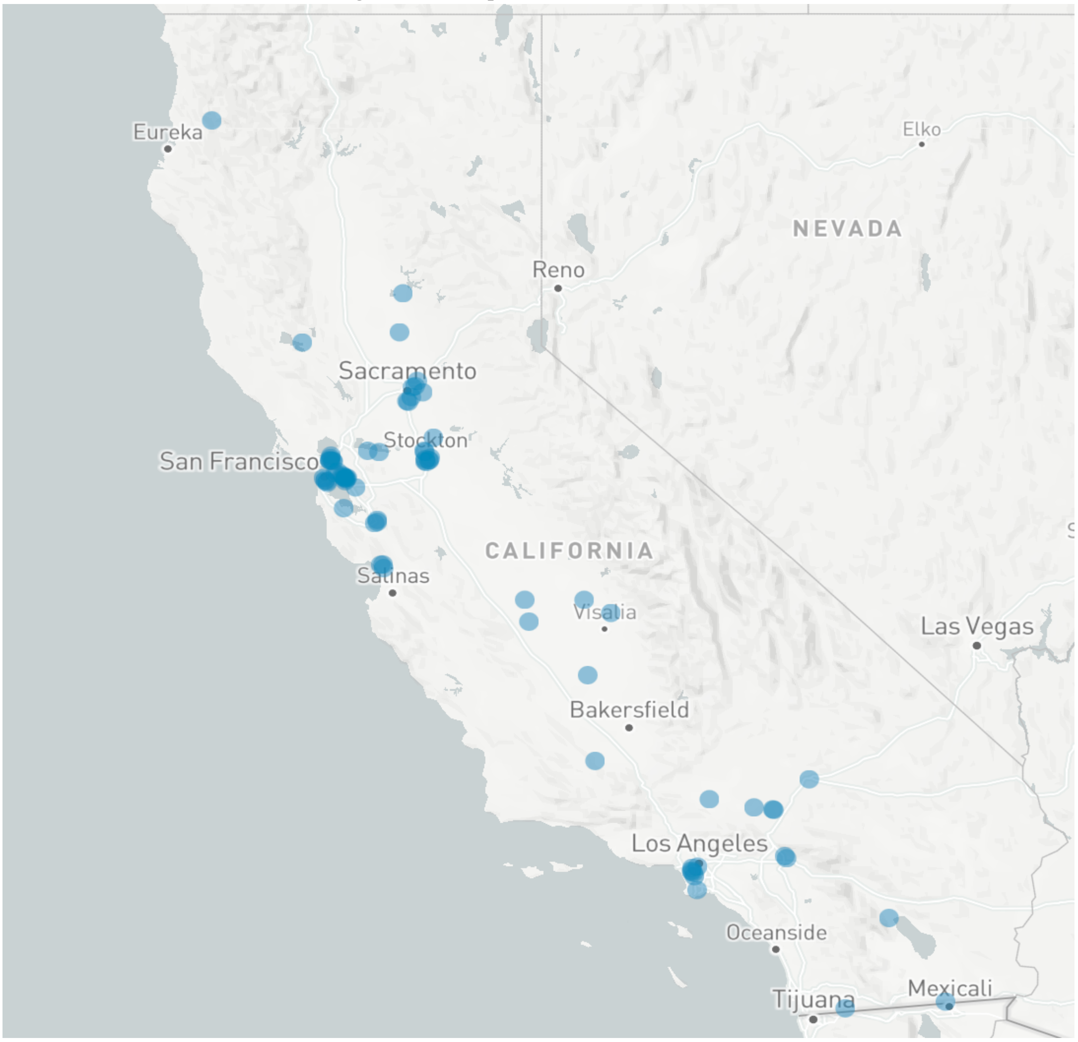


Figure A2—ESLBG Participation and Grade-3 ELA Test Scores by Baseline Assignment Variable

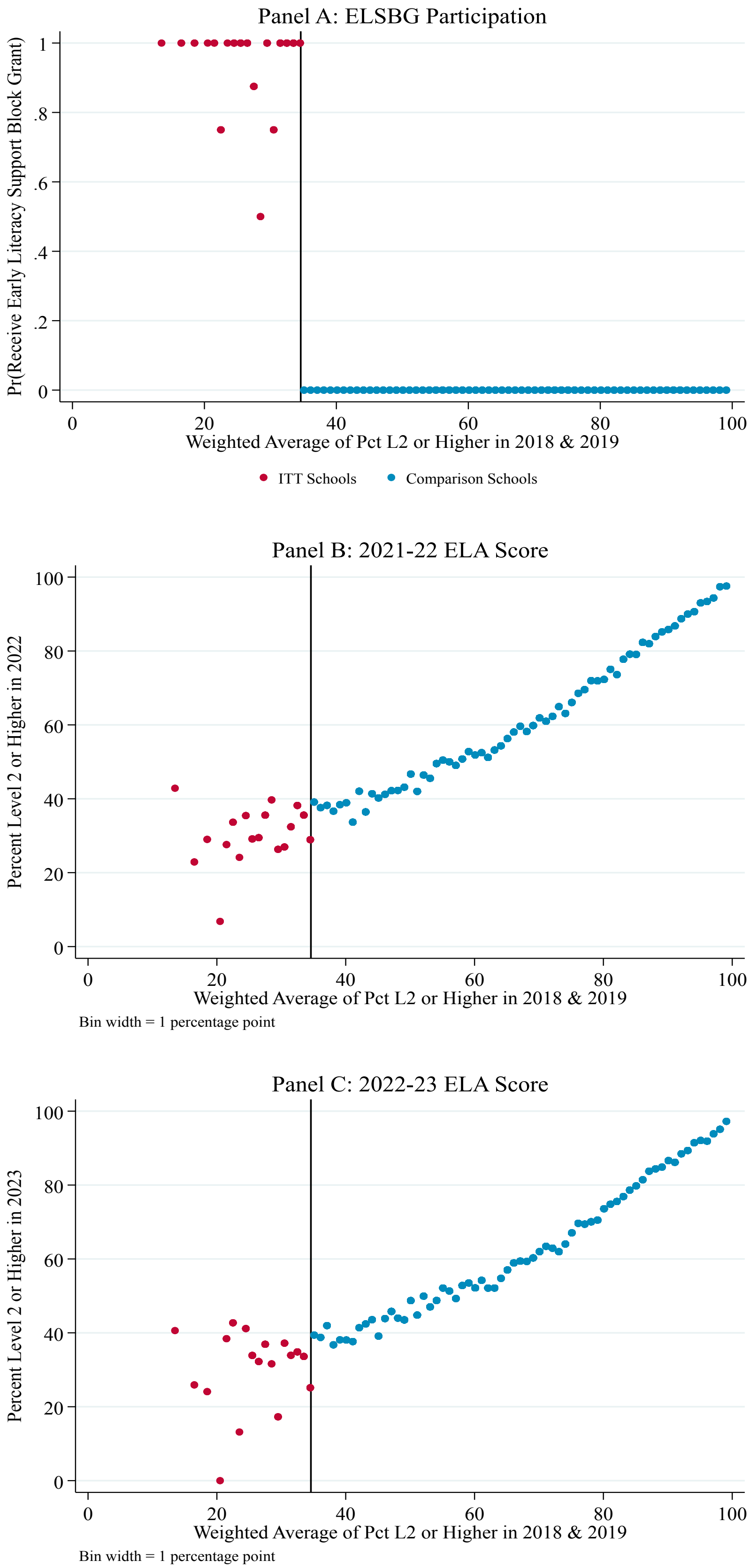
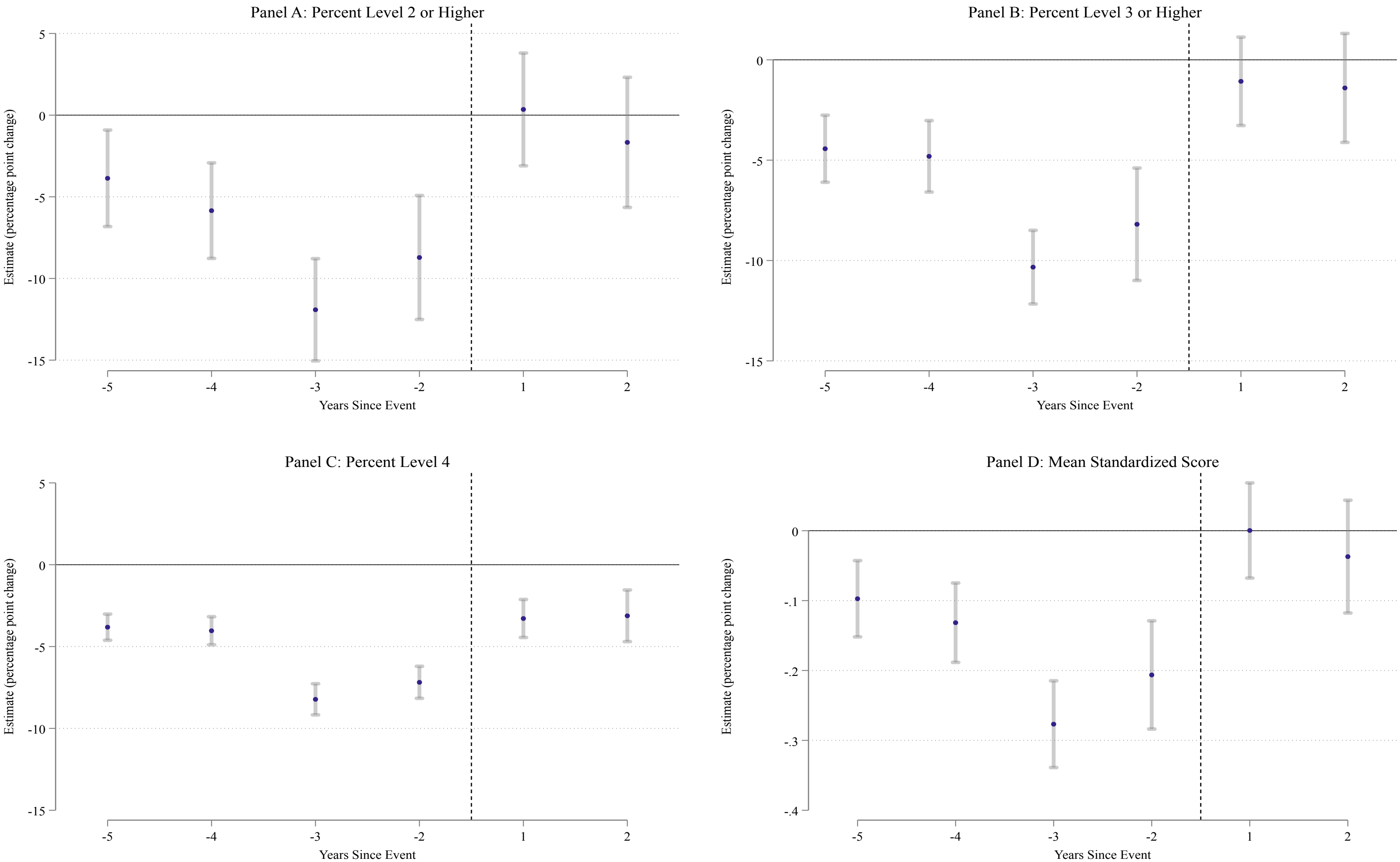


Figure A3—ELSBG Event-Study Estimates for Grade-3 ELA Test Scores, DID-TWFE Specifications



Note: Data are missing for 2019-20 (time -1) and 2020-21 (time 0) because of the COVID-19 pandemic, which led to test cancellation and limited test administration in California.

Figure A4—ELSBG Event-Study Estimates for Grade-3 Math Test Scores

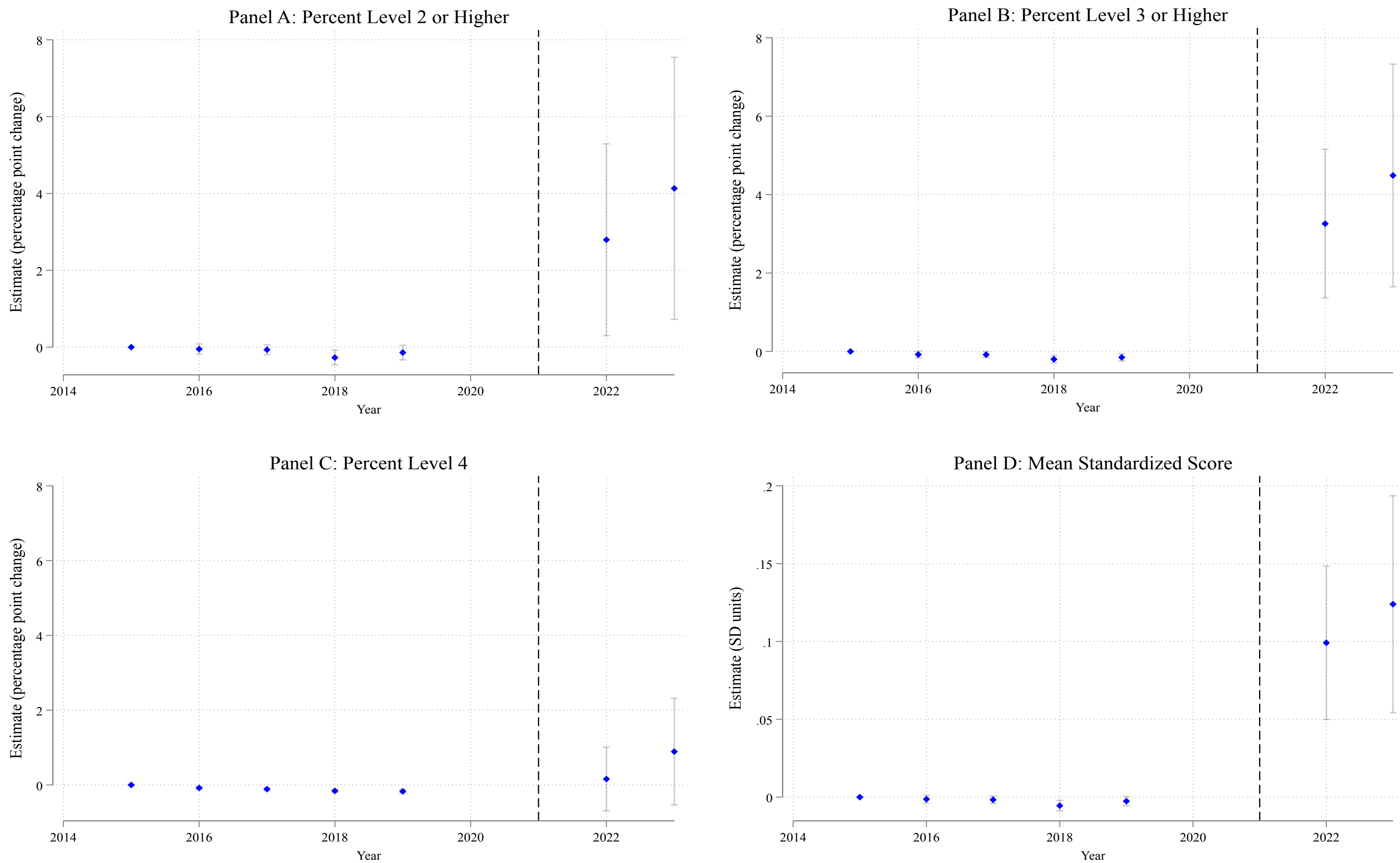


Figure A5—ELSBG Event-Study Estimates for Grade-5 ELA Test Scores

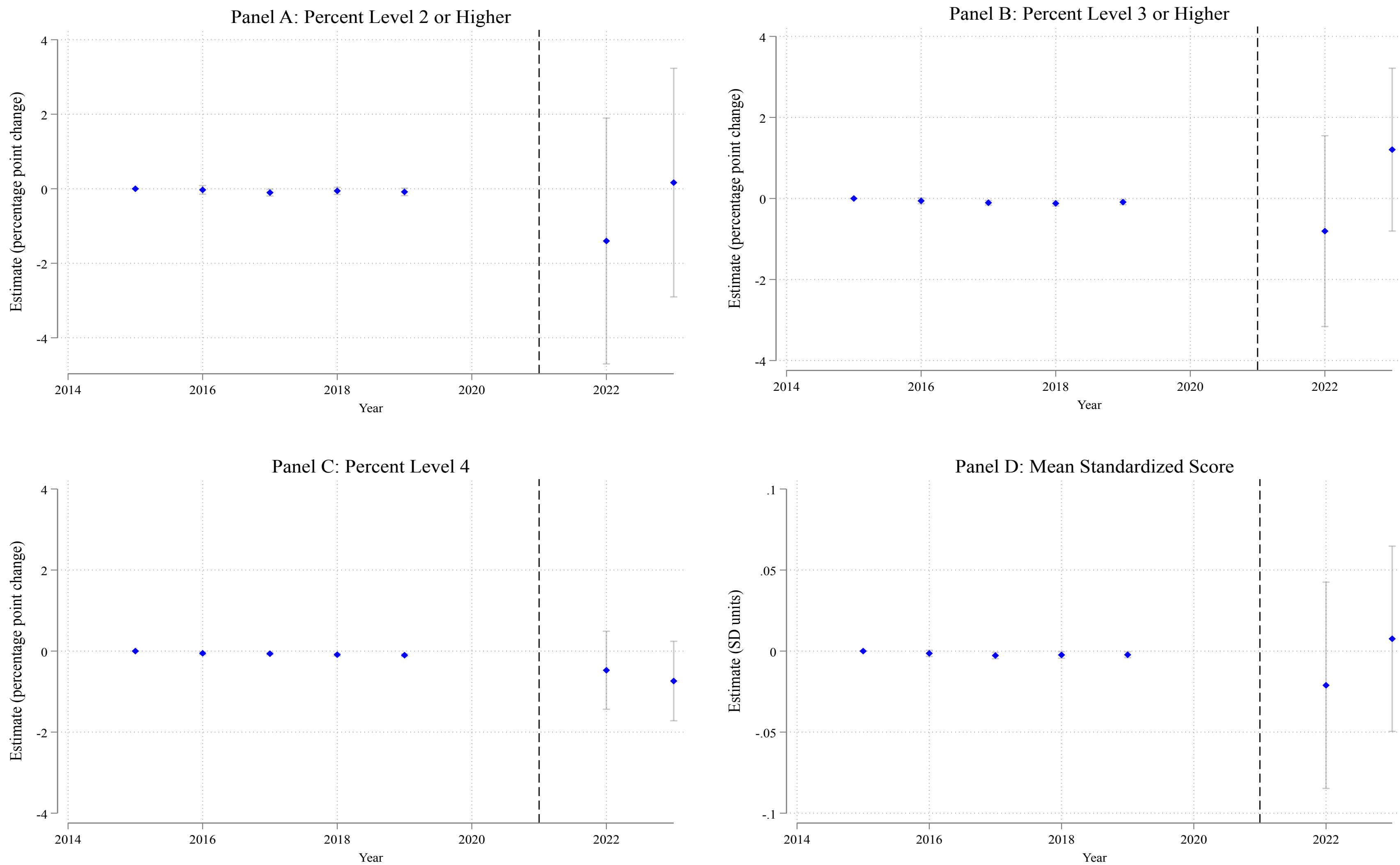


Figure A6—ELSBG Event-Study Estimates for Grade-5 Math Test Scores