Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 23-887

The Achievement Effects of Scaling Early Literacy Reforms

Sarah Novicoff
Stanford University

Thomas S. Dee
Stanford University

While legislators have implemented many "science of reading" initiatives in the last two decades, the evidence on the impact of these reforms 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 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. Our preferred specification finds 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 improvements in math.

VERSION: January 2025

Suggested citation: Novicoff, Sarah, and Thomas S. Dee. (2025). 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

The Achievement Effects of Scaling Early Literacy Reforms

Sarah Novicoff and Thomas S. Dee

Stanford University

Author Note: Correspondence regarding this article should be addressed to Sarah Novicoff at snovi@stanford.edu.

Sarah Novicoff
Stanford University
520 Galvez Mall
Stanford, CA 94305
snovi@stanford.edu
(310) 601-6315

Thomas S. Dee
Stanford University
520 Galvez Mall
Stanford, CA 94305
tdee@stanford.edu
(650) 814-3499

The Achievement Effects of Scaling Early Literacy Reforms

Abstract—While legislators have implemented many "science of reading" initiatives in the last two decades, the evidence on the impact of these reforms 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 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. Our preferred specification finds 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 improvements in math.

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), more than a third of U.S. fourth graders (i.e., 37 percent) scored Below Basic in reading (*NAEP Reading*, 2022)—this level varies little from three decades prior (i.e., 38 percent 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, 2023) over the best pedagogical approach to early literacy (i.e., the "Reading Wars"). The Reading Wars persist in public discourse despite a research consensus in psychological science about the way children learn to read: through understanding of the alphabetic principle (i.e., the idea that written letters and letter-pairs correspond to sounds) typically built with phonemic awareness and letter knowledge-building activities, through building of fluency in repeated word reading tasks, and by building both vocabulary and comprehension skills to parse and infer from more complex units of language (e.g., phrases, sentences, and paragraphs) (e.g., Castles et al., 2018; National Reading Panel, 2000). These findings from psychological sciences have motivated advocacy for the inclusion of "science of reading" practices, such as systematic phonics instruction and comprehension-building strategies like summarization and prediction, into U.S. schools. This movement has notched numerous legislative successes in the last two decades, with 40 states passing new laws since 2013 (Schwartz, 2024). These laws range from bans of curricular approaches to the introduction of new professional development programs to new requirements for teacher preparation programs.

However, the effects of these state initiatives on everyday classroom literacy practices are uncertain. A prominent example involves the three-cueing method wherein teachers encourage students to seek information from surrounding cues when they cannot determine a word. For example, three-cueing teachers encourage students to look around the word to the context of the sentence (i.e., "What word would make sense here?") rather than sound out its components. Despite recent changes in state laws—including specific bans on three-cueing in Arkansas, Indiana, and Ohio—75 percent of K-2 teachers reported using this strategy to teach reading as recently as 2019 (Kurtz et al., 2020). Educational researchers have also critiqued the science of reading movement for being too focused on fluency-building in lieu of comprehension and for presenting the science as "settled" when they believe science involves a perpetual openness to testing new hypotheses (Tierney & Pearson, 2024).

The limited evidence on the implementation and success of these initiatives in U.S. schools further complicates this debate. Most evidence currently available on these initiatives is descriptive and complicated by the presence of confounding factors, like grade retention, or fails to validate the effectiveness of the science of reading when implemented. For example, prominent claims of a "Mississippi Miracle" in reading achievement appeal to the state's distinctive test-score trends on the National Assessment of Educational Progress and attribute a 0.14 SD increase in test scores to a science of reading state initiative in 2013 (Kristof, 2023; *NAEP,* 2019; *NAEP*, 2013; Spencer, 2024). Similarly, in Michigan, grade-3 reading achievement had been declining prior to the state's "Read by Grade 3" law and then increased by 0.10 SD after the law had been implemented (Strunk et al., 2021). However, analyses like those showing improvements in Mississippi and Michigan do not necessarily account for other changes occurring over time. These potential confounds include the end of the Great Recession, changes in the test-taking population due to grade retention (e.g., changing the age at test), and the Obama Administration's Race to the Top initiative. The limited evidence from quasi-experimental designs that account for these confounds is less encouraging. For example, a recent study based on comprehensive state-by-year panel data presents event-study evidence based on "difference in differences" (DID) designs (Westall & Cummings, 2023). The study finds 0.025-0.05 SD increases in state standardized tests after the implementation of early literacy policies but find that states with retention components fully account for the effect observed.[1] In addition, a federally funded evaluation of President Bush's Reading First program, based on a regression-discontinuity (RD) design, found that this initiative significantly increased instructional time and practices aligned with the science of reading but did not significantly improve students' reading comprehension in grades one, two, or three (Gamse et al., 2008). For grade-3, the effect of Reading First was found to be 0.01 SD.

Given the paucity of literature on government-led implementation of science of reading, 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 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."

This study engages the following research question: How did the Early Literacy Support Block Grant (ELSBG) change grade-3 achievement in English Language Arts in California's lowest-performing schools? We examine the causal impact of ELSBG eligibility through a discontinuous assignment rule and several other quasi-experimental designs. We privilege the results based on the synthetic "difference-in-differences" (SDID) design and three different "difference in difference in differences" (DDD) designs. These designs leverage repeated observations of test score performance from the same set of schools over time to determine how

ELSBG-eligible school performance changed relative to their past performance and to comparison schools elsewhere in the state; DDD designs additionally leverage test score performance in non-focal grades and subjects to further isolate the effects of this program. Complementary results from this range of quasi-experimental methodologies provide alternative ways to address parallel-trends violations, concerns about mean reversion that may vex conventional difference-in-difference (DID) designs, and potential confounds related to the heterogenous learning impact of the COVID-19 pandemic. Furthermore, we examine the results of all of these quasi-experimental designs in samples restricted to include comparison schools that more closely resemble ELSBG-eligible schools (i.e., lower achievement at baseline). In brief, this study's preferred specification finds consistent evidence 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)

Program Design

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 to create 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 with a specific focus on grades K-3 (*Ella T. Settlement*, 2020).[2] ELSBG was therefore triple-targeted: targeted to a specific subject area, a specific grade range, and a specific set of low-performing schools.

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 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 K-3 literacy instruction (St. Andre, 2020). The state disbursed these small initial planning funds to eligible and interested school districts in the middle of the 2020-21 school year (i.e., the planning year). The take-up of ELSBG among eligible schools 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; when combined with resources spent at the state level, this funding supported programming for 15,541 K-3 students with an average one-year cost of $1,144 per pupil.)[3] Implementation of those plans began in July 2021 with schools receiving their first-year allocations whenever their plans were approved; funds could then be spent in the program's designated three implementation years (i.e., SY 2021-22, SY 2022-23, and SY 2023-24). 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., kindergarten through third grade).[4] Through a competitive selection process, the state 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). SCOE, in its role as the Expert Lead in Literacy, required that purchases across all four categories reinforce the site’s literacy goals.

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. This program design is distinct from past school improvement efforts like No Child Left Behind and School Improvement Grants. Those programs operated along a multi-step process: first, identify low-performing schools, and, second, mandate how the school will improve its performance. Sometimes, programs included a step three: provide resources to support the implementation of prescribed improvement strategies. The program design of ELSBG is different from past school improvement efforts in that, after identifying low-performing schools, it required and supported those schools to undertake a research and planning process to determine why their performance had lagged. While these plans accommodated flexibility, they were subject to oversight from a state-selected office and operated under broad guidelines about the tenets of effective instruction.

Program Implementation

To examine how 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. Session 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 also sponsored the participation of 32 District Grant Leads in Lexia Learning’s Language Essentials for Teachers of Reading and Spelling (LETRS) professional development, which is closely aligned with the “science of reading.” 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).

We note that engaging in SCOE programming was optional for staff at ELSBG schools, and thus the high take-up rate likely reflects enthusiasm from staff about the offerings; if teachers did not feel that sessions bore some practical import to their day-to-day instruction, teachers likely would not have participated. Responses of participants in this programming to a SCOE-administrated survey support this claim. For example, 97 percent of respondents reported that the Online Elementary Reading Academy either met or exceeded their expectations while 98 percent reported that the course helped them "learn the research on the essential components of reading instruction" (Sullivan, 2022). Administrator interviews, conducted as part of the same SCOE annual report on programming, further highlighted the value of these opportunities. One administrator said: "It has been really positive, the trainings that we go to, I feel like they are really insightful and informative." Another administrator stated: "The big thing for us was the alignment across the classrooms and really aligning practices… The biggest impact or the takeaway [was] how we are teaching kids how to read and the phonics part of it that we were missing for a long time."

To determine which actions were taken by districts themselves to supplement the SCOE offerings, 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. Staff compensation (i.e., salaries and benefits) altogether represented 71 percent of the budgeted expenditures for the ELSBG funds at the school-level. These payments could have been used to hire new staff or compensate existing staff for their time attending professional development or for additional hours on site providing intervention services for students. Among staffing expenses, we found that hiring new on-site instructional coaches or teachers on special assignment, who were trained by SCOE in the science of reading, constituted the largest expenditure category with 28 percent of budgeted expenditures (or, 40 percent of staffing expenditures). Outside professional development accounted for 12.55 percent of total ELSBG budgeted spending while another 9.69 percent was budgeted towards books, supplies, or technology. The remaining 6.64 percent was budgeted across parent outreach programming, assessment tools, district oversight, and other miscellaneous purchases.

We note heterogeneity in the breakdown of spending by school size, urbanicity, and student race/ethnicity. Small schools planned to spend the largest percentage of their budgets (36 percent) on hiring new on-site coaches compared to 28 percent in medium-sized schools and 25 percent in large schools. This may reflect economies of scale at play in larger schools; in other words, a large school and a small school may both wish to hire one instructional coach for $90,000 a year but, because larger schools received more total funds, that single salary will constitute a larger percentage of the total for the small school than it will for the large school. In addition, we observe higher spending on staffing in urban and suburban contexts (72 percent) than in towns or rural communities (63 percent). Though this may be due to different preferences for staffing across different communities, this may also be driven by higher costs of staffing due to higher costs of living in urban and suburban areas. We also observe stronger preferences for instructional coaches among majority Hispanic schools, compared to majority Black schools or those with neither a Black- nor Hispanic-majority, even when comparing only within schools with the same levels of urbanicity.

School 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 expenditure reports and 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).

For schools not in ELSBG, California encourages science of reading practices through its English Language Arts/English Language Development Framework, a document of recommendations and resources for teachers, school staff, and administrators. For example, in its chapter on kindergarten and grade-one instructional strategies, the framework emphasizes that students must "be phonemically aware" to learn to read (*English Language Arts/English Language Development Framework*, 2015). However, this framework represents recommendations to educators, rather than mandates embraced by other states. In 2024, state lawmakers proposed legislation that would mandate instructional materials adoption aligned to the science of reading (SOR) and require teacher preparation program accreditation to be determined based partially on SOR instruction, among other changes. The California Teachers Association opposed the legislation, arguing that "restricting instructional methods stifles teachers' creativity and innovation in the classroom" (Bramble, 2024). Californians Together, a statewide advocacy group for English Learners, has argued similarly, saying mandating science of reading from the state level "overlooks the importance of allowing teachers to adapt instruction to fit the unique needs of their students" (Langreo, 2024). The bill was ultimately tabled without a hearing, though its backers intend to propose an adapted version of the bill in a subsequent legislative session.

Without state mandates, the choice of whether and how to use instructional practices supported by the science of reading occurs at the district level. A 2022 study of reading programs in more than 300 California districts (enrolling 72 percent of students in the state) showed that, despite this local flexibility, two curricula were chosen by 72 percent of surveyed districts (The California Reading Curriculum Report, 2022). Those two popular curricula are listed by their publishers as aligned with SOR, but advocates for SOR often describe both curricula as “balanced literacy” instead (Tadayon, 2022). Therefore, it is difficult to determine what pedagogical approach was being used to educate students in the comparison group for this study (i.e., schools not in ELSBG) because it likely varied. However, due to high attendance rates at ELSBG-sponsored professional development sessions about SOR, even for staff in schools with prior literacy initiatives, ELSBG likely increased the exposure of teachers and staff to this instructional philosophy.

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 an ELSBG focal grade) and fifth graders as a comparison group.[5] Assessments in English Language Arts/Literacy assess four areas – reading, writing, listening, and research/inquiry – while assessments in mathematics assess four additional areas – concepts and procedures, problem solving, communicating/reasoning, and model and data. Technical documentation shows that CAASPP for both subjects across all tested grades has a reliability score above 0.86 and has been extensively tested for validity to draw inferences about a student's knowledge in the tested subject (*CAASPP Technical Reports and Studies*, 2024). English Language Arts/Literacy tests represent our primary outcome for this study. We additionally assess mathematics because both ELA and math are typically taught by the same elementary teachers, and we aim to assess the degree to which ELSBG's professional development affected teacher capacity narrowly in literacy or more broadly across both subjects. We also recognize that mathematical instruction and assessment by grade-3 increasingly relies on foundations in literacy through word problems, and thus an intervention in one subject can have spillover effects on another.

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. Notably, CAASPP was consistent during this time period and proficiency benchmarks remained consistent as well, with the same minimum score determining whether a grade-3 student Nearly Met, Met, or Exceeded the standard in English Language Arts for every year in our time period. We standardize the scale score within grade and year using the student-level means and standard deviations publicly reported by the California Department of Education.

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 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 fewer than 11 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.[6] 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 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.[7]

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

In exploratory analyses, we also use publicly available budgets submitted to the California Department of Education and the Sacramento County Office of Education at the conclusion of the planning year of the grant in June 2021. When the CDE and SCOE approved these budgets and their accompanying "Literacy Action Plans" (i.e., narrative descriptions of how the money would be spent along with justifications for those choices and more detailed timelines of implementation) between July and December 2021, school districts or charter management organizations posted them publicly on their websites where we accessed them for this study. Budgets can and often do cover multiple schools within a district if multiple schools were eligible for the grant; in total, these 36 budgets explain the planned expenditures for 72 schools.[8]

Each budget contains a list of planned expenditures for the school district. Each expenditure is listed as an expense of the planning year of the grant (2020-21) or one of the three grant implementation years (2021-22, 2022-23, and 2023-24). Expenditures were also listed with the name of the school site at which they will be spent or listed for the school district itself, in which case we divided the amount by the number of schools in the district under the grant and allocated it evenly across each school. In most cases, expenditures were then described in the reports in a few words (e.g., "contract for professional development," "reimbursement for teacher planning time") though some districts chose to be more descriptive (e.g., "0.5 FTE Program Assistant English Language Arts and English Language Development"). All expenditures were then categorized by the authors into ten categories as explained in the Appendix.

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).[9]

We proposed a regression-discontinuity (RD) design that leveraged the cross-sectional variation in a baseline school-level assignment variable (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. Intuitively, this research design functions by comparing schools eligible for ELSBG to schools ineligible for ELSBG with similar levels of the continuous assignment variable. This design rests on the assumption that California elementary schools with

33 percent of their students scoring Level 2 or higher on grade-3 ELA are comparable to California elementary schools with 37 percent of their students receiving that same designation on the same assessment except for the fact that one set of schools was eligible for an intervention and another was not. Therefore, if significant differences in test scores are observed between these two similar sets of schools, this research design assigns the cause of that difference to the intervention.

Because the state leveraged a school's Percent Level 2 or higher in the pre-treatment years to determine ELSBG eligibility, we correspondingly designated the Percent Level 2 or higher (i.e., the percent of students scoring Standard Nearly Met or higher) on the post-treatment grade-3 ELA exam as the single confirmatory outcome measure.

Difference-in-Differences (DID) Design

Because the overall results from the RD design are mixed, we also explore our results using a "difference-in-differences" (DID) design based on school-year panel data. Rather than making a comparison narrowly between schools on either side of the cutoff score in a single year as in an RD design, DID designs compare the treatment group (i.e., ELSBG-eligible schools) to a wider set of comparison schools (i.e., ELSBG-ineligible schools) across multiple years. This technique first calculates the difference between the pre- and post-treatment outcomes for the treatment group and separately for the comparison group. Then, this method subtracts the change over time for one group from the other. This two-step procedure allows researchers to control for changes over time that occurred across the ecosystem of study – in this case, California elementary schools – and to control for inherent and stable differences between the treatment and the comparison group (e.g., different racial composition).

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 a 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:

$$Y_{st} = \alpha_s + \beta_t + \tau D_{st} + \varepsilon_{st} \tag{2}$$

in which outcome Y_{st} in school s and year t is a function of school and year fixed effects (i.e., α_s, β_t), a binary indicator for ELSBG eligibility in the post-treatment years (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. Our DID results also suffer from the possibility of mean reversion, in which extreme values tend to move closer to the average over time; because ELSBG-eligible schools are drawn from the bottom of the ELA-score distribution, if ELSBG-eligible schools are compared to higher-performing schools, then mean reversion may cause an appearance of improvement due to ELSBG when those improvements may have been statistically likely to occur without the intervention.

Synthetic Difference-in-Differences Design

We address the implied internal-validity threats 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.[10] Given these weights, the SDID procedure forms an estimate of the effect of interest (i.e., $\hat{\tau}^{sdid}$) through this least-squares minimization:

$$\left(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\operatorname{argmin}}\left\{\sum_{s=1}^{N}\sum_{t=1}^{T}(Y_{st} - \mu - \alpha_s - \beta_t - D_{st}\tau)^2 * \hat{\omega}_s^{sdid} * \hat{\lambda}_t^{sdid}\right\} \tag{3}$$

To conduct statistical inference for estimates based on equation (3), 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.

We explore the credibility and robustness of SDID-based results in several ways. First, we present results with and without covariate adjustments.[11] 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. 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.)[12] These estimates also address concerns about mean reversion. Specifically, the bottom 500 schools are drawn from the bottom of the grade-3 ELA distribution, just like ELSBG-eligible schools, and are similarly likely to move closer to the average over time; by comparing ELSBG schools to only other very low-performing schools, we difference out any mean reversion that may be occurring.

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 (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 A3, 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.

Triple Differences Design

We also present results based on three different "difference-in-difference-in-differences" (DDD) designs that provide alternative approaches to addressing parallel-trend violations. DDD designs build on the logic of a DID model but incorporate an additional difference; in addition to making comparisons across time and across treatment status, DDD incorporates a comparison within a treated unit to non-treated observations. In this case, the treated unit is an ELSBG-eligible school and non-treated observations are found in non-focal subject or grades (e.g., math in grade-3, math and ELA in grade-5). The DDD approach tacitly assumes that these non-focal subject or grade test scores 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. 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. We present direct evidence on this question by presenting SDID estimates of the effect of ELSBG eligibility on these other achievement outcomes.

We note that each DDD design also relies on a parallel-trends assumption; namely that the comparative trends in a given "placebo" grade and subject (i.e., grade-3 math, grade-5 ELA and math) across treated and comparison schools provide a valid counterfactual for how *untreated* potential outcomes in grade-3 ELA would have changed across these schools. While the parallel-trends assumption is fundamentally untestable, we do find event-study evidence that is consistent with this identifying assumption. Specifically, Figures A4, A5, and A6 indicate that, as with grade-3 ELA achievement (Figure A3), achievement in the placebo grades and subjects had negative pretrends unique to ELSBG-eligible schools prior to treatment. We note that these comparative pretrends correspond uniquely well for grade-3 math.

Exploratory Analyses of School Spending

We also examine the relationship between growth in standardized test scores and planned expenditure categories of ELSBG spending, using the regression specification below:

$$Y_s = \alpha_s + \beta C_s + X_s + \varepsilon_s \tag{4}$$

Where Y_s represents the change in the percent of grade-3 students scoring Standard Nearly Met or higher in ELA between 2018 and 2019 (i.e., the last two pre-ELSBG years) and 2022 and 2023 (i.e., the first two ELSBG years). C_s represents a vector of thousands of dollars allocated across the possible categories of spending, excluding benefits because they are collinear with staffing costs. X_s represents a vector of covariates, which vary with the model but include urbanicity, racial/ethnic demographics of a school, the percent of students receiving free- or reduced-price lunch, and pre-trends in academic performance from 2015-2017. This regression is not causal because it cannot control for the endogenous role administrators played in selecting their expenditures, but it offers exploratory evidence of the role of certain expenditure choices and of the broader program structure in the academic achievement of ELSBG schools.

Results

Regression Discontinuity

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 A4)—a virtually "sharp" assignment to treatment that is represented visually in panel A of Figure A7. Furthermore, the full-sample reduced form 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 A4). 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 A4) 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 A7) does not provide clear evidence of an impact, possibly due to the lack of power from the small sample size.

Synthetic Difference-in-Differences (DID) Design

We present our main results (i.e., the estimated effects of ELSBG eligibility on the four different measures of ELA achievement) using our preferred specification in Table 2. These SDID-based estimates consistently indicate that ELSBG eligibility had positive and statistically significant effects on grade-3 ELA test scores. 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 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. We note that effects are larger for schools in urban and rural environments, compared to those in suburban communities.

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.[13] 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.

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 targeted grades.

The spillover effect of ELSBG into grade-3 math could potentially be driven by two forces. First, in third grade, math instruction and assessment typically require a strong foundation in literacy. The Common Core State Standards for grade-3 repeatedly emphasize the role that word problems play in grade-level content (*Grade 3 Operations & Algebraic Thinking*, 2010), and the California Assessment of Student Performance and Progress (i.e., the assessment used in this paper to determine academic achievement) includes numerous word problems in its practice test for grade-3 math. Examples range from the straightforward—"Megan baked 28 sugar cookies and 24 chocolate chip cookies. Enter the total number of cookies Megan baked in all."—to the more complex—"Which expression is equal to 6x3, and why?... 3x6 because the order of the numbers does not matter in multiplication" (*Online Practice and Training Tests*, 2022). If ELSBG improved the reading of students in low-performing elementary schools, this may have better enabled them to understand the word problems commonly found in grade-3 math instruction and assessment.

Second, elementary teachers typically deliver both ELA and math content to the same group of students. ELSBG's delivery of improved curriculum and professional development on literacy topics to elementary teachers in low-performing schools may have freed them up to spend additional time focusing on math instruction. Alternatively, as teachers improved their literacy instruction, this may have improved their relationship with students, which could in turn improve their math instruction and their student math performance on standardized tests.

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.[14] 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 DID estimates based on TWFE specifications (Figure A3).[15] The event-study SDID 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 A8, A9, A10). 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 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 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 A10).

Triple Differences Design

We also 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 A11). 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 SDID 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.

Exploratory Analyses of School Spending

We explore the relationship between spending choices with ELSBG funds and academic achievement on grade-3 ELA in Table A12. In our preferred model with demographic controls and the inclusion of prior academic achievement trends, we find that most categories of spending are not associated with increases in academic growth. This offers suggestive evidence that ELSBG's bundle of "science of reading"-driven professional development with additional targeted funds and state oversight—which all schools in this regression received—is associated with the academic achievement gains discussed above rather than a particular spending choice. The only category in which spending was a statistically significant predictor of growth was parental outreach. This relationship is not causal because, though choices can be predicted by observable characteristics like school urbanicity or poverty, choices also emerge from a collection of unobserved factors that cannot be controlled for and might bias the estimate. For example, if schools who choose to do outreach to families are the schools who care most deeply about family engagement, then a positive relationship between parental engagement and academic achievement could be due to this unobserved buy-in rather than the causal effect of spending on outreach itself.

We add an additional note to the interpretation of these coefficients. ELSBG's theory of change does not posit that there is a single best choice for all school sites to make to improve. It instead hinges on the ability of local stakeholders to select the best expenditures for their school site in consultation with state-selected experts. The lack of statistically significant coefficients on particular categories of spending supports this theory of change. Schools with the most improvement did not spend ESLBG funds in the same way but rather spent it in a wide variety of ways that were best for them.

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 in the United States. 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 implementing a class-size reduction akin to Project STAR would increase total expenditures per pupil by roughly a third. In the California context of 2020-21 immediately prior to ELSBG implementation, this would mean increasing per-pupil expenditures by roughly $4,790. Project STAR produced a learning gain of

0.22 SD so, this implies a return of 0.046 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 per pupil in 2021 dollars increases test scores by 0.0097 SD.[16] 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 and operates in a similar way to the increases in spending evaluated in that paper (i.e., the funds supplement, not supplant, existing spending). We note that, even when taking our most conservative estimate of ELSBG's impact (i.e., an effect size of 0.05 SD based on using grade-3 math in a DDD design; Table A11), the learning return to this investment would be five times larger than that implied by an indiscriminate spending increase. This comparison does not mean that ELSBG is necessarily the best use of funding, but that it did produce larger test-score gains than other prominent early learning interventions and than an indiscriminate increase in total 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.

Notes

1. We note that the frequent bundling of new grade-retention policies with 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 well as mean-reversion, as internal-validity threats.
2. The California Department of Education (CDE) retained $3 million of this appropriation to support its administration and oversight of the ELSBG program.
3. In 2020-21, the year prior to ELSBG implementation, the average per pupil expenditures for a California school was $14,370 (*California General Fund Expenditures*, 2021). Therefore, ELSBG's addition of $1,144 per pupil in 2021-22 represents an 8 percent increase in the typical school budget.
4. In California, County Offices of Education (COEs) are administrative units that provide various services and specific educational programs in support of area school districts. California also designates transitional kindergarten (TK) as part of kindergarten, and thus ELSBG funds could be used to support programming in TK for school sites that offered TK.
5. Each intent-to-treat (ITT) school also served grade-5 students, but grade-5 students were outside ELSBG's targeted grade levels, which makes them a meaningful comparison group within each school and year for some of our results (i.e., those based on a DDD design). Almost half of the ITT schools do not serve students in grade 6 or above, leading us to not to consider older grades for potential comparison.

6. 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.
7. 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.
8. Though ELSBG was intended to be awarded to 75 schools when designed (*Ella T. Settlement*, 2020) and was offered to 75 schools at its outset (*ELSB Grant Eligible Schools*, 2020), one school closed in June 2021 after the planning year of the grant and thus did not submit budgets for grant spending. Two additional schools declined to apply for the grant altogether and thus also did not submit budgets.
9. Our pre-registration is available at osf.io/5jgwu and dated April 26, 2022. The first year of our outcome data (i.e., spring 2022 assessments) were not publicly available until October 24, 2022.
10. See Algorithm 1 in Arkhangelsky et al. (2021) and Clarke et al. (2023) for details on the construction of these weights.

11. 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.
12. 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.
13. 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).
14. 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 A3).
15. 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.
16. 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.

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

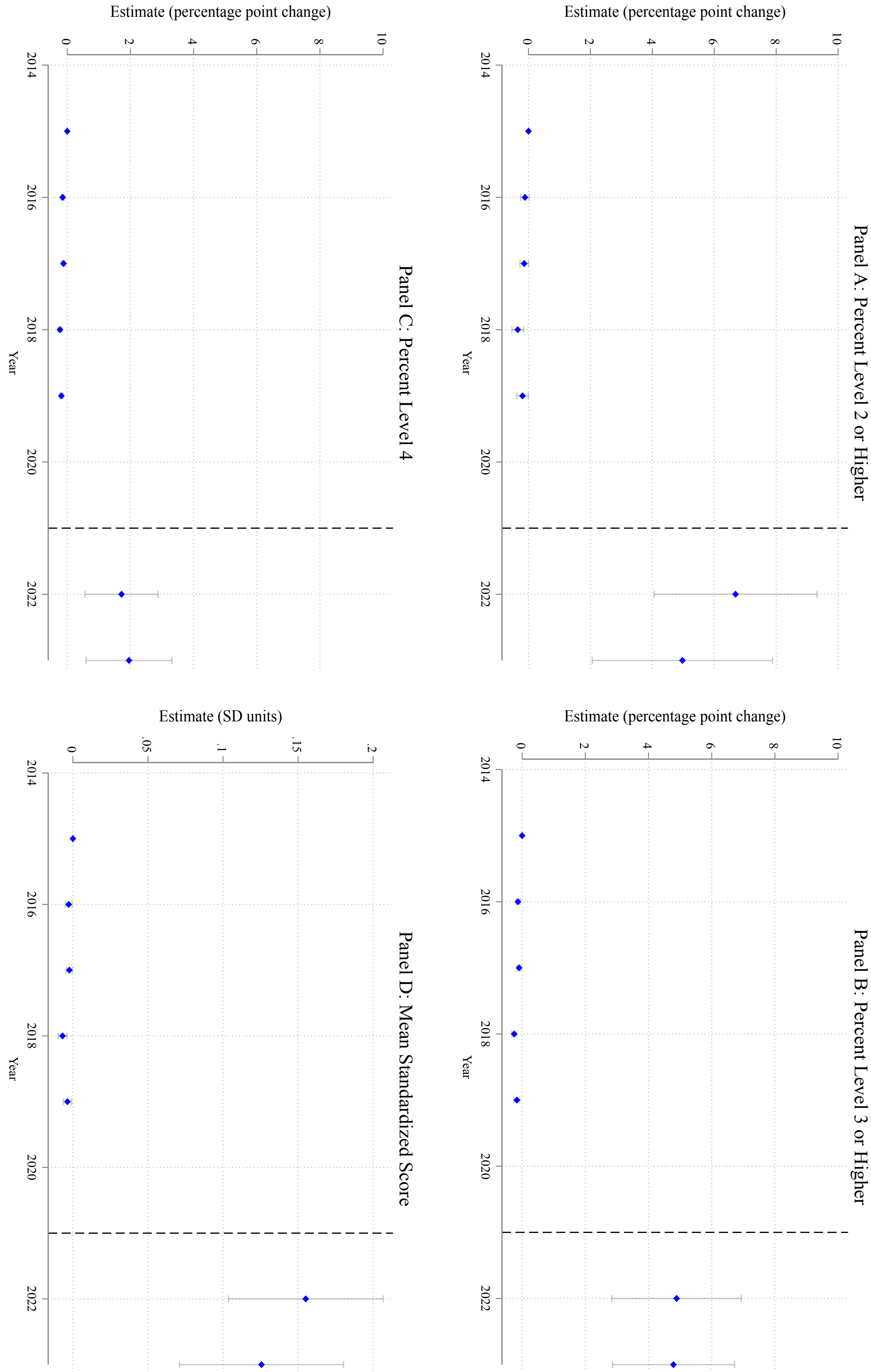
Note: These SDID intent-to-treat (ITT) estimates are based on a balanced panel of California elementary schools that reported test scores over seven years (2015, 2016, 2017, 2018, 2019, 2022, and 2023) in grade-3 ELA. Bootstrapped standard errors are in parentheses and clustered at the school level. 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



References

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

Bramble, S. (2024, March 28). *AB 2222 (Oppose)*. California Teachers Association. https://epe.brightspotcdn.com/98/1d/74c809ed497c8ee32723761d181f/ab2222.pdf

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

CAASPP Technical Reports and Studies. (2024, March 19). Educational Testing Service. https://www.cde.ca.gov/ta/tg/ca/caasspprptstudies.asp

California General Fund Expenditures, School Districts, 2020-21. (2021). Ed-Data. https://www.ed-data.org/ShareData/Html/119580

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/

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

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://assets.contentstack.io/v3/assets/blt5775cc69c999c255/blt2c7f1f36a0132938/200220-literacy-ca-ella-t-settlement-agreement.pdf

English Language Arts/English Language Development Framework. (2015, July). California Department of Education. https://www.cde.ca.gov/ci/rl/cf/documents/elaeldfwchapter3.pdf

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

Grade 3 Operations & Algebraic Thinking. (2010). Common Core State Standards Initiative.

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

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. https://epe.brightspotcdn.com/1b/80/706eba6246599174b0199ac1f3b5/ed-week-reading-instruction-survey-report-final-1.24.20.pdf

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

Langreo, L. (2024, April 17). For Now, California Won't Mandate "Science of Reading." Here's What Happened. *Education Week*. https://www.edweek.org/teaching-learning/for-now-california-wont-mandate-science-of-reading-heres-what-happened/2024/04

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 2019 Highlights. (n.d.). National Assessment of Education Progress. Retrieved June 6, 2022, from https://www.nationsreportcard.gov/highlights/reading/2019/

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

National Assessment of Student Progress (NAEP) 2013 and 2022 Reading Assessments. (2013, and 2022). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. https://www.nationsreportcard.gov/ndecore/xplore/NDE

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

Online Practice and Training Tests. (2022). California Assessment of Student Performance and Progress. https://www.caaspp.org/practice-and-training/

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, December 5). *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. (2023, 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

Schwartz, S. (2024, April 29). 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

Spencer, N. (2024). Comprehensive early literacy policy and the "Mississippi Miracle." *Economics of Education Review*, *103*, 102598. https://doi.org/10.1016/j.econedurev.2024.102598

St. Andre, E. (2020, November 17). *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, May). *Early Literacy Support Block Grant, Expert Lead in Literacy Annual Report Year 2: 2021-22*. https://drive.google.com/file/d/1cu3zRCNsNHSAMWMG24gdOVlHyVet7NRx/view

Tadayon, A. (2022, October 11). Coalition says ‘science of reading-aligned’ core curricula barely used in California. *EdSource*. https://edsource.org/2022/coalition-says-science-of-reading-aligned-core-curricula-barely-used-in-california/679596

The California Reading Curriculum Report. (2022). California Reading Coalition. https://www.careads.org/curric-report-viewer

Tierney, R. J., & Pearson, P. D. (2024). *Fact-checking the Science of Reading: Opening Up the Conversation*. Literacy Research Commons.

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

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/

Appendix: Categorization of ELSBG Planned Expenditures

We categorized each planned expenditure from each ELSBG budget into the categories below. The categories were as followed and informed by the California Department of Education's typical classification of spending with some exceptions. For example, though the California Department of Education typically categorizes all salaries to certificated staff (i.e., school employees with a license for their position such as teachers, instructional coaches, or administrators) in one category, we split them into two to allow for a difference between new hiring and additional compensation for existing teachers.

1. New On-Site Coach or Teacher on Special Assignment: This category includes the part-time or full-time hiring of instructional coach(es) whose responsibilities typically include teacher observation and support, or (less frequently) the hiring of additional trained teachers who support students with high levels of need. Example descriptions of expenditures in this category read "K-3 Early Literacy Support Coach Salary" or "Instructional (Literacy) Coach 1 FTE Average" or "Intervention Teacher 1.0 FTE."
2. Existing Teacher Time: This category compensates or covering for existing teachers for their additional hours. We note that this category is inclusive of teacher time for multiple purposes permitted by the grant, including hours spent in professional development, hours compensating a substitute to cover a class while a teacher is in professional development, providing intervention support to students, or facilitating family outreach. Example descriptions of expenditures in this category read "Stipends for CORE course completion x 21 staff members" or "Resident Substitute (accounts for 1 substitute) to cover a Kindergarten group of 4 educators to collaborate on literacy."

3. Paraprofessionals: This category compensates new or existing instructional aides, who most often provide support for struggling students or facilitate small-group instruction. Example descriptions of expenditures in this category read “Salaries for aides in K-3 classes” or “Two full time Tutors.”
4. Staff Benefits: This category includes benefits like health insurance and pension contributions for staff hired under the prior three staffing categories. In addition, because pension contributions are calculated as a percentage of the salary of an employee, if an employee earned more in a given year due to them working additional hours, then pension contributions also would have increased. Though it would be ideal to disentangle this spending and assign it to one of the prior three categories and/or to a purpose (e.g., literacy instruction or family outreach), school districts typically reported this spending as a lump sum without tagging it to a job title (e.g., “Benefits to include STRS [State Teacher Retirement System], Medicare, Unemployment, OPEB [Other Post-Employment Benefits] and health insurance”).
5. Books, Supplies, and Technology: This category covers purchases of curricular materials (e.g., “Heggerty Phonemic awareness books”), supplies (e.g., “sound spelling cards”), technology (“RAZ Kids subscriptions for teachers”) or other related costs (e.g., “Printing costs for student materials”).
6. Parent Workshops: This category includes all inputs into the organization of parent workshops. This might include the compensation of staff specifically to facilitate parent workshops, the printing of handouts for workshops, or the distribution of books for parent workshops.

7. Outside Professional Development (PD): This category covers planned expenditures to outside organizations to facilitate training sessions or coaching for teachers, or hired as consultants to help a school or school district implement their plans for improvement.
8. Assessment Systems: This category contains planned purchases of data or assessment systems to help school staff analyze student performance (e.g., "DIBELS Data Dashboard").
9. District-Level Indirect Costs: Districts often supervised the grant and its expenditures from a central office, and districts were allowed to use small percentage of the grant to cover these costs. The maximum allowed percentage varied by year and by district, and is set by the California Department of Education uniformly across all grants (i.e., is not specific to ELSBG).
10. Other: This category includes a small number of expenditures that could not fit into an existing category (e.g., "Wellness Coordinator", "Virtual field trips").

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 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 A4—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 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-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

Table A11—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-treatment period. *p<.1. **p < .05. ***p < .01

Table A12: Association between Spending (in thousands of dollars per student) in ELSBG Budget Categories and Academic Growth

Category of Spending	(1)	(2)	(3)	(4)	(5)
New On-Site Coach or Teacher on	1.57	-0.45	-0.25	-0.68	-1.62
Special Assignment	(3.57)	(3.35)	(3.24)	(3.12)	(3.37)
Existing Certificated Staff	-1.98	-2.81	-2.98	-3.19	-3.52
	(3.46)	(3.91)	(4.11)	(3.96)	(3.97)
Paraprofessionals	-1.24	-1.27	-0.71	0.09	0.18
	(3.23)	(3.11)	(3.21)	(3.37)	(3.54)
Outside PD	0.41	-0.77	-0.93	-1.47	-2.50
	(4.06)	(3.64)	(3.86)	(4.01)	(4.07)
Books, Supplies, or Technology	12.25**	7.64	6.29	6.13	5.76
	(5.28)	(5.71)	(5.49)	(5.66)	(6.34)
Parent Outreach	15.16	15.20	14.89	15.17	24.53**
	(14.33)	(14.68)	(10.35)	(12.00)	(10.98)
Assessments	3.22	12.77	5.33	5.63	-16.85
	(12.79)	(12.49)	(12.07)	(12.22)	(11.08)
Control for Urbanicity?		X	X	X	X
Control for Race/Ethnicity?			X	X	X
Control for School Poverty?				X	X
Control for Pre-Trends?					X
N Schools	67	67	67	67	64

Note: Robust standard errors reported in parentheses. Spending categories are input in thousands of dollars with Other as the reference category. Data drawn from district-level budgets submitted to the California Department of Education regarding planned expenditures under the Early Literacy Support Block Grant. School performance here is measured by a weighted average of the percent of students scoring at Level 2 (Standard Nearly Met) or Higher on ELA tests for 3rd grade for targeted years. To calculate a change in school performance, we use the difference between the weighted average in 2018 and 2019 (the ELSBG assignment variable) and 2022 and 2023 (the two post-ELSBG years of available test score data). When added, a control variable for urbanicity uses the National Center for Education Statistics classifications of Urban, Suburban, Town/Rural and uses urban schools as the reference category. Controls for race/ethnicity include the percent of a school that was Black, the percent Hispanic, and the percent White in 2020-21 (the planning year of the grant). A control for school poverty uses the percent of receiving free- or reduced-price lunch in 2020-21. Control for pre-trends adds a variable for the change in Percent Level 2 or Higher between 2017 and 2015, which is only available for 64 of the 67 schools due to data censoring when a school has fewer than 11 students in grade-3 or when a school opened during that time period. *p<.1. **p < .05. ***p < .01

Figure A1—Map of ELSBG ITT Schools

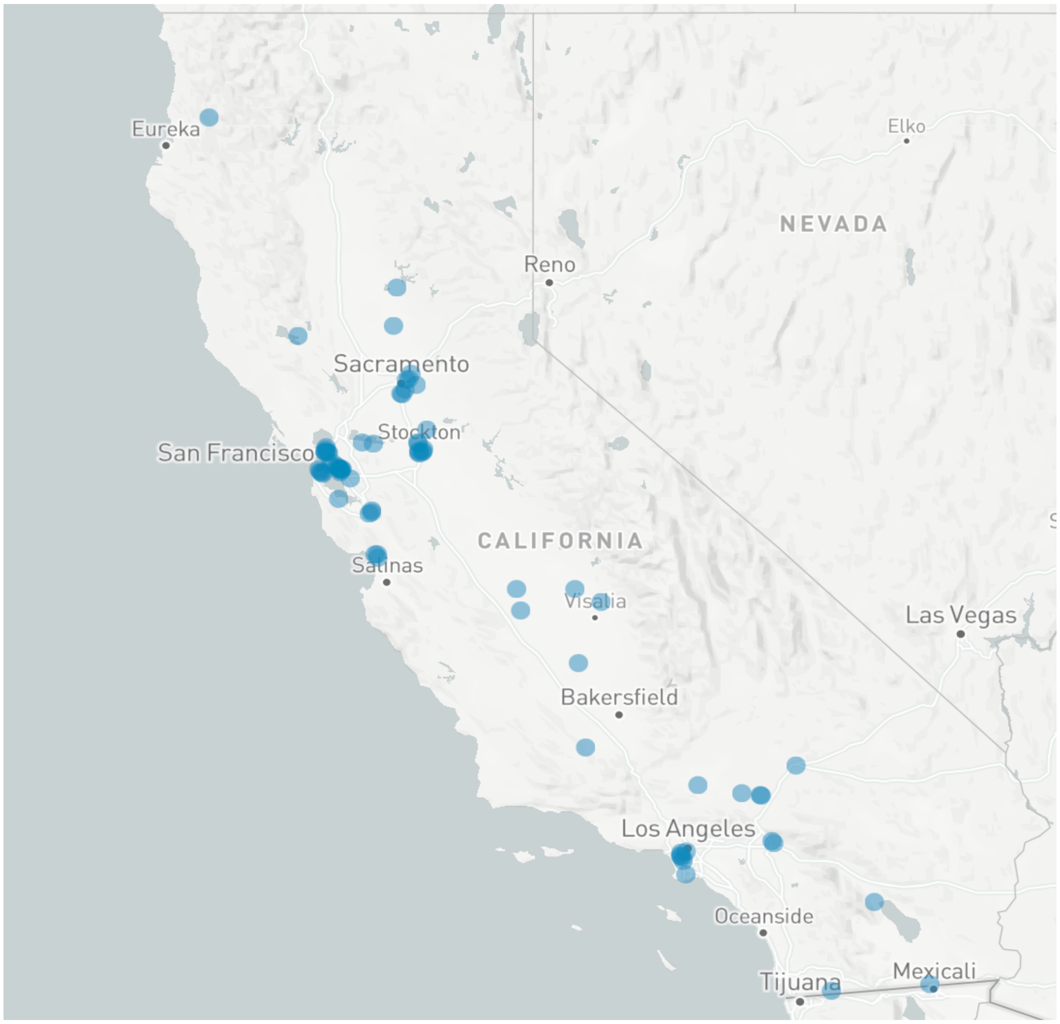


Figure A2—Visualizations of Density of the Assignment Variable

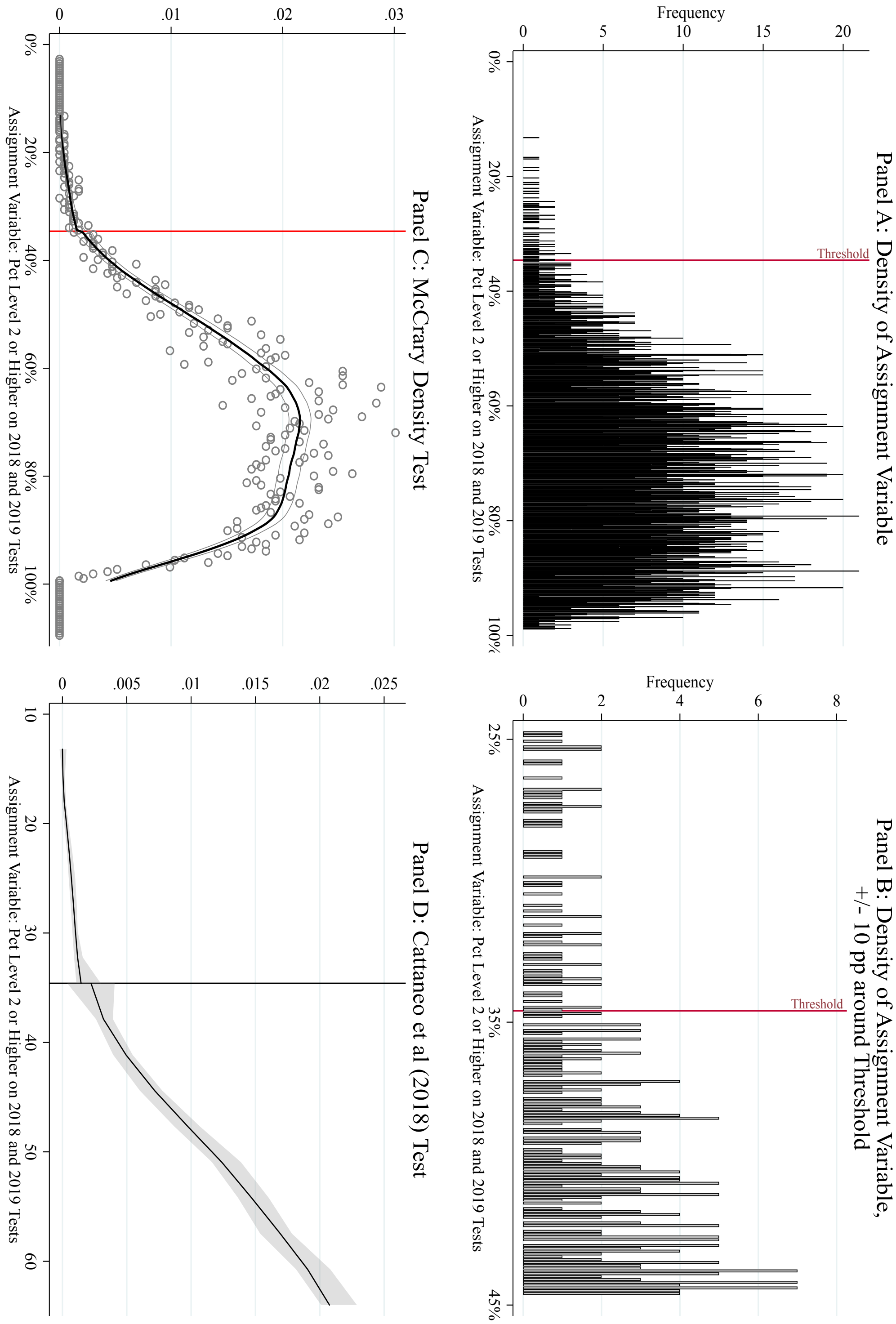
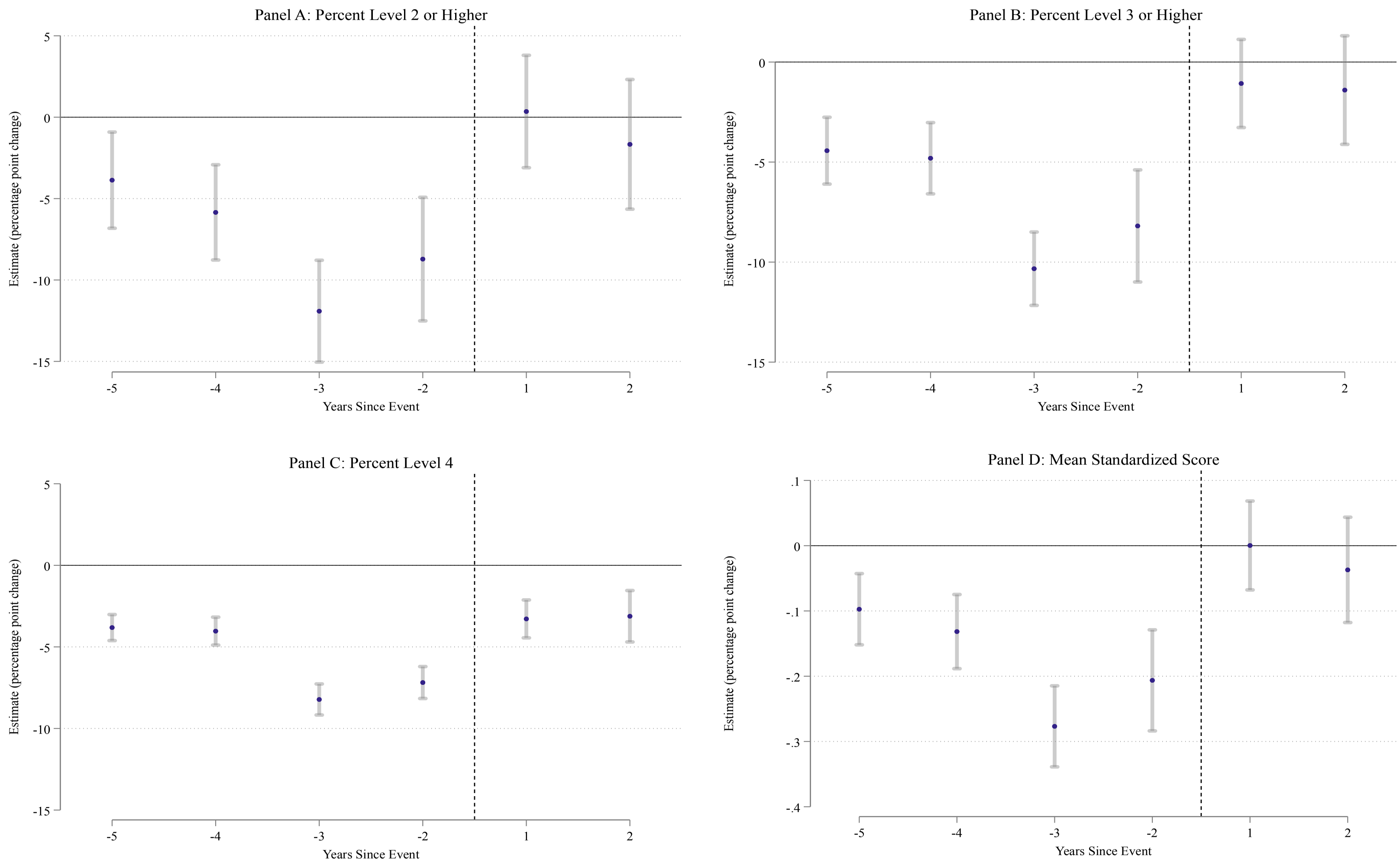
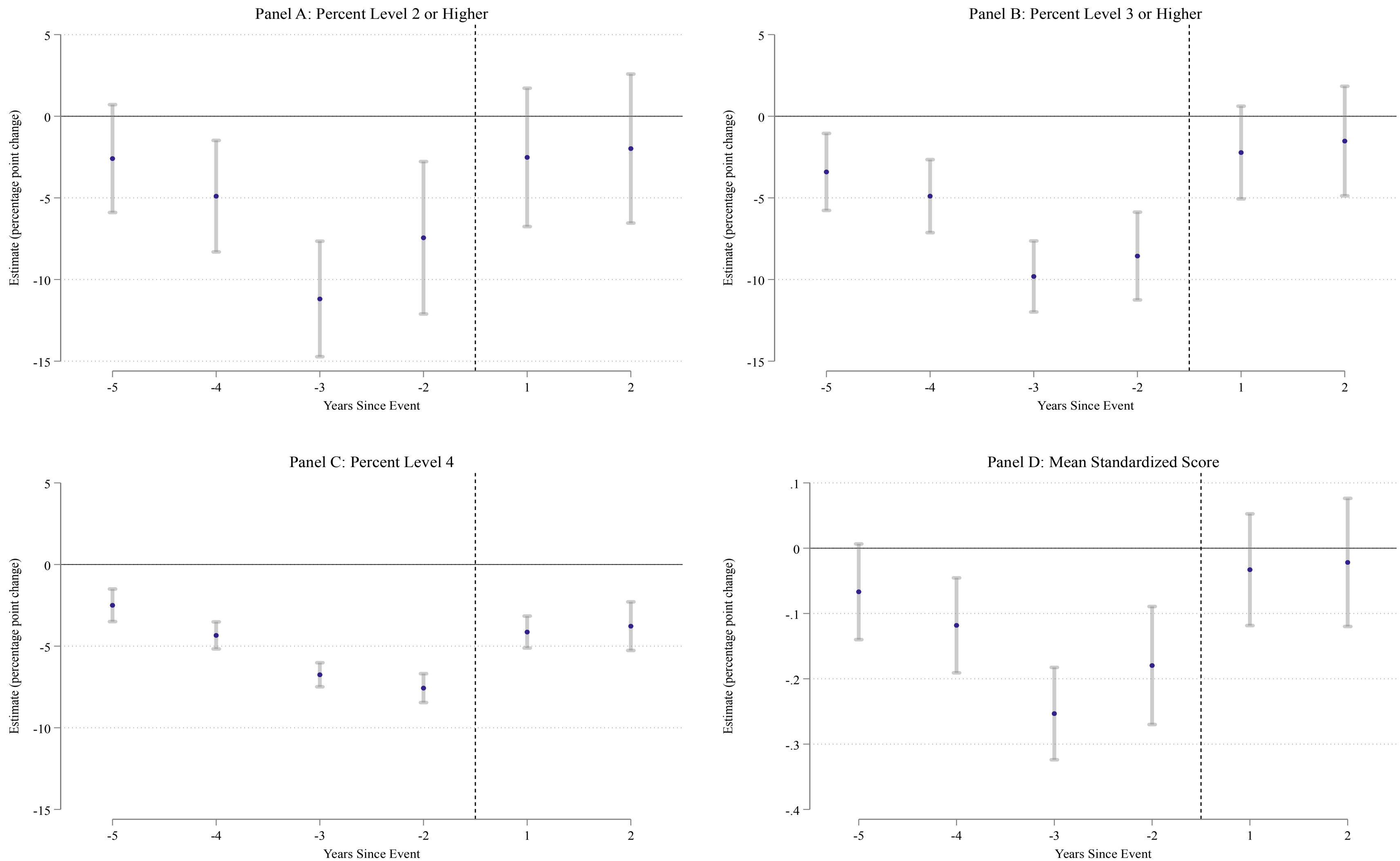


Figure A3 –ELSBG Event-Study Estimates for Grade-3 ELA Test Scores, DID-TWFE Specifications



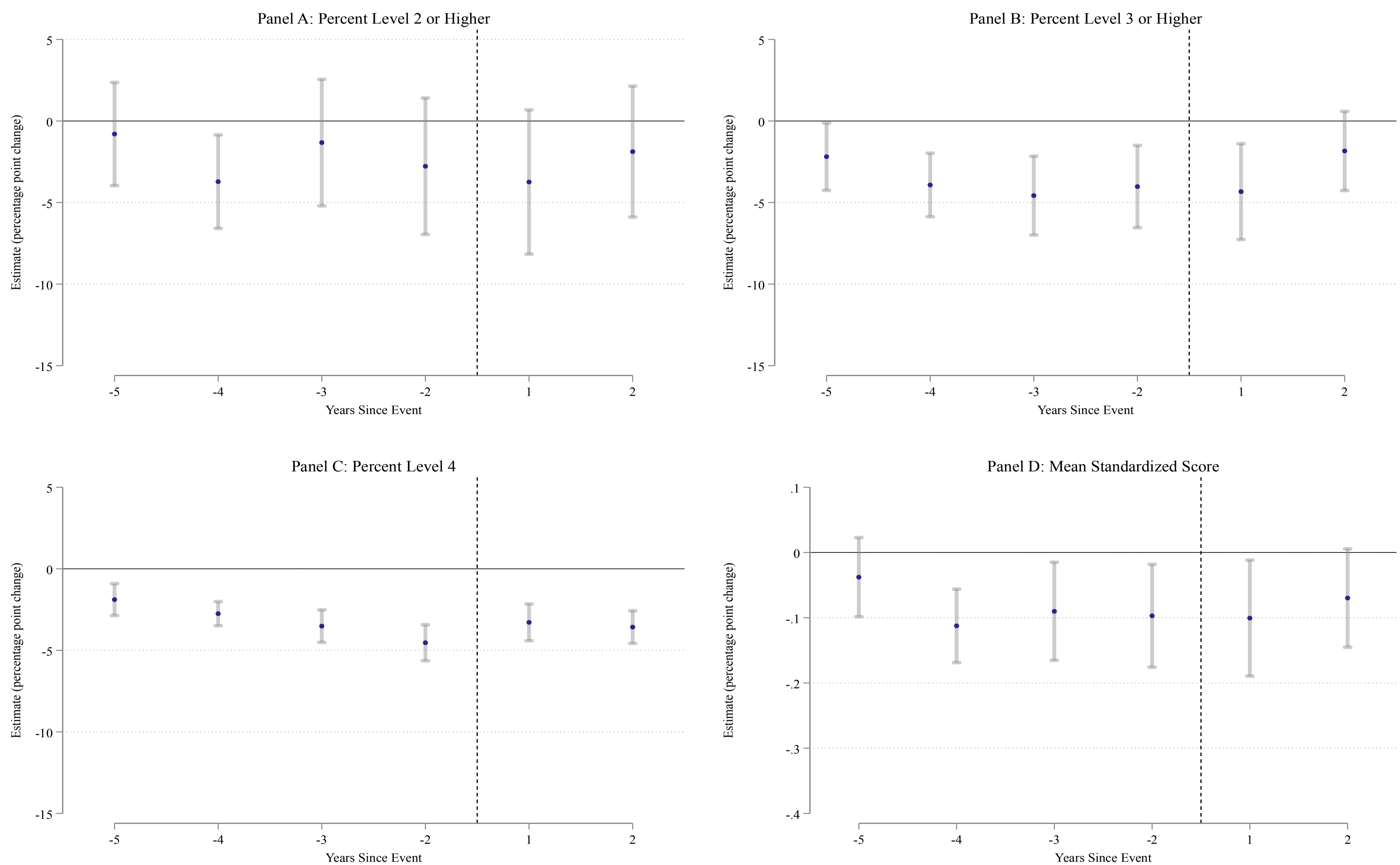
Note: Data are missing for 2019-20 (time -1) and 2020-21 (time 0) because of the COVID-19 pandemic, which led to test cancellation and limited test administration in California.

Figure A4 —ELSBG Event-Study Estimates for Grade-3 Math Test Scores, DID-TWFE Specifications



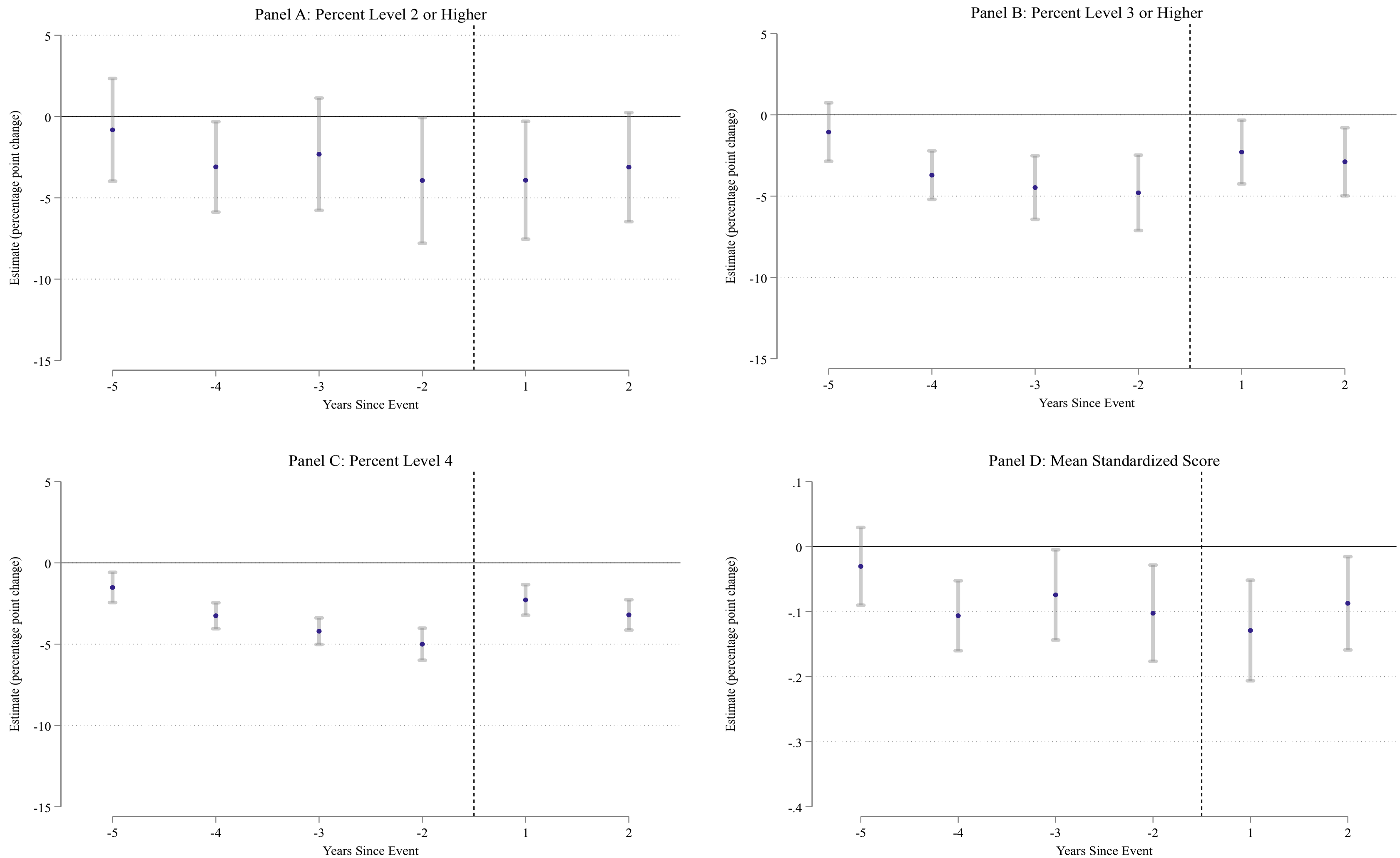
Note: Data are missing for 2019-20 (time -1) and 2020-21 (time 0) because of the COVID-19 pandemic, which led to test cancellation and limited test administration in California.

Figure .A5 –ELSBG Event-Study Estimates for Grade-5 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 A6 –ELSBG Event-Study Estimates for Grade-5 Math 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 A7 –ESLBG Participation and Grade-3 ELA Test Scores by Baseline Assignment Variable

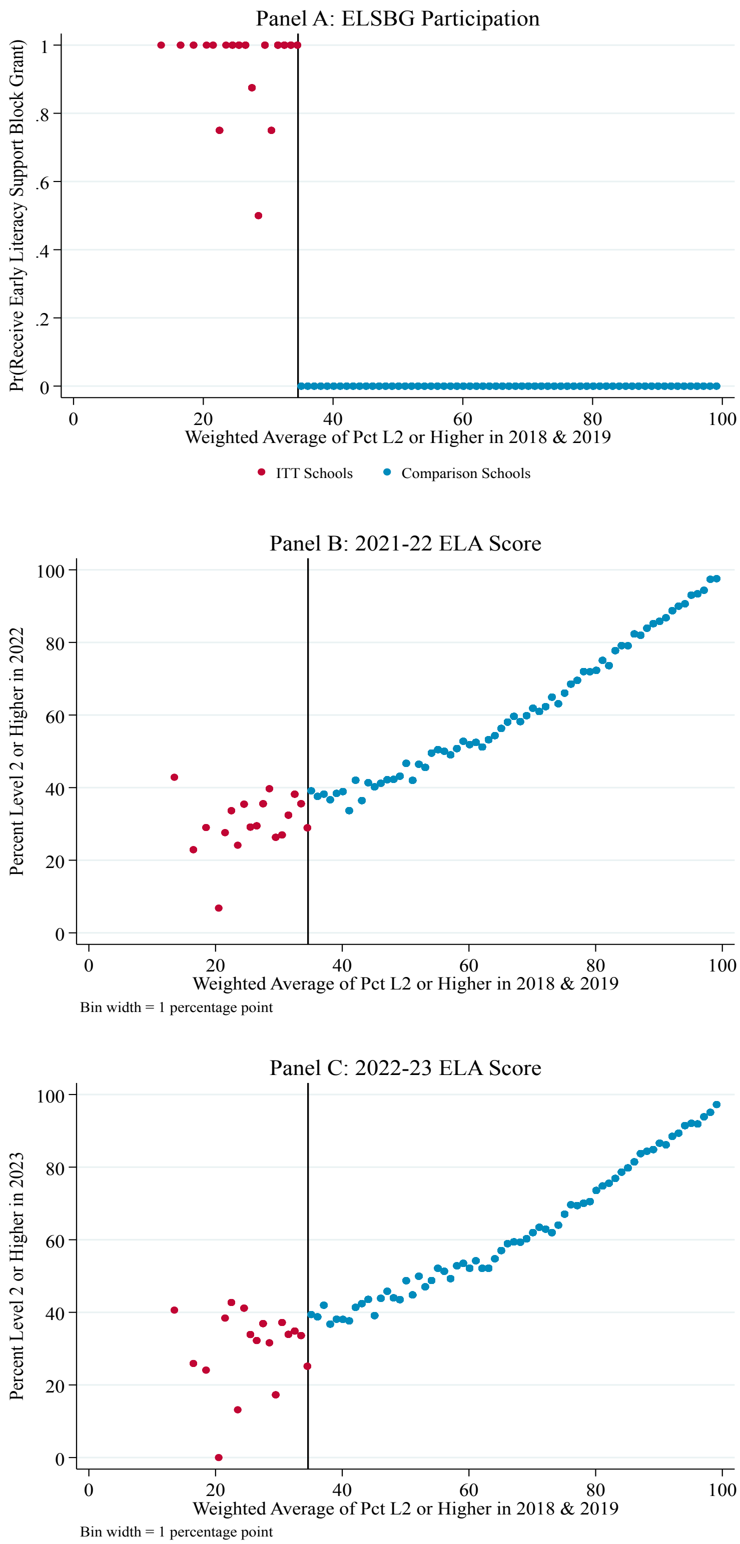


Figure A8 –ELSBG Event-Study Estimates for Grade-3 Math Test Scores

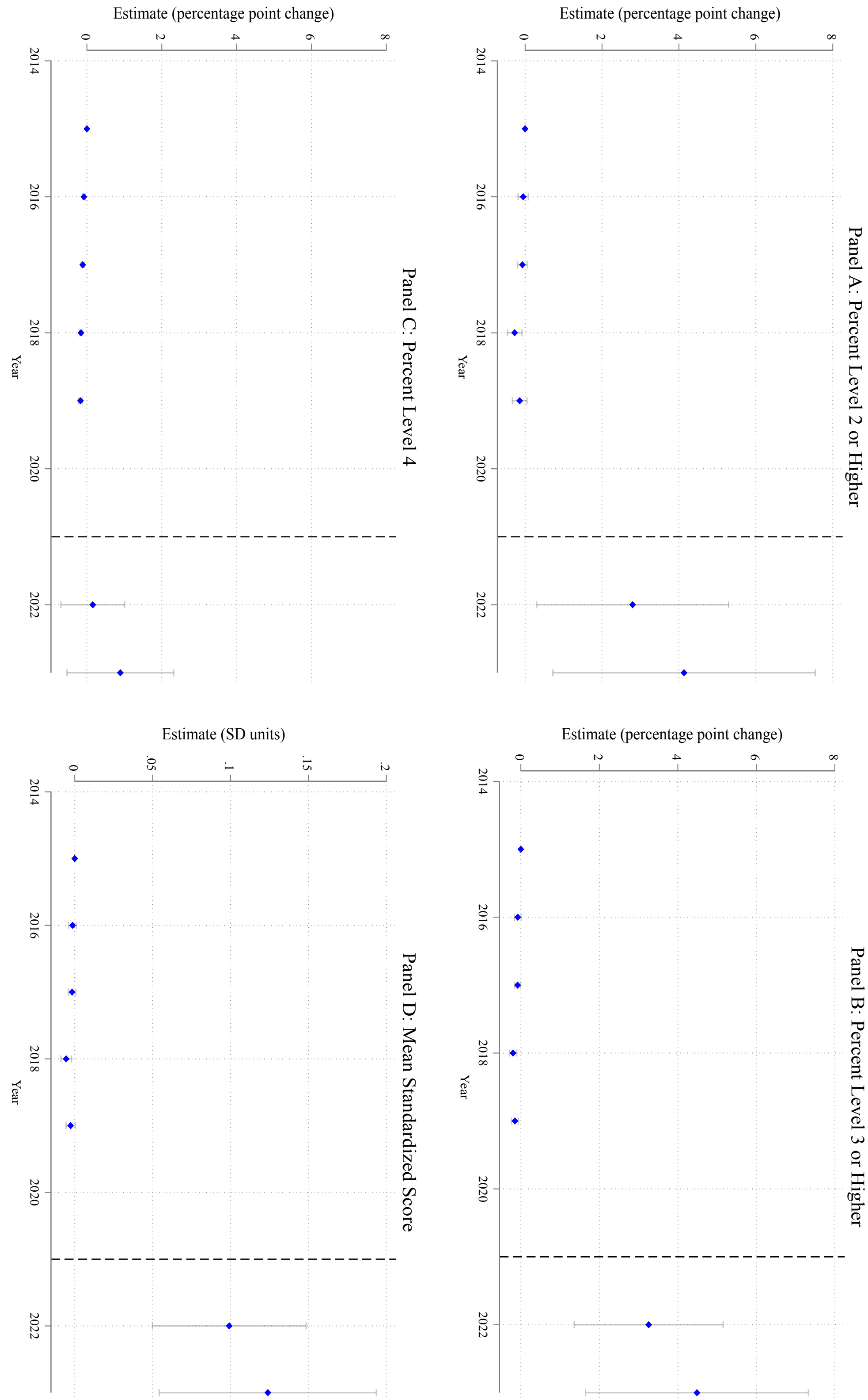


Figure A9 –ELSBG Event-Study Estimates for Grade-5 ELA Test Scores

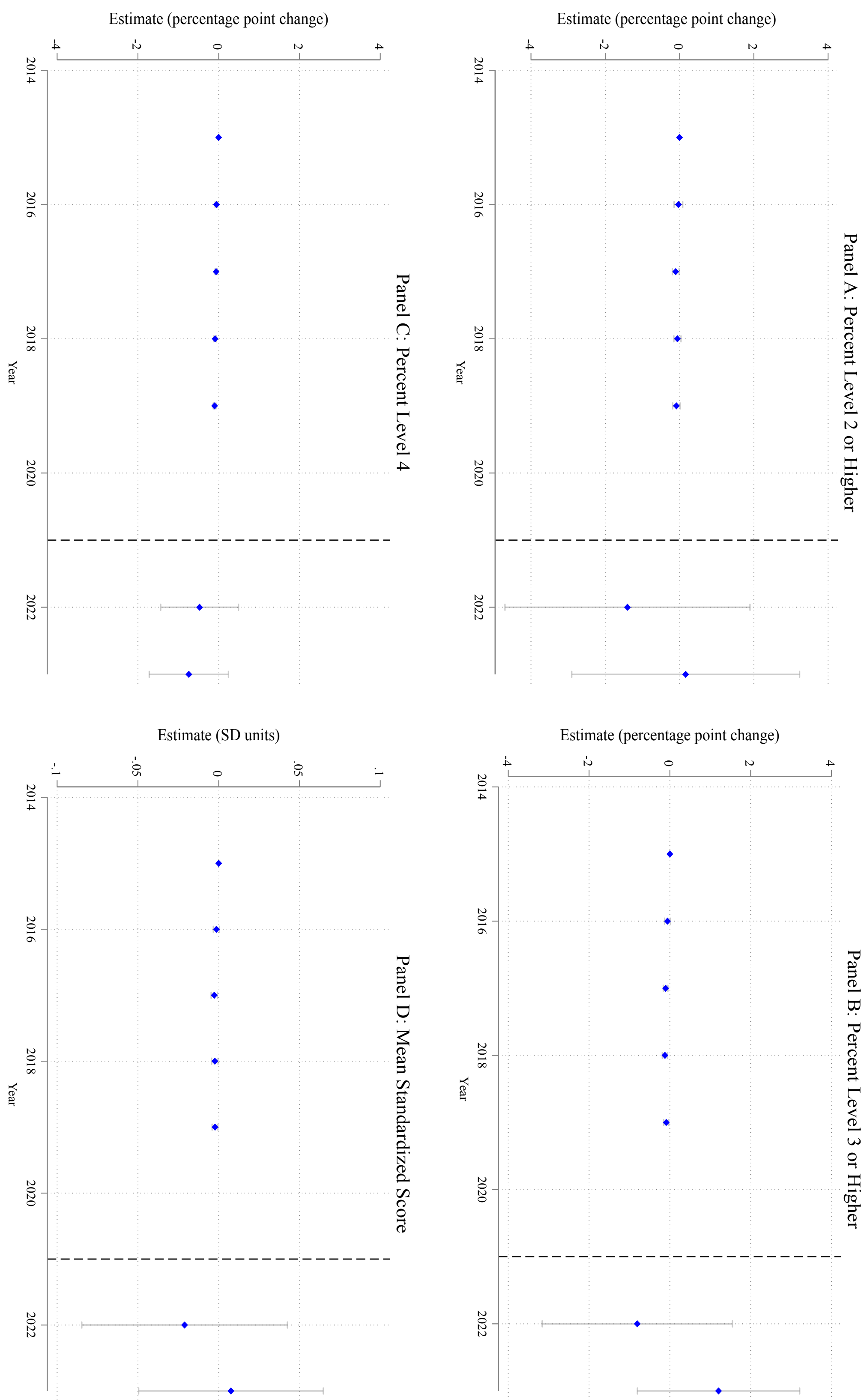


Figure A10 ELSBG Event-Study Estimates for Grade-5 Math Test Scores