

The Achievement Effects of Scaling Early Literacy Reforms

Sarah Novicoff 

Thomas S. Dec 

Stanford University

While legislators have implemented many “science of reading” initiatives in the last 2 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 \$1,000 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 English Language Arts (ELA) achievement in its first years of implementation (0.14 standard deviation) as well as smaller improvements in math.

Keywords: *achievement, economics of education, educational policy, educational reform, literacy, quasi-experimental analysis*

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 United States with the early-literacy skills relevant to realizing their academic potential. For example, in the most recent National Assessment of Education Progress, more than a third of U.S. fourth graders (i.e., 37%) scored Below Basic in reading (The Nation’s Report Card, 2022)—this level varies little from 3 decades prior (i.e., 38% in 1992). Civil-rights groups (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; Castles et al., 2018;

National Reading Panel, 2000). This process is especially important in English, an alphabetic language in which symbols correspond to sounds, in contrast to other languages like Japanese in which symbols represent syllables or Chinese where symbols can represent words or elements of meaning (Rayner et al., 2001).

Psychological studies, most involving laboratory experiments of young children and their ability to recognize sounds and words, have articulated a clear theory for how children can be taught to read (for a detailed summary, see Castles et al., 2018). When children are taught to segment words into their sounds and recognize which letters or letter-pairs correspond to those sounds, they can transfer that knowledge to new words with relative ease. Without that explicit instruction, children do not readily learn this letter-sound correspondence (Byrne, 2005; Byrne & Fielding-Barnsley, 1990). Children also benefit from learning how to extract sounds from long strings of speech (i.e., segmentation; Melby-Lervåg et al., 2012). Once these skills are in place, children can begin to decode (Ehri, 2005; Ehri & Wilce, 1985). However, reading is a mix of decoding and comprehension (Hoover & Gough, 1990). A child who can decode words but not comprehend their meaning is not “reading,” while a child who can comprehend meaning only from spoken language but not from written text is not “reading” either. Reading relies on knowledge (including vocabulary and context) and cognitive resources (like executive function and working memory) as well (Perfetti & Stafura, 2014). Key strategies for promoting reading comprehension include discussion of text with peers and teachers for summarization and prediction as well as explicit instruction on vocabulary, among other techniques (Castles et al., 2018).

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 2 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% of K–2 teachers reported using this strategy to teach reading as recently as 2019 (Kurtz et al., 2020). Education researchers have also critiqued the science of reading (SOR) 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 SOR 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 SOR state initiative in 2013 (Kristof, 2023; Spencer, 2024; The Nation’s Report Card, 2013, 2019). 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 to 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 SOR but did not significantly improve students’ reading comprehension in grades 1 to 3 (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 SOR, 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 SOR 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 SOR 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 ELSBG change grade-3 achievement in English Language Arts (ELA) 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 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% 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 SOR 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 (Joyce & Cartwright, 2020) are widely recognized as a central impediment to evidence-based reforms. ELSBG’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 (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

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 ELSBG—a targeted and multifaceted 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. v. State of California*, 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 2017–2018 and 2018–2019 school years who scored at the lowest (i.e., “Standard Not Met”) of the four levels on the state’s English Language Arts/Literacy (ELA) assessment and weighting the average by the number of test-takers in each of the years (State of California, Legislature. Senate Bill 98, 2020). Because one of the 75 eligible schools closed before the case was settled, California Department of Education (CDE) expanded eligibility to the 76th lowest-performing school based on this baseline score.

In August 2020, the CDE 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 3-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 2020–2021 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 1-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 2021–2022, SY 2022–2023, and SY 2023–2024). The state map in Supplemental Appendix Figure A1 (available in the online version of this article) shows that the ELSBG-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 and in addition to 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 (a)

high-quality literacy teaching (e.g., new instructional coaches, increased professional development), (b) support for literacy learning (e.g., diagnostic assessment tools, instructional materials), (c) pupil supports (e.g., tutoring, after-school programming), and (d) 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 (State of California, Legislature, Senate Bill 98, 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 toward 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 include 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 SOR. 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% of respondents reported that the Online Elementary Reading Academy either met or exceeded their expectations while 98% 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% 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 SOR, constituted the largest expenditure category with 28% of budgeted expenditures (or 40% of staffing expenditures). Outside professional development accounted for 12.55% of total ELSBG budgeted spending while another 9.69% was budgeted toward books, supplies, or technology. The remaining 6.64% 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%) on hiring new on-site coaches compared to 28% in medium-sized schools and 25% 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%) than in towns or rural communities (63%). 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 SOR (D’Souza & Vasquez, 2022). Other roundtable participants underscored how ELSBG increased practitioners’ understanding of and appreciation for the SOR. 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 SOR 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-1 instructional strategies, the framework emphasizes that students must “be phonemically aware” to learn to read (CDE, 2014). However, this framework represents recommendations to educators, rather than mandates as embraced by other states. In 2024, state lawmakers proposed legislation that would mandate instructional materials adoption aligned to the 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 SOR 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. In the following legislative session, in 2025, legislators reached a compromise to unanimously pass a bill that would update California’s list of recommended instructional materials to align with evidence-based teaching practices, require the state to provide a list of professional development programs that also support these teaching methods, and mandate that teacher preparation programs include this content (Fensterwald, 2025).

However, without state curricular mandates, the choice of whether and how to use instructional practices supported by the SOR still occurs at the district level in California. A 2022 study of reading programs in more than 300 California districts (enrolling 72% of students in the state) showed that, despite this local flexibility, two curricula were chosen by 72% of surveyed districts (California Reading Coalition,

2022). Those two popular curricula are listed by their publishers as aligned with SOR, but advocates for SOR 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 SOR (Gamse et al., 2008; Westall & 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 and current policy under the ESSA. The evidence on the implementation quality and impact of those reforms is at best mixed (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 2 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 .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 2014–2015 school year to the 2022–2023 school year. This implies 7 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 2 years of post-treatment test scores (i.e., those taken in spring 2022 and spring 2023) correspond to the first 2 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 (SDs) publicly reported by the CDE.

Using data from the CDE’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., SDID) 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 Supplemental Appendix Table A1 in the online version of the journal) 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 ELA test scores 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 2017–2018 and 2018–2019 standardized ELA tests) have much lower test scores in ELA than comparison schools. Specifically, in ITT schools, only 31.15% 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%). 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 DDD (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 Supplemental Appendix Table A2 (available in the online version of this article), we present the baseline (i.e., 2014–2015 to 2018–2019) 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% of students in ELSBG-eligible schools were also eligible for free or reduced-price lunches and nearly 20% were Black. The corresponding averages in ELSBG-ineligible schools were 61% and 5%, respectively.

In exploratory analyses, we also use publicly available budgets submitted to the CDE and the SCOE 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

TABLE 1

Descriptive Statistics for Test Scores by Grade, Subject, and Intent-to-Treat

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

Note. School-year test data are based on the 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 seven school years from 2014–2015 to 2022–2023, excluding SY 2019 to 2020 when tests were not administered and SY 2020 to 2021 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. CAASPP=California Assessment of Student Performance and Progress; ELA=English Language Arts; ITT=intent-to-treat status.

July and December 2021, school districts or CMOs 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 (2020–2021) or one of the three grant implementation years (2021–2022, 2022–2023, and 2023–2024). 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 10 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 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 2018–2019 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% of their students scoring Level 2 or higher on grade-3 ELA are comparable to California elementary schools with 37% 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.

DID Design

Because the overall results from the RD design are mixed, we also explore our results using a 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% Level 2 or higher; Table 1) while the comparison schools have a right-shifted test-score distribution (i.e., 68% 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:

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

in which outcome Y_{st} in school s and year t is a function of school and year fixed effects (i.e., α_s, 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 Supplemental Appendix Figure A3 in the online version of the journal) provide evidence inconsistent with this assumption. Specifically, across all four 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 DID Design

We address the implied internal-validity threats to this DID-based approach in several ways. Our main approach is to rely on a 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, $\hat{\omega}_s^{\text{sdid}}$, that optimally align pre-treatment trends across treated and comparison units. The SDID procedure also introduces time-specific weights, $\hat{\lambda}_t^{\text{sdid}}$, 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., $\hat{\tau}^{\text{sdid}}$) through this least-squares minimization:

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

To conduct statistical inference for estimates based on Equation 2, we rely on a block-bootstrap procedure (Arkhangelsky et al., 2021; Algorithm 2), which, though computationally intensive, performs well particularly in settings like ours where the number of treated units is large (Clarke et al., 2023). The central appeal of the SDID procedure is that it addresses internal-validity concerns by focusing on comparisons between treated units and similar comparison units (i.e., a type of “localness” noted by Arkhangelsky et al., 2021). The simultaneous use of both unit and time weights also enhances this localness by giving this procedure “a type of double robustness property” that reduces the influence of potential biases related to any one weight that may be misspecified (Arkhangelsky et al., 2021; Liu et al., 2024). 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).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% 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 Supplemental Appendix Table A1 (available in the online version of this article), 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 Supplemental Appendix Table A3 (available in the online version of this article), 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 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 TWFE (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 though that these DDD estimates 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, Supplemental Figures A4 to A6 indicate that, as with grade-3 ELA achievement (Supplemental Appendix Figure A3 in the online version of the journal), 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:

$$Y_s = \alpha_s + \beta C_s + X_s + \varepsilon_s, \quad (3)$$

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 to 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 (Supplemental Appendix Table A4 in the online version of the journal)—a virtually “sharp” assignment to treatment that is represented visually in panel A of Supplemental Appendix Figure A7 (available in the online version of this article). 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 $< .01$; column (2) in Supplemental Appendix Table A4 in the online version of the journal). 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 Supplemental Appendix Table A4 in the online version of the journal) 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 Supplemental Appendix Figure A7 in the online version of the journal) does not provide clear evidence of an impact, possibly due to the lack of power from the small sample size.

Synthetic 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% of a student-level SD (i.e., $0.14SD$, $p < .01$). Given that the annual reading-achievement gains of children between grades 2 and 3 are, on average, $0.60SDs$ (Hill et al., 2008), this effect size associated with ELSBG eligibility implies a gain of

TABLE 2

Estimated Effect of ELSBG on Third Grade ELA Test Scores

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

Note. These SDID ITT estimates are based on a balanced panel of California elementary schools that reported test scores over 7 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 receiving free- or reduced-price lunch (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. ELA=English Language Arts; ELSBG=Early Literacy Support Block Grant; ITT=intent-to-treat; SDID=synthetic “difference-in-differences.”

* $p < .1$. ** $p < .05$. *** $p < .01$.

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 $< > .01$). To put this estimated effect into perspective, we note that it constitutes a 20% 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 and 1.82 $< >$ percentage points, respectively—are also statistically significant ($p < .01$). 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% change in the percent of students at or above Level 3 and a 59% 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 *SDs*—is equivalent to 12% 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 fifth 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 (Common Core State Standards Initiative, 2010), and the CAASPP (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 6×3 , and why? . . . 3×6 because the order of the numbers does not matter in multiplication” (CAASPP,

TABLE 3

Estimated Effect of ELSBG on Other Grade-Subject Test Scores

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

Note. These SDID ITT estimates are based on a balanced panel of California elementary schools that reported the grade-subject test scores over 7 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. ELA=English Language Arts; ELSBG=Early Literacy Support Block Grant; ITT=intent-to-treat; SDID=synthetic “difference-in-differences.”

* $p < .1$. ** $p < .05$. *** $p < .01$.

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 (Supplemental Appendix Figure A3 in the online version of the journal).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 (Supplemental Appendix Figures A8–A10 in the online version of the journal). 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 (Supplemental Appendix Tables A6–A9 in the online version of the journal). 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 Tables 2 and 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 (Supplemental Appendix Table A10 in the online version of the journal).

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

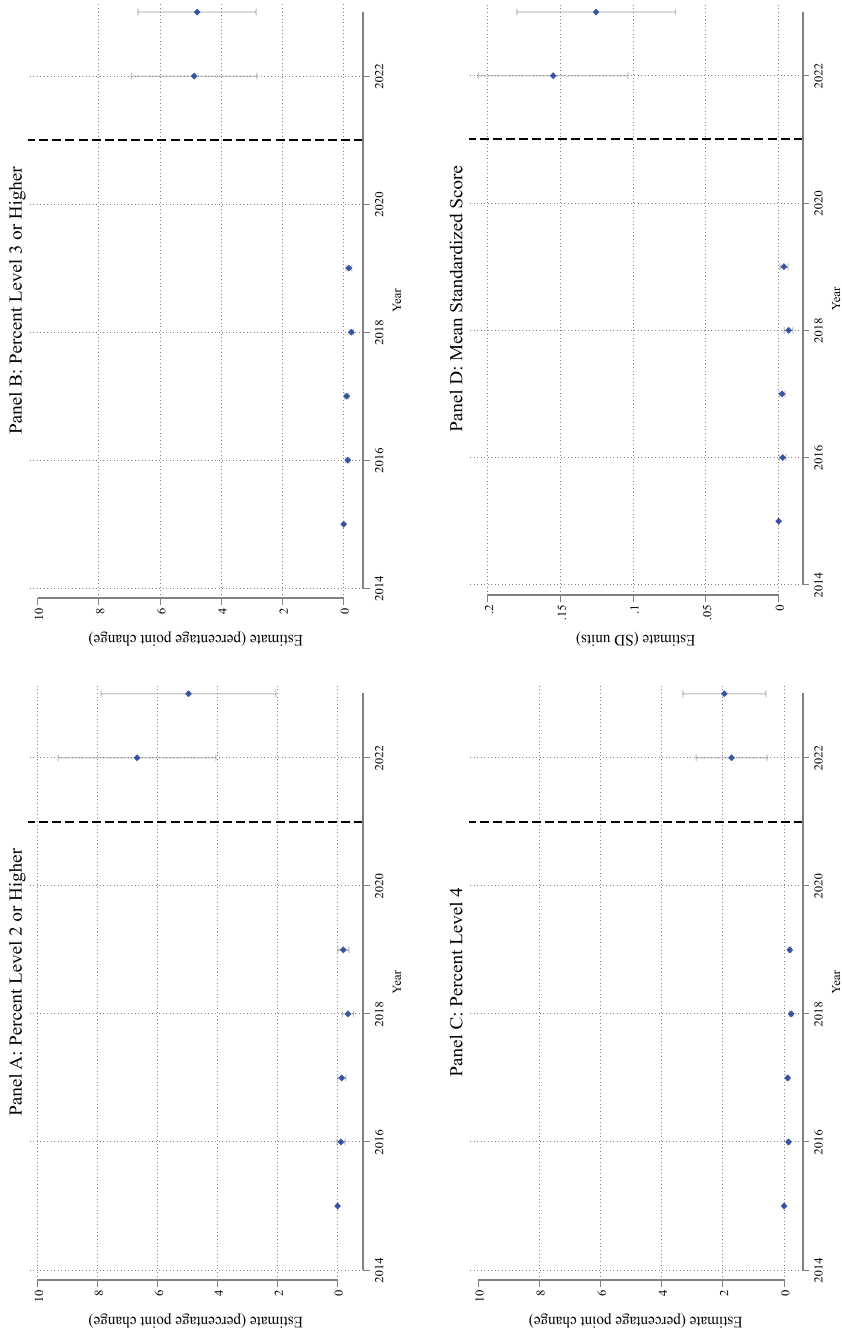


FIGURE 1. *ELSBG event-study estimates for grade-3 ELA test scores.*
 Note. ELA = English Language Arts; ELSBG = Early Literacy Support Block Grant.

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 (Supplemental Appendix Table A11 in the online version of the journal). The DDD estimates are smallest (i.e., effect size=0.05; $p < .01$) 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 Supplemental Appendix Table A12 (available in the online version of this article). 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 ELSBG 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 *SDs* among the more than 7,000 third graders served by the targeted schools over the first 2 years of the grant. This is a larger effect size than almost 90% 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% (i.e., a 6.00 percentage-point gain relative to a pre-treatment baseline of 30.56%). We also find that this initiative also led to smaller gains in grade-3 math achievement (i.e., 0.11 *SDs*) and, as

intended, had no effects among grade-5 students outside the program's focus.

These results occurred in schools serving high percentages of English learner students—specifically, the population of ELSBG schools was 42.91% English learners in the year before the program's implementation. This is of particular note due to the ongoing debate in California and other states about whether the SOR movement has adequately accounted for the diverse needs of multilingual learners in its implementation (Bramble, 2024; Schwartz, 2022).

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 SCOE 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 2 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 2021–2022 school year) scored similarly to 2022–2023 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 SCOE as the Expert Lead in Literacy, and \$1 million spent by the CDE). These resources supported the program among 15,541 K–3 students in 75 schools, implying an average 1-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 (2,021 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 to 2021 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 and Mackevicius (2021) suggest that a 1-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; Supplemental Appendix Table A11 in the online version of the journal), 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 ELSBG 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.

Acknowledgments

We thank staff members at the California Department of Education (CDE) and the Sacramento County Office of Education (SDOE), in particular Becky Sullivan, for their contributions to our understanding of this program.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sarah Novicoff  <https://orcid.org/0009-0007-4105-1962>

Thomas S. Dee  <https://orcid.org/0000-0001-7524-768X>

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 CDE retained \$3 million of this appropriation to support its administration and oversight of the ELSBG program.

3. In 2020–2021, the year prior to ELSBG implementation, the average per pupil expenditures for a California school was \$14,370 (Ed-Data, 2021). Therefore, ELSBG's addition of \$1,144 per pupil in 2021 to 2022 represents an 8% increase in the typical school budget.

4. In California, County Offices of Education 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 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 2018–2019 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 2021–2022 school year. Five schools remained open from 2015 to 2023 but had their test score data censored in one of the 7 years of our study because they had fewer than 11 third graders take the ELA test that year. These 10 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. v. State of California, 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% 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 RD design (i.e., column 2 in Supplemental Appendix Table A2 in the online version of the journal).

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 (Supplemental Tables A1 and A3 in the online version of the journal).

15. Interestingly, despite the evidence of pre-trends in TWFE-based estimates, the impact estimates based on that approach (Supplemental Appendix Table A5 in the online version of the journal) are similar to those reported in Table 2.

16. Jackson and Mackevicius (2021) show that the average \$1,000 (in 2018 dollars) increase in per-pupil public school spending over 4 years (i.e., roughly a \$4,000 increase per student) increases test scores by 0.0352SD. To make this comparable with ELSBG cost estimates for the first year of programming in SY 2021 to 2022, we use the Consumer Price Index to adjust for inflation and divide by four to obtain an estimate for a single year.

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_DlgfNdkljd/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>
- Byrne, B. (2005). Theories of learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 104–119). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470757642.ch6>
- Byrne, B., & Fielding-Barnsley, R. (1990). Acquiring the alphabetic principle: A case for teaching recognition of phoneme identity. *Journal of Educational Psychology*, *82*(4), 805–812. <https://doi.org/10.1037/0022-0663.82.4.805>
- CAASPP Technical Reports and Studies. (2024, March 19). *Educational testing service*. <https://www.cde.ca.gov/ta/tg/ca/caaspprtstudies.asp>
- California Assessment of Student Performance and Progress. (2022). *Online practice and training tests*. <https://www.caaspp.org/practice-and-training/>
- California Department of Education. (2014, July 9). *English language arts/English language development framework*. <https://www.cde.ca.gov/ci/rl/cf/documents/elaeldfwchapter3.pdf>
- California Reading Coalition. (2022). *The California reading curriculum report*. <https://www.careads.org/curric-report-viewer>
- 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., Pailaňir, D., Athey, S., & Imbens, G. (2023). *Synthetic difference in differences estimation* (arXiv:2301.11859). arXiv. <https://doi.org/10.48550/arXiv.2301.11859>
- Common Core State Standards Initiative. (2010). National Governors Association Center for Best Practices and Council of Chief State School Officers. *Grade 3 operations & algebraic thinking*.
- 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://edsources.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://edsources.org/broadcasts/early-literacy-for-all-in-california>
- Ed-Data. (2021). *California general fund expenditures, school districts, 2020-21*. <https://www.ed-data.org/ShareData/Html/119580>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? On JSTOR. *Reading Research Quarterly*, 20(2), 163–179.
- Ella T. v. State of California. (2020, February 20). *Settlement implementation agreement*. 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>
- ELSB Grant Eligible Schools. (2020). *California department of education*. <https://www.cde.ca.gov/pd/ps/elsbeligibility.asp>
- Fensterwald, J. (2025, April 30). Legislative deal on reading instruction reached in the nick of time. *EdSource*. <https://edsources.org/2025/legislative-deal-on-reading-instruction-reached-in-the-nick-of-time/731682>
- 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/pubinfo.asp?pubid=NCEE20094038>
- Hemelt, S. W., & Jacob, B. (2017). *Differentiated accountability and education production: Evidence from NCLB waivers* (Working Paper No. 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>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Huerta-Price, B., & Sanchez, C. (2022). *CUSD ELSB grant annual report Y1*. Calexico Unified School District. https://docs.google.com/document/d/1WI6EZrkyk3CaEoG3VczclQ_ITYsBLGrJ/edit
- Jackson, C. K., & Mackevicius, C. (2021). *The distribution of school spending impacts*. 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*, 52(3), 183–187. <https://doi.org/10.3102/0013189X231155154>
- Kristof, N. (2023, May 31). Opinion | Mississippi is offering lessons for America on education. *The New York Times*. <https://www.nytimes.com/2023/05/31/opinion/mississippi-education-poverty.html>
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2), 497–532.
- Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, 51(7), 500–506. <https://doi.org/10.3102/0013189X221109178>
- Kurtz, H., Lloyd, S., Harwin, A., Chen, V., & Furuya, Y. (2020). *Early reading instruction: Results from a National Survey*. EdWeek Research Center.

- <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://edsources.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. (2024). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*. Advance online publication. <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.
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352. <https://doi.org/10.1037/a0026744>
- 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>
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- 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.mofa.com/documents/171205-ellat-t-v-california-complaint.pdf>
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2(2), 31–74. <https://doi.org/10.1111/1529-1006.00004>
- Schwartz, S. (2022, April 21). The "Science of Reading" and English-Language Learners: What the research says. *Education Week*. <https://www.edweek.org/teaching-learning/the-science-of-reading-and-english-language-learners-what-the-research-says/2022/04>
- 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>
- State of California, Legislature. Senate Bill 98. (2020). *SB 98 authorization of the ELSB grant*. <https://www.cde.ca.gov/pd/ps/elsbgrantsb98.asp>
- Strunk, K., Wright, T., Kilbride, T., Zhu, Q., Cummings, A., West, J., Turner, M., & De Voto, C. (2021). *Michigan's read by grade three law: Year one report*. Education Policy Innovation Collaborative at Michigan State University. <https://epicedpolicy.org/rbg3-year-one-report/>
- 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/1cu3zRCNsNHSAMWGM24gdOVlHyVet7NRx/view>
- Tadayon, A. (2022, October 11). Coalition says "science of reading-aligned" core curricula barely used in California. *EdSource*. <https://edsources.org/2022/coalition-says-science-of-reading-aligned-core-curricula-barely-used-in-california/679596>
- The Nation's Report Card. (2019). *NAEP reading 2019 highlights*. National Assessment of Education Progress. <https://www.nationsreportcard.gov/highlights/reading/2019/>
- The Nation's Report Card. (2013). *National Assessment of Student Progress (NAEP) 2013 and 2022 reading assessments*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. <https://www.nationsreportcard.gov/ndecore/xplore/NDE>

- The Nation's Report Card. (2022). *NAEP reading: National achievement-level results*. National Assessment of Education Progress. <https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4>
- 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*, 35(1), 1–6. <https://eric.ed.gov/?id=EJ848235>
- Westall, J., & Cummings, A. (2023). *The effects of early literacy policies on student achievement* (Ed Working Papers 23-788). <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/>

Authors

SARAH NOVICOFF, MA, is a PhD candidate in educational policy and economics of education at the Stanford University Graduate School of Education. Her research uses quantitative and mixed-methods to understand how policy can improve the educational experiences of students from traditionally underserved backgrounds.

THOMAS S. DEE, PhD, is the Barnett Family Professor at Stanford University. His research focuses largely on the use of data and quantitative methods to inform contemporary issues of education policy and practice.

Manuscript received July 19, 2024

First revision received December 23, 2024

Second revision received March 27, 2025

Accepted May 9, 2025