IES Institute of Education Sciences

Appendix

Did the Striving Readers Comprehensive Literacy Grant Program Reach Its Goals?

An Implementation Report

NCEE 2024-003a
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation at IES

This report was prepared for IES under Contract No. 91990018C0020 by the American Institutes for Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

May 2024

Garet, M. S., Le Floch, K. C., Hubbard, D., Carminucci, J., & Goodson, B. (2024). *Did the Striving Readers Comprehensive Literacy Grant Program reach its goals? An implementation report* (NCEE 2024-003r). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee

This report is available on the Institute of Education Sciences website at http://ies.ed.gov/ncee.

Did the Striving Readers Comprehensive Literacy Grant Program Reach Its Goals?

Appendix

May 2024

Michael S. Garet
Kerstin Carlson Le Floch
Daniel Hubbard
Joanne Carminucci
American Institutes for Research

Barbara Goodson
Abt Associates

NCEE 2024-003a
U.S. DEPARTMENT OF EDUCATION

Contents

Exhibits

INTRODUCTION

Boosting literacy among school-age children remains a national priority. Nearly one third of students in the United States have not developed the foundational reading skills needed to succeed academically, with students living in poverty, students with disabilities, and English learners especially at risk. Starting in 2010, Congress invested more than $1 billion for state literacy improvement efforts through the Striving Readers Comprehensive Literacy (SRCL) program. SRCL was intended to focus funding on disadvantaged schools, encourage schools to use evidence-based practices, and support schools and teachers in providing comprehensive literacy instruction. These efforts were expected to lead to improved literacy outcomes for students. This study assesses how well SRCL implementation was aligned with these goals, using information collected from states, districts, and schools in all 11 states awarded grants in 2017. This document provides supporting details on the SRCL program and how the study was carried out, including supplementary analyses for the findings presented in the report.

APPENDIX A. BACKGROUND ON STRIVING READERS COMPREHENSIVE LITERACY GRANT PROGRAM

This introductory appendix provides background on the Striving Readers Comprehensive Literacy (SRCL) grant program. It begins with an overview of the federal program guidelines (section A.1). Next it lists the states that received awards (section A.2) and provides descriptive information about the use of SRCL funds, by state (section A.3). Finally, the appendix summarizes each state's approach to SRCL (section A.4).

A.1. Program Background

The SRCL program was authorized as part of the fiscal year 2010 Consolidated Appropriations Act (Pub. L. No. 111-117) under the Title I demonstration authority (Part E, Section 1502 of the Elementary and Secondary Education Act). The State Funding Allocations requirement was from Title III of Division H of the Consolidated Appropriations Act, 2016 (Pub. L. 1s14-113).

The purpose of the SRCL discretionary grants was to create a comprehensive literacy program within states to advance literacy skills–including preliteracy skills, reading, and writing–for students from birth through grade 12, including English learners and students with disabilities. Research has found that a comprehensive literacy approach can have long-lasting effects on student reading performance (see, for example, Ferguson et al., 2011). To guide how these comprehensive programs would be developed, the 2017 SRCL Notice Inviting Applicants (NIA) included one absolute priority that states were required to plan for and implement, and two competitive preference priorities for which they would receive extra points in the application scoring if they appropriately described and committed to implement them as part of their state program. These priorities, listed in Exhibit A.1, are a signal of the U.S. Department of Education's goals for SRCL at the time that the NIA was written.

Exhibit A.1. Excerpted priorities, 2017 SRCL NIA

Description of Absolute and Competitive Preference Priorities for SRCL
Absolute Priority: Interventions and Practices Supported by Moderate or Strong Evidence
Under this priority, a State educational agency (SEA) must ensure that evidence plays a central role in the SRCL subgrants. Specifically, in its high-quality plan, an SEA must assure that (1) it will use an independent peer review process to prioritize awards to eligible subgrantees that propose high-quality comprehensive literacy instruction programs that are supported by moderate evidence or strong evidence, where evidence is applicable and available, and (2) the comprehensive literacy instruction program proposed by eligible subgrantees will align with the State's comprehensive literacy plan as well as local needs.
Competitive Preference Priority 1: Serving Disadvantaged Children
Under this priority, an SEA must describe in its application a high-quality plan to award subgrants that will serve the greatest numbers or percentages of disadvantaged children, including children living in poverty, English learners, and children with disabilities.

Description of Absolute and Competitive Preference Priorities for SRCL
Competitive Preference Priority 2: Alignment within a Birth through Fifth Grade Continuum
Under this priority, an SEA must describe in its application a high-quality plan to align, through a progression of approaches appropriate for each age group, early language and literacy projects supported by this grant that serve children from birth to age five with programs and systems that serve students in kindergarten through grade five to improve school readiness and transitions for children across this continuum.
Definition of comprehensive literacy instruction
Comprehensive literacy instruction means instruction that - (a) Includes developmentally appropriate, contextually explicit, and systematic instruction, and frequent practice, in reading and writing across content areas; (b) Includes age-appropriate, explicit, systematic, and intentional instruction in phonological awareness, phonic decoding, vocabulary, language structure, reading fluency, and reading comprehension; (c) Includes age-appropriate, explicit instruction in writing, including opportunities for children to write with clear purposes, with critical reasoning appropriate to the topic and purpose, and with specific instruction and feedback from instructional staff; (d) Makes available and uses diverse, high-quality print materials that reflect the reading and development levels, and interests, of children; (e) Uses differentiated instructional approaches, including individual and small group instruction and discussion; (f) Provides opportunities for children to use language with peers and adults in order to develop language skills, including developing vocabulary; (g) Includes frequent practice of reading and writing strategies; (h) Uses age-appropriate, valid, and reliable screening assessments, diagnostic assessments, formative assessment processes, and summative assessments to identify a child's learning needs, to inform instruction, and to monitor the child's progress and the effects of instruction; (i) Uses strategies to enhance children's motivation to read and write and children's engagement in self-directed learning; (j) Incorporates the principles of universal design for learning; (k) Depends on teachers' collaboration in planning, instruction, and assessing a child's progress and on continuous professional learning; and (l) Links literacy instruction to the State's challenging academic standards, including standards relating to the ability to navigate, understand, and write about complex subject matters in print and digital formats.

A.2. States Awarded/Grant Amounts

SRCL grant applications were due in July 2017 and funds were awarded in October 2017. States then held subgrant competitions and disbursed funds in spring 2018. Seven states awarded 2-year subgrants and four states awarded 3-year subgrants. Thus, unless subgrantees received extensions from states, the 2019-20 study data represent the second and last implementation year for subgrantees in seven states and the second of 3 implementation years for subgrantees in the remaining four states.

Exhibit A.2 presents information on the SRCL grants and subgrants awarded for the 2018–19 cohort. Subgrants may be awarded to single school districts, stand-alone early childhood centers, and consortia of multiple districts. Kansas, Louisiana, New Mexico, and Ohio awarded some of their subgrants to consortia; the "Subgrants" column counts each consortium once, rather than counting each district in a consortium, whereas the "Districts Receiving Subgrants" column counts all districts in a consortium separately. Not all subgrantees within a state necessarily received the same amount of funding. For example, some states gave larger subgrants to applicants serving more students, whereas others gave larger subgrants to applicants serving more students who are disadvantaged. Other aspects of subgrant applications also may have affected subgrant sizes. Subgrantees were not obligated to provide SRCL funds to all schools in the district. Subgrantee applications listed which schools the applicants proposed to fund.

Exhibit A.2. SRCL grants and subgrants, by state

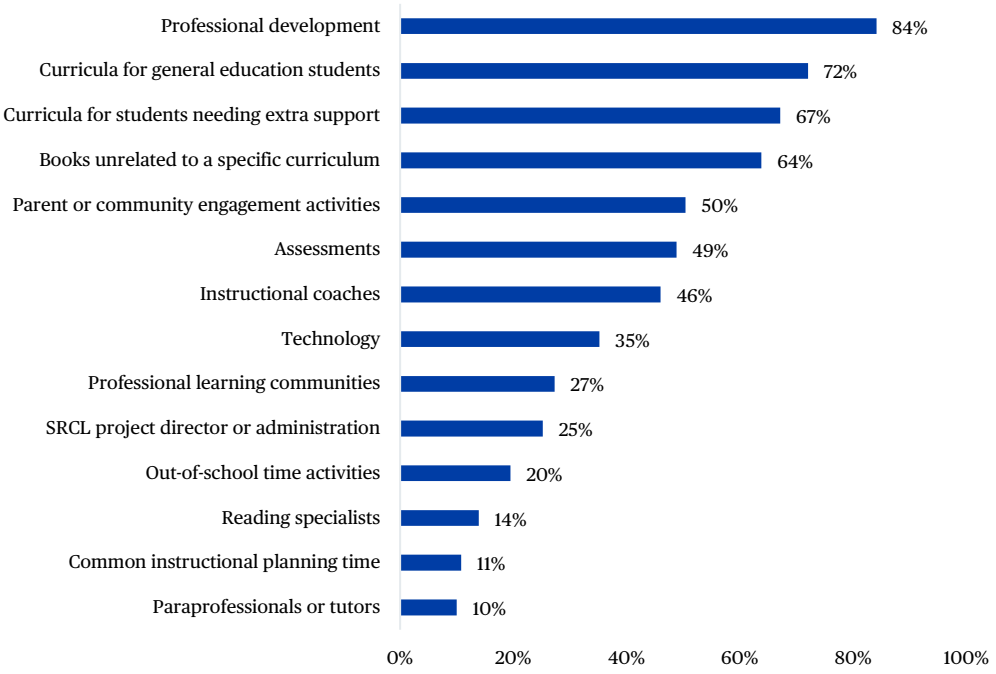
Grantee	Grant Size	Sub-grants	Districts Receiving Subgrants[a]	Total Districts in the State	Total Schools in the State	Schools Funded by SRCL Subgrants[b]	Early Childhood Centers Funded by SRCL Subgrants	Average Subgrant Size	Average per School
Georgia	$61,579,800	39	38	232	2,309	262	33	$1,578,969	$235,037
Kansas	$27,042,948	8	32	311	1,314	188	25	$3,380,369	$143,845
Kentucky	$24,933,147	41	40	186	1,534	206	4	$608,126	$121,035
Louisiana	$55,500,000	38	84	204	1,383	276	26	$1,460,526	$201,087
Maryland	$45,000,000	24	23	24	1,418	1,033	85	$1,875,000	$43,562
Minnesota	$21,150,093	23	23	564	2,430	57	10	$919,569	$371,054
Montana	$24,111,419	21	28	487	823	64	0	$1,148,163	$376,740
New Mexico	$20,000,000	11	13	146	882	104	3	$1,818,182	$192,308
North Dakota	$28,800,000	15	15	225	531	110	27	$1,920,000	$261,818
Ohio	$35,000,000	46	114	1,043	3,566	446	38	$760,870	$78,475
Oklahoma	$21,323,575	25	25	597	1,807	79	38	$852,943	$269,919
All	**$364,440,982**	**291**	**435**	**4,019**	**17,997**	**2,825**	**289**	**$1,252,375**	**$129,006**

[a] Excludes early childhood centers, libraries, private schools, and other entities not included in the Common Core of Data.
Source: Office of Elementary and Secondary Education list of grant awardees, 2018–19 Common Core of Data. Lists of funded schools provided by state education agencies.
Note: The number of districts receiving subgrants is greater than the number of subgrants if a state awarded subgrants to consortia consisting of multiple districts. The number of districts receiving subgrants is less than the number of subgrants if some subgrants were awarded to entities unaffiliated with school districts, including independent preschools, libraries, and museums.

A.3. Use of SRCL Funds

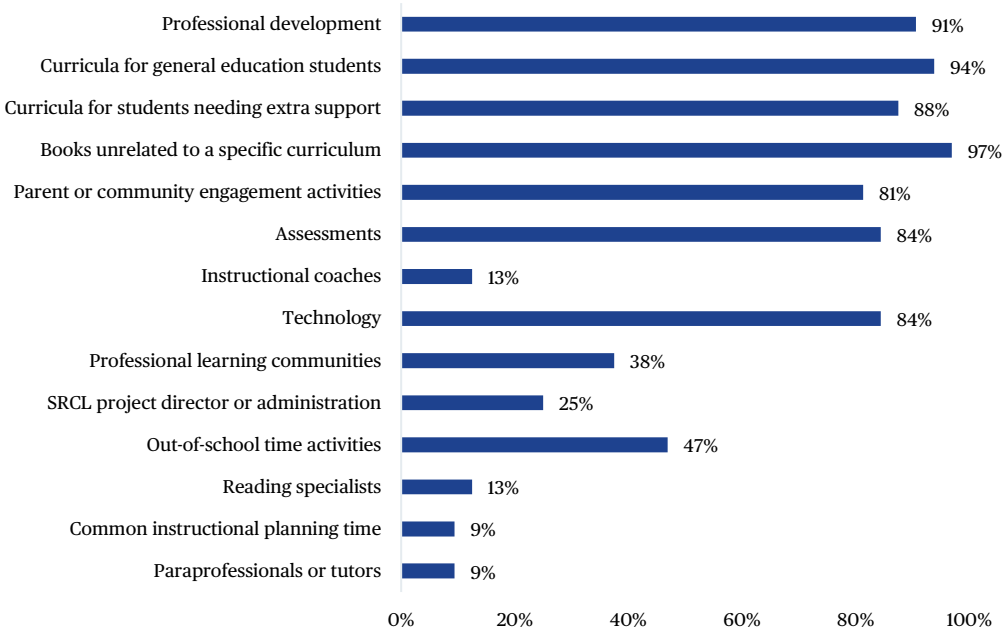
Exhibit A.3 presents the percentage of SRCL districts that spent SRCL funds on each of 14 resources in 2018–19, based on the 389 districts that responded to a study-administered survey of the 435 districts that were awarded SRCL subgrants. Exhibits A.4–A.14 present the same percentages separately by state (see Exhibit B.3 for information on state-specific sample sizes and response rates). The most common use of SRCL funds was professional development, which was purchased by 84 percent of SRCL districts. More than half of the SRCL districts also purchased general-purpose curricula, curricula for struggling readers, and books that were not part of any specific curriculum; 64 percent ranked the purchase of literacy curricula, either for general education students, students needing extra support, or both, among their top three uses of SRCL funds. Fewer than half of the SRCL districts spent SRCL funds on instructional coaching; however, this varied substantially across states. For instance, Minnesota made the use of SRCL funds to support instructional coaching an explicit priority in their requests for subgrant applications, and 100 percent of districts in Minnesota spent SRCL funds on instructional coaching. Other applications of SRCL funds, such as out-of-school time, reading specialists, common planning time, and tutoring, were not commonly purchased, with fewer than 20 percent of the SRCL districts reporting paying for them with SRCL funds.

Exhibit A.3. Percentage of SRCL districts using SRCL funds on types of resources, 2018–19



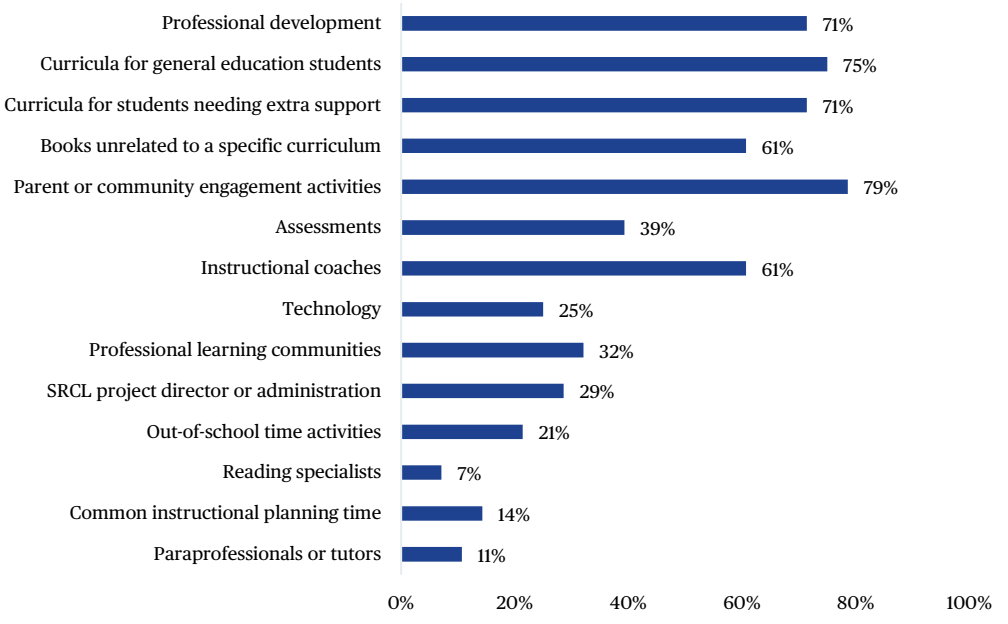
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 389 districts.
Note: 82 percent of responding districts spent SRCL funds on curricula for general education students, curricula for students needing extra support, or both.

Exhibit A.4. Percentage of Georgia SRCL districts using SRCL funds on types of resources, 2018–19



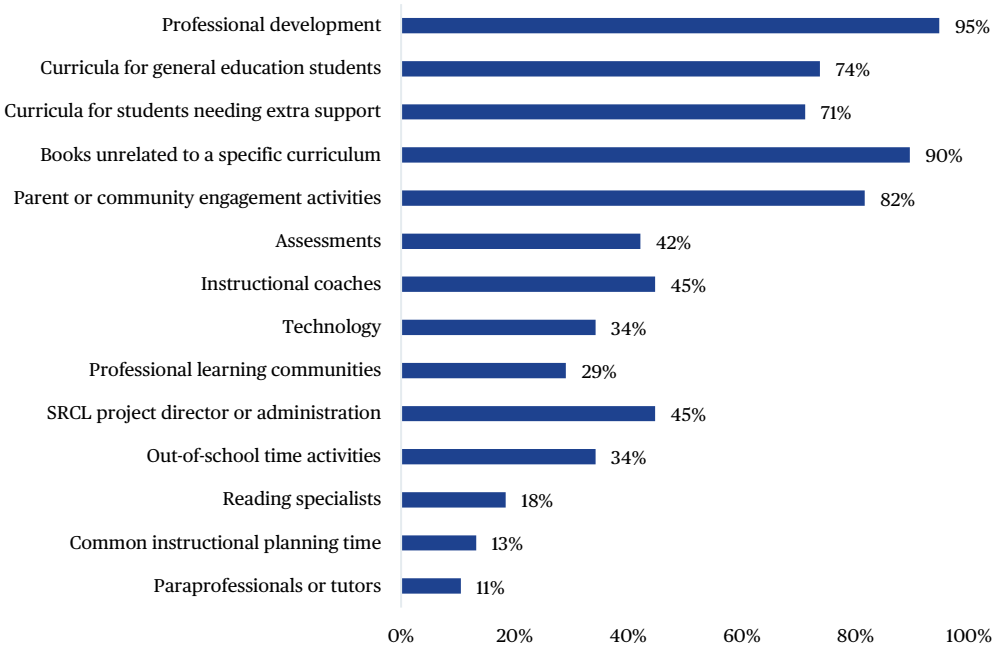
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 32 districts in Georgia.

Exhibit A.5. Percentage of Kansas SRCL districts using SRCL funds on types of resources, 2018–19



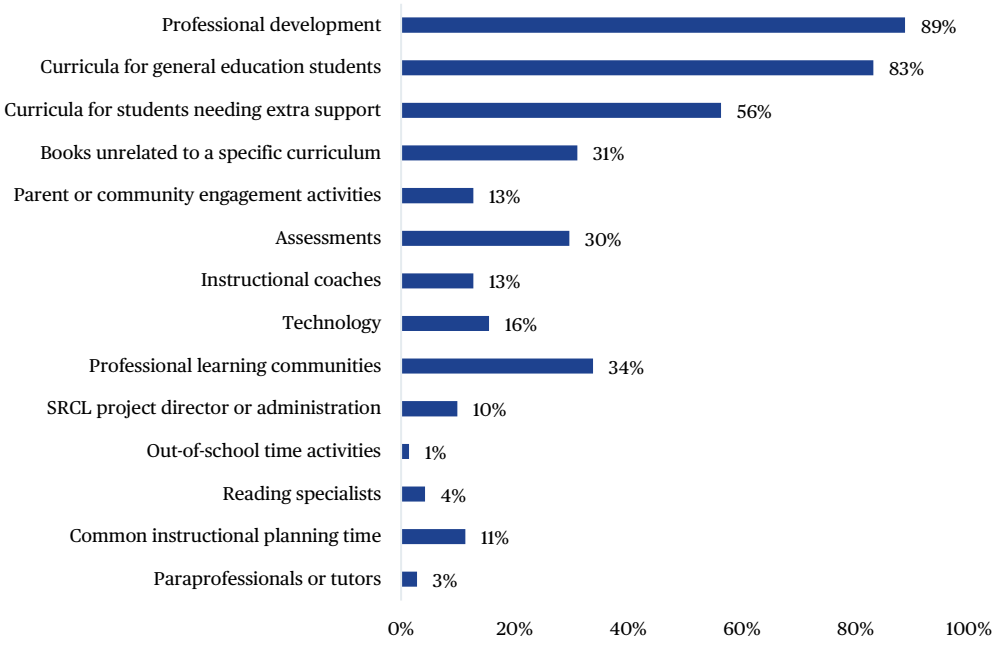
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 28 districts in Kansas.

Exhibit A.6. Percentage of Kentucky SRCL districts using SRCL funds on types of resources, 2018–19



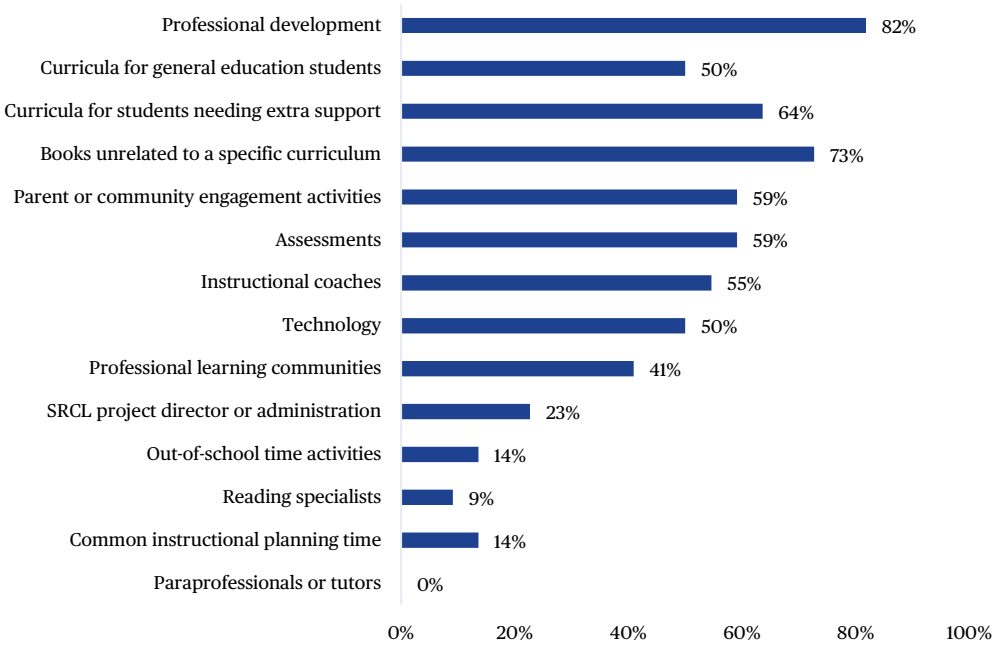
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 38 districts in Kentucky.

Exhibit A.7. Percentage of Louisiana SRCL districts using SRCL funds on types of resources, 2018–19



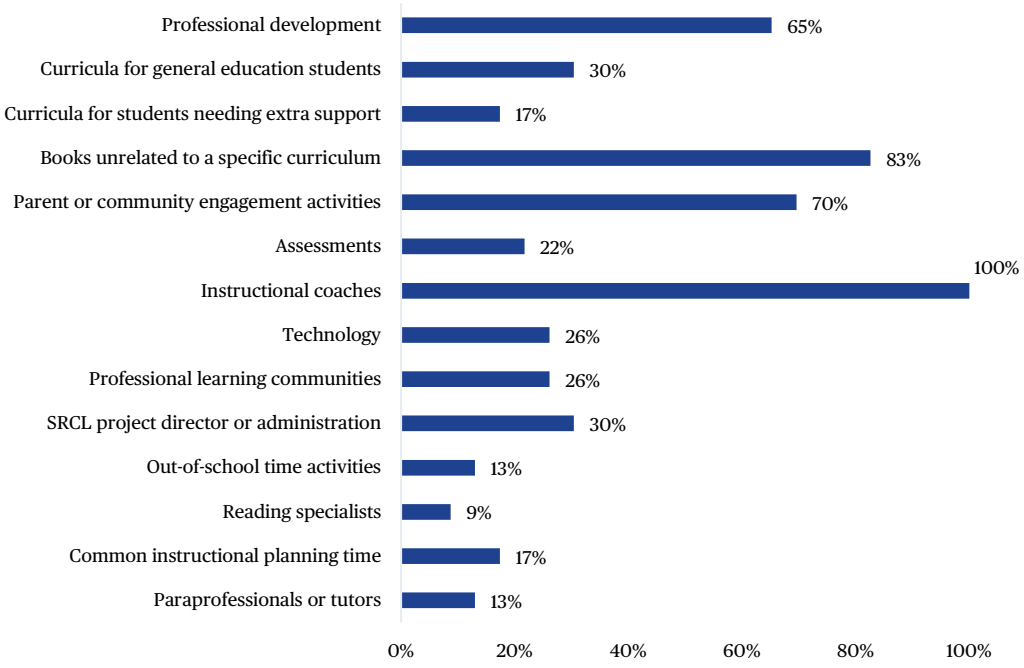
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 71 districts in Louisiana.

Exhibit A.8. Percentage of Maryland SRCL districts using SRCL funds on types of resources, 2018–19



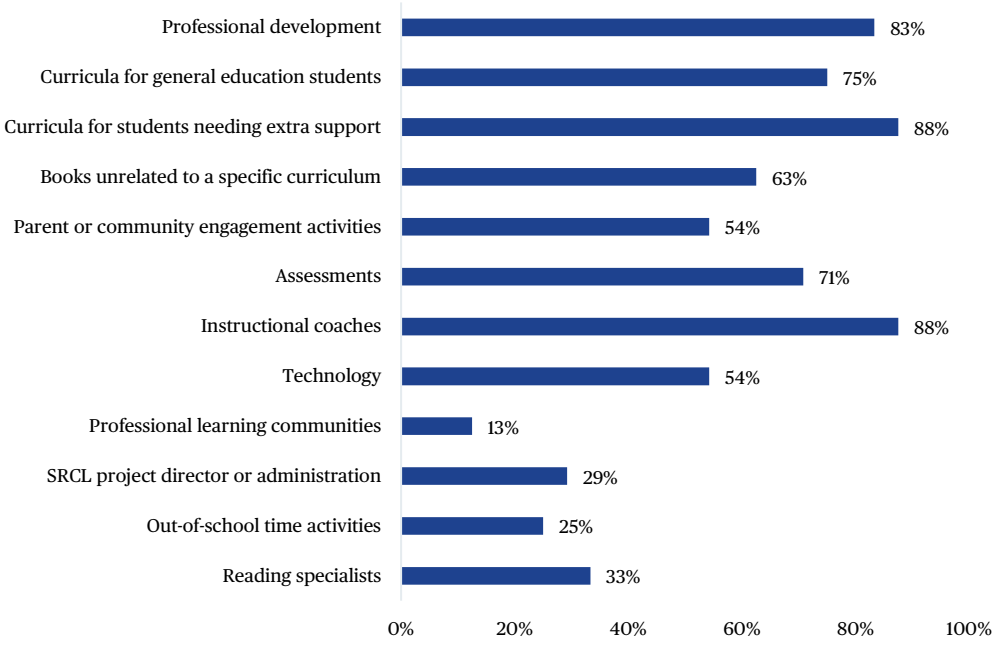
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 22 districts in Maryland.

Exhibit A.9. Percentage of Minnesota SRCL districts using SRCL funds on types of resources, 2018–19



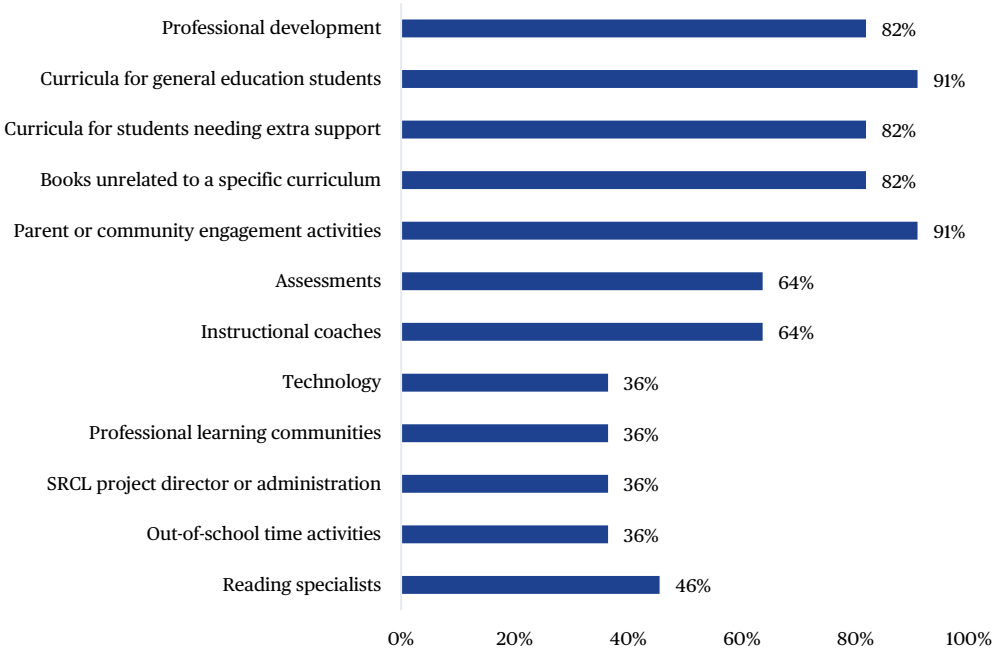
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 23 districts in Minnesota.

Exhibit A.10. Percentage of Montana SRCL districts using SRCL funds on types of resources, 2018–19



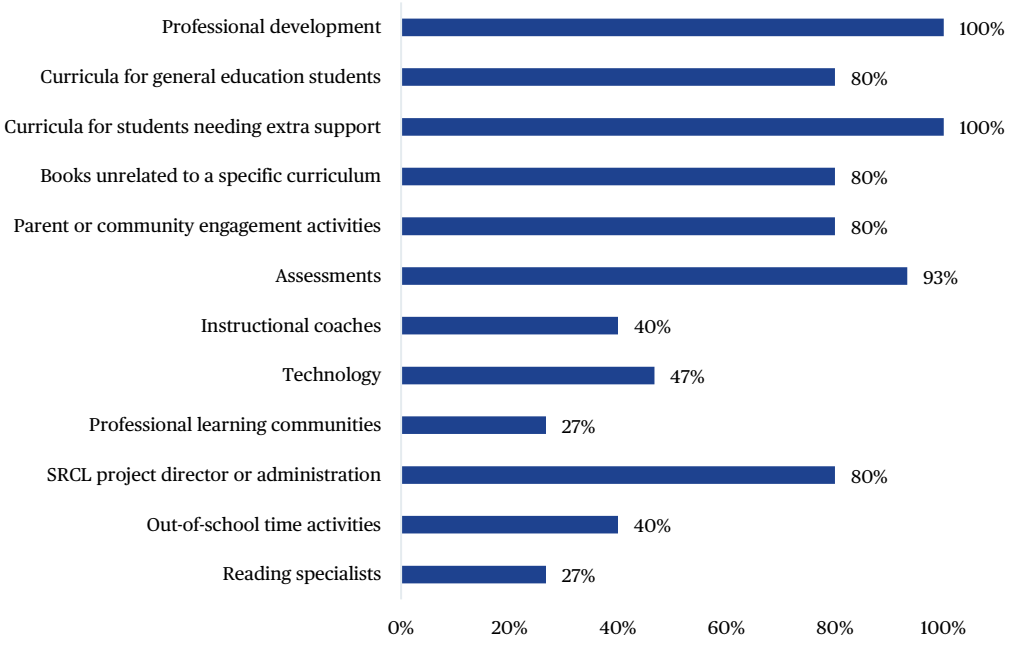
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 24 districts in Montana.

Exhibit A.11. Percentage of New Mexico SRCL districts using SRCL funds on types of resources, 2018–19



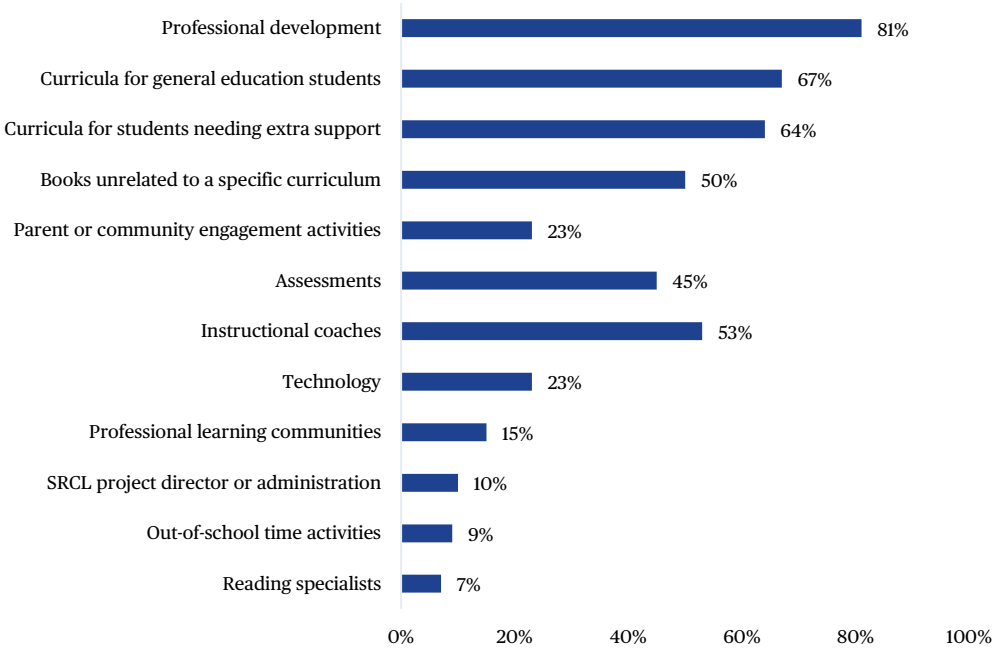
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 11 districts in New Mexico.

Exhibit A.12. Percentage of North Dakota SRCL districts using SRCL funds on types of resources, 2018–19



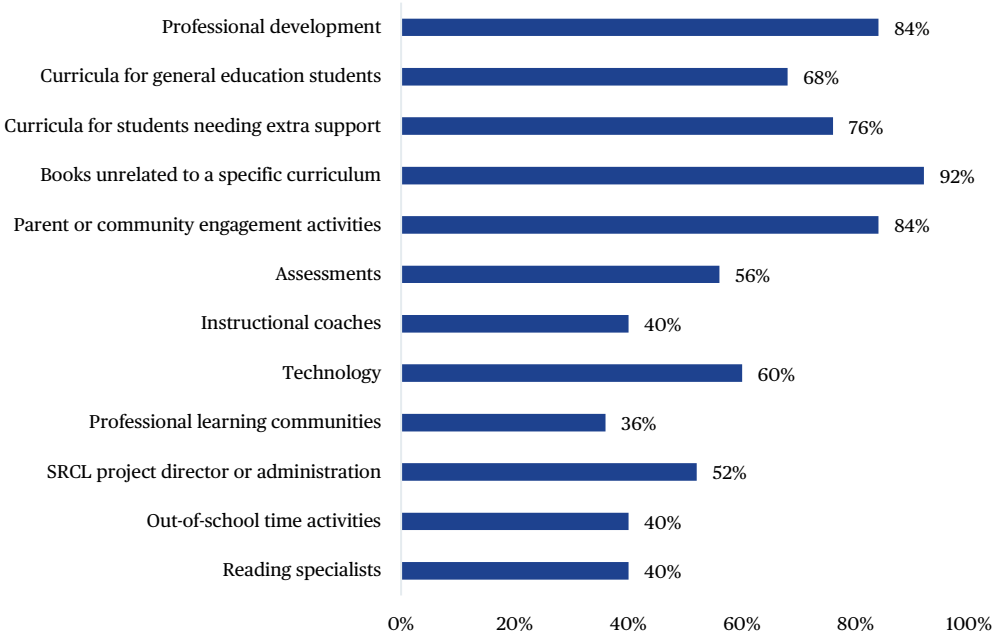
Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 15 districts in North Dakota.

Exhibit A.13. Percentage of Ohio SRCL districts using SRCL funds on types of resources, 2018–19



Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 100 districts in Ohio.

Exhibit A.14. Percentage of Oklahoma SRCL districts using SRCL funds on types of resources, 2018–19



Source: Survey of SRCL districts, spring 2019, question 13. The unweighted sample size is 25 districts in Oklahoma.

A.4. Supplementary Information on State Approaches to SRCL

The following tables provide information on the following topics: background on each state's SRCL grant (Exhibit A.15), state requirements for targeting disadvantaged students (Exhibit A.16), state emphasis on use of evidence, state approaches to use of evidence (Exhibit A.17), and state technical assistance in support of comprehensive literacy instruction (Exhibit A.18). The information in these tables was derived from interviews with state administrators and state policy documents, including the state SRCL grant applications and state requests for subgrant applications.

Exhibit A.15. Background on state SRCL grants

State	State Had a Prior SRCL Award	Percentage of Districts in State Receiving SRCL Funds	State Had Requirements for Use of SRCL Funds	If Yes, What Are the Requirements?
Georgia	Yes	17%	Yes	Subgrantees were required to: (1) participate in grantee-provided professional learning and (2) implement professional learning communities, peer coaching, and collaborative planning time.
Kansas	Yes	11%	Yes	Subgrantees were required to: (1) provide training on effective technology use for literacy development for parents, teachers, principals, and early learning program leaders and (2) provide instructional coaching for educators. In addition, subgrantees were required to hire an external evaluator.
Kentucky	Yes	22%	Yes	Subgrantees were required to: (1) implement at least one literacy program from the state-approved list; (2) provide professional learning aligned to the selected literacy program(s); and (3) use an assessment system aligned to state standards.
Louisiana	Yes	19%	Yes	Subgrantees were allowed to spend SRCL funds only on the following program components: (1) a curriculum from the state-approved list, (2) professional development from the state-approved list of vendors for each curriculum), and (3) assessments related to the chosen curriculum.
Maryland	No	100%	No	Not applicable
Minnesota	Yes	4%	Yes	Subgrantees were required to: (1) hire literacy coaches, and (2) implement instructional practices from the state's list of evidence-based Tier 1 instructional practices.

State	State Had a Prior SRCL Award	Percentage of Districts in State Receiving SRCL Funds	State Had Requirements for Use of SRCL Funds	If Yes, What Are the Requirements?
Montana	Yes	4%	Yes	Subgrantees were required to: (1) send school leadership teams to two statewide workshops, (2) use a walkthrough system, (3) form a district leadership team to meet with an instructional consultant affiliated with the state education agency, (4) identify time for teacher team meeting with a state-provided instructional consultant, and (5) use a set of required assessments.
New Mexico	Yes	10%	No	Not applicable
North Dakota	Yes	7%	No	Not applicable
Ohio	Yes	14%	No	Not applicable
Oklahoma	Yes	4%	No	Not applicable

Note: The percentage of districts in the state receiving SRCL funds is calculated using the total number of districts in the state, not the number of districts that would have been eligible based on student characteristics. If the number of districts in a consortium was unknown, the study team assumed three districts per consortium.
Kansas: 8 subgrantees comprising 4 consortia (consisting of 28 districts) and 4 district subgrantees
New Mexico: 11 subgrants comprising 10 districts and 1 consortium
Ohio: 46 subgrants comprising 25 districts or early childhood providers and 21 consortia
Source: State interviews with 11 state grant directors; state SRCL grant applications; state SRCL subgrant Request for Applications.

Exhibit A.16. State requirements for targeting disadvantaged students

State	Eligibility Requirement for Subgrant Applicants	Competitive Priority for Subgrant Applicants	Threshold for Percentage of Disadvantaged Students in Applicant Districts	Threshold for Percentage of Disadvantaged Students in Schools of Subgrant Applicants
Georgia	No	Yes	Competitive priority points were awarded to subgrantee applications based on the rank ordering of schools based on student growth rates on 3rd grade reading and the percentage of students who were low-income, English learners, or had disabilities.	No threshold specified for schools
Kansas	Yes	No	Subgrantees were required to meet at least one of the following criteria: at least 49 percent low-income students, at least 15 percent students with disabilities, or at least 10 percent English learners.	No threshold specified for schools
Kentucky	Yes	Yes	To be eligible for an award, subgrantees were required to serve at least 65 percent low-income students (defined as receiving free or reduced-price lunch). Competitive priority points were awarded using the following criteria: (1) serving a high percentage of disadvantaged students relative to other applicants, (2) alignment of Birth-5th grade, and (3) exceeding 125 percent of the state's rate of child homelessness.	No threshold specified for schools
Louisiana	Yes	Yes	Subgrantees were required to meet the threshold of 50 percent disadvantaged students across categories: low-income students, special education, English learners, and other types of disadvantage including migrant students and students in foster care. There was a competitive priority for districts with higher percentages of disadvantaged students.	No threshold specified for schools
Maryland	No	Yes	All districts in the state received a subgrant. The state used the percentage of disadvantaged students, including low-income students, English learners, and students with disabilities, in determining the funding amounts for each district.	No threshold specified for schools
Minnesota	No	Yes	The state prioritized subgrantees serving the greatest percentages of disadvantaged students, defined as those eligible for free or reduced-price lunch, students with disabilities, and English learners.	No threshold specified for schools

State	Eligibility Requirement for Subgrant Applicants	Competitive Priority for Subgrant Applicants	Threshold for Percentage of Disadvantaged Students in Applicant Districts	Threshold for Percentage of Disadvantaged Students in Schools of Subgrant Applicants
Montana	Yes	No	Subgrantees had to include (1) 50 percent or more students eligible for free or reduced-price meals, or (2) 15 percent or 1,000+ students identified with disabilities.	Each SRCL-funded school needed to enroll 40 percent or more students eligible for free or reduced-price meals, and either (a) less than 50% of students proficient on the ELA state assessment or (b) more than 20 percent of students identified as English learners.
New Mexico	No	Yes	The state prioritized subgrantees meeting at least one threshold for serving disadvantaged students. These thresholds were defined as (1) serving 75 percent low-income students, (2) serving 16 percent students receiving special education, and (3) serving 15 percent English learners. The state set these thresholds based on statewide averages.	No threshold specified for schools
North Dakota	Yes	Yes	Subgrantees were required to meet the threshold of serving 40 percent economically disadvantaged students, as well as serving high proportions of Native American students at risk of school failure and students with low literacy performance. The state awarded competitive priority points to subgrant applicants serving higher percentages of disadvantaged students.	No threshold specified for schools
Ohio	No	Yes	Subgrantee applicants were required to identify the percentage of children served who were English learners, economically disadvantaged, and developmentally delayed across three age bands. The state then allocated subgrant funding in accordance with (1) the regional share of statewide disadvantaged students, (2) the number of disadvantaged students to be served through the grant proposal, and (3) the state-level grade band distribution as defined by the U.S. Department of Education.	No threshold specified for schools

State	Eligibility Requirement for Subgrant Applicants	Competitive Priority for Subgrant Applicants	Threshold for Percentage of Disadvantaged Students in Applicant Districts	Threshold for Percentage of Disadvantaged Students in Schools of Subgrant Applicants
Oklahoma	Yes	Yes	Subgrantees were required to meet the following criteria: (1) at least 50 percent of students eligible for free or reduced lunch, and (2) either 20 percent of students identified for special education services, or at least 10 percent of students identified as English learners.	Subgrantees received additional competitive points if all of the elementary schools, the middle school, or the high school served through the grant met the 50 percent free or reduced-price lunch requirement.

Source: State interviews with 11 state grant directors; state SRCL grant applications; state SRCL subgrant Request for Applications.

Exhibit A.17. State approaches to use of evidence

State	Study Rating of State Emphasis on ESSA Strong or Moderate Evidence[a]	Required State List	Focused Primarily on Practices Rather Than Programs	State Allowed ESSA Tiers 3 and/or 4	State Reported Verifying Evidence Supporting Districts' Program Purchases Before Award or Funding	State Reported Helping Subgrantees Select Programs That Met ESSA Tiers 1 or 2 Post-Award	Directed Districts to Non-WWC Clearing-houses or Resources	If Yes, List Resources
Georgia	Medium	No	No	Yes, explicitly allowed	No; subgrantees were not required to propose programs in their applications.	Yes	Yes	REL Southeast, Promising Practices Network, Blueprints for Violence Prevention, and Social Programs That Work.
Kansas	Medium	No	Yes	Yes, implicitly allowed	Yes	No	Yes	EdReports, CORE Achieve (from the Consortium on Reaching Excellence in Education), and Learning Forward.
Kentucky	Medium	Yes[b]	No	Yes, implicitly allowed	No; required selection from state list	Yes	No	Not applicable
Louisiana	Minimal	Yes[c]	No	Yes, implicitly allowed	No; required selection from state list	No	No	Not applicable
Maryland	Medium	No	No	Yes, implicitly allowed	Yes	Yes	Yes	EdReports
Minnesota	Minimal	Yes[d]	Yes	Yes, implicitly allowed	No; required selection from state list	No	No	Not applicable
Montana	Major	No	No	No, restricted to Tiers 1 and 2	Yes	Yes	Yes	EdReports and a list of evidence-based programs from Colorado.

State	Study Rating of State Emphasis on ESSA Strong or Moderate Evidence[a]	Required State List	Focused Primarily on Practices Rather Than Programs	State Allowed ESSA Tiers 3 and/or 4	State Reported Verifying Evidence Supporting Districts' Program Purchases Before Award or Funding	State Reported Helping Subgrantees Select Programs That Met ESSA Tiers 1 or 2 Post-Award	Directed Districts to Non-WWC Clearing-houses or Resources	If Yes, List Resources
New Mexico	Minimal	No	No	Yes, implicitly allowed	No	No	Yes	EdReports
North Dakota	Major	No	No	No, restricted to Tiers 1 and 2	Yes	Yes	No	Not applicable
Ohio	Medium	No	No	Yes, explicitly allowed	Yes	Yes	Yes	REL Midwest and Ohio's Evidence-Based Clearinghouse.
Oklahoma	Medium	No	Yes	Yes, explicitly allowed	No	Yes	Yes	State created an online tool to help subgrantees select literacy programs and/or practices.

[a] For the major/medium/minimal ratings in the first column, states were categorized based on a rubric with the following elements: Overall state goal regarding strong or moderate evidence, approach to evidence use in subgrant application process, and approach to evidence use after subgrant awards. States were allocated 0, 1, or 2 points for each indicator, which were then summed. Overall scores of 0–1 = minimal, 2–4 = medium, 5–6 = major. Ratings for the indicators are as follows:
Focus on ESSA Strong or Moderate Evidence: Scores of 2 = state aimed to only fund interventions or practices with strong or moderate evidence. Scores of 1 = state encouraged interventions or practices with strong or moderate evidence. Scores of 0 = state's focus on evidence was not related to ESSA's definitions of strong or moderate evidence.
Approach to ESSA strong/moderate evidence at application stage: Scores of 2 = state attempted to require strong or moderate evidence for all interventions or practices proposed in application. Scores of 1 = state required strong or moderate evidence for at least one intervention or practice proposed in subgrantee applications or gave more points if proposed interventions and practices had strong or moderate evidence. Scores of 0 = strong or moderate evidence was not a factor in award decisions.
Approach to ESSA strong/moderate evidence at postaward stage: Scores of 2 = state worked with subgrantees to restrict purchases to only interventions or practices with strong or moderate evidence. Scores of 1 = state encouraged subgrantees to choose or provided support aimed at helping subgrantees choose interventions or practices with strong or moderate evidence. Scores of 0 = state did not actively guide subgrantees toward interventions or practices with strong or moderate evidence.
[b] Kentucky developed a matrix of approved services and supports—a list of 11 endorsed programs that the state considered to be evidence-based. Subgrantees had to agree to use at least one endorsed program for each age category. Although the evidence base was the initial reason for including each of these structured programs on Kentucky's list, the state prioritized programs that had sustainable, embedded frameworks for schools and communities. Each program was associated with expert coaching and professional supports. Each program had the capacity for statewide implementation in multiple sites and to provide connected comprehensive literacy instruction.
[c] Louisiana identified 20 curricula based on an evaluation of each program, which included, alignment with state content standards, age-appropriateness, coherence of tasks, and a

focus on foundational skills. Additional indicators of quality included supports for parent participation, scaffolding and support for diverse learners, range and volume of texts, and writing. Subgrantee applicants had to propose using programs on the state's predetermined list or they had to provide a justification for using other programs based on existing evidence of effectiveness.

[d] Minnesota provided a list of evidence-based practices (not literacy programs) focused on Tier 1 core instruction and required subgrantees to agree to implement practices from this list when they applied. The sources Minnesota used for these practices are documents on essential instructional practices in prekindergarten, K-3, and 4-5 prepared by the Michigan Association of Intermediate School Administrators General Education Leadership Network. Minnesota discouraged subgrantees from using SRCL funds to purchase literacy interventions or curricula.

Source: State interviews with 11 state grant directors; state SRCL grant applications, state SRCL subgrant Request for Applications.

Exhibit A.18. State technical assistance after subgrant award

State	State Provided Direct Coaching/TA to Subgrantee-Selected Coaches	State Provided TA on Comprehensive Literacy Instruction	Further Information About Technical Assistance
Georgia	No	Yes	In-person institutes to discuss statewide findings and to improve evidence-based practice areas that district data suggested needed additional emphasis. Online courses integrated course offerings from the state Comprehensive Reading Solutions website and from Georgia's institutions of higher education.
Kansas	Yes	Yes	Technical assistance provided in collaboration with the Center for Research on Learning at the University of Kansas and the Kansas affiliate of the National Council of Teachers of English. Individual technical assistance site visits in first year of the grant and convened subgrantee leaders monthly by phone, for opportunities to share best practices, network, and collaborate. Four thematic communities of practice covering adolescent literacy, family engagement (especially for families of English learners), instructional learning, and early learners. Online video coaching model and video library of model instructional practices. Approximately twice per month, teachers videotaped their instruction and uploaded the videos to an online data management platform; teachers then engaged in self-reflection and collaboration with their coach and received individualized, web-based coaching.
Kentucky	No	Yes	Initial kickoff meeting for subgrantees on literacy, which included literacy experts and breakouts on specific literacy topics. The state relied on professional development provided by developers of the endorsed programs or by the state system of regional educational cooperatives, established to provide comprehensive educational services to subgrantees in Kentucky.
Louisiana	No	Yes	In-person quarterly meetings with each SRCL school's teacher leader to provide training on components of comprehensive literacy instruction; teacher leaders were then responsible for supporting their peers' literacy instruction. Annual 3-day literacy summit for teacher leaders. Eight days per year of professional development for content leaders. Two state literacy networks provided on-site technical assistance and support.
Maryland	No	Yes	Full-day workshops for school-based staff on literacy topics. State engaged partners to provide technical assistance on specific topics. State professional development focused on two evidence-based programs: LETRS for early reading and Reading Apprenticeship for adolescents.

State	State Provided Direct Coaching/TA to Subgrantee-Selected Coaches	State Provided TA on Comprehensive Literacy Instruction	Further Information About Technical Assistance
Minnesota	Yes	Yes	Local coaches worked in small groups to create "practice profiles" for specific evidence-based practices, identifying core components and research-based "teacher moves" associated with each practice to use in coaching teachers.
Montana	No	Yes	Technical assistance provided directly to subgrantees by working with school leadership team. Annual session on evidence-based interventions and instructional coaching. Online learning hub for teachers. State-supported instructional consultants to provide professional development to subgrantees.
New Mexico	Yes	Yes	One-day summit that covered family engagement, identifying and implementing high quality instructional materials, demonstrations of possible assessments of literacy instruction. Talked about Critical Friends--defining reflective practice through a consultancy process.
North Dakota	No	Yes	SEA provided professional development on literacy transitions, multitiered systems of support, and topics tailored to subgrantees' reported needs. Seven state regional education centers provided professional development on literacy interventions and implementation fidelity.
Ohio	No	Yes	Two adolescent literacy specialists conducted in-person technical assistance. Created an adolescent literacy pilot in which a national expert worked with four schools and an adolescent literacy specialist focused on creating tools and building capacity beyond the life of the grant. Provided three full days of training on emergent literacy to state support teams, who then delivered training locally. Offered an online literacy academy consisting of recorded online sessions.
Oklahoma	No	Yes	Facilitated five communities of practice about literacy topics and one book study. Provided newsletters with tips and information about literacy and provided an online hub of resources about literacy. Visited each subgrantee three times, conducted webinars on literacy-related topics, registered grantees for other state-provided trainings (e.g., on curricula), and engaged support from the University of Oklahoma.

Source: State interviews with 11 state grant directors, state SRCL grant applications, state SRCL subgrant Request for Applications.

APPENDIX B. METHODS

The implementation study of the Striving Readers Comprehensive Literacy (SRCL) grant program was designed to address the following research questions:

1. To what extent do SRCL grantees and subgrantees carry out their efforts in ways that are aligned with the program's goals, including targeting disadvantaged students, using literacy programs supported by research evidence, and providing high-quality comprehensive literacy instruction?

2. How do trends in reading achievement differ for SRCL-funded schools versus similar non-funded schools?

This appendix describes the data sources for the study (section B.1), the methods used to determine the level of evidence supporting programs purchased with SRCL funds (section B.2), the approach to establishing and measuring benchmarks of comprehensive literacy instruction (section B.3), approaches for computing state averages and significance testing (section B.4), analysis of targeting of SRCL funds to districts and schools with disadvantaged students (section B.5), analysis of evidence use (section B.6), analysis of comprehensive literacy instruction (section B.7), and analysis of achievement trends in states that received SRCL funds (section B.8).

B.1. Data Sources

The following sections provide details on the data sources used in this study, starting with a general overview of data collection. The next section summarizes the state director interviews. A lengthier section provides details on the surveys, including questionnaire development, sampling, and weights. The final sections summarize the study's use of clearinghouses of evidence reviews, demographic data, and state assessment data.

B.1.1. Overview of Data Sources

The analyses conducted for this report primarily draw on data collected through surveys the study team administered during spring and summer 2019 to SRCL districts and to a representative sample of the leaders in SRCL schools (principals of K–12 schools and directors of early childhood centers, or ECCs for short). Surveys were also administered to samples of teachers in SRCL schools in two waves: spring and summer 2019 and spring 2020, just prior to school closures associated with the pandemic. Most analyses of the teacher survey data use the spring 2020 administration of the survey, as this was the final year of SRCL funding for most subgrantees. Both administrations of the survey are used in analyses of changes in teacher practice over time.

In addition, the study team interviewed state administrators with primary responsibility for SRCL in the fall of 2018, 2019, and 2020. A few analyses in this report draw on other information, including state SRCL grant applications submitted to the U.S. Department of Education, Requests for Applications (RFAs) through which states solicited subgrants, and subgrant applications.

Exhibit B.1 summarizes all data collection activities; additional information about these data sources and the surveys is provided in the narrative that follows. State administrative data and state aggregated test scores are discussed in section B.8.

Exhibit B.1. Data sources, including the sample, timing of data collection, response rate, and information obtained from each source

Data Source	Sample	Timing of Data Collection	*N*	Information Obtained
State director interviews	All 11 states	Fall 2018, fall 2019, fall 2020	11	Subgrant application processes, rationale for state approach to SRCL, state provision of technical assistance, selection of approved interventions, use of strong or moderate evidence
District survey	All 435 funded districts	Spring 2019	389	Use of SRCL funds, literacy programs purchased with SRCL funds, information sources regarding evidence of effectiveness, usefulness of technical assistance, selection of schools for receipt of SRCL funds
School leader survey	506 randomly sampled SRCL schools and early childhood centers	Spring 2019	449	Literacy programs used with struggling readers, organization of instruction, professional development in reading, perceived changes associated with SRCL
Teacher survey, wave 1	3,095 teachers in sampled schools and early childhood centers	Spring 2019	2,404	Implementation of comprehensive literacy instruction, professional development in literacy, teacher certification
Teacher survey, wave 2	3,097 teachers in sampled schools and early childhood centers	Spring 2020	2,831	Implementation of comprehensive literacy instruction, professional development in literacy, teacher certification
Evidence review	What Works Clearinghouse database search; Evidence for ESSA clearinghouse search; scan of academic journal databases	2020, 2021	N/A	Level of evidence supporting programs that district leaders reported purchasing with SRCL funds or teachers reported using in SRCL schools
Common Core of Data	All public schools in SRCL states	Spring 2019	30,319	Urban-centric locale code, grade levels served, percentage of students per school by race and eligibility for free or reduced-price lunch

Data Source	Sample	Timing of Data Collection	*N*	Information Obtained
Stanford Education Data Archive	All public schools in SRCL states with standardized test scores	Spring 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016	14,514	School-level average student achievement in reading/English language arts, percentage of students per school by English learner status and special education status
Document review	All 11 SEAs	Fall 2018	11	State literacy plans, state SRCL grant applications, states' request for SRCL subgrant applications, authorized use of SRCL funds, plans for targeting funds to disadvantaged students, priorities for use of evidence
State student administrative data	All students in grades 3–8 in 9 SEAs (all except Kansas and North Dakota)	2014–15 through 2018–19	17,677,587	Standardized test scores, race and ethnicity, eligibility for free or reduced-price lunch, English learner status, student with disability status, school enrollment
State aggregated test scores	All schools serving grades 3–8 in Kansas and North Dakota	2014–15 through 2018–19	30,311	Standardized test scores (overall and by demographic group)

Note: N/A is not applicable; SEA is state education agency.

B.1.2. State Director Interviews

The evaluation drew on qualitative data from three rounds of annual telephone interviews (fall 2018, 2019, and 2020) with administrators for state education agencies (SEAs) and reviews of key documents related to the state operation of the program. These sources together provided information on state grantees in four focal priority areas that reflected key expectations of the state grantees set out by the U.S. Department of Education program office. These areas included (a) how the state approached the grant priority for interventions and practices with strong or moderate evidence of effectiveness and how they evaluated the evidence base for the interventions proposed by subgrantees; (b) other aspects of how states approached the subgrant application process, including how states targeted districts with disadvantaged students; (c) provision of technical assistance to subgrantees on developing their grant applications and, post award, on implementing their planned literacy interventions in district schools; and (d) implementation by the state of a continuous improvement process.

Each round of interviews followed a set of structured, closed-ended questions with open-ended follow-up questions that provided state directors the opportunity for more expansive answers to the questions. The interviews were conducted with the state project directors and other state leadership identified with assistance from the state directors. These state staff were then interviewed as a team by two members of the evaluation team. Interviews were audio-recorded and each transcript was

systematically coded by at least two analysts. Coded data were organized in an Excel-based template. The analysis focused on identifying distinguishing features of each state's approach to each of the four focal priority areas (listed above) and determining metrics for reporting on state-to-state variation. Exhibit B.2 lists the focus questions about state grantees, the data sources used to answer the questions, and the data collection schedule.

Exhibit B.2. Timeline of state director interviews

Interview Year	Interview Timeframe	SRCL Grant Years Covered	Focus Areas
1	Fall 2018	2017–18	1. How did SEAs approach making subgrant awards? 2. How did SEAs prioritize subgrant applications, including the alignment of the subgrantees' plans with demonstrated local needs, evidence of effectiveness, and targeting districts with disadvantaged students? 3. How did SEAs review subgrant applications? 4. What were the amounts and types of technical assistance that each SEA provided to subgrantees prior to and after grant awards?
2	Fall 2019	2017–18, 2018–19	5. What topics did SEAs cover in technical assistance or support they provided to subgrantees? In what formats were the technical assistance or support provided? 6. How do SEAs use continuous improvement and evaluate the effectiveness of subgrantees' projects? What data do they collect?
3	Fall 2020	2017–18, 2018–19, 2019–20	7. How do SEA grantees' approaches to SRCL fit into the SEAs' overall approaches to literacy? 8. How did grantees address the program priority for interventions and practices supported by strong or moderate evidence?

Note: SEA is state education agency.

The primary purpose of the state director interviews was to understand the ways in which states evaluated and allocated subgrant awards and supported subgrantees' implementation of interventions. The study team used the states' interviews in concert with reviews of state grant applications and state RFAs for subgrantees. The grantee interview protocols were designed to enable the study team to verify or clarify elements of the applications, explore the rationale for grantee approaches, and collect additional information about subgrantee award procedures and evidence supporting selected interventions. Additionally, the study team used the documents and states' prior interviews to tailor interview questions and identify areas to clarify in subsequent interview years.

During the first round of interviews in 2018, the interviews focused on the application review process for awarding grants and whether the review process aligned with the SRCL program expectations for

states in prioritizing funding based on the characteristics of the literacy programs proposed by applicants. These include characteristics such as proposed use of literacy interventions or practices with evidence of effectiveness; literacy plans aligned with local needs assessments and the state literacy plan; and targeting funds to a high proportion of disadvantaged children. In the first year, the interview also asked about grantee plans to provide technical assistance to subgrantees and plans to collect and use data to monitor and support continuous improvement of subgrantees.

During the second year, 2019, the state grantee interviews focused on technical assistance or support provided to SRCL subgrantees and program improvement or feedback processes. The interviews asked state grantees about technical assistance or support provided directly to individual SRCL subgrantees or groups of subgrantees, or resources to which SRCL subgrantees were directed that focused on the topics that states covered in the technical assistance. The interviews also asked about any program improvement or feedback processes that states used—particularly whether states established feedback loops with subgrantees or collected data for continuous improvement purposes from subgrantees, and if so, how they approached using the data for continuous improvement.

In the third and final round of interviews, in 2020, the interviews focused on each state's approach to supporting evidence-based programs and practices through SRCL, including how they addressed the program's absolute priority of using interventions and practices supported by strong or moderate evidence. Additionally, the interviews asked about state grantees' overall approaches to improving literacy outcomes for disadvantaged students and how their state approaches affected each state's approach to SRCL.

State Interview Participants

State grantees participated in interviews via telephone or video conference during each of the 3 grant years. Grant directors from each of the 11 SRCL state grantees were asked to participate in the interviews and to invite additional members of their SRCL staff to the interviews, which many grantees did. During the first round of interviews (2018), only 9 of the 11 state grantees were interviewed, because federal approval to collect data from more than nine respondents had not yet been obtained.[1] In the subsequent 2 years of interviews (2019 and 2020), all 11 grantees participated in the interviews. For the two state grantees that were not interviewed in 2018, the study team tailored the year 2 interview protocols to include questions from both the year 1 and year 2 interview protocols.

[1] Data collection activities conducted under a federal contract must be approved by the Office of Management and Budget (OMB) if the number of respondents is greater than nine. The study team wanted to collect data on the early phases of the SRCL implementation process prior to OMB approval and hence could only do so with nine states.

State Interview Procedures

During an introduction to the evaluation given as a webinar at the start of the evaluation to state grantees, the study team informed states that they would be asked to participate in telephone interviews on an annual basis during the grant period. During each year in which interviews were conducted, the study team contacted state grant directors via e-mail to ask them to participate in interviews. In the e-mail communications, the study team invited the state grantees to include additional members of the state SRCL team.

In the e-mail communications with SEA grantees, the study team provided a high-level overview of topics that would be covered in each interview. Additionally, the study team shared a written informed consent document with the grant directors prior to participation. At the beginning of each interview, the study team provided information to participants letting them know that responses to the data collection would be used only for descriptive research purposes and that, as part of the study, data on state policies and supports may be reported by state. Thus, while personally identifiable information about individual respondents would not be released as part of the study, data displayed by state could be attributed to the state agency or possibly an individual respondent.

Each interview was approximately 60 minutes in length, except the interviews conducted in year 2 (2019) with the two grantees who did not participate in year 1 interviews. The year 2 interviews for these two grantees were approximately 90 minutes in length, and covered content from both the year 1 and year 2 interview protocols. Interviews were recorded with the permission of SEA participants for use in creating detailed transcriptions for coding and analysis.

B.1.3. Surveys

As described in Exhibit B.1, the study administered surveys to districts, school leaders, and teachers in spring and summer 2019 (wave 1) and to teachers in spring 2020 (wave 2). The surveys were administered by NORC, in collaboration with the American Institutes for Research (AIR).

Questionnaire Development

To ensure the quality of the data collection instruments, the study team conducted an initial set of informational interviews, pilot tested the draft instruments, and convened a Technical Working Group (TWG) to review the instrument priorities. As the first step in this instrument development process, the study team engaged in a series of informational interviews with nine current and former SRCL grantees. These informational interviews helped refine the organization and content of the survey questions and ensure the items would result in high-quality data to address the study's research questions. Whenever possible, the study team sought to use items from existing, validated, national surveys. Items related to comprehensive literacy instruction on the teacher survey drew from the Early Childhood Longitudinal Study teacher survey, grades 1, 2, and 3 (ECLS); the National Assessment of Education Progress, 8th-grade teacher survey (NAEP); the Reading First Impact Study teacher survey (Gamse et al., 2008); and the Common Core in the Schools: A First Look at Reading Assignments teacher survey (Shanahan & Duffett, 2013).

The second phase of the instrument development process involved piloting the instruments. The study team conducted piloting interviews with a small set of district officials, principals, and teachers to pilot the survey as a whole, and to solicit feedback on specific items. The study team asked participants to complete the survey prior to the interview, keeping track of the time required to finish the survey, and documenting any challenges. When interviewing the respondent about the survey, the study team followed a structured protocol to learn about respondent burden time, and survey items or instructions that were difficult to understand, poorly worded, or had other problems. The study team also asked for suggestions – for example, with regard to the time frame that would provide for optimal respondent recall. Pilot participants were also asked to send the completed survey so the study team could gauge variation in responses. Immediately after the interview, the interviewers recorded comments and responses in an Excel-based data capture form, using the audio recording, as necessary. Final decisions regarding item revisions were made after the interviews were completed.

Sampling

Beginning in fall 2018, the study team requested lists of all SRCL-funded districts, schools, and early childhood centers in each state to establish the sampling frame for the surveys. After a review of these lists, follow-up questions were sent to confirm school type, addresses, or National Center for Education Statistics (NCES) identification information. After the list for each state was considered final by the study team, it was provided to the team statistician for school sampling.

District (subgrantee) sampling. SRCL provided grants to 11 states: Georgia, Louisiana, Maryland, Ohio, North Dakota, Kansas, Kentucky, Montana, Oklahoma, Minnesota, and New Mexico. SRCL states primarily awarded subgrants to school districts or consortia of districts, although a small number of subgrants were awarded to nonprofit early childhood centers. Because only eight subgrantees were early childhood centers, and because the nature of their SRCL activities was substantively different from districts serving students in primary and secondary school, the study restricted the sample of subgrantees to districts in which students were enrolled. The lists provided by the 11 grantee states included 501 districts, of which 435 were determined to have received SRCL funds. (Sixty-six of the 501 entities listed did not actually receive SRCL funds or were duplicates and were determined to be ineligible for study data collection.) Some SRCL subgrantees were consortia of multiple districts. The study surveyed all eligible SRCL districts whether they participated in consortia or were stand-alone subgrantees. Hence, when describing the sample and respondents, the text refers to districts rather than subgrantees. Districts were asked to designate the person most knowledgeable about SRCL grant funding to complete the survey.

Exhibit B.3 summarizes the district sampling and survey response rates by state.

Exhibit B.3. District sample and response rate by state

State	Total Districts Listed by SRCL States	Eligible Districts	Total Responses	Response Rate
Georgia	38	38	32	84%
Kansas	32	32	28	88%
Kentucky	41	40	38	95%
Louisiana	122	84	71	85%
Maryland	24	23	22	96%
Minnesota	23	23	23	100%
Montana	32	28	24	86%
New Mexico	15	13	11	85%
North Dakota	13	15	15	100%
Ohio	136	114	100	88%
Oklahoma	25	25	25	100%
Total	501	435	389	89%

School sampling. The study also sampled schools and early childhood centers funded by SRCL subgrantees. For simplicity, this is generally referred to as the school sample. The school sample was drawn from the full population of schools that received funds from SRCL districts.

As shown in Exhibit A.2, the number of schools and early childhood centers funded by SRCL varied substantially across the 11 SRCL states. For example, the number of supported schools ranged from 57 in Minnesota to 1,033 in Maryland. Thus, a sample drawn in proportion to the number of funded schools in each state would give much more emphasis to states like Maryland, which chose to spread funds across many schools, than to states like Minnesota. To avoid this, the study sampled the same number of schools in each state, and results based on the survey weight each state equally. Hence, results describe the average across SRCL states, giving equal weight to each state.

The study drew an initial sample of 506 schools, 46 from each state. The sample size was chosen primarily to achieve the desired precision for the teacher survey (see below). In addition, it was designed to achieve a margin of error of plus or minus 5 percentage points for results for a binary variable based on the school leader survey, combining the 11 SRCL states. For each state separately, the school leader survey was expected to achieve a margin of error of plus or minus 15 percentage points.

The school sample was restricted to schools that received SRCL funding with nonzero enrollment. The school sample drawn for wave 1 of the study was retained for wave 2. The schools in each state were stratified by school level, producing five mutually exclusive strata: early childhood centers, elementary schools without early childhood centers (ESO), elementary schools with early childhood centers (ESW), middle schools (MS), and high schools (HS). Schools were sorted within strata to ensure balance by district and urbanicity. Exhibit B.6 gives a breakdown of the total sample by urbanicity and school stratum.

In addition to drawing an initial sample of 506 schools, NORC selected a reserve sample of schools which could be used if all efforts to obtain a response from schools in the initial sample failed, to ensure targets could be met (Exhibit B.4).

Exhibit B.4. Initial and reserve school sample, by stratum, for an example state

Stratum	Initial Sample	Reserve Sample	Total Sample
Early childhood centers	14	3	17
Elementary schools without early childhood programs	9	2	11
Elementary schools with early childhood programs	5	1	6
Middle schools	9	2	11
High schools	9	2	11
Total	46	10	56

A review of each state's school lists indicated that in some instances, stratum sample size targets could not be met because funding for that state did not include schools in certain strata (e.g., early childhood centers), or the number funded in a stratum was lower than the target for the stratum. In those instances, additional schools were drawn from other strata in order to achieve the initial 46 schools within the state. Exhibit B.5 summarizes the initial and realized sample size by stratum.

Exhibit B.5. Initial and realized school sample by stratum

Stratum	Initial Sample	Realized Sample	Percent of Realized Sample from Reserve Sample
Early childhood centers	154	111	25.2%
Elementary schools without early childhood programs	99	95	6.3%
Elementary schools with early childhood programs	55	102	9.8%
Middle schools	99	99	7.1%
High schools	99	99	8.1%
Total	506	506	11.7%

Exhibit B.5a. shows the number of SRCL districts in each state that had schools in the initial and realized school sample. The resulting sample of schools were located in 252 of the 435 funded districts; no schools were located in the other 183 districts.

Exhibit B.5a. Districts with sampled schools

	Districts with Schools in the Initial Sample	Districts with Schools in the Realized Sample
Georgia	29	29
Kansas	21	21
Kentucky	27	26
Louisiana	33	30
Maryland	20	18
Minnesota	21	21
Montana	29	25
New Mexico	13	13
North Dakota	14	14
Ohio	34	32
Oklahoma	23	23
Total	264	252

In wave 1, NORC contacted a sample of 599 schools (506 in the initial sample and 93 in the reserve sample), and 506 agreed to participate. Overall, 11.7% of the realized sample was obtained from the reserve sample. Exhibit B.6 describes the realized sample by urbanicity and stratum.

Exhibit B.6. Realized school sample by urbanicity and stratum

Stratum	Data on Urbanicity Not Available	City	Rural	Suburban	Town	Total
Early childhood centers	99	†	4	†	4	111
Elementary schools without early childhood programs	1	18	19	18	39	95
Elementary schools with early childhood programs	0	14	47	18	23	102
Middle schools	0	†	46	†	23	99
High schools	1	15	34	26	23	99
Total	101	57	150	86	112	506

Note: Data on urbanicity were obtained from the Common Core of Data (CCD). Urbanicity data are missing for the 99 stand-alone early childhood centers that do not appear in the CCD. A dagger (†) corresponds to a cell in which there are too few cases to report.

School leaders in each sampled school were asked to complete a survey or designate someone who was most knowledgeable about SRCL grant funds to do so. From the 506 schools, 111 program directors of early childhood centers and 395 principals of schools serving at least one grade between kindergarten and 12th grade were included in the sample. Exhibit B.7 summarizes the school leader sample and the

number of completed surveys by school stratum. The response rate for the total sample was 89.4%, computed as the number of responding school leaders (449) divided by the number of eligible school leaders in the sample (506 school leaders minus 4 ineligible school leaders, for a total of 502).

Exhibit B.7. Wave 1 sample and response rate for the school leader survey, by stratum

Stratum	School Leaders	Responding School Leaders	Ineligible School Leaders[a]	Eligible School Leader Response Rate
ECC	111	95	4	88.8%
ESO	95	89	0	93.7%
ESW	102	89	0	87.3%
MS	99	90	0	90.9%
HS	99	86	0	86.9%
Total	506	449	4	89.4%

[a] School leaders were classified as ineligible if their center closed or merged with another center.
Note: ECC is early childhood center; ESO is elementary schools without early childhood programs; ESW is elementary schools with early childhood programs; MS is middle schools; HS is high schools.

In wave 2, NORC returned to the 506 wave 1 schools. Of the 506 in the wave 1 sample, 496 remained eligible, and 10 were deemed ineligible (including the four early childhood centers referenced in Exhibit B.7 that were determined to have closed or merged at the time of the wave 1 school leader survey). Reasons for ineligibility included closure or that they were no longer receiving funds.

Teacher sampling. After selecting and recruiting the school sample, NORC requested teacher rosters from the 506 schools that agreed to participate, which created the sampling frame for teachers. The teacher sample included

- all full-time regular classroom teachers who taught reading or English language arts (ELA) as part of their teaching load; and
- teachers who provided extra support to students in reading or writing, such as reading specialists or interventionists.

No part-time, social studies, math, or science teachers were included unless they also provided literacy instruction. No special subject teachers (such as art, music, or physical education), paraprofessionals, teachers' aides, speech therapists, social workers, or any non-classroom faculty (other than those who provide extra support to students in reading or writing) were included. Each teacher in the sampling frame was classified into one of five strata based on grade taught:[2]

[2] Because teachers were sampled within sampled schools, the teacher strata were selected to align with common school grade-level structures as far as possible (pre-K, kindergarten–grade 5, grades 6–8, and grades 9–12). Results on comprehensive literacy instruction based on the teacher survey were reported for four grade bands—pre-K, kindergarten–grade 3, grades 4–8, and grades 9–12—reflecting grade-level differences in the focus of literacy instruction.

- prekindergarten (pre-K);
- lower elementary grades (kindergarten-grade 2, also referred to as K-2);
- upper elementary grades (grades 3-5);
- middle school (grades 6-8); or
- high school (grades 9-12).

Sampling of teachers at each school was conducted on a rolling basis, as schools agreed to participate and provided teacher rosters.

The teacher survey was designed to achieve a margin of error of plus or minus 5 percentage points for results by grade band (pre-K, K=3, 4-8, and 9-12), combining the 11 states.[3] Analyses by grade band are the primary focus of the teacher analysis. For each state separately, the sample was expected to produce a margin of error of plus or minus 7.5 percentage points, combining teachers across grade bands. To achieve the desired precision, NORC targeted the selection of teachers as follows for each school:

- pre-K ($N = 3$);
- lower elementary grades (K-2; $N = 4$);
- upper elementary grades (grades 3-5, $N = 4$);
- middle school (grades 6-8, $N = 8$); and
- high school (grades 9-12, $N = 8$).

Based on the total intended sample of 506 schools and the expected number of teachers per school by grade and school level, the sampling strategy was expected to yield a sample of 3,430 teachers. Assuming an 85 percent response rate for teachers, this would achieve 2,918 total completed teacher surveys (Exhibit B.8).

Exhibit B.8. Targeted number of schools and teachers by grade span stratum

Grade Span Stratum	Expected Number Of Schools With Teachers in Grade Span	Targeted Number of Teachers Per School	Targeted Number of Teachers	Expected Number of Completed Teacher Surveys
Pre-K	202	3	606	516
Grades K-2	151	4	604	514
Grades 3-5	151	4	604	514
Grades 6-8	101	8	808	687
Grades 9-12	101	8	808	687
Total	506		3,430	2,918

[3] The required sample size was based on an assumed intraclass correlation of .10 and an assumed variance inflation factor due to weighting of 1.1.

In the event that a teacher was deemed ineligible after selection (no longer at the school, on leave from the school and would not return prior to the end of the school year, or did not teach the grade initially reported), a new teacher was selected to replace the original teacher when possible.

Of the 506 schools sampled, 503 provided teacher rosters in wave 1, listing 6,147 teachers. Of these, 3,095 teachers were sampled. Exhibit B.9 summarizes the teacher sample by school stratum.[4]

Exhibit B.9. Wave 1 teacher sample and response rate by stratum

Stratum	Sampled Teachers	Eligible Teachers	Responding Teachers	Ineligible Teachers	Eligible Teacher Response Rate
ECC	280	263	209	17	79.5%
ESO	649	613	530	36	86.5%
ESW	848	761	661	87	86.9%
MS	702	626	522	76	83.4%
HS	616	568	482	48	84.9%
Total	3,095	2,831	2,404	264	84.9%

Note: ECC is early childhood center; ESO is elementary schools without early childhood programs; ESW is elementary schools with early childhood programs; MS is middle schools; HS is high schools.

For wave 2, 3,097 teachers were sampled from a total of 6,256 on the rosters provided by the schools. Exhibit B.10 summarizes the wave 2 teacher sample by school type.

Exhibit B.10. Wave 2 teacher sample and response rate by stratum

Stratum	Sampled Teachers	Eligible Teachers	Responding Teachers	Ineligible Teachers	Eligible Teacher Response Rate
ECC	266	241	228	25	94.6%
ESO	693	670	668	23	99.7%
ESW	834	784	771	50	98.3%
MS	720	652	633	68	97.1%
HS	584	546	531	38	97.3%
Total	3,097	2,893	2,831	204	97.9%

Note: ECC is early childhood center; ESO is elementary schools without early childhood programs; ESW is elementary schools with early childhood programs; MS is middle schools; HS is high schools.

[4] To encourage the respondents to complete the web survey, an in-kind gift of a $25 Amazon gift code was offered to teachers. The gift codes were given to any teachers who completed the entire survey. To facilitate communication with each sampled school, a school coordinator was identified as the primary point of contact for NORC. These school coordinators were provided an incentive for their assistance ($50). The school coordinator for each of the 503 schools that provided rosters was sent the incentive. In a few instances, one school coordinator provided rosters for multiple schools; that coordinator received the incentive for each of those schools.

Exhibit B.11 presents the number of respondents to each survey by state. Note that some analyses based on survey data have fewer observations per state than the number that appears in Exhibit B.11, depending on whether the relevant items were completed by all survey respondents. For instance, districts that did not purchase programs are not included in most analyses related to evidence use, and teachers who did not assign novels are not included in analyses of the age-appropriateness of assigned novels.

Exhibit B.11. Survey sample sizes by state

State	District Survey (2018–19)	School Leader Survey (2018–19)	Teacher Survey (2018–19)	Teacher Survey (2019–20)
Georgia	32	29	237	268
Kansas	28	27	204	213
Kentucky	38	36	299	307
Louisiana	71	27	194	231
Maryland	22	31	221	265
Minnesota	23	34	233	276
Montana	24	42	226	255
New Mexico	11	30	181	233
North Dakota	15	40	251	301
Ohio	100	27	190	251
Oklahoma	25	31	168	231
Total	389	354	2,404	2,831

Note: "Teacher Survey" combines the pre-K teacher survey and the K-12 teacher survey, which asked slightly different questions. The sample sizes for specific outcomes derived from survey data may differ slightly from these sample sizes because not all questions were relevant to all respondents.

Sampling Weights

NORC developed weights so that responses to the district, school leader, and teacher survey items in wave 1 and the teacher survey in wave 2 could be adjusted to reasonably reflect the responses for the full set of educators involved in SRCL. Prior to beginning weight construction, each sample member was assigned one of the following statuses:

- ineligible;
- eligible nonrespondent; or
- eligible respondent

District (subgrantee) survey weights. NORC computed the sampling weight for responding district *k* in states as:

$$W_{1sk} = \frac{D_{s1}}{d_{s1}}$$

where D_{s1} is the count of eligible districts in state *s*, and d_{s1} is the count of eligible districts in states that completed the district survey. The weight was set to missing for all ineligible districts, and to zero for all nonresponding districts.

School weights. The school sample was selected within the 11 grantee states and five nonoverlapping school strata defined by school level (ECC, ESW, ESO, MS, and HS). Overall, there were 55, 11 × 5, explicit sampling strata for schools. Within each stratum, schools were selected into the sample using equal probability systematic sampling from the population list of schools sorted by urbanicity and district. N^h was the frame count of schools in stratum h.

As previously noted, NORC drew an initial sample of 506 schools, with a reserve sample of 93 schools. Ineligible schools and nonresponding eligible schools were replaced by draws from the reserve sample. If replacements were found to be ineligible or failed to respond, they too were replaced. NORC then computed the base weight for school *i* in stratum *h* defined by:

$$S_{hi} = \frac{N_h}{n'_h}$$

Here, N_h is the total number of SRCL-funded schools in stratum *h*, and n'_h is the final released sample size of eligible schools, including all of the eligible schools encountered throughout the replacement process. The base weight is an approximation of the reciprocal of the probability of selection of the school. The base weights were calibrated to population counts of schools by urbanicity.

NORC then computed a nonresponse adjusted school weight. Two versions were created: one for the school leader survey weights, and one for the teacher weights. (For the former, schools were considered respondents if they agreed to participate; for the latter, schools were considered respondents if they agreed to participate and supplied a teacher roster.) For the teacher survey, NORC then computed the nonresponse adjusted weight for school *i* in stratum *h* defined by:

$$W_{2hi} = \frac{\sum_{j \in I_h} S_{hj}}{\sum_{j \in C_h} S_{hj}} S_{hi}$$

where I_h is the set of eligible schools in the sample in stratum *h*, and C_h is the set of eligible schools in the sample in stratum *h* that agreed to participate and supplied the teacher roster. The weight W_{2hi} was set to missing for ineligible schools, and to 0 for eligible schools that declined to participate or failed to supply the roster. For the school leader survey, NORC computed the nonresponse adjusted weight W_{hi} analogously, where I_h is the set of eligible schools in the sample in stratum *h*, and C_h is the set of eligible schools in the sample in stratum *h* that agreed to participate.

School leader survey weights. The school leader weight was computed as:

$$P_{hi} = W_{hi} \frac{\sum_{j \in C_h} W_{hj}}{\sum_{j \in C_h} W_{hj} r_{hj}}$$

where W_{hi} is as defined above, and r_{hi} is the response indicator (= 1 for respondent, = 0 for nonrespondent) for the leader of the i^{th} school in stratum h. The school leader weight was set to 0 for all school leaders who failed to complete the survey.

Teacher survey weights. The teacher sampling frame was stratified as follows:

- pre-K;
- lower elementary grades (K–2);
- upper elementary grades (grades 3–5);
- middle school grades (grades 6–8); and
- high school grades (grades 9–12).

Recognizing that some teachers teach in more than one grade, each teacher was classified into the first stratum in which they qualified. For example, if a teacher taught in both grades 5 and 6, then the teacher was classified into the upper elementary (grades 3–5) stratum. Thus, each teacher in the sampling frame was classified into one and only one stratum.

When sampling, NORC targeted the selection of three pre-K teachers, four lower elementary teachers (K–2), four upper elementary teachers (grades 3–5), eight middle school teachers (grades 6–8), and eight high school teachers (grades 9–12) per school. NORC sampled teachers within each nonempty teacher stratum within a school. Thus, the number of teachers selected varied from school to school, depending on the grades actually present at the school. (The actual sample size of teachers within a stratum was capped at the population size of teachers within the stratum.)

Within eligible and cooperating school *hi* and teacher stratum *g*, M_{hig} is the number of teachers in the population in the sampling frame for teachers in grade band stratum g, and is the number of teachers selected into the sample. NORC computed the base weight for teacher *j* in teacher stratum *g* in school *hi* as:

$$T_{higj} = W_{2hi}\frac{M_{hig}}{m_{hig}}$$

Nonresponse adjustment cells were defined by school stratum *h* and teacher stratum *g*. Prior to calculating the final weights, NORC tabulated the sample sizes and the number of teacher completes by cell, collapsing cells as necessary. Then, NORC computed the nonresponse adjusted teacher weight defined by:

$$W_{3higj} = r_{higj}T_{higj}\frac{\sum_{(hi'gj')\in A_{hg}} T_{hi'gj'}}{\sum_{(hi'gj')\in A_{hg}} r_{hi'gj'}T_{hi'gj'}}$$

where A_{hg} is the set of teachers who were selected into the sample in the specific teacher's school stratum *h* and teacher stratum *g*, and who were classified as eligible; and r_{higj} is a response indicator (= 1 for respondents, = 0 for nonrespondent or ineligible teachers) for the $higj^{th}$ teacher.

The final teacher weights were set to missing for all teachers in the selected sample who were classified as ineligible, and to 0 for teachers who failed to respond.

B.1.4. Evidence Review

The study team conducted a review of the evidence supporting the programs purchased by SRCL districts, based on the What Works Clearinghouse (WWC) and the Evidence for ESSA database. Section B.2 presents information on the procedures followed in the evidence review.

B.1.5. Demographic Data for Targeting Analyses

Information on school demographic characteristics was taken from the 2018-19 Common Core of Data (CCD), collected by the National Center for Education Statistics, Institute of Education Sciences, at the U.S. Department of Education. This year of data represents the first year in which schools received SRCL funds. For the analyses reported in Exhibit 1 in the report and Exhibits C.1, C.2, C.3, C.5, and C.6, the CCD provided data used to classify schools as disadvantaged based on the percentage of students eligible for free or reduced-price lunch.

B.1.6. Prior Achievement Data for Targeting Analyses

The study drew on data from the Stanford Education Data Archive (SEDA) for the analyses reported in Exhibit 1 in the report and Exhibits C.1-C.4a, C.5, and C.6. The SEDA provides average test scores at the school level on a common scale across states. This study used version 3.0 of the SEDA, containing data through 2015-16, standardized at the grade-cohort level, with scores pooled across grades, years, and subjects from 2008-09 through 2015-16.[5] Rather than reporting average scores separately for math and ELA, the SEDA reports the average of the two subjects and the difference between the two; the study team approximated the pooled average ELA score by taking the average score across math and ELA and subtracting half of the average gap between math and ELA scores. This resulting variable, unique at the school level, is used as an indicator of disadvantage in the analysis of how funds are directed toward disadvantaged students (Exhibit 1 in the report, and Exhibits C.1, C.2, C.4a, C.5, and C.6). Additionally, the SEDA reported schoolwide percentages of students who were English learners and students with disabilities as of 2015-16, which were used in classifying schools as disadvantaged in Exhibit 1 in the report and Exhibits C.1, C.2, C.4a, C.5, and C.6).

B.2. Methods Used to Determine the Level of Evidence Supporting Literacy Programs Purchased by SRCL Districts and Used by SRCL Teachers

Supporting the use of evidence-based interventions is an important component of the Every Student Succeeds Act (ESSA) overall, not just for SRCL, and the statute provides a definition of tiers of evidence in section 8101(21)(A), as follows:

"… the term 'evidence-based' when used with respect to a State, local educational agency, or school activity, means an activity, strategy, or intervention that -

[5] For more technical details about the collection and standardization of the SEDA data, consult the SEDA 3.0 technical manual: https://stacks.stanford.edu/file/druid:mk782vn8293/SEDA_documentation_v30_09212019.pdf.

(i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on –

(I) *strong evidence* from at least one well-designed and well-implemented experimental study;

(II) *moderate evidence* from at least one well-designed and well-implemented quasi-experimental study;

(III) *promising evidence* from at least one well-designed and well-implemented correlational study with statistical controls for selection bias; or

(ii) (I) *demonstrates a rationale* based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes; and

(II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention."

These statutory definitions apply to all programs authorized by the Elementary and Secondary Education Act (ESEA), as reauthorized by ESSA. The *Education Department General Administrative Regulations* (EDGAR) include evidence definitions aligned with ESSA, but specify additional details to help the Department assess evidence for its discretionary grant programs.[6] The EDGAR definitions are outlined in Box B.1. Two information clearinghouses—the WWC and Evidence for ESSA—have applied the statutory and regulatory standards to studies of educational interventions and informed analyses in this report. The WWC's evidence tier designations are consistent with EDGAR. As the Evidence for ESSA ratings do not include the large, multisite sample criterion used in the EDGAR definitions of strong and moderate evidence, these ratings are not consistent with EDGAR but are consistent with the statutory definitions. Both are described in the section that follows.

[6] See *Education Department General Administrative Regulations*, 34 C.F.R. § 77.1.

Box B.1. EDGAR Regulatory Definitions of Strong and Moderate Evidence

Tier 1 (strong evidence): At least one study	Tier 2 (moderate evidence): At least one study
• Uses an experimental (randomized controlled trial) design and is carried out in ways that meet WWC standards without reservations • Has at least one statistically significant favorable effect (no overriding negative effects) • Is based on a study sample with more than one site and at least 350 individuals • Is based on a study sample that overlaps with the characteristics of the population **and** setting of the entity proposing to implement the program	• Uses a quasi-experimental design and is carried out in ways that meet WWC standards with reservations • Has at least one statistically significant favorable effect (no overriding negative effects) • Is based on a study sample that includes more than one site and at least 350 individuals • Is based on a study sample that overlaps with the characteristics of the population **or** setting of the entity proposing to implement the program

The district survey asked respondents to identify the literacy and reading interventions that the districts adopted and paid for through SRCL funds, with separate items for programs used for all students in a classroom (core instructional program) and for programs used specifically with struggling readers (supplementary programs). The 82 percent of districts that purchased at least one program were then asked to list the grade levels of students who use each program. The teacher and school leader surveys also asked for the names of literacy programs used in SRCL classrooms, although the programs listed by respondents may not have been purchased with SRCL funds.

The study team compiled a list of all literacy programs that district, principal, and teacher respondents entered in their survey responses. Some entries were excluded because they were determined not to be literacy programs within the scope of this analysis–for example, screening assessments or after-school activities.

For each core and supplemental program listed by districts, the evidence review team reviewed the WWC's website, intervention report summaries, and the Data from Individual Studies database to determine whether the program had ever been reviewed by the WWC. If programs had been reviewed, the study team recorded the standards used in the review, the date the review was performed, the protocol that guided the review, and the rating provided by the WWC for the quality and amount of evidence supportive of the program.

The evidence review process started with the WWC's Data from Individual Studies database, retrieved from the WWC website. This database is organized by program, providing the names of studies on each

program, and the findings for each outcome, sample, and subsample. [7] The database lists only those studies and findings that meet WWC standards.[8] For programs that are listed in the database and have findings flagged as having ESSA tier 1 or ESSA tier 2 evidence, the review recorded the relevant grade level for that finding, the standards and protocol used in the review, and the ESSA rating.

The next step in the review process was a search in the WWC website's "intervention report" search field for each program that was identified by subgrantees. An intervention report contains a summary of all findings that met WWC standards at the time the report was completed. From each intervention report that described studies of the program that met WWC standards, the evidence review team recorded the date of the report's publication, the version of WWC standards used, the number of studies that met WWC standards, the grade levels of the students in the studies, the reading-related outcome, and the WWC rating and effect size for the outcome.

The review process drew on the intervention reports because they provide a synthesis of all findings that meet standards on an intervention, and thus provide a more comprehensive view than findings from individual studies. In particular, programs can meet the "350 students" and "multiple settings" criteria for ESSA tier 1 or tier 2 evidence, drawing on multiple studies that meet WWC standards in an intervention report, even if the standards are not met based on an individual study.

These sources informed a summary of evidence on the impact of the program on reading-related outcomes according to the WWC. Programs were classified as one of the following:

1. Never reviewed by the WWC;
2. reviewed by the WWC, but no studies were found that met WWC standards;
3. reviewed by the WWC, and evidence suggests a negative impact on students' reading-related outcomes;
4. reviewed by the WWC, and evidence suggests no impacts or mixed impacts on students' reading-related outcomes; or
5. reviewed by the WWC, and evidence suggests a positive impact on students' reading-related outcomes.

Literacy programs that were not reviewed by the WWC were the subject of two additional queries. First, the study team reviewed the Evidence for ESSA clearinghouse, which conducts systematic evidence

[7] The WWC refers to programs as one kind of intervention. Other types of interventions include practices and policies, which were not relevant to this evidence review.

[8] The database includes only 31 findings that do not meet WWC standards and eight studies that do not meet standards. There are, in fact, thousands of studies that have been reviewed by the WWC that do not meet WWC standards, but only a small subgroup of those are included in the Data from Individual Studies database. We do not know why only a portion of studies or findings that do not meet WWC standards are listed.

reviews that are similar but not identical to the WWC's reviews.[9] As noted above, Evidence for ESSA ratings do not include the large, multisite sample criterion in the regulatory definitions of strong and moderate evidence used by the WWC. The study team identified programs that had not been reviewed by the WWC but which met the highest evidence standards associated with Evidence for ESSA.

For all remaining programs (that is, those that were not the subject of a review by the WWC or Evidence for ESSA), the study team sought to identify any published studies on the program, regardless of the level of rigor. The following resources were searched for each of the literacy programs: EBSCO, ERIC, Academic Search Premier, ProQuest, Education Collection, and PsycInfo. If no publications were identified through these resources, the study team e-mailed the program developer and reviewed the program website for mentions of research. Publications that did not present research (such as blog posts, opinion pieces, promotional materials) were excluded. As a result of this review, the study can distinguish between literacy programs for which published studies exist (with no judgment regarding the level of rigor) and programs for which no published research could be identified.

The categories of evidence used to classify literacy programs purchased with SRCL funds are listed below. The first two are combined in the body of the report, while the appendix exhibits retain this distinction.

1. *Strong or moderate review rating in WWC*. Programs had research that met the regulatory strong or moderate definition, as confirmed by an official WWC review.
2. *Strong or moderate review rating in Evidence for ESSA*. Programs had research not yet reviewed by the WWC, or that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence (which Evidence for ESSA did not include in their ratings).
3. *Promising evidence review rating*. Programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence.[10]
4. *Null or negative findings review rating.* Programs had research that met WWC standards but did not find significant positive effects.

[9] From the Evidence for ESSA website: "In most respects, Evidence for ESSA follows the *WWC Standards and Procedures Guide, Version 3.0*, and therefore usually comes to similar or identical conclusions. However, there are differences in our standards, agreed to by the Evidence for ESSA Technical Working Group (TWG). For example, Evidence for ESSA does not accept measures made by developers or researchers, it does not accept after-the-fact (post-hoc) designs, and it requires a minimum study duration of twelve weeks. A link to the Evidence for ESSA Standards and Procedures can be found in the first FAQ" (Center for Research and Reform in Education).

[10] Evidence for ESSA includes a category called *promising*. The study team classified programs that Evidence for ESSA places in this category as *promising*, except when they have moderate or strong evidence in WWC.

5. *Existing studies reviewed but do not meet standards of rigor*. Programs had research reviewed by the WWC or Evidence for ESSA but either did not meet standards or were based on an ineligible study design.[11]
6. *No review of existing studies*. Programs have been evaluated in published research that has not been reviewed by the WWC or Evidence For ESSA.
7. *No published research identified*. Programs for which no published research could be found.

B.3. Approach to Establishing and Measuring Benchmarks of Comprehensive Literacy Instruction

To measure teachers' implementation of comprehensive literacy instruction (CLI), the study team used data from the survey of teachers conducted in spring and summer 2019 and spring 2020. This section begins with an introduction to the study team's general approach to measuring CLI, followed by details for the analytic approach to each feature of CLI.

B.3.1. Overview of Approach for Measuring CLI

As noted in the report, the study focused on six features of comprehensive literacy instruction that are highlighted in the ESSA definition of CLI, and that evidence suggests are especially salient in promoting literacy outcomes such as achievement on English language arts assessments: (1) emphasis on key domains of literacy, (2) use of explicit instructional strategies, (3) differentiated instruction, (4) use of formative assessment, (5) frequent opportunities to practice, and (6) assignment of age-appropriate reading materials. See Appendix E for a full list of the 12 components of CLI included in the ESSA definition. For the purposes of the evaluation, the study collapsed 8 of the 12 components into six features, as described in Appendix E. The evaluation did not measure 4 of the 12 components defined by ESSA, because they were not readily measured based on survey data and involved aspects of teacher practice other than instruction (e.g., collaboration).

As described in section B.1, whenever possible, the study's survey of teachers used items from existing, validated, national surveys to measure comprehensive literacy instruction. The study's survey contained multiple items designed to measure each of the six features. The study team combined the survey items for each feature into a single indicator of teacher practice for that feature, rather than reporting on each survey item separately. The approach for each feature is described in detail below. Drawing on the items used to measure each feature, the team calculated the percent of teachers reportedly engaging in each feature at a benchmark level, indicating satisfactory implementation of the feature.

[11] "No studies met standards" is the term used to describe programs reviewed by WWC for which "no studies were found that fell within the scope of the [...] protocol and met WWC evidence standards. Therefore, the WWC is unable to draw any research-based conclusions about the effectiveness or ineffectiveness of [program] to improve outcomes in this area." Similarly, "No studies met inclusion criteria" is language that Evidence for ESSA uses for programs not meeting the analogous standards for that clearinghouse.

Setting the benchmarks relied on the following principles:

- **Making use of reading expert judgment about how to combine survey items and establish benchmarks.** The study team drew on five experts in reading research to guide the selection of items to measure each feature, and to set the cut points used to assess whether teachers reached the benchmark level of performance. These five experts represent a subset of the TWG for the study. The expert panel consulted in creating the benchmarks included:
  - Tim Shanahan, Ph.D., professor emeritus of education at the University of Illinois-Chicago, past president of International Literacy Association
  - Kymyona Burk, Ed.D., senior policy fellow at ExcelinEd, former state literacy director at the Mississippi Department of Education
  - Barbara Foorman, Ph.D., professor of education at Florida State University, director emeritus of Florida Center for Reading Research
  - Jimmy Kim, Ed.D., professor of education at Harvard University, director of READS Lab (Research Enhances Adaptations Designed for Scale in Literacy)
  - Sharon Vaughn, Ph.D., professor of education at the University of Texas, former editor-in-chief of *Journal of Learning Disabilities*

  The discussion of each feature below provides more detail on how expert advice was incorporated into the measures.
- **Distinguishing the level of use of the practices included in each feature.** The survey items used to measure the six features of comprehensive literacy instruction asked teachers to indicate the level of emphasis they gave to each practice or the frequency with which they used the practice. The benchmark for these features was set based on emphasis or frequency—not just whether practices were used—with the details of the method differing for the six features. For the first two features (emphasis on key domains and explicit instruction), greater use of each practice is not necessarily better, and thus the desired level for each practice was set by asking experts to respond based on the available research for each specific grade or grade band.[12] For the third and fourth features (differentiated instruction and use of formative assessment), the desired level was set by requiring responses at the higher end of the survey item scale (e.g., often or always). For the fifth feature (frequent opportunities to practice), the benchmark was set based on total use across the measured opportunities for practice. The sixth feature (age-appropriate reading materials) focused on the proportion of novels assigned that were age-appropriate.
- **Requiring at least 60 percent overlap with specified or recommended practices**. For each feature, a cut point was used to determine if teachers met the benchmark for a given feature, based on the sum of the number of points the teachers received on the survey items used to measure the feature. For four of the six features, this cut point was set at **60 percent or higher.** That is, to meet the benchmark, one would expect to see teachers using at least

[12] In determining the appropriate levels of emphasis on key domains and time spent on explicit instruction, experts responded separately for kindergarten, first grade, second grade, and third grade; experts gave a combined response for grades 4–5, for grades 6–8, and grades 9–12.

60 percent of the measured activities for the feature at the desired level. For example, if there were five items used to measure a feature, and a teacher could receive a score between 0 and 1 on each item, then they would need a score of at least 3 to meet the benchmark.[13]

The study team tested the sensitivity of reported results to key aspects of the rules used to assess whether teachers met the benchmarks for each feature, and the general pattern of results was unchanged. For example, the study team tested whether the results differed if the cutoff was set at 50 percent or 70 percent, rather than 60 percent, of the possible points for the feature (see Exhibits D.12–D.15). The study did not compute conventional reliability indices for the measures, because the practices included in each measure take class time to implement. Thus, a teacher who devotes substantial time to some of the practices included in a measure could not devote as much time to others. Thus, the study team would not necessarily anticipate that all the items tapping a measure would be positively correlated, as assumed in typical reliability indices.

Finally, the study team drew on the results for the six features to determine the percent of teachers who met all of the benchmarks, five of the six, etc. The analysis reported in Exhibit 5 is based on teachers with complete data on all features. Exhibit B.12 reports on the percentage of teachers with missing data by feature. The feature that is most often missing data is assignment of age-appropriate reading materials, for teachers in grades 4-8 and 9-12. Teachers lack data for this feature if they did not assign a novel to the whole class in the two weeks prior to taking the survey, or if the assigned novel did not have a Lexile level associated with it. Section B.3.7 contains more information about the approach and data availability for this feature.

To assess the sensitivity of the results to the fact that many teachers in grades 4-8 and 9-12 lacked data for this feature, the team conducted an analysis focusing on results for the other five features only. Results are reported in Appendix D, Exhibits D.8–D.11.

[13] Two features were not based on the 60 percent rule. For differentiated instruction, the benchmark required one specific practice plus one other; and for frequent opportunities for practice, the benchmark was based on the total number of days that opportunities for practice were offered.

Exhibit B.12. Percentage of teachers with missing data by feature of comprehensive literacy instruction

	Pre-K		Grades K-3		Grades 4-8		Grades 9-12	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Emphasis on Key Domains	0%	0%	0%	0%	0%	0%	0%	0%
Explicit Instruction	0%	0%	0%	0%	0%	0%	0%	0%
Differentiated Instruction	16%	14%	4%	5%	6%	5%	15%	16%
Formative Assessment	0%	0%	0%	0%	0%	0%	0%	0%
Frequent Opportunities to Practice Writing	0%	0%	4%	6%	5%	5%	2%	2%
Frequent Opportunities to Practice Reading	0%	0%	4%	6%	n/a	n/a	n/a	n/a
Age-Appropriate Reading Materials	n/a	n/a	n/a	n/a	63%	67%	57%	60%

B.3.2. Feature: Emphasis on Key Domains of Literacy

To measure whether teachers appropriately emphasized key domains of literacy instruction, the SRCL teacher survey included a question adapted from the Early Childhood Longitudinal Study. The question was the same for all teachers on the K-12 survey (Exhibit B.14). The pre-K teacher survey (Exhibit B.13) did not include the items on critical analysis of literary or informational text.

Exhibit B.13. Items included for pre-K

In your instruction this school year, how much emphasis have you given to each of the following topics? Check one per row.

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Phonological awareness (sounds of letters and syllables)	☐	☐	☐	☐
b. Phonics	☐	☐	☐	☐
c. Reading fluency	☐	☐	☐	☐
d. Vocabulary	☐	☐	☐	☐
e Reading comprehension (e.g., comprehension strategies)	☐	☐	☐	☐
f. Oral language (listening comprehension, speaking)	☐	☐	☐	☐
g. Writing	☐	☐	☐	☐
h. Grammar, punctuation, and capitalization (conventions of English)	☐	☐	☐	☐
i. Spelling	☐	☐	☐	☐

Exhibit B.14. Items included for grades K–12

In your instruction in this class this school year, how much emphasis have you given to each of the following skills or concepts? Check one per row.

		No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a.	Decoding (e.g., phonological awareness, phonics skills)	☐	☐	☐	☐
b.	Reading fluency	☐	☐	☐	☐
c.	Vocabulary	☐	☐	☐	☐
d.	Reading comprehension (e.g., comprehension strategies)	☐	☐	☐	☐
e.	Oral language (listening comprehension, speaking)	☐	☐	☐	☐
f.	Writing	☐	☐	☐	☐
g.	Handwriting	☐	☐	☐	☐
h.	Grammar, punctuation, and capitalization (conventions of English)	☐	☐	☐	☐
i.	Spelling	☐	☐	☐	☐
j.	Critical analysis of literary text	☐	☐	☐	☐
k.	Critical analysis of informational text	☐	☐	☐	☐

A substantial body of research suggests that different domains should be emphasized at different grade levels (see, for example, Chall, 1983). To set benchmarks for the desired emphasis on these domains by grade, the study team requested that the expert panel (listed in section B.3.1) indicate—by responding to the survey items—how much emphasis they believed should be given to each domain for students in Pre-K, K-3, 4-8, and 9-12, or how much time should be devoted to each instructional practice. After a first round of rating, we convened the expert panel to discuss any concerns they had with the rating process. While they acknowledged that there were areas of disagreement, on balance, the level of disagreement did not raise concerns. However, they did express concerns about responding for grades K-3 as a set given the very important grade level differences in literacy instruction in kindergarten and grades 1, 2, and 3. Hence, the expert panel suggested that they provide ratings for each of those grades separately, a suggestion that the study team implemented. Thus, the experts reported the emphasis that should be given to each domain for teachers teaching pre-K, kindergarten, grade 1, grade 2, grade 3, grades 4–5, grades 6–8, and grades 9–12, basing their judgment on their analysis of the available research evidence.

When soliciting expert ratings of the survey measures of comprehensive literacy instruction, the study did not force agreement among the experts. This approach is grounded in the assumption that forcing agreement may give a misleading sense of the extent to which there is consensus in the field. The approach used was adapted from the approach used in analyses of the Survey of Enacted Curriculum (SEC), which has been widely used to assess the alignment between standards, as coded by experts, assessments, and instruction (see Porter et al., 2002). The SEC has been used to assess the alignment between state assessments and state content standards (Polikoff, Porter, & Smithson, 2011) as well as the alignment between textbooks and standards (Polikoff, Zhou, & Campbell, S., 2015). Consistent with

this approach, the scoring system allocates more points when teachers agree with more experts, but credit teachers whose responses are consistent with any experts.

While experts generally agreed on the emphasis that should be given to each domain, there was not complete agreement. To measure the degree of agreement among experts, the study team computed the percent of experts who agreed on the most common rating for each domain. The agreement was 71 percent for pre-K, 73 percent for kindergarten, 67 percent for grade 1, 67 percent for grade 2, 60 percent for grade 3, 64 percent for grades 4–5, 65 percent for grades 6–8, and 65 percent for grades 9–12.

To measure whether a teacher's reported level of emphasis on each domain was consistent with the expert ratings, the study team assigned points, giving more points when the teacher's rating agreed with more experts.[14] The points were assigned as follows:

- one point if the teacher response matched the most common rating given by the experts;
- less than 1 point if the teacher response matched a response given by at least one expert, but not the most common response; or
- no points if the teacher response matched no experts.

If experts gave non-consecutive responses for a particular domain, a teacher would receive zero points for a domain if their answer was within the range of experts' responses for that domain. For instance, if some experts recommended a little emphasis on a given domain, and other experts recommended major emphasis, with no experts recommending moderate emphasis, a teacher who reported placing moderate emphasis on a domain would receive zero points for that domain. In practice, there were very few gaps in the experts' responses. The experts were asked about 86 combinations of grade bands and domains; only four of those combinations had any gaps in the experts' responses, and in all grade bands except for kindergarten and grade 3, there were no gaps at all. Given the fact that gaps were very rare, it is unlikely that this aspect of the rating process had an appreciable effect on the results.

Exhibit B.15 shows the number of points assigned to a teacher, given the teacher's response **and** the number of experts the teacher agreed with. For example, the first panel shows that if all five experts agreed, a teacher would receive 1 point if the teacher responded with the level of emphasis chosen by the most experts (five) and 0 points otherwise. The second panel shows that if four experts agreed, then a teacher would receive 1 point if the teacher responded with the level of emphasis chosen by the most experts (four), ¼ point if the teacher responded with the level of emphasis chosen by the second-most experts (one), and 0 points otherwise.

[14] Several other approaches to assigning points were also considered. For example, the study team considered assigning a teacher points based on the proportion of raters that assigned the rating the teacher selected. This would give more emphasis to domains on which the raters agreed, but because a teacher would be unable to get a full point unless all raters agreed, the total points possible would depend on the expert ratings, making the benchmark hard to set and interpret.

Exhibit B.15. Assignment of points to a teacher based on expert ratings of key domains of literacy

Expert Agreement	Number of Experts Teacher Matched	Points Assigned
If all 5 experts chose the same level	5	1
	0	0
If 4 experts chose the same level, 1 chose a different level	4	1
	1	1/4
	0	0
If 3 experts chose the same level, 2 chose another level	3	1
	2	2/3
	0	0
If 3 experts chose the same level, 2 chose different levels	3	1
	1	1/3
	0	0
If 2 experts chose the same level, 2 others chose another level, 1 chose a third level	2	1
	1	1/2
	0	0
If 2 experts chose the same level, 3 others chose different levels	2	1
	1	1/2
	0	0

The ratings given for decoding for kindergarten through grade 12 were doubled, because decoding includes two essential elements of reading instruction: phonemic awareness and phonics. (The ECLS item on which the survey was based combined these elements.)

In kindergarten through grade 12, a teacher who provided responses perfectly aligned with the experts would receive 2 points on decoding and 1 point each on all other domains, or 12 points total. In pre-K teachers could receive up to 9 points. At the opposite extreme, a teacher who was not aligned with the experts on any domain would receive 0 points. A teacher was determined to have met the benchmark if they received at least 60 percent of the possible points across domains (7.2 of the 12 possible points in kindergarten through grade 12, and 5.4 points in pre-K).

Exhibits B.16–B.23 provide the scoring for teachers in pre-K, kindergarten, grade 1, grade 2, grade 3, grades 4–5, grades 6–8, and grades 9–12.

Exhibit B.16. Scoring (number of points assigned) for emphasis on key domains of literacy, pre-K

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Phonological awareness (sounds of letters and syllables)	0	0	1/4	1
b. Phonics	1/2	1	1	0
c. Reading fluency	1	1/4	0	0
d. Vocabulary	0	0	1/4	1
e. Reading comprehension (e.g., comprehension strategies)	1	2/3	0	0
f. Oral language (listening comprehension, speaking)	0	0	0	1
g. Writing	0	1	1/4	0
h. Grammar, punctuation, and capitalization (conventions of English)	1	2/3	0	0
i. Spelling	2/3	1	0	0
Scoring rule: Benchmark met if sum of points ≥ 5.4.				

Exhibit B.17. Scoring (number of points assigned) for emphasis on key domains of literacy, kindergarten

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)[a]	0	0	0	1
b. Reading fluency	1/3	1	0	1/3
c. Vocabulary	0	0	1/4	1
d. Reading comprehension (e.g., comprehension strategies)	1/2	1	1	0
e. Oral language (listening comprehension, speaking)	0	0	0	1
f. Writing	0	1	1	1/2
g. Handwriting	0	1	1/3	1/3
h. Grammar, punctuation, and capitalization (conventions of English)	1/3	1	1/3	0
i. Spelling	1/3	1	0	1/3
j. Critical analysis of literary text	1	0	0	0
k. Critical analysis of informational text	1	0	0	0
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.18. Scoring (number of points assigned) for emphasis on key domains of literacy, grade 1

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	0	0	0	1
b. Reading fluency	0	0	2/3	1
c. Vocabulary	0	0	1/4	1
d. Reading comprehension (e.g., comprehension strategies)	0	1	1	1/2
e. Oral language (listening comprehension, speaking)	0	0	1/4	1
f. Writing	0	0	1	2/3
g. Handwriting	0	1/4	1	0
h. Grammar, punctuation, and capitalization (conventions of English)	1/3	1/3	1	0
i. Spelling	0	1/3	1	1/3
j. Critical analysis of literary text	1	2/3	0	0
k. Critical analysis of informational text	1	2/3	0	0
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.19. Scoring (number of points assigned) for emphasis on key domains of literacy, grade 2

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	0	0	1/4	1
b. Reading fluency	0	0	0	1
c. Vocabulary	0	0	0	1
d. Reading comprehension (e.g., comprehension strategies)	0	1/2	1	1
e. Oral language (listening comprehension, speaking)	0	0	1/4	1
f. Writing	0	0	1	1/4
g. Handwriting	0	2/3	1	0
h. Grammar, punctuation, and capitalization (conventions of English)	1/2	1/2	1	1/2
i. Spelling	0	1/2	1	1
j. Critical analysis of literary text	2/3	1	0	0
k. Critical analysis of informational text	2/3	1	0	0
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.20. Scoring (number of points assigned) for emphasis on key domains of literacy, grade 3

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	0	1	1	1/2
b. Reading fluency	0	0	1/4	1
c. Vocabulary	0	0	0	1
d. Reading comprehension (e.g., comprehension strategies)	0	0	2/3	1
e. Oral language (listening comprehension, speaking)	0	0	1	2/3
f. Writing	0	0	1/4	1
g. Handwriting	2/3	0	1	0
h. Grammar, punctuation, and capitalization (conventions of English)	1/2	0	1	1
i. Spelling	0	1/3	1/3	1
j. Critical analysis of literary text	1	1/2	1	0
k. Critical analysis of informational text	1	1/2	1	0
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.21. Scoring (number of points assigned) for emphasis on key domains of literacy, grades 4–5

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	1/2	1	1	0
b. Reading fluency	0	0	1	1/4
c. Vocabulary	0	0	0	1
d. Reading comprehension (e.g., comprehension strategies)	0	1/3	1/3	1
e. Oral language (listening comprehension, speaking)	0	1	1	1/2
f. Writing	0	0	1/4	1
g. Handwriting	1	1/2	1	0
h. Grammar, punctuation, and capitalization (conventions of English)	0	1/4	1	0
i. Spelling	0	1/3	1	1/3
j. Critical analysis of literary text	0	2/3	1	0
k. Critical analysis of informational text	0	2/3	1	0
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.22. Scoring (number of points assigned) for emphasis on key domains of literacy, grades 6–8

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	1	1/4	0	0
b. Reading fluency	0	2/3	1	0
c. Vocabulary	0	0	1/4	1
d. Reading comprehension (e.g., comprehension strategies)	0	0	2/3	1
e. Oral language (listening comprehension, speaking)	0	1	1	1/2
f. Writing	0	0	0	1
g. Handwriting	1	2/3	0	0
h. Grammar, punctuation, and capitalization (conventions of English)	0	2/3	1	0
i. Spelling	0	2/3	1	0
j. Critical analysis of literary text	0	0	2/3	1
k. Critical analysis of informational text	0	0	2/3	1
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

Exhibit B.23 Scoring (number of points assigned) for emphasis on key domains of literacy, grades 9–12

	No Emphasis	A Little Emphasis	Moderate Emphasis	Major Emphasis
a. Decoding (e.g., phonological awareness, phonics skills)	1	1/4	0	0
b. Reading fluency	1/2	1	1	0
c. Vocabulary	0	0	2/3	1
d. Reading comprehension (e.g., comprehension strategies)	0	0	1	2/3
e. Oral language (listening comprehension, speaking)	0	1	1/4	0
f. Writing	0	0	1/4	1
g. Handwriting	1	2/3	0	0
h. Grammar, punctuation, and capitalization (conventions of English)	0	2/3	1	0
i. Spelling	0	1	1	1/2
j. Critical analysis of literary text	0	0	1/4	1
k. Critical analysis of informational text	0	0	1/4	1
Scoring rule: Benchmark met if sum of points ≥ 7.2 (with points for decoding doubled).				

B.3.3. Feature: Use of Explicit Instructional Strategies

To examine the use of explicit instructional strategies in literacy, teacher survey respondents were asked how frequently they used specific explicit instructional strategies within a subset of the literacy domains using existing items from surveys conducted with large, nationally representative samples of teachers (ECLS and NAEP). The subset of domains measured was selected based on expert advice about which should be given moderate or major attention in particular grade spans.[15]

For each domain, the study team identified existing items from ECLS (pre-K and grades K-3) and NAEP (grades 4-12) that tap instructional practices that teachers might use in the domain (exhibits B.24-B.31.) The items ask about the frequency with which teachers use the practices. In particular, the pre-K items asked about the number of times each week a teacher used a strategy, the items for grades K-3 asked about the number of days each year, and the items for grades 4-12 asked how often teachers used a given strategy (sometimes, often, etc.).

The members of the expert panel were asked to indicate how frequently they believed a teacher should engage in each practice, based on the available evidence base, using the survey's response metric, separately for pre-K, kindergarten, grade 1, grade 2, grade 3, grades 4-5, grades 6-8, and grades 9-12.[16] Each teacher was assigned points based on the level of agreement between the teacher's responses and the expert judgments, using a method related to the method used for emphasis on key domains, described above. A teacher was assigned 1 point for an item if the teacher's response agreed with the response chosen by a majority of the experts; 0.5 points if the response agreed with at least one expert, but not the majority, or if the response was not chosen by any experts but was between responses chosen by at least one expert; and 0 points if the response did not agree with any of the experts. To measure the degree of agreement among experts, the study team computed the percentage of experts who agreed on the most common rating for each explicit instructional practice. The agreement was 64 percent for pre-K, 60 percent for kindergarten, 54 percent for grade 1, 63 percent for grade 2, 60 percent for grade 3, 75 percent for grades 4-5, 61 percent for grades 6-8, and 69 percent for grades 9-12.

The method used for explicit instruction differs from the method used for emphasis on key domains in two respects: the score assigned to teachers whose response agreed with some experts, but not the majority, and the treatment of responses that were chosen by no experts but were between responses chosen by at least one expert. The difference in the method is due to the difference in the number of response categories—four for emphasis on key domains, but five or six for explicit instruction. Because of the larger number of response categories, the expert responses tended to be more spread out for explicit instruction. Also, on occasion, experts respond in nonconsecutive categories—for example, some experts might respond "2-3 times a month," and some might respond "3-4 times a week," but

[15] For pre-K, the items focused on comprehension of spoken text, vocabulary, and phonics/phonemic awareness. For grades K-3, the items focused on comprehension of spoken and written text, vocabulary, and phonics/phonemic awareness. For grades 4-12, the items focused on reading comprehension and critical analysis of written text.

[16] Four of the five experts who provided ratings for emphasis on key domains provided ratings for explicit instruction.

none might respond "1-2 times a week," which lies in between. We assigned 0.5 points for "1-2 times a week" in this case. For a few items, there was no agreement among experts on the appropriate response. For these, the study team assigned 0 points for all responses.

A teacher was determined to have met the benchmark if they received at least 60 percent of the possible points. Because items on which no two experts agreed provided 0 possible points, those items were effectively excluded from the analysis.

Exhibits B.24-B.31 provide the items and scoring for teachers in pre-K, kindergarten, grade, 1, grade 2, grade 3, grades 4-5, grades 6-8, and grades 9-12.

Exhibit B.24. Items and scoring (number of points assigned) for explicit instruction, pre-K

In the current school year, how often are children in your class asked to do the following early literacy activities? Check one response for each activity.

	Never	Once a Month or Less	2-3 Times a Month	1-2 Times a Week	3-4 Times a Week	Daily
Comprehension of spoken text[a]						
Listen to you read stories but they don't see the print	0	0	0	0	0	1
Retell stories	0	0	0	0.5	1	0
Answer open-ended questions about books you have read to them	0	0	0	0	1	1
Answer factual questions about a story that you have read to them	0	0	0	0	1	1
Vocabulary[a]						
Discuss new or difficult vocabulary	0	0	0	0.5	0.5	1
Phonics/phonemic awareness[a]						
Work on learning names of letters	0	0	0	0	0.5	1
Learn about rhyming words and word families	0	0	0	0.5	0.5	1
Work on phonics	0	0	0	0.5	0.5	1
Read text with strong phonetic patterns	0	0.5	0.5	1	0.5	0.5
Scoring rule: Benchmark met if sum ≥ 5.4 points.						

[a] This heading did not appear in the items presented to teachers.

Exhibit B.25. Items and scoring (number of points assigned) for explicit instruction, kindergarten

From the first day of school until today, please indicate how many days each of the following reading/ELA skills and concepts has been covered in your class? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Not all skills/concepts may apply to your grade level. Check one per row.

	Not Taught	On 1-10 Days	On 11-20 Days	On 21-40 Days	On 41-80 Days	On More Than 80 Days
Comprehension of spoken and written text[a]						
Predicting what might occur next in the text	0	0.5	0.5	0.5	1	0.5
Identifying main ideas in a story	0	0	0.5	0.5	0.5	1
Retelling stories, including main ideas and details	0	0	0	0.5	0.5	1
Describing characters, settings, and major events in a story	0	0	0	0	1	0.5
Using text features to locate key facts or information (for example, captions, bold print, subheadings, indexes)	1	0.5	0.5	0.5	0.5	0
Comparing and contrasting the themes, setting, and plots of two similar stories	1	0.5	0.5	0.5	0	0
Asking and/or answering questions to demonstrate understanding of key details in a text	0	0	0	0	0.5	1
Vocabulary[a]						
Using text features such as glossaries and other references to learn word meanings	1	0.5	0.5	0.5	0.5	0
Determining the meaning of words and phrases as they are used in a text	0	0	0	0.5	1	0
Phonics/phonemic awareness[a]						
Segmenting words into phonemes	0	0	0	0	0.5	1
Manipulating phonemes to form new words	0	0	0	0.5	0.5	1
Breaking spoken words into sounds	0	0	0	0	0	1
Blending sounds to form words	0	0	0	0	0.5	1
Reading irregularly spelled words	0	0	0	1	1	0
Decoding regularly spelled two-syllable words	0	0	0	0	0	0
Decoding multi-syllabic words	1	0.5	0.5	0.5	0.5	0
Identifying and knowing the meaning of common prefixes or suffixes	0.5	1	0.5	0.5	0	0
Distinguishing long and short vowels in one-syllable words	0	0	0	0	0	0
Scoring rule: Benchmark met if sum ≥ 10.8 points						

[a] This heading did not appear in the items presented to teachers.

Exhibit B.26. Items and scoring (number of points assigned) for explicit instruction, grade 1

From the first day of school until today, please indicate how many days each of the following reading/ELA skills and concepts has been covered in your class? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Not all skills/concepts may apply to your grade level. Check one per row.

	Not Taught	On 1-10 Days	On 11-20 Days	On 21-40 Days	On 41-80 Days	On More Than 80 Days
Comprehension of spoken and written text[a]						
Predicting what might occur next in the text	0	0	0	0.5	0.5	1
Identifying main ideas in a story	0	0	0.5	0.5	0.5	1
Retelling stories, including main ideas and details	0	0	0	0	0.5	1
Describing characters, settings, and major events in a story	0	0	0	0	0.5	1
Using text features to locate key facts or information (for example, captions, bold print, subheadings, indexes)	0.5	0.5	0.5	1	0.5	0
Comparing and contrasting the themes, setting, and plots of two similar stories	0	0	0	0	0	0
Asking and/or answering questions to demonstrate understanding of key details in a text)	0	0	0	0	0	1
Vocabulary[a]						
Using text features such as glossaries and other references to learn word meanings	0	0	0	0	0	0
Determining the meaning of words and phrases as they are used in a text	0	0	0	0.5	0.5	1
Phonics/phonemic awareness[a]						
Segmenting words into phonemes	0	0	0	0	1	1
Manipulating phonemes to form new words	0	0	0	0.5	1	0.5
Breaking spoken words into sounds	0	0	0	0	1	1
Blending sounds to form words	0	0	0	0	0.5	1
Reading irregularly spelled words	0	0	0	1	0.5	1
Decoding regularly spelled two-syllable words	0	0	0.5	0.5	0.5	1
Decoding multi-syllabic words	0	0	0	0	0	0
Identifying and knowing the meaning of common prefixes or suffixes	0	0	0.5	1	0.5	0
Distinguishing long and short vowels in one-syllable words	0	0	0	1	0.5	1
Scoring rule: Benchmark met if sum ≥ 10.8 points						

[a] This heading did not appear in the items presented to teachers.

Exhibit B.27. Items and scoring (number of points assigned) for explicit instruction, grade 2

From the first day of school until today, please indicate how many days each of the following reading/ELA skills and concepts has been covered in your class? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Not all skills/concepts may apply to your grade level. Check one per row.

	Not Taught	On 1-10 Days	On 11-20 Days	On 21-40 Days	On 41-80 Days	On More Than 80 Days
Comprehension of spoken and written text[a]						
Predicting what might occur next in the text	0	0	0	0	1	1
Identifying main ideas in a story	0	0	0	0	0	1
Retelling stories, including main ideas and details	0	0	0	0.5	0.5	1
Describing characters, settings, and major events in a story	0	0	0	0	0.5	1
Using text features to locate key facts or information (for example, captions, bold print, subheadings, indexes)	0	0.5	0.5	0.5	1	0.5
Comparing and contrasting the themes, setting, and plots of two similar stories	0	0	0	0	0	0
Asking and/or answering questions to demonstrate understanding of key details in a text)	0	0	0	0	0	1
Vocabulary[a]						
Using text features such as glossaries and other references to learn word meanings	0.5	0.5	0.5	0.5	1	0.5
Determining the meaning of words and phrases as they are used in a text	0	0	0	0	0	1
Phonics/phonemic awareness[a]						
Segmenting words into phonemes	0	0	0	0	0	0
Manipulating phonemes to form new words	0	1	0.5	0.5	0.5	0.5
Breaking spoken words into sounds	0	1	0.5	1	0	0
Blending sounds to form words	0	0.5	0.5	1	0.5	0.5
Reading irregularly spelled words	0	0	0	0	1	0.5
Decoding regularly spelled two-syllable words	0	0	0	0.5	1	0.5
Decoding multi-syllabic words	0	0	0	0.5	0.5	1
Identifying and knowing the meaning of common prefixes or suffixes	0	0	0	0	1	0.5
Distinguishing long and short vowels in one-syllable words	0	0	0.5	1	0.5	0
Scoring rule: Benchmark met if sum ≥ 10.8 points						

[a] This heading did not appear in the items presented to teachers.

Exhibit B.28. Items and scoring (number of points assigned) for explicit instruction, grade 3

From the first day of school until today, please indicate how many days each of the following reading/ELA skills and concepts has been covered in your class? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Not all skills/concepts may apply to your grade level. Check one per row.

	Not Taught	On 1-10 Days	On 11-20 Days	On 21-40 Days	On 41-80 Days	On More Than 80 Days
Comprehension of spoken and written text[a]						
Predicting what might occur next in the text	0	0	0	0.5	1	0
Identifying main ideas in a story	0	0	0	0	0	1
Retelling stories, including main ideas and details	0	0	0	0.5	0.5	1
Describing characters, settings, and major events in a story	0	0	0	0	1	1
Using text features to locate key facts or information (for example, captions, bold print, subheadings, indexes)	0	0	0.5	0.5	0.5	1
Comparing and contrasting the themes, setting, and plots of two similar stories	0	0	0	0	0	0
Asking and/or answering questions to demonstrate understanding of key details in a text)	0	0	0	0	0	1
Vocabulary[a]						
Using text features such as glossaries and other references to learn word meanings	0	0	0.5	0.5	0.5	1
Determining the meaning of words and phrases as they are used in a text	0	0	0	0	0.5	1
Phonics/phonemic awareness[a]						
Segmenting words into phonemes	1	0.5	0.5	0.5	0	0
Manipulating phonemes to form new words	1	0.5	0.5	0.5	0	0
Breaking spoken words into sounds	1	0.5	0.5	0.5	0	0
Blending sounds to form words	1	0.5	0.5	0.5	0	0
Reading irregularly spelled words	0.5	0.5	0.5	0.5	1	0
Decoding regularly spelled two-syllable words	0	0	0.5	0.5	1	0
Decoding multi-syllabic words	0	0	0	0	0.5	1
Identifying and knowing the meaning of common prefixes or suffixes	0	0	0	0	0.5	1
Distinguishing long and short vowels in one-syllable words	0	0	0	0	0	0
Scoring rule: Benchmark met if sum ≥ 10.8 points						

[a] This heading did not appear in the items presented to teachers.

Exhibit B.29. Items and scoring (number of points assigned) for explicit instruction, grades 4–5

When reading a story, article, or other passage with your students, how often do you ask your students to do the following? Check one per row.

	Never or Hardly Ever	Once in a While	Sometimes	Often	Always or Almost Always
Reading comprehension[a]					
Summarize the passage	0	0	0	0	1
Interpret the meaning of the passage	0	0	0	0	1
Identify the main ideas of the passage	0	0	0	0	1
Identify the themes of the passage	0	0	0	0.5	1
Evaluate the main evidence in a persuasive/argument passage	0	0	0.5	0.5	1
Critical analysis of written text[a]					
Question the motives or feelings of the characters	0	0	0	1	0.5
Analyze two or more texts on the same topic	0	0	1	1	0
Analyze the author's organization of information in a passage	0	0	0.5	1	0
Critique the author's craft or technique	0	0	1	1	0
Scoring rule: Benchmark met if sum ≥ 5.4 points					

[a] This heading did not appear in the items presented to teachers.

Exhibit B.30. Items and scoring (number of points assigned) for explicit instruction, grades 6–8

When reading a story, article, or other passage with your students, how often do you ask your students to do the following? Check one per row.

	Never or Hardly Ever	Once in a While	Sometimes	Often	Always or Almost Always
Reading comprehension[a]					
Summarize the passage	0	0	0	1	1
Interpret the meaning of the passage	0	0	0	0.5	1
Identify the main ideas of the passage	0	0	0	0.5	1
Identify the themes of the passage	0	0	0	0.5	1
Evaluate the main evidence in a persuasive/argument passage	0	0	0	0.5	1
Critical analysis of written text[a]					
Question the motives or feelings of the characters	0	0	0	1	1
Analyze two or more texts on the same topic	0	0	0	1	1
Analyze the author's organization of information in a passage	0	0	0	1	1
Critique the author's craft or technique	0	0	0.5	0.5	1
Scoring rule: Benchmark met if sum ≥ 5.4 points					

[a] This heading did not appear in the items presented to teachers.

Exhibit B.31. Items and scoring (number of points assigned) for explicit instruction, grades 9–12

When reading a story, article, or other passage with your students, how often do you ask your students to do the following? Check one per row.

	Never or Hardly Ever	Once in a While	Sometimes	Often	Always or Almost Always
Reading comprehension[a]					
Summarize the passage	0	0	0	1	1
Interpret the meaning of the passage	0	0	0	0.5	1
Identify the main ideas of the passage	0	0	0	0.5	1
Identify the themes of the passage	0	0	0	0	1
Evaluate the main evidence in a persuasive/argument passage	0	0	0	0.5	1
Critical analysis of written text[a]					
Question the motives or feelings of the characters	0	0	0	1	1
Analyze two or more texts on the same topic	0	0	0	0.5	1
Analyze the author's organization of information in a passage	0	0	0	0.5	1
Critique the author's craft or technique	0	0	0	1	1
Scoring rule: Benchmark met if sum ≥ 5.4 points					

[a] This heading did not appear in the items presented to teachers.

B.3.4. Feature: Differentiated Instruction

The pre-K and K–12 teacher surveys included items asking about differentiated instructional strategies for struggling readers, based on items used in the Reading First Impact Study teacher survey (Gamse et al., 2008; exhibit B.32.) The items tap aspects of differentiation that apply across grade levels, rather than practices related to instruction in particular domains of literacy, which might differ across grade bands. Given the focus of the SRCL program on disadvantaged students, including students performing below grade level, the survey items asked about differentiated approaches for struggling students. For pre-K teachers, the items were asked of all teachers, assuming they would be teaching at least some students performing below age expectations, but teachers who reported not teaching any students below age expectations were excluded from the measure of differentiated instruction. For K–12 teachers, the items were asked only of teachers who reported teaching struggling readers.[17]

[17] Less than 7 percent of teachers reported that they did not teach struggling readers.

Exhibit B.32. Items and scoring rule for reported use of differentiated instructional strategies for struggling readers, all grades

During this school year, how often do your students who are performing below age expectations in early literacy receive each of the following supports?[18]

	Never or Hardly Ever	Once in a While	Sometimes	Often	Always or Almost Always
a. Separate instructional activities or materials as part of regular instruction (e.g., readings that are on a different level)				☑	
b. Extra instructional time with the lead teacher (i.e., with me or a second lead teacher)					
c. Extra instructional time with an assistant teacher (Note: This item was omitted if teachers had indicated in a previous item that their classroom did not have an assistant teacher.)					
d. Extra instructional time with a paraprofessional, tutor, or volunteer (Note: This item was omitted if teachers had indicated in a previous item that their classroom did not have a paraprofessional or volunteer.)				☑	
e. Extra instructional time from a reading specialist or interventionist (push-in or pull-out)					
Scoring rule: Benchmark met if ☑ in row a + at least one ☑ in rows d-e.					

The study's panel of reading experts advised that use of differentiated materials is a critical component of differentiated instruction, and thus use of such materials often, always, or almost always was required to meet the threshold; simply having another adult in the classroom was insufficient.

B.3.5. Feature: Use of Formative Assessment

This feature was measured by the extent to which teachers reported using the results of progress testing to identify students who needed additional support or remediation, to target instruction and group students by skill level, and to identify gaps in the curriculum. The items were developed for the survey, drawing on the Reading First Impact Study teacher survey (Gamse et al., 2008; Exhibit B.33.)

A teacher met the benchmark if they scored at least 3 out of 5 (that is, reported using at least 60 percent of the practices moderately or extensively), indicating that the teacher used formative assessment for a range of purposes.

[18] The item stem for pre-K teachers differed, asking about students who are performing below age expectations in early literacy rather than struggling readers (students performing below grade level).

Exhibit B.33. Items and scoring rule for use of formative assessment, all grades

To what extent have you used the results from screening, progress, or proficiency tests in early literacy for the following activities?

	Did Not Use in This Way	Used Minimally	Used Moderately	Used Extensively
a. Identify individual students who need additional instructional support, such as tutoring				☑
b. Identify individual students for remedial services/supports				☑
c. Tailor instruction to individual students' needs				☑
d. Identify and correct gaps in the curriculum or program				☑
e. Group students for instruction				☑
Scoring rule: Benchmark met if at least three ☑ in rows a-e.				

B.3.6. Feature: Frequent Opportunities to Practice

Writing

To assess the frequency of opportunities to practice writing skills, the survey asked different questions to teachers of pre-K, grades K–3, and grades 4–12, reflecting the typical nature of writing assignments in these grade bands. The items for pre-K and grades K–3 were drawn from ECLS, and the items for grades 4–12 were drawn from NAEP.

Although the three sets of items focused on frequency, the specific response options differed. The items for pre-K and grades K–3 asked about the number of times activities occurred each week, while the items for grades 4–12 asked about the number of days each year. To set a common benchmark across grades, each response option was translated into a number of days, using the low end of the range for each response option. For example, "once a month or less" was converted to 1 day (the lowest end of the range), and "11 to 20 days" was converted to 11 days.

To meet the benchmark for frequent opportunities to practice writing skills, a teacher must report devoting a total of at least 80 days to the activities relevant to their grade band, implying that practice takes place during a substantial portion of a typical school year (usually 180 days).

Exhibits B.34–B.36 display the survey items, response scale, and estimated number of days corresponding to each response option.

Exhibit B.34. Items and scoring rule for frequent opportunities for practice in writing, pre-K

In the current school year, how often do children in your class do each of the following writing activities?

Survey Response Scale	Never	Once a Month or Less	2–3 Times a Month	1–2 Times a Week	3–4 Times a Week	Daily
Number of Days Added to Score	**0**	**1**	**18**	**36**	**72**	**100**
Write words from dictation to improve spelling						
Write with encouragement to use invented spelling, if needed						
Compose and write stories or reports						
Write stories in a journal						
Write letters to other children or to family members						
Write their name						
Scoring rule: Benchmark met if sum of number of days ≥ 80						

Exhibit B.35. Items and scoring rule for frequent opportunities for practice in writing, grades K-3

In the current school year, how often do children in your class do each of the following writing activities?

Survey Response Scale	Not Taught	On 1–10 Days	On 11–20 Days	On 21–40 Days	On 41–80 Days	On More Than 80 Days
Number of Days Added to Score	**0**	**1**	**11**	**21**	**41**	**80**
Writing an opinion piece, giving reasons for the opinion						
Writing an information piece that contains a clear topic with supporting details						
Writing an information piece that develops the topic with facts, definitions, and/or details						
Writing a narrative with two or more appropriately sequenced events						
Using temporal words or phrases to signal the order of events						
Scoring rule: Benchmark met if sum of number of days ≥ 80						

Exhibit B.36. Items and scoring rule for frequent opportunities for practice in writing, grades 4-12

In this class this year, how often do you give the following writing assignments to your students?

Survey Response Scale	Never	About Once or Twice a Year	About Once or Twice a Month	About Once or Twice a Week	Every Day or Almost Every Day
Number of Days Added to Score	**0**	**1**	**9**	**36**	**100**
Assignments that students have to complete within one session					
Assignments that are designed for students to work on over extended periods of time (e.g., several class periods)					
Assignments that students have to complete together with other students					
Assignments of at least a couple of pages (e.g., a paper or report)					
Scoring rule: Benchmark met if sum of number of days ≥ 80					

Reading

Frequency of opportunities to practice reading skills was assessed only among teachers of pre-K and grades K-3. Similar items for grades 4-12 were not included, because instruction at those grade levels is typically focused on reading comprehension rather than reading practice. The items used for this measure are based on ELCS.

Exhibit B.37 displays the items used for pre-K. The scoring system was similar to the system used for pre-K writing, described above.

Exhibit B.37. Items and scoring rule for frequent opportunities for practice in reading, pre-K

In the current school year, how often are children in your class asked to do the following early literacy activities? Check one response for each activity.

Survey Response Scale	Never	Once a Month or Less	2–3 Times a Month	1–2 Times a Week	3–4 Times a Week	Daily
Number of Days Added to Score	**0**	**1**	**18**	**36**	**72**	**100**
Listen to you read stories where they see the print						
Read books they have chosen for themselves						
Read text with patterned or predictable text						
Scoring rule: Benchmark met if sum of number of days ≥ 80						

Exhibit B.38 displays the item used for grades K–3, which was based on ECLS.

Exhibit B.38. Item and scoring rule for frequent opportunities for practice in reading, grades K-3

From the first day of school until today, how many days have students in your class been asked to read and reread passages orally with guidance on pacing, intonation, and expression as part of reading/ELA instruction? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Check one.

Survey Response Scale	Not Taught	On 1–10 Days	On 11–20 Days	On 21–40 Days	On 41–80 Days	On More Than 80 Days
Read and reread passages orally with guidance on pacing, intonation, and expression as part of reading/ELA instruction?						☑
Scoring rule: Benchmark met if ☑						

B.3.7. Feature: Assignment of Age-Appropriate Reading Materials

To examine the use of age-appropriate instructional strategies in literacy, the survey asked teacher respondents in grades 4–12 to write in the titles of novels assigned to students within the last 2 weeks, and to indicate whether each text was assigned to the entire class or to student subgroups (Exhibit B.39). The analysis for this feature focused on grades 4–12 because students in earlier grades are less likely to be assigned novels as part of regular instruction.

Exhibit B.39. Item for age-appropriate reading materials, grades 4–12

Were any students in this class required to read a novel over the past two weeks? (Include a novel students read over the past two weeks, even if it was originally assigned earlier.)

(If Yes) Please provide the author and title of each novel that students were required to read over the past two weeks.

Author	Title	Was This Assigned for All Students or Only Some of the Students in the Class? Please select one per row.
		☐ All ☐ Some
		☐ All ☐ Some
		☐ All ☐ Some
		☐ All ☐ Some

Teachers could include up to four novels in their responses. Out of 1,457 teachers in grades 4–12, 750 (51 percent) had assigned at least one novel to the entire class in the last 2 weeks. The remaining teachers did not assign novels to the entire class and were not included in the analysis of age-appropriateness. Only 1 percent of teachers assigned a novel to some of the students in the class but not to the entire class.

For each novel assigned to the entire class, the study team compared the Lexile level—a numerical measure of text complexity—to the typical Lexile range associated with a teacher's grade. The survey item and approach drew on the work of Shanahan and Duffett (2013), which used Lexiles to assess age-appropriate reading assignments in grades 4–12.[19]

A Lexile level is a readability measure developed by MetaMetrics, Inc. that aims to measure reading comprehension via text features such as vocabulary and sentence complexity. A student who is reading a book with a Lexile level that aligns with the student's Lexile reading measure would, on average, see a 75 percent comprehension rate—not too frustrating, but challenging enough to promote reading growth. While Lexile levels allow us to analyze a text's difficulty, this method of analysis does not include a measure of content sophistication or maturity, factors that might otherwise contribute to the age-appropriateness of a novel.

Lexile levels are available only for prose, so teachers who most recently assigned plays or poetry do not have Lexile levels available. Lexiles are also restricted to fiction, so nonfiction assignments were not included.

Of the 750 teachers who assigned at least one novel, 559 (75 percent) provided the name of at least one novel for which the study team was able to determine the Lexile level. (Some novels were not listed in the MetaMetrics database and so could not be included in the analysis.)

[19] Shanahan and Duffett (2013) provided teachers a list of novels and other readings and asked them to indicate which were assigned over a full year, rather than asking about a 2-week period.

The study team compared the Lexile levels of the novels teachers listed to the typical grade ranges as described by MetaMetrics. If a novel's Lexile level fell within the appropriate range for the grade the teacher taught, it was determined that the novel was age-appropriate. MetaMetrics provides two different grade-range/Lexile maps: an "old" Lexile grade map, and a new "Common Core" grade map. The Common Core map is more challenging and is intended to be aspirational–that is, it sets the Lexile range for each grade band somewhat higher than the old range. For this study, the older range was used. The Lexile levels associated with different grade bands appear in Exhibit B.40.

Exhibit B.40. "Old" and CCR Lexile level ranges by grade band

Text Complexity Grade Band	Old Lexile Ranges	Lexile Ranges Aligned to CCR Expectations
Grades 4 and 5	645-845	740-1010
Grades 6-8	860-1010	925-1185
Grades 9-10	960-1115	1050-1335
Grades 11-CCR	1070-1220	1185-1385

Note: CCR is College and Career Ready Standards.
Source: lexile.com/educators/measuring-growth-with-lexile/college-and-career-readiness/.

The benchmark for age-appropriate novel assignments was set at 60 percent. That is, a teacher was determined to have met the benchmark if 60 percent or more of the novels they assigned fell within the typical Lexile range for the grade (Exhibit B.41.)

Exhibit B.41. Scoring for assignment of age-appropriate reading materials

Number of Novels Assigned by Teacher for Which Lexile Level Was Available	Percent of Teachers Assigning the Indicated Number of Novels	Benchmark Met if Number of Age-Appropriate Novels Was at Least:
1	79	1
2	12	2
3	6	2
4	3	3

Teachers for whom no Lexile levels were available, either because they did not assign novels or because Lexile levels were not available for the novels assigned, were excluded from the denominator in calculating the percentage who met the benchmark for assignment of age-appropriate texts. They were also excluded from the calculation of the percentage of teachers who met the benchmarks for all six features of CLI measured by the study. Taking these exclusions into account, 38 percent of teachers in grades 4-12 were included in the analysis for the assignment of age-appropriate materials.

B.4. Approaches to Computing State Averages and Significance Testing

The following section presents a summary of the methods used in compiling overall findings across states and testing for significant differences across groups such as years and grade bands.

B.4.1. Computation of State Averages

Except when otherwise noted, exhibits presenting data for all SRCL states combined are created by first obtaining the relevant value for each of the 11 SRCL states, and then computing the simple average of the state-level values to calculate the overall value. This method was chosen because SRCL awarded grants to states. To the extent there are meaningful policy differences between states, the simple average gives equal weight to each state, rather than assigning greater importance to states with more SRCL districts (for analyses in which the unit of analysis is a district) or more survey respondents (for analyses in which the unit of analysis is a principal or school leader or a teacher).[20] For instance, in Exhibit 2 in the report, the blue dot in the row labeled "All," representing the overall percentage of SRCL schools classified as disadvantaged, is the average of the values in the 11 state-specific blue dots. In analyses of survey data, each observation is assigned its respective survey weight when calculating state-specific values. See section B.1 for more details on how the weights are created.

B.4.2. Statistical Significance Testing

Some analyses examine differences in outcomes between groups, such as differences in comprehensive literacy benchmarks by grade band (Exhibit 6 in the report), differences in evidence use by core and supplemental curriculum (Exhibits C.14a and C.14b), differences in comprehensive literacy benchmarks by SRCL implementation year (Exhibits C.21–C.24), and differences in comprehensive literacy benchmarks by exposure to professional development on features of comprehensive literacy instruction (Exhibits D.21–D.28). The study team computed the mean of the outcome of interest separately for each state for each category (e.g., separately for the first and second year of implementation, or for core and supplemental curricula). The team then averaged the state-level means within a category to obtain the equally-weighted average outcome for the category. All standard errors account for the survey design, including the school strata and the clustering of teachers within schools. Standard errors were computed using the Taylor Series method, using the *svy* prefix in Stata.

To test the statistical significance of the difference between categories (e.g., between the first and second year of implementation), the team computed the difference between the categories separately for each state, along with the variance of the difference, computed as the sum of the variances of the estimates for the categories being compared. The variance of the average difference across states was computed as the sum of the state-specific variances, divided by the square of the number of states (11^2 = 121). This was used as the basis for a *t*-test, with degrees of freedom equal to the number of primary sampling units (schools) minus the number of strata.

While this method assumes that observations in the two categories being compared are independent, this may not always be true in practice. For instance, teachers in different grade bands (e.g., K-3 and 4-6) were sampled from the same schools and districts, so there is likely to be some dependence between those teachers. Similarly, teachers in the two implementation years were sampled from the same

[20] For instance, 100 districts in Ohio responded to the district survey, while 25 districts in Oklahoma responded, primarily reflecting differences in the number of districts receiving SRCL funds in the two states. The averaging methods used in this study give Ohio and Oklahoma equal weight, rather than having Ohio implicitly count for four times as much weight.

sample of schools. Given that the ignored dependence involves whatever effects teachers in the same schools share, it is likely to lead to standard errors that are too large, thus reducing the chance of obtaining statistically significant differences. This would be the case, for example, in comparing teachers in the K-3 and 4-6 grade bands, if schools in which teachers in grades K-3 are especially likely to meet benchmarks are those in which teachers in grades 4-6 are also especially likely to meet benchmarks. However, it is possible that ignoring this dependence between the two groups being compared might lead to standard errors that are too small, leading to a false indication of statistical significance.

B.5. Analyses of Targeting of SRCL Funds to Districts and Schools With Disadvantaged Students

The analysis presented in Exhibit 1 in the report displays the percentage of SRCL and non-SRCL schools in each SRCL state that are in the most disadvantaged quartile in their state on one or more of the following four measures: the percentage of students eligible for free or reduced-price lunch, the percentage of students who are English learners, the percentage of students with disabilities, and the school's average score on state standardized English language arts exams. The average test score is from 2008–09 through 2015–16 in the SEDA; the percentages of students who are English learners and students with disabilities are from the 2015–16 SEDA; and the percentage of students eligible for free or reduced-price lunch is from the 2018–19 CCD.

Schools missing any of the four measures were dropped from the analyses of targeting reported in Exhibit 2. Thus, the total percentage of schools classified as disadvantaged across measures exceeds the percentage disadvantaged for any single measure. For instance, a school with no available data on its percentage of students with disabilities was excluded from both the numerator and the denominator in calculating the percentages in Exhibit 2 in the report and in Exhibit C.1. These schools were dropped before the quartiles were created; the schools classified as disadvantaged based on the percentage of students who are English learners have percentages of English learners among the highest 25 percent *among schools with nonmissing data on all four measures*.

Schools with missing data on one or more indicators were excluded from the calculation in Exhibit 2 because they have fewer pathways to meet at least one of the study's criteria for classification as disadvantaged. If more SRCL schools than non-SRCL schools have missing data, then the inclusion of schools with missing data could make SRCL schools seem less disadvantaged, simply because they had fewer pathways to meet the criteria. Similarly, if more non-SRCL than SRCL schools have missing data. Including schools with missing data could make non-SRCL schools seem more disadvantaged.

For the study's definition of disadvantage, the percentage of non-SRCL schools in a state that are disadvantaged on at least one of the four measures is subtracted from the percentage of SRCL schools in the same state that are disadvantaged on at least one of the four measures, to obtain a single difference between SRCL and non-SRCL schools for that state; if this single difference is at least 15 percentage points, then that state is said to exhibit substantial targeting. As shown in Exhibit 2 in the report, the characterization of all states' targeting would remain the same if the threshold for substantial targeting was set to 20 percentage points, 25 percentage points, or 30 percentage points.

Exhibit C.1 presents the data from Exhibit 2 in the report in tabular form. Exhibit C.2 presents data for each of the four measures individually. Again, the samples used in these exhibits drop schools that are missing any of the four measures from the numerator, the denominator, and the construction of the quartiles. Exhibit C.3 quantifies the prevalence of missing data by measure and state.

To assess the sensitivity of the decision to base the analysis in Exhibit 2 on schools with data on all four indicators, Exhibit C.4a replicates Exhibit 2 but with schools missing some but not all measures included in the sample. Exhibit C.4b presents the percentage of schools classified as disadvantaged on the percentage of students eligible for free or reduced-price lunch, with schools missing some but not all measures included in the sample.

Exhibit C.5 replicates Exhibit 2 in the report (and Exhibit C.6 replicates Exhibit C.1) using states' own criteria for determining whether schools are disadvantaged, which may or may not overlap exactly with the common criteria the study used to assess school disadvantage across states. While the analysis in Exhibit 2 is the primary analysis, because it assesses all states by the same criteria, the analysis in Exhibit C.15 sheds light on whether states implemented the decision rules outlined in their grant applications. Exhibit B.42 presents states' own thresholds for classifying schools as disadvantaged when reviewing subgrant applications, as determined by document reviews and interviews with state grant administrators. Some states used these thresholds as *eligibility requirements*, meaning that only applicants that met these criteria would be considered for subgrants; others used them as *competitive preference priorities*, meaning that applicants that met these criteria would be more likely to receive subgrants, but applicants that did not meet these criteria would still be considered. Georgia and North Dakota had more complex criteria and were excluded from the analysis in Exhibit 5.

Exhibit B.42. States' criteria for classifying schools as disadvantaged

State	Criteria
Georgia	First cutoff was requirement to get at least 75 points as a district to move on to points for competitive priorities. Assigned competitive priority points based on rank ordering of the schools on 2 factors: use of new state mandated assessment; poverty. Had a matrix that laid out schools from high poverty/low growth, all the way to low poverty/high growth, and created a rank ordering system based on that, and then the study team assigned points to segments of those schools. The most possible points being 20 all the way down to 1. District has to establish a need based on Georgia Department of Education analysis of third-grade reading growth for students who are economically disadvantaged, students with disabilities, and English learners.
Kansas	At least 49 percent of students eligible for free or reduced-price lunch, at least 10 percent of students classified as English learners, at least 15 percent of students classified as students with disabilities, or more than one of the above.
Kentucky	At least 65 percent of students eligible for free or reduced-price lunch.
Louisiana	The sum of the percentages of students eligible for free or reduced-price lunch, English learners, and students with disabilities in a school must be at least 50 percent.

State	Criteria
Maryland	The sum of the percentages of students eligible for free or reduced-price lunch, English learners, and students with disabilities in a school must be greater than the statewide mean of that sum.
Minnesota	At least 40 percent of students eligible for free or reduced-price lunch.
Montana	Two or more of the following criteria must be met: at least 50 percent of students eligible for free or reduced-price lunch, at least 20 percent of students classified as English learners, and at least 15 percent of students classified as students with disabilities.
New Mexico	At least 74 percent of students eligible for free or reduced-price lunch, at least 15 percent of students classified as English learners, at least 16 percent of students classified as students with disabilities, or more than one of the above.
North Dakota	At least 40 percent of students eligible for free or reduced-price lunch; additional points for higher percentages of Native American students at risk of school failure and higher percentage of students not meeting literacy standards in grades 3, 5, and 8 on state ELA assessments.
Ohio	At least 50 percent of students eligible for free or reduced-price lunch, at least 5 percent of students classified as English learners, at least 10 percent of students classified as students with disabilities, or more than one of the above.
Oklahoma	At least 50 percent of students eligible for free or reduced-price lunch, *and either* (a) at least 10 percent of students classified as English learners, or (b) at least 20 percent of students classified as students with disabilities, or both.

Source: Reviews of state requests for subgrant applications and interviews with state grant administrators. The rules for Georgia and North Dakota could not be replicated with available data and those states are excluded from Exhibit C.5, but Georgia and North Dakota are included in Exhibit B.42 for reference.

B.6. Analyses of Evidence Use

This appendix section provides additional detail on the evidence use analyses presented in the report, with supporting exhibits in Appendix C. It also provides details about other analyses related to evidence use presented in Appendix D.

The prevalence of evidence-based programs in SRCL-funded districts and schools was assessed in four ways in the report. Each of these analyses approaches the issue from a different angle. The sections that follow describe each of these analyses presented in Appendix C, along with related analyses presented in Appendix D.

All analyses are weighted as described in section B.1.2, except Exhibit 3 in the report, and Exhibits C.8 and C.10 in Appendix C. Weights are not salient for the analyses in Exhibits 3 and C.8 because they focus on the program as the unit of analysis rather than the district. Weights are not used in Exhibit C.10 because the focus is on the number of districts purchasing each program, and incorporating weights would result in fractional numbers, which would make the results more challenging to interpret. In all analyses of survey-reported program usage, a name standardization process was used to ensure that different spellings and abbreviations of the same program were recoded to the same name.

B.6.1. Literacy Programs Purchased by SRCL Districts

Given SRCL's emphasis on adopting programs supported by strong or moderate evidence, the report includes analyses that examine the evidence supporting newly-adopted programs purchased with SRCL grant funds. This first analysis characterizes the evidence supporting the 236 distinct programs purchased across all SRCL-funded districts (Exhibit 3 in the report and Exhibit C.8 in Appendix C) regardless of how many districts purchased them. The sample includes programs that districts indicated were used with all students (core curricula), as well as programs used only with struggling readers (supplemental programs). Each program is included in the calculation once, regardless of the number of districts that purchased it.

B.6.2. Literacy Program Purchases

Because some programs were purchased by more than one SRCL district, the second analysis takes into account the number of districts that purchased each of the 236 programs (Exhibit 4 in the report and Exhibit C.9 in Appendix C). This analysis characterizes the level of evidence supporting each district program purchase decision and assesses whether districts' decisions were aligned with SRCL's expectation to prioritize programs supported by strong or moderate evidence.

Specifically, Exhibit 4 in the report presents the percentage of program purchases in each evidence rating category. The unit of analysis is a district-program pair. For instance, if three districts in North Dakota purchased the same program, it would be counted three times in the associated evidence category. Similarly, if one district in Maryland purchased three programs, each with a different evidence rating, that district would be counted once in each of the three evidence categories in Exhibit 4. As in Exhibit 3 in the report, the analysis includes programs used with all students as well as programs used only with struggling readers. The percentages shown in Exhibit 4 were calculated separately by state, using the survey weights described in section B.1 to account for nonresponse. The "All" percentage presented in Exhibit 4 is the simple average of the 11 state-level percentages, as described in section B.4.

To provide context, Exhibit C.10 presents a plot of the frequency with which programs were purchased; a program that was purchased by three districts, for instance, is counted once in the bar with the value 3 on the horizontal axis.

B.6.3. Districts That Purchased Programs Supported by Strong or Moderate Evidence

The analyses discussed in the previous sub-sections describe the extent to which programs purchased with SRCL funds were supported by strong or moderate evidence, combining purchases across districts. But they do not shed light on how many districts purchased programs supported by strong or moderate evidence or how consistently each district did so.

Exhibits C.11 and C.12 in Appendix C present the results of additional analyses that focus on the district as the unit of analysis. The analysis in Exhibit C.11 assigns each district the highest evidence rating category (see section B.2) among the programs it purchased with SRCL funds. This analysis provides information on the percent of districts that purchased at least one program supported by strong or

moderate evidence. Exhibit C.12 replicates Exhibit C.11 but includes an additional category for districts that did not purchase any programs with SRCL funds.

B.6.4.Teachers' Use of Literacy Programs Supported by Evidence

The first three analyses are restricted to districts that purchased programs with SRCL funds, but some districts did not use their SRCL funds for this purpose. It is possible that districts already using programs supported by strong or moderate evidence opted to use their SRCL funds on other types of resources. The fourth analysis assesses the level of evidence supporting all literacy programs that teachers reported using, regardless of funding source or when it was first adopted. This provides a broader picture of the evidence supporting literacy programs used in SRCL schools.

This analysis counts each teacher once whether they reported using one program or multiple programs in their classroom. Exhibits C.13a and C.13b present the highest evidence rating category (see section B.2) among the programs teachers reported using on the teacher survey, separately for 2018–19 and 2019–20. For instance, if a teacher in Kentucky used two programs, one with promising evidence and one with a null or negative result, that teacher would be counted only in the "Promising evidence" section of the bar for Kentucky. Teachers who did not report using externally developed programs are excluded from the analysis. To explore whether teachers were more likely to use core or supplemental programs supported by strong or moderate evidence, Exhibit C.14a and C.14b separate the "All" column from Exhibits C.13a and C.13b into supplemental and core programs, separately for 2019–20 (Exhibit C.14a) and 2018–19 (Exhibit C.14b).

B.6.5. Additional Analyses

Two exhibits in Appendix C contain additional context for the evidence use analyses presented in the report. Exhibit C.16 presents the percentage of districts that considered various factors when choosing which programs to purchase with SRCL funds; these factors are not mutually exclusive, so districts could list as many factors as were relevant to their purchasing processes. Exhibit C.15 displays the percentage of districts that reported being offered technical assistance to help them select programs supported by strong or moderate evidence, broken down by whether they received the assistance and whether it was very, somewhat, or not useful. The label at the top of each bar displays the percentage of districts that were offered technical assistance; the remainder of the districts were not offered technical assistance.

Appendix D includes three additional analyses related to the use of literacy programs supported by strong or moderate evidence by SRCL districts and schools. Exhibit D.1 displays principals' reports of literacy programs, by the highest level of evidence, similar to Exhibit C.11. Next, the analysis in Exhibit D.2 applies a process for characterizing district-level use of evidence that is more stringent than the analysis in Exhibit C.11. This analysis shows the percentage of districts that purchased *only* programs supported by strong or moderate evidence.

The results in Exhibit C.11 indicate that there is variation across states in the percent of districts purchasing programs supported by strong or moderate evidence. One potential explanation is the number of programs purchased, because districts that purchased more programs would have had more opportunity to buy at least one program supported by strong or moderate evidence, even if by chance.

B.7. Analyses of Comprehensive Literacy Instruction

The analyses of comprehensive literacy instruction reported in Exhibits 5 and 6 in the report were based on the teacher survey administered in spring 2020, using the items discussed in section B.3 of this appendix. Because the measures of comprehensive literacy instruction differ depending on teachers' grade bands, it was necessary to assign each teacher who responded to the survey to a single grade band for the purposes of analysis. As such, teachers who taught in more than one of the grade bands (kindergarten, grade 1, grade 2, grade 3, grades 4–5, grades 6–8, and grades 9–12) were assigned to the grade band corresponding to the earliest grade they taught.[21]

Exhibit 5 in the report presents the survey-weighted percentage of teachers meeting zero to six benchmarks for features of comprehensive literacy instruction in 2019–20. Only teachers with no missing data on any of the six features were included in the analysis. For instance, a teacher in grades 4–8 or grades 9–12 who did not assign any novels with Lexile levels had missing data on the assignment of age-appropriate reading materials and was excluded from the analysis.

Exhibit 6 in the report presents the survey-weighted percentages of teachers in each grade band who met each benchmark in 2019–20. Teachers with missing data for a given feature were dropped from the analyses of that feature and from the analysis of the number of benchmarks met in Exhibit 5 in the report, but were not dropped from the sample for other features for which they have nonmissing data. Exhibits C.21–C.24 present these results side by side for 2018–19 and 2019–20, with separate figures for each grade band. The procedure described in section B.4 is used to test for differences across years or grade bands in the percentage of teachers meeting each benchmark.

B.8. Analyses of Achievement Trends

This appendix section provides additional detail on the achievement trends analysis described on page 11 of the report, including an overview of data sources, matching procedures, and analysis methods. The analysis compared trajectories in SRCL schools and a set of matched comparison schools in each SRCL state. To obtain the comparison schools, each SRCL school was matched to schools that did not receive SRCL funding in the same state, based on grade span, demographics, location, school size, and test scores in the years prior to the SRCL grant. In each of the 11 SRCL states, a comparative interrupted time series (CITS) model was estimated for the years 2014–15 through 2018–19, comparing student standardized test scores in SRCL schools and matched comparison schools during the 4 years preceding the disbursement of SRCL funds and the first year of the SRCL program.

B.8.1. Data

For 9 of the 11 SRCL states, the CITS analysis is based on administrative data, with one observation per student per year. The administrative datasets list the student's grade, race, gender, and eligibility for free or reduced-price lunch; whether the student is an English learner or has a disability; and the

[21] As described in section B.3, the benchmarks for emphasis on key domains and explicit instruction were assessed by comparing teacher responses with the responses of experts. Experts provided separate responses for teachers in pre-K, kindergarten, grade 1, grade 2, grade 3, grades 4–5, grades 6–8, and grades 9–12. Six teachers taught in more than one of these bands, and they were dropped from the analysis.

school that the student attended at the time of achievement testing, as well as scale scores on the English language arts and math tests, though the latter were not analyzed. The analysis includes data from 2014–15 through 2018–19; the observations from 2018–19 correspond to the period in which SRCL schools received SRCL funds.[22]

All achievement scale scores were normalized within grade, subject, year, and state to have a mean of zero and a standard deviation of 1, using a linear transformation.

In eight of the states participating in the study, student-level data were provided by the respective state education agencies; data from Kentucky were provided by the Kentucky Center for Statistics. For 2 of the 11 SRCL states—Kansas and North Dakota—state education agencies do not share individual-level data with researchers outside their departments. Rather than student-level data, Kansas and North Dakota provided the number of students at each of the four proficiency levels by school, grade, year, and subject. These proficiency distributions were provided for all students and for populations of interest, such as students eligible for free or reduced-price lunch, English learners, and students with disabilities. The ordered probit procedure outlined by Reardon et al. (2017) was used to convert these proficiency distributions into school average scores, assuming a scale with a mean of 0 and a standard deviation of 1. No student-level covariates were included in the model for these states; the CITS for these states controlled for school covariates only.

B.8.2. Matching Procedure

Relying on student-level administrative data for the universe of public schools in SRCL states for which student-level data were available, plus school-level data for the universe of public schools in SRCL states that did not provide student-level data, the study team used propensity score matching methods to identify a set of comparison schools similar to the SRCL schools in demographics, geography, size, and student achievement during the pre-SRCL years.

In principle, comparison schools could be chosen from among unfunded schools in subgrantee districts or schools in districts that did not receive SRCL funds. Although the first of these options would ensure that matched comparison schools share the same district context as SRCL schools, many subgrantee districts chose to fund all or nearly all schools. Thus, potential comparison schools were drawn from districts that did not receive SRCL funds.[23]

Within each state, for each SRCL school, a set of potential comparison schools was identified based on each school's propensity score, which represents a school's predicted probability of being in the SRCL

[22] If a student had multiple test scores in the same school, grade, year, and subject, the higher score was kept. If a student had test scores from multiple grades or schools in the same year and subject, both scores were dropped.

[23] Schools in Maryland were excepted from this rule; all traditional public school districts in Maryland received SRCL funds, so there are no unfunded districts available as matches. Treatment schools in Maryland are instead matched to unfunded schools in districts that received SRCL subgrants, though matched pairs do not have to be in the same district.

group.[24] Additionally, schools were classified as elementary schools, middle schools, or other grade configurations, based on the lowest and highest grades offered in the respective school; SRCL schools could only be matched to comparison schools in the same grade configuration category. The propensity score was estimated using a separate logit regression in each state with the following independent variables:

- number of students enrolled in grades K–12 in the school;
- percentage of students in the school who are eligible for free or reduced-price lunch;
- indicator variables for the school's one-digit urban-centric locale code (urban, suburban, town, or rural);
- percentage of students in a school who are Black (with no other race reported, and not Hispanic);
- percentage of students in a school who are Hispanic (regardless of race);
- percentage of students in a school who are Native American (with no other race reported, and not Hispanic);
- percentage of students in a school who are Asian (with no other race reported, and not Hispanic);
- percentage of students in a school who are English learners;
- percentage of students in a school classified as students with disabilities; and
- school average state standardized ELA test score across all tested grades from the 2015–16 academic year.

Matching was conducted using a propensity score caliper of 0.2 standard deviations. That is, for each SRCL school, all nonfunded schools with a propensity score within 0.2 standard deviations of the SRCL school's propensity score were considered a potential match. Up to five comparison schools were selected for each SRCL school. If a SRCL school had no potential comparison schools within its propensity score caliper, it was dropped from the CITS analysis sample. Matching was carried out without replacement, meaning that each potential comparison school could be matched to only a single SRCL school. Exhibit B.43 presents the results of the school matching process in each state in the study. In most states, with the exceptions of Maryland, North Dakota, and Ohio, very few schools were dropped from the matching process because they did not have a comparison school within 0.2 standard deviations of them. Additionally, in most states, almost all schools that had a comparison school within 0.2 standard deviations of them were able to complete the matching process; as expected, in Maryland, where the number of available comparison schools was relatively low because so many schools were funded, many SRCL schools were unable to find matches. As a result, in the states where fewer SRCL schools matched successfully, including Maryland, North Dakota, and Ohio, the SRCL schools in the sample may not be

[24] Schools in all unfunded districts were included as potential matches, although some unfunded districts might not have been suitable candidates for SRCL subgrants because they do not serve many disadvantaged students. Schools in these districts would be unlikely matches for SRCL schools based on propensity score matching. Selecting comparison schools from unfunded districts will avoid differences between SRCL and comparison schools that are attributable to within-district selection, and selecting based on propensity scores should reduce differences in the observable characteristics of students enrolled.

representative of the full population of SRCL schools in the respective states, and results for those states may not generalize effectively to each respective state as a whole.

Exhibit B.43. Numbers of matched and unmatched SRCL schools and matched comparison schools, by state

State	All SRCL Schools	SRCL Schools With No Matches in Caliper	SRCL Schools With Potential Matches	SRCL Schools Matched Successfully	Percentage Matched	Matched Comparison Schools
Georgia	175	0	175	175	100%	621
Kansas	120	15	105	87	73%	295
Kentucky	101	4	97	91	90%	273
Louisiana	99	6	93	76	77%	208
Maryland	835	135	700	139	17%	156
Minnesota	35	0	35	34	97%	141
Montana	27	0	27	27	100%	88
New Mexico	57	1	56	44	77%	122
North Dakota	63	21	42	21	33%	34
Ohio	215	98	117	97	45%	351
Oklahoma	46	6	40	36	78%	101

The matching algorithm matched the most similar schools first; that is, given a set of SRCL schools $\{t_1, \ldots, t_N\}$ and a set of potential comparison schools $\{c_1, \ldots, c_M\}$, if out of all possible matched pairs, the pair (t_j, c_i) has the smallest difference in propensity scores, t_j was matched to c_i, c_i could not be chosen as a match for any other SRCL school, and t_j could not be matched to another unfunded school until all other SRCL schools in $\{t_1, \ldots, t_N\}$ were either matched or dropped.[25] Once every SRCL school was matched to one unfunded school or dropped for having no matches available, the process was repeated to allow each SRCL school a second match, a third match, and so on, until all unfunded schools within a caliper of any SRCL school were matched.

Exhibit B.44 presents concordance statistics on the propensity-score matches. For each state, the exhibit presents the percentage of schools that received SRCL funds, the percentage of schools with predicted probabilities of treatment above 0.5 that received funds, and the percentage of schools with predicted probabilities of treatment less than or equal to 0.5 that did not receive funds. It also presents parallel results using the state-specific proportion of schools receiving funding as the cutoff for the probability of treatment, rather than 0.5. In all states except Maryland, more than 85% of schools with

[25] If a SRCL school was included in the matching process but all the comparison schools in its caliper matched to other SRCL schools first, then the school in question ended up with no available matches and was dropped from the sample. This happened frequently in Maryland because more than half of Maryland schools were funded, and thus there were not enough comparison schools to match each SRCL school. This sometimes happened in other states as well. See Exhibit A.2.

low predicted probabilities of treatment were not funded, regardless of whether 0.5 or the proportion of funded schools was used as the threshold value. In six states, more than 80% of schools with predicted probabilities of treatment above 0.5 were funded. The states in which this percentage was low were states that funded relatively small proportions of schools. Overall, the propensity-score model is effective at filtering out schools that are unlikely to be awarded SRCL funding, while it also demonstrates that not every school with the most common characteristics of SRCL schools necessarily received SRCL funds, which leaves a plausible pool of comparison schools that are similar to the SRCL schools in terms of demographics and student achievement.

Exhibit B.44. Concordance statistics for propensity-score model

State	Pct. Funded	p = 0.5		p = Proportion Funded	
		Accurate Positives	Accurate Negatives	Accurate Positives	Accurate Negatives
Georgia	12%	100%	88%	23%	95%
Kansas	18%	83%	87%	34%	90%
Kentucky	15%	53%	86%	29%	95%
Louisiana	15%	100%	85%	24%	92%
Maryland	82%	87%	76%	92%	36%
Minnesota	4%	n/a	96%	11%	98%
Montana	9%	20%	91%	24%	97%
New Mexico	14%	15%	86%	18%	94%
North Dakota	50%	87%	86%	87%	86%
Ohio	21%	86%	91%	62%	94%
Oklahoma	6%	0%	94%	17%	99%

Note: p refers to the predicted probability that a school receives SRCL funds, based on that state's propensity score model. "Accurate positives" refers to the percentage of schools with predicted probabilities of SRCL receipt greater than the indicated value of p (0.5, or the proportion that received SRCL funds). "Accurate negatives" refers to the percentage of schools with predicted probabilities of SRCL receipt less than or equal to the indicated value of p (0.5 or the proportion that did not receive SRCL funds). No schools in Minnesota had predicted probabilities above 0.5; fewer than 10 schools in Georgia, Louisiana, Montana, and Oklahoma had predicted probabilities above 0.5.

Exhibit B.45 presents baseline equivalence information on the variables included in the outcome analyses. For each student-level variable, the Treatment Mean and Comparison Mean columns present the average of the state-level average values of that variable among students who were enrolled in a SRCL school in 2017-18 and enrolled in a comparison school that year. The means of student-level variables are adjusted for school random effects, and comparison observations are weighted by the number of comparison schools matched to the same treatment school; for instance, if treatment school A matched to comparison schools B, C, and D, all observations in schools B, C, and D would be weighted by one third. For each school-level variable, the Treatment Mean and Comparison Mean columns present the average of the state-level average values of that variable among SRCL schools in

2017–18 and comparison schools in that year; comparison observations are weighted by the number of comparison schools matched to the same treatment school. The Difference column subtracts the comparison mean from the treatment mean, while the Effect Size column presents the average of the state-level differences that were standardized into Hedges' *g* terms (for continuous variables) or Cox's index terms (for binary variables). Kansas and North Dakota are not included in calculations involving student-level variables. All effect sizes at baseline are below 0.25, putting them in the acceptable range for statistical adjustment.

Exhibit B.45. Baseline equivalence information for students in trend analysis sample

Variable	Treatment Mean	Comparison Mean	Difference	Effect Size
Student standardized test score	-0.104	-0.091	-0.013	-0.016
Student is Black	0.207	0.182	0.025	0.074
Student is Hispanic/Latino	0.205	0.187	0.018	0.080
Student is Native American	0.066	0.075	-0.009	0.032
Student is Asian	0.026	0.022	0.004	-0.005
Student is eligible for free or reduced-price lunch	0.714	0.697	0.017	0.082
Student is in special education	0.149	0.145	0.004	0.018
Student is an English learner	0.105	0.098	0.007	0.068
School is in an urban locale	0.198	0.164	0.034	0.134
School is in a town locale	0.260	0.278	-0.018	-0.090
School is in a rural locale	0.303	0.322	-0.019	-0.055
School percentage of students with disabilities	0.152	0.149	0.003	0.051
School percentage of English learners	0.101	0.095	0.006	0.073
School percentage of students eligible for free or reduced-price lunch	0.650	0.647	0.003	0.020
School percentage Black	0.174	0.157	0.017	0.090
School percentage Hispanic/Latino	0.247	0.183	0.064	0.069
School percentage Native American	0.059	0.067	-0.008	-0.020
School size	469.2	465.5	3.7	0.011

Exhibit B.46 duplicates Exhibit B.45, but for disadvantaged students only. School-level variables are not shown in the table because all schools in the CITS had at least one disadvantaged student, so the averages of school-level variables are identical to those shown in Exhibit B.45. As in the full sample, all baseline differences among disadvantaged students have effect sizes below 0.25, putting them in the acceptable range for statistical adjustment.

Exhibit B.46. Baseline equivalence information for students in trend analysis sample, disadvantaged students only

Variable	Treatment Mean	Comparison Mean	Difference	Effect Size
Student standardized test score	-0.264	-0.261	-0.003	-0.005
Student is Black	0.225	0.207	0.018	0.076
Student is Hispanic/Latino	0.226	0.209	0.017	0.062
Student is Native American	0.069	0.077	-0.008	0.021
Student is Asian	0.025	0.023	0.002	-0.020
Student is eligible for free or reduced-price lunch	0.928	0.923	0.005	0.103
Student is in special education	0.218	0.221	-0.003	-0.005
Student is an English learner	0.127	0.121	0.006	0.048

B.8.3. Analysis Methods

The team conducted separate grade-specific CITS analyses for each state, for students in grades 3–8.

States with student-level data. The following model was used for states that provided student-level data. The model takes into account the clustering of students within years and years within schools:

$$Y_{its} = \alpha + \beta Time_t + \gamma Treat_s + \varphi \mathbb{I}(t = 2019) + \theta Time_t Treat_s + \delta Treat_s \mathbb{I}(t = 2019) + \boldsymbol{X}_i'\Theta + \boldsymbol{Z}_s'\Psi + \omega_s + \nu_{st} + \varepsilon_{its}$$

where

Y_{its} is the standardized ELA test score for student i in year t at school s.

$\mathbb{I}(\cdot)$ represents the indicator function, taking the value 1 if the expression in parentheses is true and 0 if it is false.

Time is a continuous measure of time, equal to t - 2015.

Treat takes value 1 if a school is a SRCL school and 0 otherwise.

$\boldsymbol{X}$ is a vector of student-level covariates. The set of student covariates $\boldsymbol{X}$ includes indicators for whether the student is Black, Hispanic, Native American, Asian, female, eligible for free or reduced-price lunch, enrolled in special education, and classified as an English learner.

$\boldsymbol{Z}$ is a vector of school-level covariates. The set of school covariates $\boldsymbol{Z}$ includes school size; urbanicity; and the percentage of students who are Black, Hispanic, Native American, Asian, English learners, enrolled in special education, and eligible for free or reduced-price lunch.

Errors are decomposed into school random effects ω, school-by-year random effects ν, and residual components ε.

In the specification, the coefficients for time (β), treatment (γ), and the interaction of time and treatment (θ) represent the pre-interruption linear trend, which may differ for treatment and control schools. The coefficient for time=2019 (φ) represents the post-interruption deviation from the linear trend for the control group. The coefficient for the interaction of time=2019 with treatment (δ), represents the difference in the post-treatment deviation between the treatment and control - the

treatment effect. The model represents a difference in difference design, assuming a linear pre-interruption trend.

Student-level observations were weighted based on the number of comparison schools matched to a single SRCL school. If one SRCL school was matched to *J* comparison schools, observations in that SRCL school received a weight of 1, whereas observations in each comparison school matched to that SRCL school received a weight of $\frac{1}{J}$. This ensures that the comparison sample is similar in the distribution of propensity scores to the SRCL sample, barring large fluctuations in school size.

States that did not provide student-level data. For Kansas and North Dakota, each observation is the school-level average state standardized ELA examination score in a given year; scores are normalized using the mean and standard deviation of student scores for the corresponding year, grade, subject, and state. Student-level demographic controls, such as race, eligibility for free or reduced-price lunch, and English learner status, are replaced with their school-level averages, such as the percentages of students in the given school who are Black, Hispanic, eligible for free or reduced-price lunch, and English learners. Each SRCL school received a weight of 1, whereas each comparison school received a weight equal to the inverse of the number of comparison schools matched to the same SRCL school ($\frac{1}{J}$, in the framework with one SRCL school matched to *J* comparison schools).

Pooling analyses across states. The analyses generated 66 estimates, one for each grade in each state that received SRCL funds. A fixed-effects meta-analysis was used to synthesize the state-by-grade-level findings into a single overall effect size. Although the primary estimate of interest is the overall relationship between SRCL and student English test scores across grades and states, similar syntheses estimate this relationship in each grade (across states) and in each state (across grades). The analysis was based on the fixed-effect meta-analysis procedure described in the *What Works Clearinghouse Procedures Handbook, Version 4.1* (pp. H2–H3), which assigns each state-by-grade-level finding a weight equal to the reciprocal of the variance of the effect size, and the meta-analytic average is equal to the sum of each finding multiplied by its weight divided by the sum of the weights. The standard error of this meta-analytic average is equal to the square root of the reciprocal of the sum of weights. Following the convention in Pigott and Polanin (2020) and Hedges et al. (2010), robust variance estimation was used to account for the dependence in the effect sizes because the same schools appear in several grade-by-state datasets.

APPENDIX C. SUPPLEMENTAL EXHIBITS AND INFORMATION ON STUDY FINDINGS

This appendix includes supporting statistical details and additional analyses related to the targeting of Striving Readers Comprehensive Literacy (SRCL) funds to districts and schools with high percentages of disadvantaged students (section C.1), use of evidence in selecting programs to be funded through SRCL grants (section C.2), and evidence of comprehensive literacy instruction in SRCL classrooms (section C.3).

C.1. Targeting

The analyses in Exhibit 2 in the report define a school as "disadvantaged" if it is in the highest quartile in its state for the percentage of students who are eligible for free or reduced-price lunch, the percentage of students who are English learners, the percentage of students who have disabilities, or any combination of the three, or if it is in the lowest quartile in its state for average standardized English scores. Thus, for each of the study's four measures, 25 percent of schools in each state are classified as disadvantaged. This implies that for each measure of disadvantage, the average of the percentage disadvantaged for SRCL schools and non-SRCL schools in a state, weighted by the proportion of schools in the state that received SRCL funds, is 25 percent. The exception is New Mexico. In New Mexico, 100 percent of students are eligible for free or reduced-price lunch in more than 25 percent of the schools. Thus, in New Mexico, any school in which 100 percent of students are eligible for free or reduced-price lunch is counted as being in the "highest quartile" for the purposes of this analysis.

Because schools are classified as disadvantaged if they are in the relevant (top or bottom) quartile on one or more of the study's measures, and the four measures are not perfectly correlated, more than 25 percent of the schools in each state are classified as disadvantaged in total. The exact percentage of all schools in a state classified as disadvantaged varies by state, depending on the correlation among the study's four measures of disadvantage. For instance, if most English learners in state A qualify for free or reduced-price lunch, while fewer English learners qualify in state B, the same schools that are classified as disadvantaged due to English learner status might be classified based on the basis of eligibility for free and reduced-price lunch in State A, while different schools might be classified as disadvantaged on each criterion in state B, which would result in state B having more total schools classified as disadvantaged. If a state prioritized the targeting of disadvantaged students, disadvantaged schools should be more likely than non-disadvantaged schools to receive SRCL funding, and the percentage of SRCL schools that are disadvantaged should be higher than the percentage of non-SRCL schools that are disadvantaged.

Exhibit C.1 presents the results of Exhibit 2 in the report in a tabular format, and also lists the percentage of schools in each state that are excluded from the calculation due to missing data.

Exhibit C.1. Percentage of SRCL and non-SRCL schools in the most disadvantaged quartile on one or more indicators, 2018–19, by state

Grantee	Percentage in Most Disadvantaged Quartile on One or More Indicators: SRCL Schools	Percentage in Most Disadvantaged Quartile on One or More Indicators: Non-SRCL Schools	Difference	Percentage of Schools Missing Data on One or More Measures
Georgia	70%	60%	10 p.p.	12%
Kansas	62%	48%	14 p.p.	35%
Kentucky	67%	54%	13 p.p.	27%
Louisiana	63%	58%	5 p.p.	32%
Maryland	51%	54%	-3 p.p.	3%
Minnesota	86%	47%	39 p.p.	30%
Montana	96%	52%	44 p.p.	43%
New Mexico	96%	62%	34 p.p.	30%
North Dakota	51%	58%	-7 p.p.	66%
Ohio	83%	41%	42 p.p.	55%
Oklahoma	83%	51%	32 p.p.	36%
All	73%	53%	**20 p.p.**	**33%**

Note: p.p. = percentage points. Schools missing data on any of the four measures were excluded from all calculations to create a consistent sample across analyses.
Source: 2018–19 Common Core of Data and 2015–16 Stanford Education Data Archive. The unweighted sample size is 7,972 schools.

Exhibit C.2 presents the percentages of schools in each state that are disadvantaged on each of the single measures separately, as well as on the study's overall disadvantage definition, based on the population of schools with data on all four characteristics. In general, states that exhibit substantial targeting overall also exhibit substantial targeting on several of the single measures, while states that do not exhibit substantial targeting overall exhibit substantial targeting on one or fewer of the single measures and negative targeting on at least one. As shown in Exhibit C.2, all states that do not exhibit substantial targeting according to the study's definition have at least one measure on which SRCL schools are less likely to be disadvantaged than non-SRCL schools are. Only one state with substantial targeting according to the study's definition, New Mexico, has any measures on which SRCL schools are less likely to be disadvantaged than non-SRCL schools are. While SRCL schools in New Mexico are slightly less likely than non-SRCL schools to be in the top quartile in the state in terms of their percentage of students with disabilities, almost all SRCL schools in New Mexico are in the top quartile in terms of their percentage of students eligible for free or reduced-price lunch, meaning that overall targeting in New Mexico still meets the study definition of substantial.

Exhibit C.2. Percentages of SRCL schools and non-SRCL schools identified as being disadvantaged, by measure and state

State	FRPL		ELs		SWD		Test Scores		Study Definition of Disadvantage			
	SRCL	Non-SRCL	SRCL	Non-SRCL	SRCL	Non-SRCL	SRCL	Non-SRCL	SRCL	Non-SRCL	Diff.	Targeting
GA	33%	24%	23%	25%†	23%	25%†	29%	23%	70%	60%	10 p.p.	Some
KS	29%	24%	46%*	20%	19%	26%†	30%	23%	62%	48%	14 p.p.	Some
KY	39%*	23%	24%	25%†	33%	24%	33%	23%	67%	54%	13 p.p.	Some
LA	36%	23%	16%	26%†	27%	25%	32%	23%	63%	58%	5 p.p.	Some
MD	21%	42%†	29%*	6%	22%	37%†	22%	32%†	51%	54%†	-3 p.p.	Negative
MN	66%*	23%	57%*	24%	29%	25%	54%*	23%	86%*	47%	39 p.p.	Substantial
MT	56%*	22%	33%	24%	56%*	22%	52%*	21%	96%*	52%	44 p.p.	Substantial
NM	95%*	47%	32%	24%	23%	25%†	18%	25%†	96%*	62%	34 p.p.	Substantial
ND	30%	19%	30%	19%	17%	32%†	22%	26%†	51%	58%†	-7 p.p.	Negative
OH	72%*	12%	41%*	21%	48%*	19%	68%*	12%	83%*	41%	42 p.p.	Substantial
OK	36%	24%	64%*	23%	33%	24%	50%*	23%	83%*	51%	32 p.p.	Substantial
All	47%*	26%	**36%**	22%	**30%**	26%	**37%**	23%	**73%***	**53%**	**20 p.p.**	**Substantial**

Note: Analysis includes schools with data on all four characteristics used to determine disadvantage. A state exhibits substantial targeting, denoted with an asterisk in the "SRCL" column, when SRCL schools are at least 15 percentage points more likely to be disadvantaged than non-SRCL schools; a state exhibits some targeting when SRCL schools are more likely to be disadvantaged than non-SRCL schools, but by less than 15 percentage points; a state exhibits negative targeting, denoted with a dagger in the "non-SRCL" column, when non-SRCL schools are more likely to be disadvantaged than SRCL schools.
Source: 2018-19 Common Core of Data and 2015-16 Stanford Education Data Archive. The unweighted sample size is 7,972 schools.

There are several states that did not demonstrate substantial targeting on the study's overall measure, including Kansas, Maryland, and North Dakota, which did substantially target funds to schools with higher percentages of English learners. However, these states did not substantially target schools with higher percentages of students eligible for free or reduced-price lunch, higher percentages of students with disabilities, or lower average test scores. In some cases, the percentage of SRCL schools classified as disadvantaged on some of these measures is lower than the percentage of non-SRCL schools, such as schools with high percentages of English learners in Kansas, schools with low average test scores in Maryland, and schools with high percentages of students with disabilities in North Dakota. As such, the

overall targeting of disadvantaged schools in those states does not meet the study's definition of substantial.

Another question that can be asked about the study's approach to the analysis of state targeting of SRCL funds is whether restricting the sample to schools with data on all four measures has an appreciable effect on the results. Information about the patterns of missing data by measure and state is found in Exhibit C.3, which presents the percentage of schools in each state that are missing data on each of the four measures as well as overall. Overall, about one third of schools in the grantee states are missing data on at least one of the measures of disadvantage. The rates of data missingness are highest in North Dakota (66%) and Ohio (55%).

Exhibit C.3. Percentages of schools missing data required for measuring targeting, by measure and state

State	FRPL	ELs	SWD	Test Scores	One or More Measures
Georgia	1%	9%	1%	4%	12%
Kansas	0%	30%	1%	12%	35%
Kentucky	0%	26%	1%	2%	27%
Louisiana	< 0.5%	26%	3%	10%	32%
Maryland	0%	< 0.5%	< 0.5%	3%	3%
Minnesota	0%	27%	1%	6%	30%
Montana	3%	1%	1%	43%	43%
New Mexico	1%	20%	14%	15%	30%
North Dakota	0%	58%	2%	31%	66%
Ohio	25%	40%	2%	6%	55%
Oklahoma	0%	31%	1%	6%	36%
All	**5%**	24%	**2%**	**9%**	**33%**

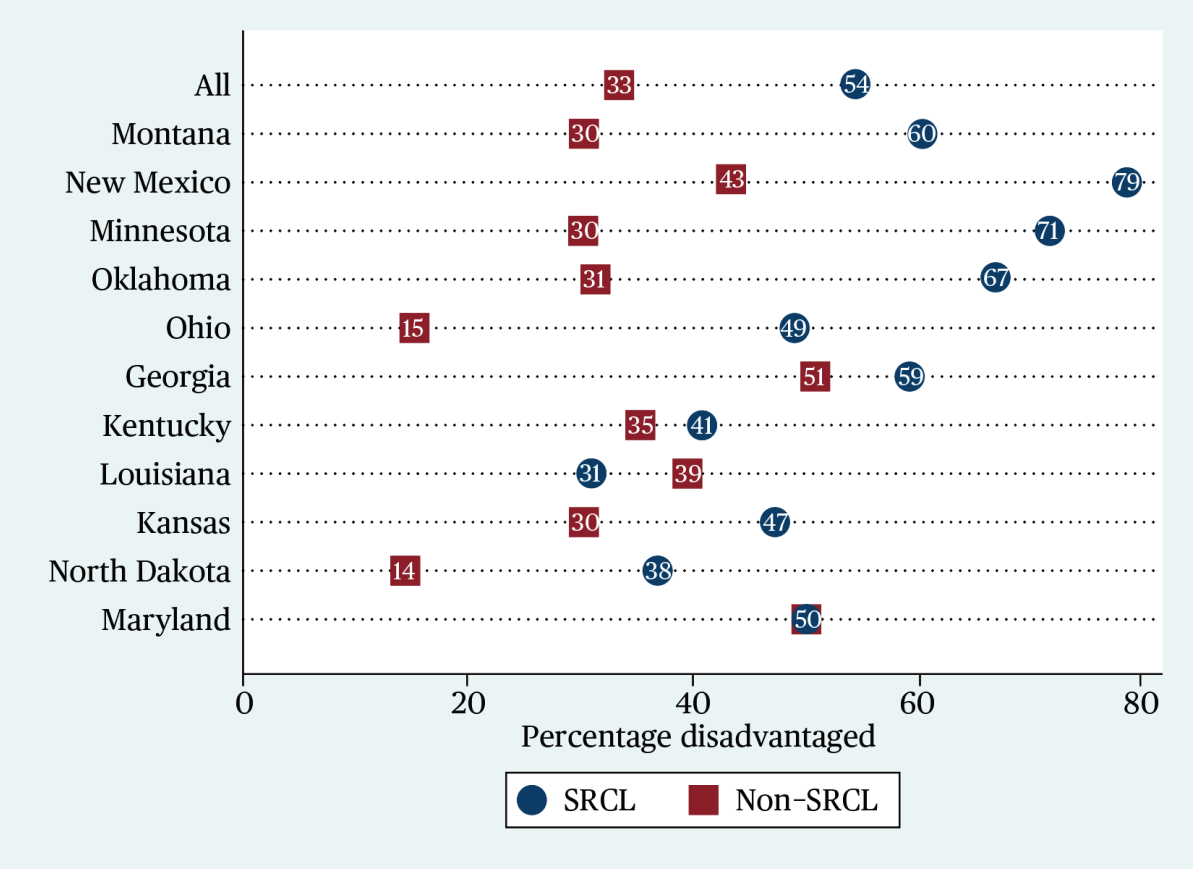
Source: 2018-19 Common Core of Data and 2015-16 Stanford Education Data Archive. The unweighted sample size is 11,816 schools.

The analyses of targeting in Exhibit 2 in the report and Exhibit C.1 is based on a common sample, in which only schools with data on all four of the study's measures of disadvantage were included. Schools with any missing data were excluded so that each school in the analysis would have the same number of chances to be classified as disadvantaged. As shown in Exhibit C.1, the percentage of schools missing data on at least one of the four measures varied across states, from 3 percent in Maryland to 66 percent in North Dakota. Although an approach that includes schools missing data on one or more measures would include more schools in the analysis, it would also lead to differences across states in the average number of measures included, and thus the expected percentage of schools classified as disadvantaged. Including schools with only some of the measures may also lead to differences in the number of measures included for SRCL and non-SRCL schools.

To estimate how sensitive the results in Exhibit 2 in the report are to the exclusion of schools missing one or more of the measures of disadvantage the study used, Exhibit C.4a below replicates the analysis

in Exhibit 2 but with all schools with data on at least one measure of disadvantage included in the sample.

Exhibit C.4a. Percentage of SRCL and non-SRCL schools in the highest quartile on at least one measure of disadvantage, by state, 2018–19, without excluding schools missing data



Note: A school is counted as disadvantaged if it is among the highest 25 percent of schools in its state on any of the four measures (average English language arts scores, percentage of students eligible for free or reduced-price lunch, percentage of English learners, or percentage of students with disabilities). States are ordered to match their order in Exhibit 2. Schools that are missing data on between zero and three of the measures are included in the sample. The percentage disadvantaged in Maryland is 50% among both SRCL and non-SRCL schools; accordingly, both markers for Maryland are plotted at the same place on the line, creating the appearance of a single marker.
Source: 2018-19 Common Core of Data, 2015-16 Stanford Education Data Archive and state education agency lists of SRCL-funded schools. The unweighted sample size is 11,816 schools.

Overall, the percentage of schools classified as disadvantaged is lower in Exhibit C.4a than in Exhibit 2. For example, in Exhibit C.4a, 54 percent of all SRCL schools are classified as disadvantaged, compared with 73 percent in Exhibit 2; and 33 percent of non-SRCL schools are classified as disadvantaged, compared with 53 percent in Exhibit 2. However, the average difference across states in percentage disadvantaged between SRCL schools and non-SRCL schools remains almost identical in percentage-point terms; the difference is 21 percentage points in Exhibit C.4a and 20 percentage points in Exhibit 2. Also, the general pattern across states is similar. In particular, Montana, New Mexico, Minnesota,

Oklahoma, and Ohio—the five states with substantial targeting in Exhibit 2 (a difference of at least 15 percentage points between SRCL and non-SRCL schools)—also show substantial targeting in Exhibit C.4a.

While the characterization of targeting in most states is similar regardless of whether schools with some missing data are included, there are some differences for specific states. The largest is for North Dakota, the state with the largest percentage of schools missing at least one measure (66 percent). In Exhibit 2, which includes only schools with complete data, the percentage of schools in North Dakota classified as disadvantaged is lower for SRCL schools than for non-SRCL schools. In Exhibit C.4a, when schools missing one or more of the four measures are included, the difference between SRCL schools and non-SRCL schools classified as disadvantaged is 24 percentage points (which exceeds the study's definition of substantial targeting). In addition, Kansas demonstrates substantial targeting when the schools with partial missing data are included in the sample but not when the sample is limited to schools with data on all four measures. Finally, Louisiana switches from positive targeting (but not substantial targeting) to negative targeting when schools with partial missing data are included.

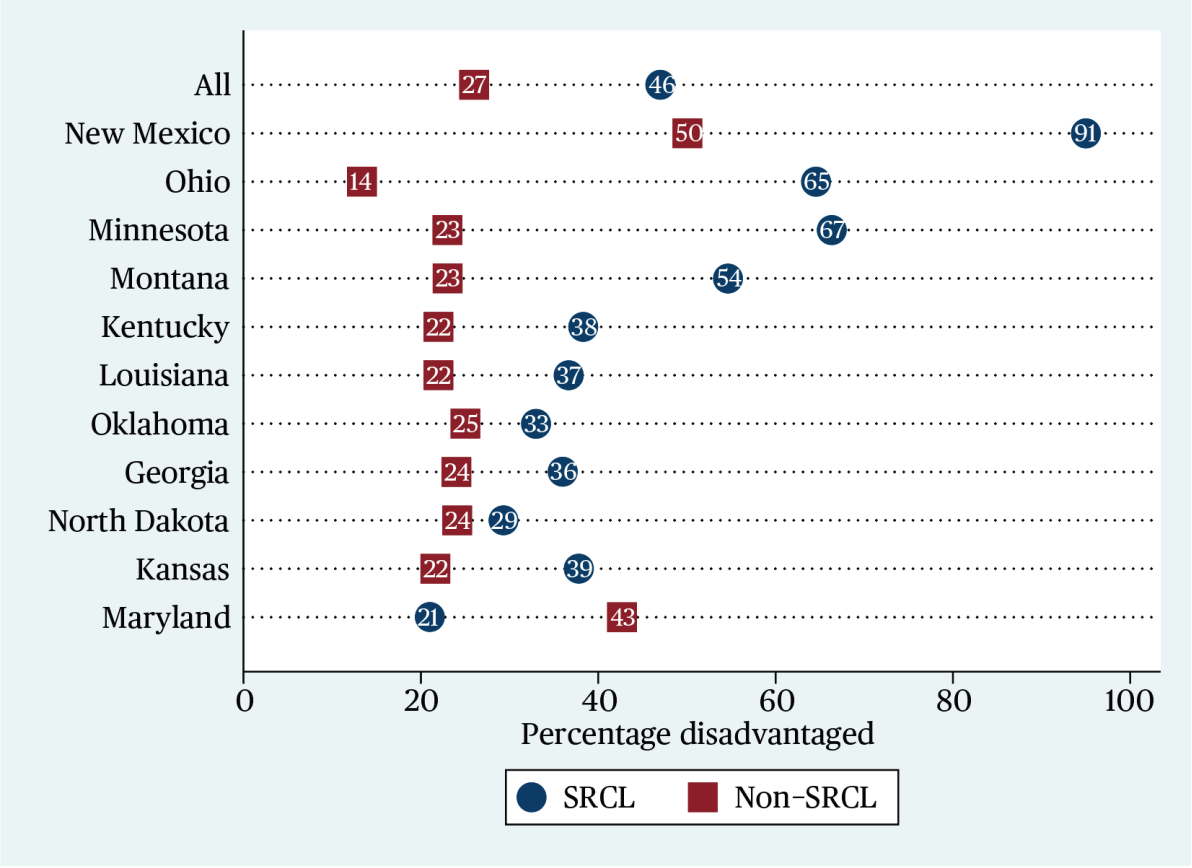
One additional question remains: whether schools that are missing data on some, but not all, of the four measures of disadvantage have a different probability of being disadvantaged on the measures for which they have data. If this were true, the targeting results would change based on the choice to exclude schools with some missing data. For a comparison between schools with some missing measures and no missing measures that does not advantage the schools with more measures available, we examine targeting patterns in a single measure in Exhibit C.4b, without restricting the sample to schools that also have data on all three of the other measures. While the probability of having at least one measure in the top quartile changes depending on the number of non-missing measures, a school should have an equal probability, 25%, of having any single measure in the top quartile regardless of whether its other measures are missing. If the results for a single measure change dramatically depending on whether schools missing other measures are included in the analysis, this uncovers a pattern in how data are missing. For instance, if the percentage of SRCL schools in a state that are disadvantaged on a particular measure increases when schools missing other measures are added, that implies that disadvantaged SRCL schools are more likely to be missing data. To explore the patterns of missing data, Exhibit C.4b presents the percentage of SRCL and non-SRCL schools in each state that are disadvantaged on the percentage of students eligible for free or reduced-price lunch, but with schools missing data on one or more of the other three measures included in the calculation. Eligibility for free or reduced-price lunch was chosen for this sensitivity analysis because it, or an equivalent measure of economic disadvantage, was included in every state's targeting plan and because most students who meet at least one category of disadvantage are eligible for free or reduced-price lunch.[26]

For the most part, the ordering of states in Exhibit C.4b matches the patterns in Exhibit C.2, indicating that the results in Exhibit C.2 are not affected substantially by the exclusion of schools with missing data. The main exception to the similarity in patterns is Kansas; when schools missing data on other indicators are excluded, 29% of SRCL schools would be classified as disadvantaged based on the

[26] Approximately 92% of students in SRCL states who are either eligible for free or reduced-price lunch, enrolled in special education, or classified as English learners, are eligible for free or reduced-price lunch. Of course, many students fit into more than one of these categories.

percentage of students eligible for free or reduced-price lunch, which would rank Kansas tenth out of 11 SRCL states. However, when schools missing data on other indicators are included, 39% of SRCL schools in Kansas would be classified as disadvantaged on the percentage of students eligible for free or reduced-price lunch, ranking Kansas fifth out of 11 SRCL states. In other words, disadvantaged SRCL schools are disproportionately more likely to be missing data in Kansas, but this does not appear to be true in other states.

Exhibit C.4b. Percentage of SRCL and non-SRCL schools in the highest quartile on the percentage of students eligible for free or reduced-price lunch, by state, 2018-19, without excluding schools missing data on other measures of disadvantage

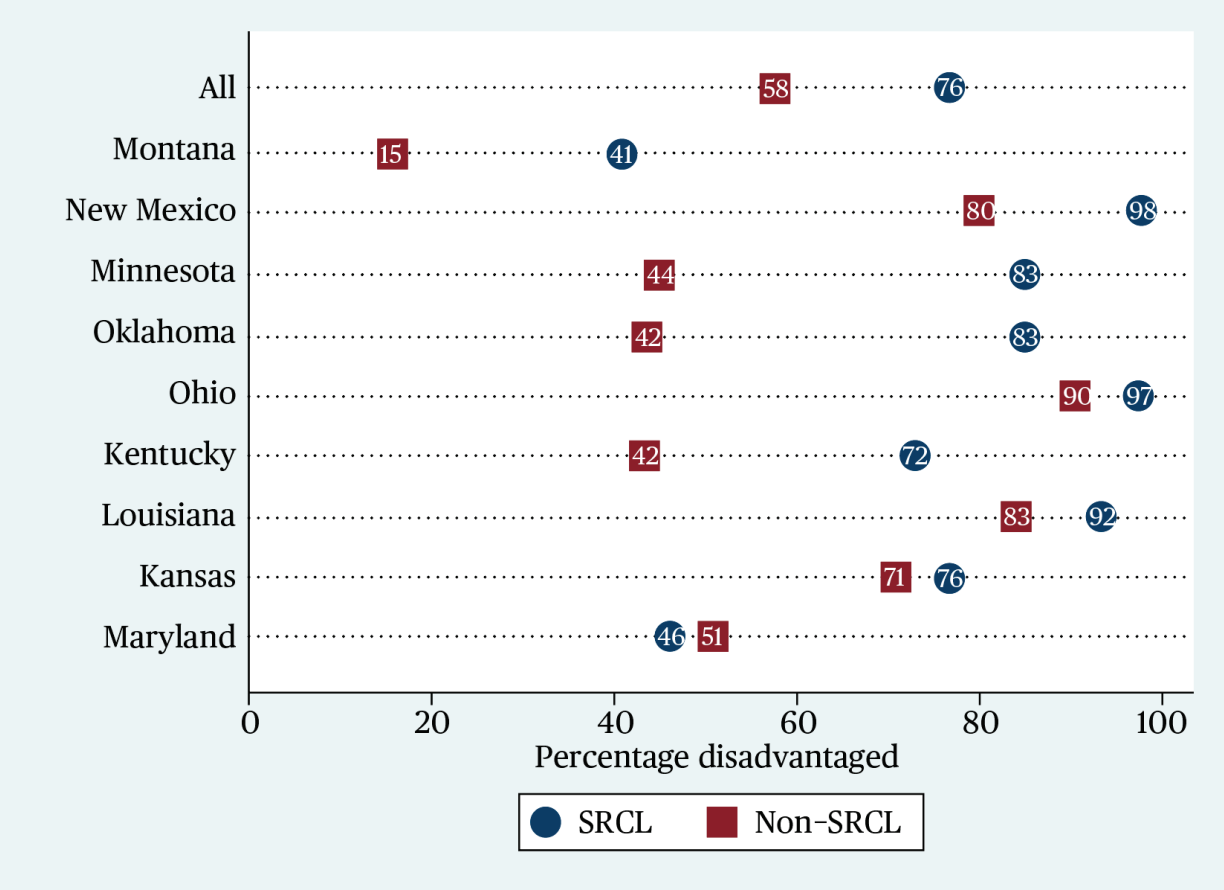


Note: A school is counted as disadvantaged if it is among the highest 25 percent of schools in its state on the percentage of students eligible for free or reduced-price lunch. States are ordered to match their order in Exhibit C.2. Schools that are missing data on between zero and three of the measures are included in the sample.
Source: 2018-19 Common Core of Data and state education agency lists of SRCL-funded schools. The unweighted sample size is 11,213 schools.

Although the SRCL program specified several categories of student disadvantage, it did not limit states to these categories, nor did it specify a minimum threshold of disadvantaged students served by funded schools. All definitions of disadvantage used in Exhibit 2 in the report and Exhibits C.1, C.2, and C.4a in the appendix are created for the purposes of the study and do not reflect federally mandated thresholds for defining or measuring disadvantage. States' definitions of disadvantaged schools varied

greatly, but the analyses in Exhibit 2 establish a standard definition to allow state-by-state comparisons. In contrast, Exhibit C.5 presents the percentages of SRCL and non-SRCL schools classified as disadvantaged under the states' own definitions. Exhibit C.6 presents supporting detail. Some states, such as Ohio and Louisiana, set thresholds under which most schools in the state were considered disadvantaged, so the percentage of disadvantaged SRCL schools differs little from the percentage of non-SRCL disadvantaged schools. Other states, including Montana, Minnesota, Oklahoma, and Kentucky, substantially targeted schools that met the state-specific definition of disadvantaged. Schools could not be classified based on Georgia and North Dakota's definitions of disadvantage with publicly available data, and those states are excluded from Exhibits C.5 and C.6. On balance, this sensitivity analysis using the state definitions leads to similar conclusions as those presented in Exhibit 2, except for Kentucky and Ohio. Kentucky achieved some targeting according to the study's measures but achieved substantial targeting when evaluated according to its own criteria. Ohio achieved substantial targeting according to the study's measures, but only some targeting based on its own criteria.

Exhibit C.5. Percentage of SRCL and non-SRCL schools classified as disadvantaged according to states' definitions, overall and by state, 2018–19



Note: p.p. is percentage points. A school is counted as disadvantaged if it meets the definition of disadvantaged used by its state agency in evaluating subgrant applications, as shown in Exhibit B.42. Schools missing data on any of the four measures (average English language arts scores, percentage of students eligible for free or reduced-price lunch, percentage of English learners, or percentage of students with disabilities) are excluded from the analyses. Sufficient data are not available in Georgia or North Dakota.

Source: 2018-19 Common Core of Data, 2015-16 Stanford Education Data Archive, and state education agency lists of SRCL-funded schools. The unweighted sample size is 6,345 schools.

Exhibit C.6. Percentage of SRCL and non-SRCL Schools classified as disadvantaged according to states' definitions, overall and by state, 2018-19, supporting detail

Grantee	Percentage Disadvantaged, SRCL	Percentage Disadvantaged, Non-SRCL	Difference	Percentage of Schools Missing Data on Measure
Kansas	76%	71%	5 p.p.	35%
Kentucky	72%	42%	30 p.p.	27%
Louisiana	92%	83%	9 p.p.	32%
Maryland	46%	51%	-5 p.p.	3%
Minnesota	83%	44%	39 p.p.	30%
Montana	41%	15%	26 p.p.	43%
New Mexico	98%	80%	18 p.p.	30%
Ohio	97%	90%	7 p.p.	55%
Oklahoma	83%	42%	41 p.p.	36%
All	**76%**	**58%**	**18 p.p.**	**32%**

Note: A school is counted as disadvantaged if it meets the definition of disadvantaged used by its state agency in evaluating subgrant applications, as shown in Exhibit B.42. To ensure a consistent sample across exhibits, schools missing data on any of the four measures included in Exhibit 2 in the report (average English language arts scores, percentage of students eligible for free or reduced-price lunch, percentage of English learners, or percentage of students with disabilities) are excluded from the analyses, regardless of which measures were included in a state's definition. "Percentage of schools missing data on measure" refers to the percentage of schools statewide missing data on the measure shown in the table, regardless of SRCL receipt or availability of data on other measures of disadvantage. p.p. = percentage points. Sufficient data are not available in Georgia or North Dakota.
Source: 2018-19 Common Core of Data, 2015-16 Stanford Education Data Archive, and state education agency lists of SRCL-funded schools. The unweighted sample size is 6,345 schools.

C.2. Evidence Use

Embedded in the SRCL program was a requirement for rigorous evidence to play a central role in selecting literacy programs and practices, specifically "strong or moderate evidence," as defined by ESSA. While the SRCL program did not require subgrantees to use SRCL funds to purchase literacy programs, 80 percent of all SRCL districts did purchase at least one program (Exhibit C.7).

This section includes analyses of the use of evidence that support and extend the analyses in the report and described in appendix sections B.6.1, B.6.2, B.6.3, and B.6.4. Section C.2.1 begins with contextual information about the percentage of districts that used SRCL funds to purchase literacy programs, followed by the percentage of programs supported by strong or moderate evidence, and the number of programs purchased by more than one district. Next, section C.2.2 focuses on district program purchases; that is, how many times a given program was purchased by SRCL districts. Section C.2.3 then turns to analyses of the percentage of districts purchasing at least one literacy program supported by strong or moderate evidence. Then, at the teacher level, section C.2.4 examines the percentage of

programs teachers reported using that are supported by strong or moderate evidence, overall and by core and supplemental programs (see appendix section B.2 for a description of the study's approach to determining the evidence level for each program purchased with SRCL funds or used in SRCL schools). The section concludes with analyses related to state technical assistance (TA) regarding use of evidence-based programs.

Although the exhibits in the report combine ratings from the What Works Clearinghouse (WWC) with those from Evidence for ESSA, the exhibits in this appendix (and associated text) distinguish between these two clearinghouses to provide additional context.

C.2.1. Literacy Programs Purchased by SRCL Districts

Although the SRCL program did not require districts to purchase literacy programs, this was a common use of SRCL funds (Exhibit A.3). In all states, at least some SRCL districts purchased literacy programs with SRCL funds, ranging from all districts in North Dakota to 26 percent of SRCL districts in Minnesota (Exhibit C.7).

Exhibit C.7. Purchase of literacy programs by all SRCL districts

	Percentage of All SRCL Districts That Purchased Literacy Programs	Percentage of All SRCL Districts That Did Not Purchase Literacy Programs
North Dakota	100%	0%
Georgia	94%	6%
Montana	92%	8%
New Mexico	91%	9%
Kentucky	84%	16%
Louisiana	82%	18%
Oklahoma	80%	20%
Kansas	79%	21%
Ohio	78%	22%
Maryland	77%	23%
Minnesota	26%	74%
All	80%	20%

Source: SRCL district survey, Q14. The unweighted sample size is 389 districts.

As discussed in section B.6.1, the analysis presented in Exhibit 3 in the report shows the percentage of the 236 *distinct literacy programs* purchased with SRCL funds, by evidence rating category (see section B.2 for an explanation of the evidence rating categories). As a reminder, Exhibit 3 in the report indicates that few programs newly purchased with SRCL funds (12 percent) were determined to be supported by strong or moderate evidence by the WWC or Evidence for ESSA. Exhibit C.8 provides additional details about this analysis.

Exhibit C.8. Percentage of literacy programs purchased with SRCL funds, by evidence level, 2018–19

Classifications	Number of Programs	Percentage
Strong or moderate evidence review rating in WWC	12	5%
Strong or moderate evidence review rating in Evidence for ESSA	16	7%
Promising evidence review rating	8	3%
Null or negative findings review rating	16	7%
Existing studies reviewed but do not meet standards for rigor	43	18%
No review of existing studies	31	13%
No published research identified	110	47%
Total	*236*	

Note: ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. Programs that received an evidence review rating of strong or moderate evidence by both the WWC and Evidence for ESSA were counted in the WWC row only. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL district survey, Q14–15 and review of WWC and Evidence for ESSA. The unweighted sample size is 236 unique programs listed by district survey respondents.

C.2.2. Literacy Program Purchases

The analysis presented in Exhibit 4 in the report takes into account the fact that some literacy programs purchased with SRCL funds were purchased by multiple districts. Exhibit 4 reports that of all the literacy program *purchase decisions* made by SRCL districts, only 22 percent were for programs supported by strong or moderate evidence.[27] Exhibit C.9 provides additional details about this analysis.

[27] The percentages in Exhibit C.9 are rounded to the nearest whole number. In the table, the percentage of district program purchases supported by strong or moderate evidence in WWC is 13.52 while the percentage of those supported by strong or moderate evidence in Evidence for ESSA is 8.74.

Exhibit C.9. Percentage of district program purchases by evidence level, 2018–19, by state

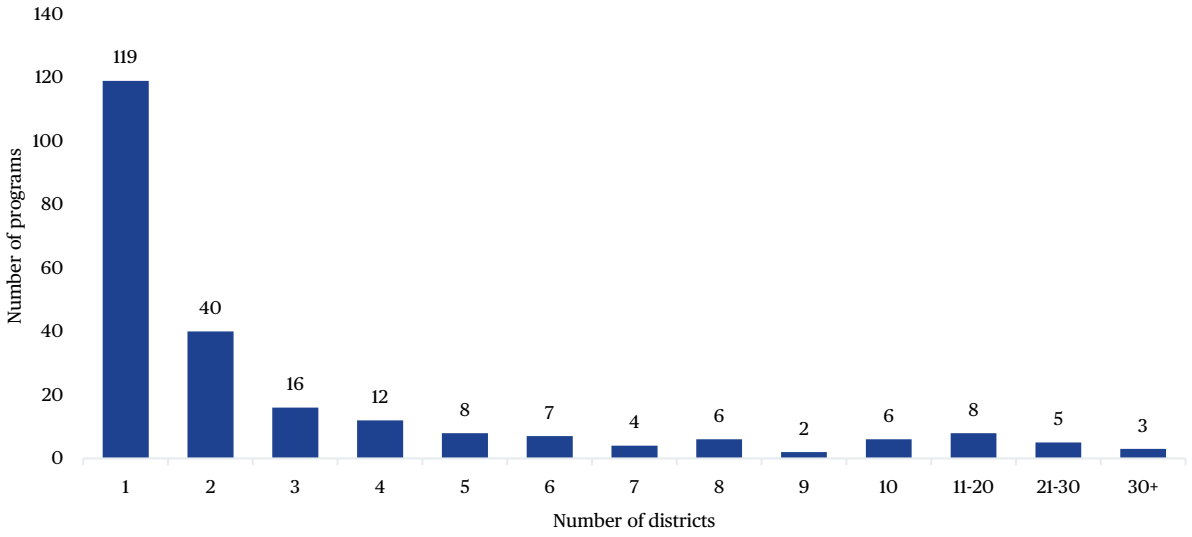
	Strong or Moderate Evidence Review Rating in WWC	Strong or Moderate Evidence Review Rating in Evidence for ESSA	Promising Evidence Review Rating	Null or Negative Findings Review Rating	Existing Studies Reviewed But Do Not Meet Standards for Rigor	No Reviews of Existing Studies	No Published Research Identified
North Dakota	40%	13%	10%	15%	8%	5%	10%
Maryland	20%	12%	6%	12%	24%	10%	14%
Georgia	18%	12%	4%	13%	25%	7%	22%
Montana	17%	10%	8%	10%	14%	13%	28%
Oklahoma	16%	20%	4%	4%	24%	7%	24%
Kentucky	13%	8%	1%	15%	18%	15%	29%
Ohio	12%	8%	2%	13%	19%	4%	42%
Kansas	9%	2%	7%	12%	14%	25%	32%
Louisiana	4%	1%	1%	10%	49%	10%	26%
New Mexico	0%	11%	0%	26%	37%	5%	21%
Minnesota	0%	0%	0%	29%	29%	0%	43%
All	14%	9%	4%	15%	24%	9%	26%

Note: ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report, as were "No reviews of existing studies" and "No published research." Programs that received an evidence review rating of strong or moderate evidence by both the WWC and Evidence for ESSA were counted in the WWC column only. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA. Percentages have been rounded to the nearest whole number, resulting in slight differences with Exhibit 4.
Source: SRCL district survey, Q14-15, and review of WWC and Evidence for ESSA. The unweighted sample size is 775 program purchases listed by 310 districts that purchased at least one program with SRCL funds.

Although district survey respondents reported purchasing 236 distinct programs in total, many of these programs were purchased by few districts. The majority of programs (187 of the 236 programs) were purchased by fewer than five districts each, with just over half of all district-cited programs (119) being purchased by only one district (Exhibit C.10). On the other hand, the three most commonly purchased programs were purchased by a combined 26 percent of responding districts that purchased at least one program.

Of the 16 programs purchased by more than 10 districts, 2 were supported by strong or moderate evidence in the WWC, and another 2 were supported by strong or moderate evidence as rated by the Evidence for ESSA clearinghouse; of the 4 programs purchased by more than 10 districts and supported by strong or moderate evidence in either the WWC or the Evidence for ESSA clearinghouse, 2 were among the 3 most commonly-purchased programs with SRCL funds.

Exhibit C.10. Number of programs purchased by one, two, three, or more districts, 2018–19



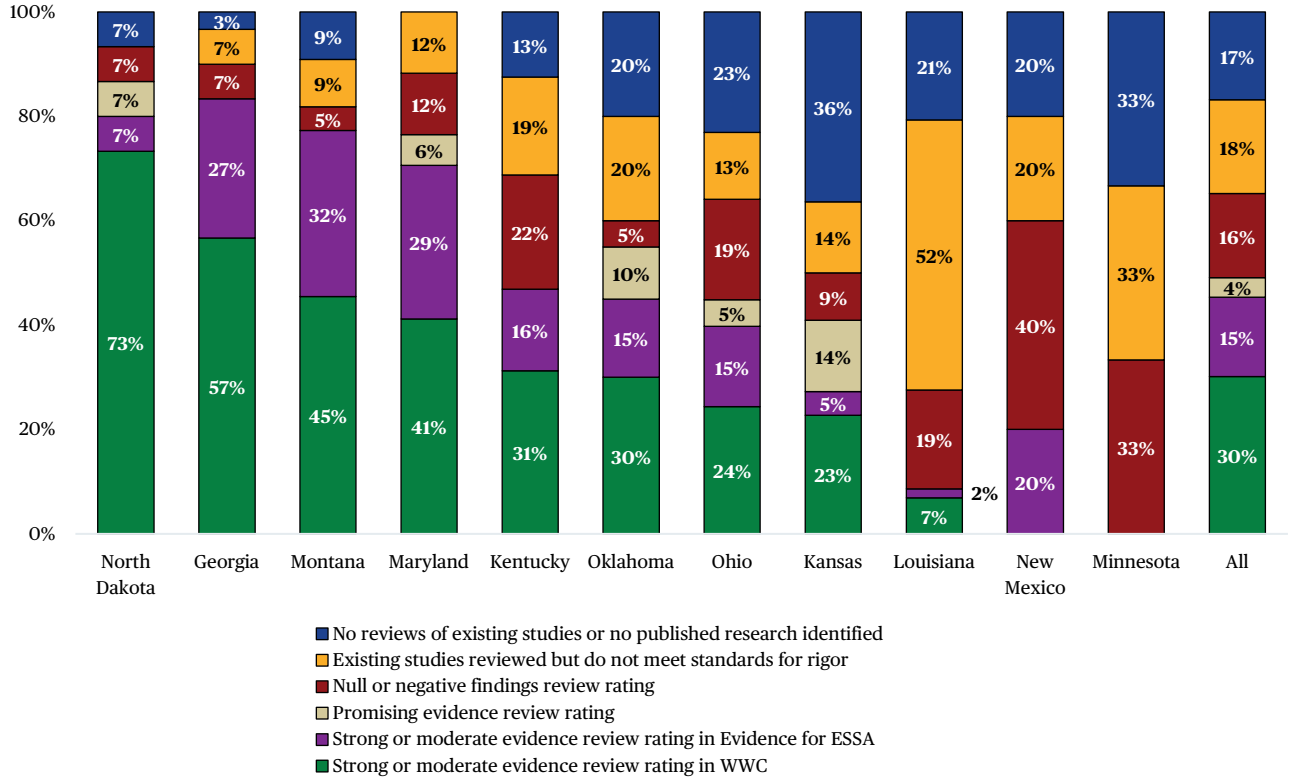
Note: This analysis is unweighted. Districts that did not purchase at least one literacy program with SRCL funds are excluded.
Source: SRCL district survey, Q14. The unweighted sample size is 310 districts that purchased at least one program with SRCL funds. Of these districts, 62 percent purchased two or more programs.

C.2.3. Districts That Purchased Programs Supported by Strong or Moderate Evidence

The analyses discussed in the previous sub-sections describe the extent to which programs purchased with SRCL funds were supported by strong or moderate evidence, combining purchases across districts. But they do not shed light on how many districts purchased programs supported by strong or moderate evidence.

Exhibit C.11 presents results of an analysis that categorized each district according to the highest evidence level across all of the programs that it purchased with SRCL funds (see section B.2 for a description of the evidence level rating categories). The results indicate that 30 percent of all districts purchased at least one program supported by strong or moderate evidence. This is a relatively weak standard for evidence use because a district gets credit for purchasing even one program supported by strong or moderate evidence. In this analysis, for instance, if a district in Oklahoma purchased one curriculum with strong or moderate evidence according to the WWC and another curriculum without published research, that district would count toward the section of Oklahoma's bar corresponding to programs with strong or moderate evidence from the WWC (dark green). The rank ordering of states did not change substantially from Exhibit 4 in the report; North Dakota remains the state in which the most districts purchase programs with strong or moderate evidence from the WWC, and the bottom five states remain identical (Ohio, Kansas, Louisiana, New Mexico, and Minnesota, in order of the most use of WWC strong or moderate evidence).

Exhibit C.11. Percentage of districts by highest evidence level among programs purchased with SRCL funds, 2018–19, by state

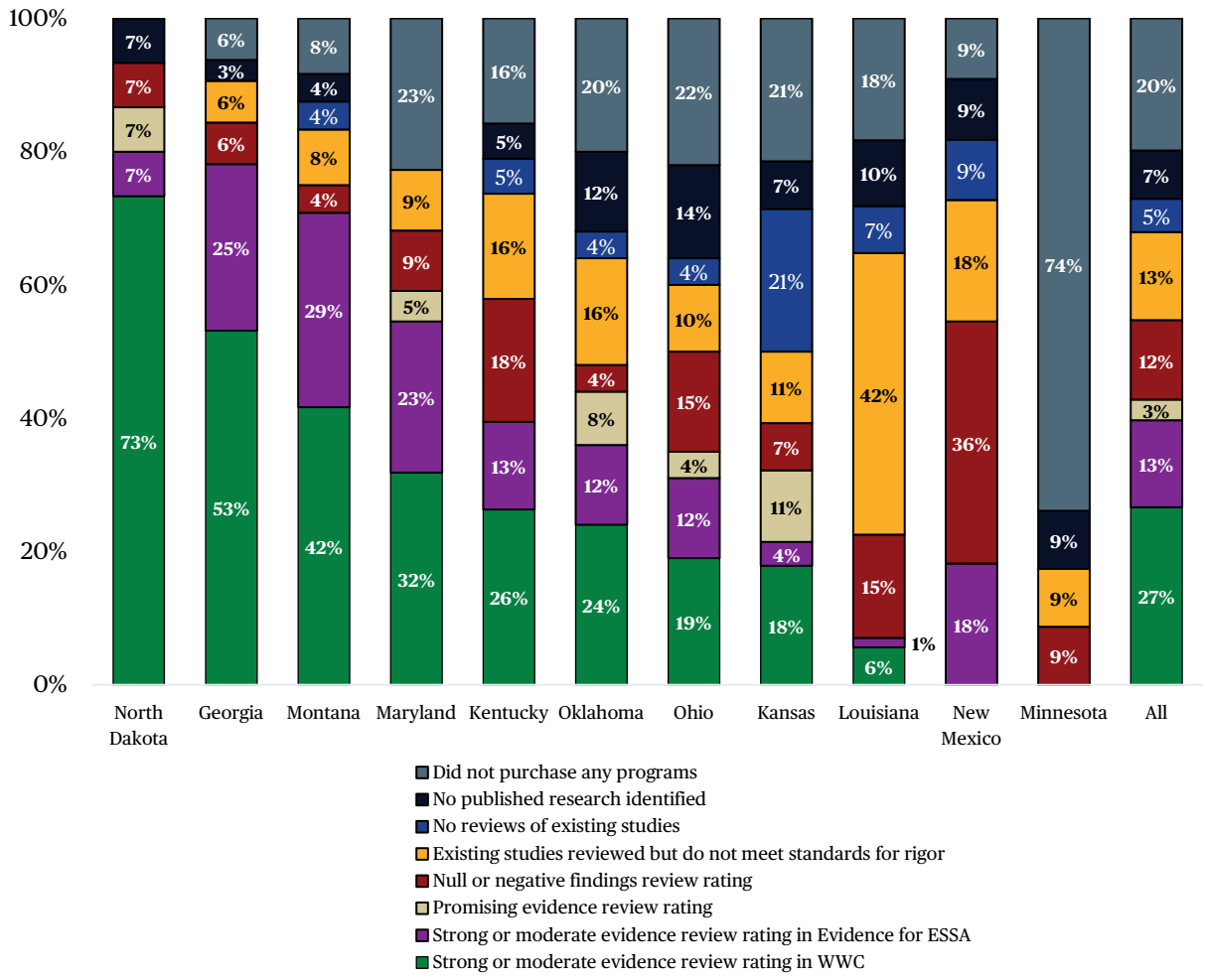


Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. In this exhibit, districts are categorized according to the highest evidence level across all of the programs it purchased with SRCL funds. Districts that did not purchase at least one literacy program with SRCL funds are excluded. This exhibit combines the categories of "no reviews of existing studies" and "no published research identified" from Exhibit 3 in the report. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL district survey, Q14-15 and review of WWC and Evidence for ESSA. The unweighted sample size is 310 districts that purchased at least one program with SRCL funds.

Although most districts purchased at least one program with SRCL funds, there was no federal requirement for them to do so, and some states had other priorities on how the funds were spent. Analyses in the report, as well as Exhibit C.11, are restricted to districts that purchased at least one program using SRCL funds. Because some states had few districts that purchased literacy programs, analyses that include nonbuyer districts conveys complementary information about SRCL program activities in these states. Thus, Exhibit C.12 replicates exhibit C.11 but includes the districts that did not buy programs. The rank order of states for the percentage of districts that purchased at least one program supported by strong or moderate evidence did not change when districts that did not purchase any programs were included in the denominator. In all states except Minnesota, at least three

quarters of districts purchased at least one program with SRCL funds, including all districts in North Dakota and at least 90 percent of the districts in Georgia, Montana, and New Mexico.

Exhibit C.12. Percentage of districts by highest evidence level among programs purchased with SRCL funds, including districts that did not buy programs, 2018–19, by state



Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. The categories "no reviews of existing studies" and "no published research identified" were also combined in Exhibit 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL district survey, Q14-15 and review of WWC and Evidence for ESSA. The unweighted sample size is 389 districts.

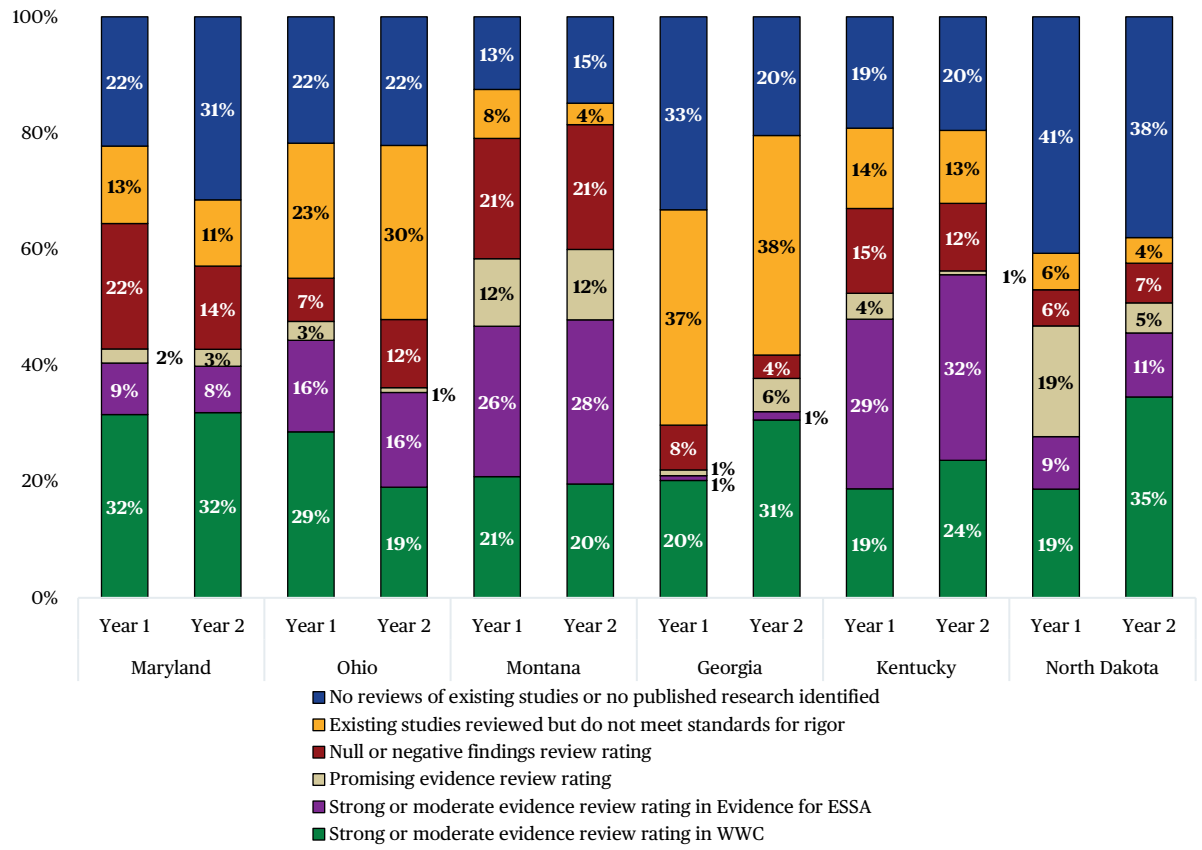
C.2.4. Teacher Use of Programs Supported by Evidence

Teachers in SRCL schools were asked to list literacy programs used in the classroom, both core curricula and supplementary programs. Consistent with districts' infrequent purchase of programs supported by strong or moderate evidence in the WWC (Exhibit 4 in the report), only 19 percent of teachers in SRCL schools reported using such a program in year 2 of SRCL implementation (Exhibits

C.13a and C.13b). The survey did not ask teachers whether their classroom programs were funded by SRCL. Thus, these percentages reflect programs purchased through all sources, not just SRCL funds.

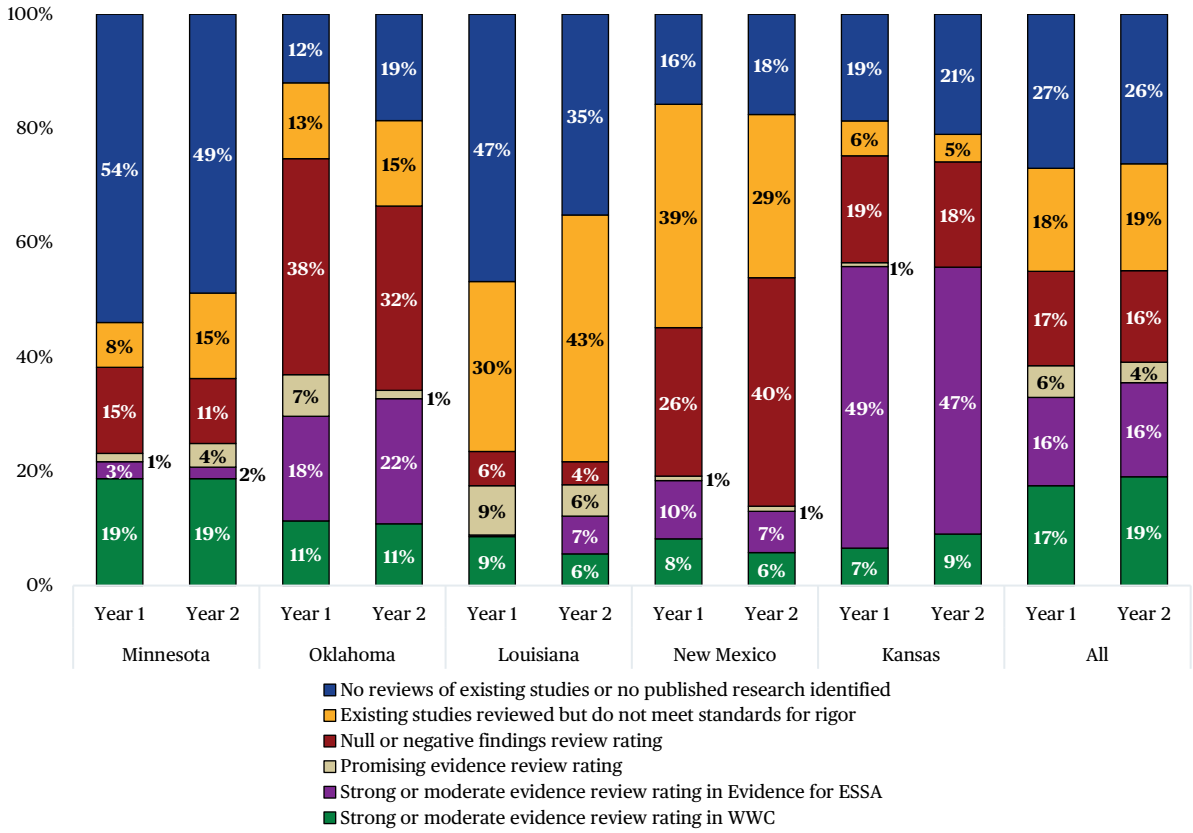
Exhibits C.13a and C.13b present the percentage of teachers in SRCL schools by the highest evidence level for programs used in their classrooms in 2018–19 and 2019–20. The results indicate little change overall; for all states, 17 percent of teachers reported using a program with strong or moderate evidence in 2018–19 (year 1), compared with 19 percent in 2019–20 (year 2), a difference that was not statistically significant.

Exhibit C.13a. Percentage of teachers in SRCL schools by highest evidence level among programs used in their classrooms, 2018–19 and 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA. Only teachers who answered that they used a program or curriculum "often," "almost daily," or "daily," provided the name of that program, and had the name of that program confirmed as a literacy program by the research team, are included in the analysis.
Source: SRCL K-12 teacher survey, 2018-19, Q19, Q21; SRCL K-12 teacher survey, 2019-20, Q19, Q21; and review of WWC and Evidence for ESSA. The unweighted sample size is 1,637 teachers who reported using at least one literacy program "often," "almost daily," or "daily" (2018-19) and 1,931 teachers who reported using at least one literacy program "often," "almost daily," or "daily" (2019-20).

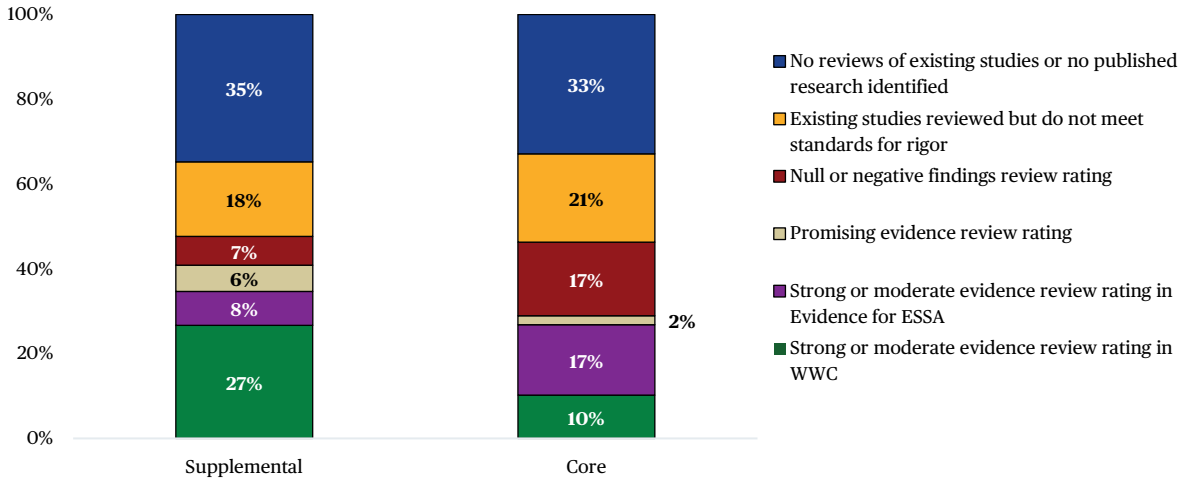
Exhibit C.13b. Percentage of teachers in SRCL schools by highest evidence level among programs used in their classrooms, 2018–19 and 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL K-12 teacher survey, 2018-19, Q19, Q21; SRCL K-12 teacher survey, 2019-20, Q19, Q21; and review of WWC and Evidence for ESSA. The unweighted sample size is 1,637 teachers who reported using at least one literacy program "often," "almost daily," or "daily" (2018-19) and 1,931 teachers who reported using at least one literacy program "often," "almost daily," or "daily" (2019-20).

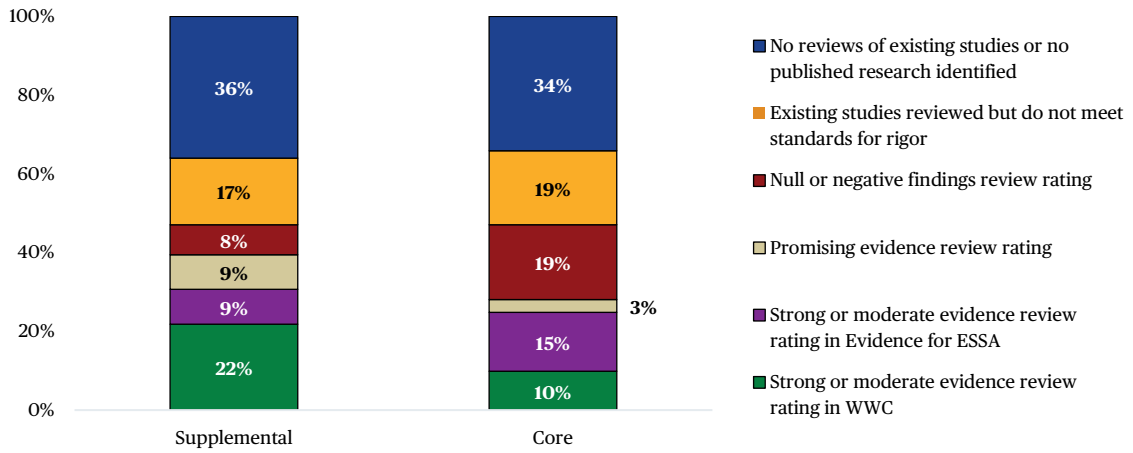
In both years of the study (2018-19 and 2019-20), the teacher survey asked respondents to list literacy programs used in the classroom, separately for core and supplemental programs. (Teachers reported programs and curricula regardless of whether or not they were purchased with SRCL funds.) Exhibit C.14a reports on the percentage of teachers using a core or supplemental literacy program supported by strong or moderate evidence in 2019-20 (Year 2). Exhibit C.14b shows the percentage of teachers listing such supplemental or core programs in 2018-19. The results are similar to the results for 2019-20. The supplemental programs teachers in SRCL schools reported using were statistically significantly more likely to have strong or moderate evidence than the core program ($p < .001$).

Exhibit C.14a. Percentage of teachers in SRCL schools by highest evidence level among core and supplemental programs used in their classrooms, 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL K-12 teacher survey, 2019-20, Q19, Q21 and review of WWC and Evidence for ESSA. The unweighted sample size is 1,768 teachers using a core literacy program and 722 teachers using a supplemental literacy program.

Exhibit C.14b. Percentage of teachers in SRCL schools by highest evidence level among core and supplemental programs used in their classrooms, 2018–19



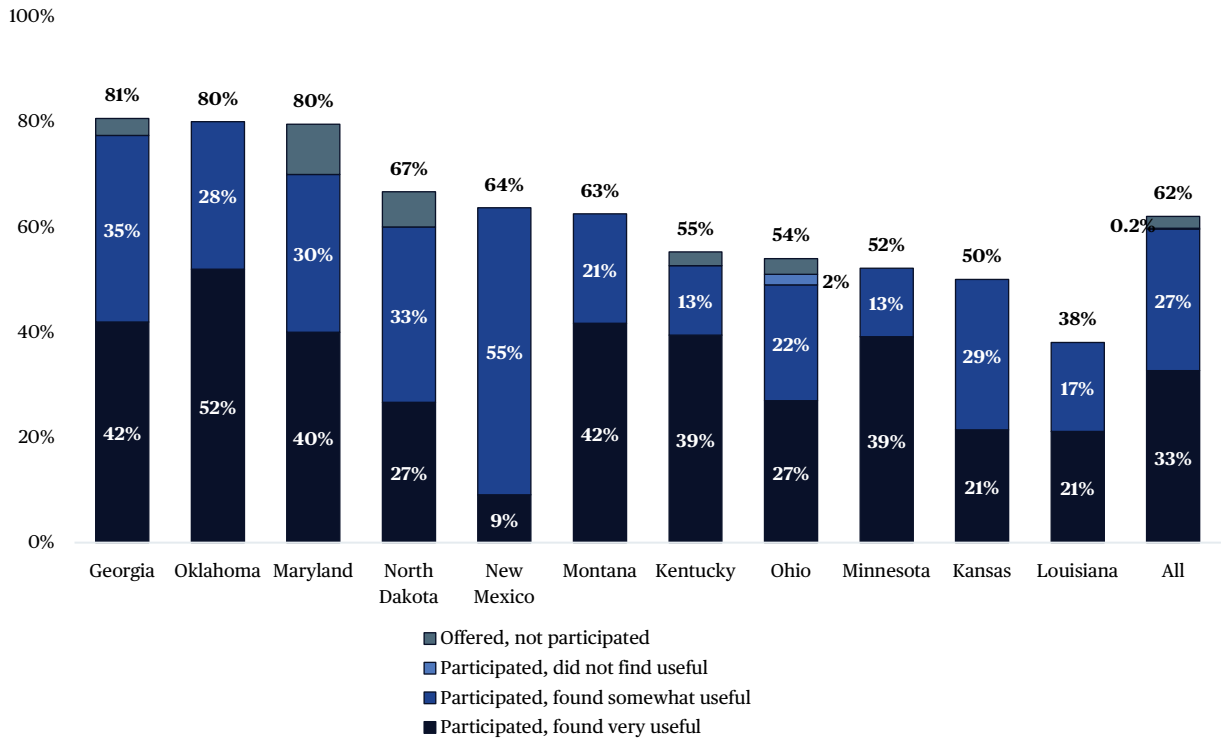
Note: The order of data in the exhibit matches the order of labels in the legend.
ESSA is Every Student Succeeds Act; WWC is What Works Clearinghouse. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. "Promising" programs had research that met WWC standards but not the large, multisite sample requirement for the regulatory definition of strong or moderate evidence or were identified as "promising" in Evidence for ESSA.
Source: SRCL K-12 teacher survey, 2018-19, Q19, Q21. The unweighted sample size is 1,495 teachers using a core literacy program and 604 teachers using a supplemental literacy program.

C.2.5. District Perceptions of Evidence-Related TA Provided to Districts

SRCL districts were asked about the availability of TA relating to "identifying and selecting reading/ELA interventions or practices that are supported by strong or moderate evidence," as well as their own participation and whether or not they found it to be useful. Most districts that indicated that this TA was offered also described it as somewhat or very useful.

States with higher percentages of purchases of programs supported by strong or moderate evidence tended to have higher percentages of district leaders reporting that evidence-related TA was offered to them. While the rank ordering differs slightly, five of the six highest ranking states in the percentage of districts buying at least one program supported by strong or moderate evidence (Exhibit C.11) are among the six highest-ranking states in the percentage of districts participating in TA and finding it useful. Similarly, four of the five highest ranking states in the percentage of program purchases supported by strong or moderate evidence (Exhibit 4 in the report) were among the five highest-ranking states in terms of participating in TA and perceptions of utility (Exhibit C.15).

Exhibit C.15. Participation and perceived utility of technical assistance provided to districts, 2018–19, by state



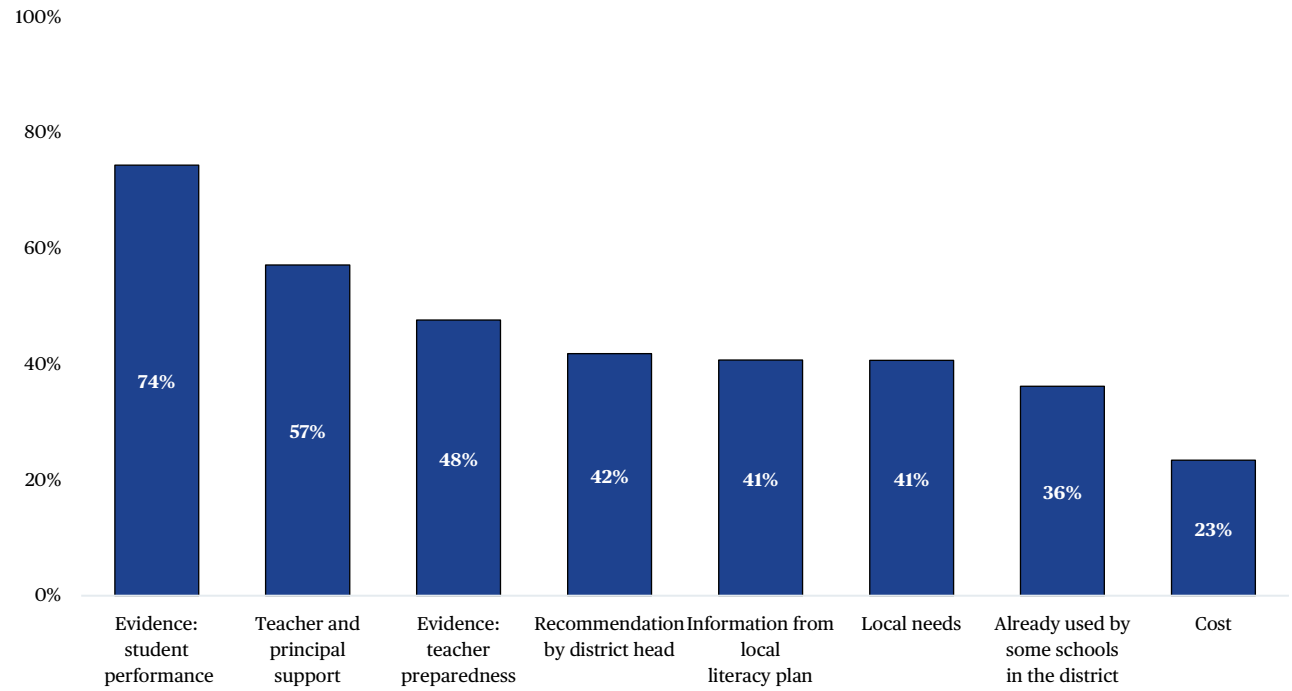
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL subgrantee survey, Q37. The unweighted sample size is 386 districts.

C.2.6. District Considerations When Selecting Literacy Programs

For each core or supplemental program purchased, districts were asked to indicate the factors they considered in the decision. A large majority of districts that purchased literacy programs (74 percent)

indicated that they considered evidence of effectiveness relating to student achievement when selecting them. Other commonly cited factors included teacher and principal support (57 percent) and evidence of effectiveness for improving teacher practice (48 percent; Exhibit C.16).

Exhibit C.16. Percentage of districts citing factors in program selection



Source: SRCL district survey, Q14-15. The unweighted sample size is 310 districts that purchased at least one program with SRCL funds.

C.3. Comprehensive Literacy Instruction

Few teachers reported engaging in all six features of comprehensive literacy instruction measured by the study, although most engaged in at least half of the features. Exhibits C.17–C.20 present the percentages of teachers in each grade band meeting study-defined benchmarks for comprehensive literacy instruction by state. These exhibits provide back-up information for Exhibit 5 in the report, which presents the percentage of teachers in each grade band meeting from 0 to 6 study-defined benchmarks for comprehensive literacy instruction.

Exhibit 6 in the report presents the percentage of teachers meeting the benchmark for each feature of comprehensive literacy instruction, by grade band, for the second year of SRCL implementation (2019-20). Exhibits C.21–C.24 check whether the percent of teachers meeting the benchmarks increased from the first to the second year of SRCL (2018-19 and 2019-20). None of the differences between years are statistically significant. Exhibits D.21–D.30 examine professional development as a potential explanation for differences across teachers in meeting the benchmarks.

Exhibit C.17. Percentage of pre-K teachers meeting benchmarks for features of comprehensive literacy instruction, 2019–20, by state

State	Percentage Meeting 0 Benchmarks	Percentage Meeting 1 Benchmark	Percentage Meeting 2 Benchmarks	Percentage Meeting 3 Benchmarks	Percentage Meeting 4 Benchmarks	Percentage Meeting 5 Benchmarks	Percentage Meeting 6 Benchmarks
Georgia	0%	1%	1%	6%	24%	57%	10%
Kansas	0%	12%	6%	16%	38%	24%	4%
Kentucky	0%	0%	0%	2%	24%	64%	10%
Louisiana	0%	0%	12%	3%	56%	27%	2%
Maryland	0%	1%	6%	14%	30%	46%	3%
Minnesota	0%	0%	34%	10%	25%	28%	4%
Montana	0%	0%	35%	15%	19%	31%	0%
North Dakota	0%	3%	12%	8%	43%	35%	0%
New Mexico	0%	0%	7%	12%	31%	44%	6%
Ohio	0%	1%	4%	7%	17%	59%	12%
Oklahoma	0%	6%	2%	12%	33%	38%	9%
All	0%	2%	11%	9%	31%	41%	5%

Source: SRCL 2019-20 pre-K teacher survey, Q19, Q21, Q25, Q27, Q29, Q32, Q33. The unweighted sample size is 283 teachers of pre-K who responded to the survey items for all six benchmarks.

Exhibit C.18. Percentage of K–3 teachers meeting benchmarks for features of comprehensive literacy instruction, 2019–20, by state

State	Percentage Meeting 0 Benchmarks	Percentage Meeting 1 Benchmark	Percentage Meeting 2 Benchmarks	Percentage Meeting 3 Benchmarks	Percentage Meeting 4 Benchmarks	Percentage Meeting 5 Benchmarks	Percentage Meeting 6 Benchmarks
Georgia	0%	2%	4%	12%	24%	41%	16%
Kansas	1%	4%	17%	33%	33%	7%	6%
Kentucky	0%	2%	13%	25%	29%	24%	7%
Louisiana	0%	14%	4%	35%	19%	14%	14%
Maryland	1%	0%	12%	27%	31%	23%	5%
Minnesota	3%	7%	18%	22%	24%	15%	10%
Montana	1%	2%	11%	13%	43%	21%	10%
North Dakota	1%	5%	15%	17%	13%	40%	8%
New Mexico	0%	1%	8%	18%	37%	17%	20%
Ohio	0%	0%	5%	14%	47%	24%	11%
Oklahoma	0%	12%	9%	32%	16%	23%	7%
All	1%	4%	10%	23%	29%	23%	11%

Source: SRCL 2019-20 K-12 teacher survey, Q19, Q21, Q25, Q27, Q29, Q32, Q33. The unweighted sample size is 729 teachers of grades K-3 who responded to the survey items for all six benchmarks.

Exhibit C.19. Percentage of grade 4–8 teachers meeting benchmarks for features of comprehensive literacy instruction, 2019–20, by state

State	Percentage Meeting 0 Benchmarks	Percentage Meeting 1 Benchmark	Percentage Meeting 2 Benchmarks	Percentage Meeting 3 Benchmarks	Percentage Meeting 4 Benchmarks	Percentage Meeting 5 Benchmarks	Percentage Meeting 6 Benchmarks
Georgia	0%	0%	7%	28%	48%	18%	0%
Kansas	0%	0%	17%	30%	17%	36%	0%
Kentucky	0%	4%	6%	34%	38%	18%	0%
Louisiana	0%	0%	8%	28%	37%	16%	10%
Maryland	0%	0%	10%	21%	46%	22%	1%
Minnesota	0%	7%	21%	25%	30%	12%	5%
Montana	0%	0%	9%	15%	50%	24%	2%
North Dakota	8%	2%	17%	44%	22%	8%	0%
New Mexico	0%	4%	2%	32%	38%	18%	5%
Ohio	3%	0%	20%	26%	34%	12%	4%
Oklahoma	3%	12%	24%	22%	30%	5%	4%
All	1%	3%	13%	28%	35%	17%	3%

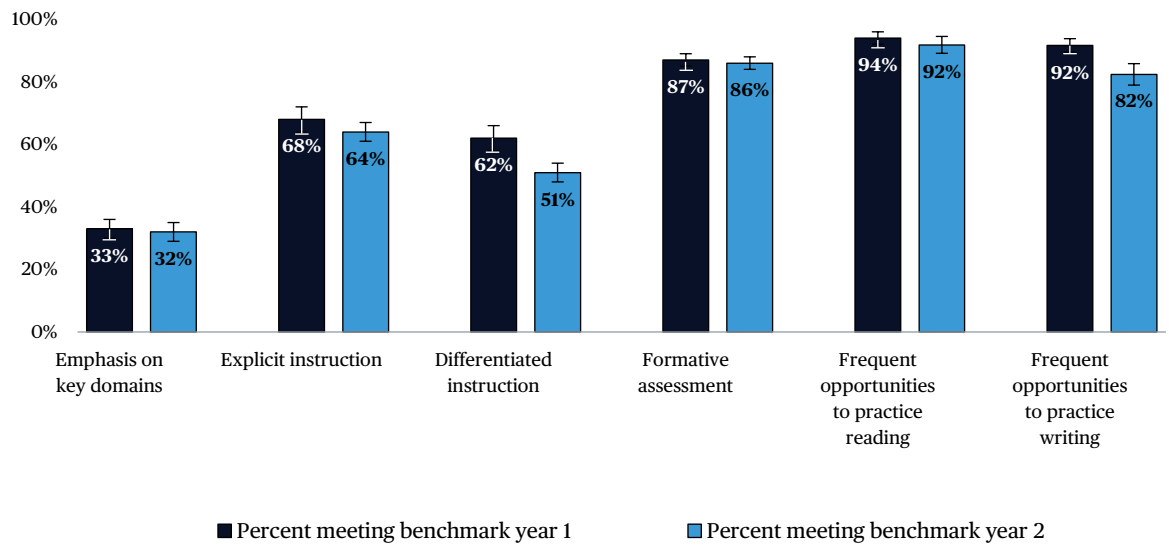
Source: SRCL 2019-20 K-12 teacher survey, Q19, Q21, Q25, Q27, Q29, Q32, Q33. The unweighted sample size is 351 teachers of grades 4-8 who responded to the survey items for all six benchmarks.

Exhibit C.20. Percentage of grade 9–12 teachers meeting benchmarks for features of comprehensive literacy instruction, 2019–20, by state

State	Percentage Meeting 0 Benchmarks	Percentage Meeting 1 Benchmark	Percentage Meeting 2 Benchmarks	Percentage Meeting 3 Benchmarks	Percentage Meeting 4 Benchmarks	Percentage Meeting 5 Benchmarks	Percentage Meeting 6 Benchmarks
Georgia	0%	6%	11%	32%	37%	13%	0%
Kansas	0%	13%	37%	26%	24%	0%	0%
Kentucky	0%	17%	25%	35%	11%	11%	0%
Louisiana	0%	10%	13%	67%	10%	0%	0%
Maryland	9%	7%	9%	12%	43%	14%	6%
Minnesota	0%	36%	13%	36%	9%	6%	0%
Montana	0%	24%	32%	16%	20%	8%	0%
North Dakota	5%	5%	43%	36%	11%	0%	0%
New Mexico	8%	15%	14%	49%	0%	15%	0%
Ohio	5%	0%	12%	19%	47%	16%	0%
Oklahoma	0%	26%	26%	31%	17%	0%	0%
All	3%	14%	21%	33%	21%	8%	1%

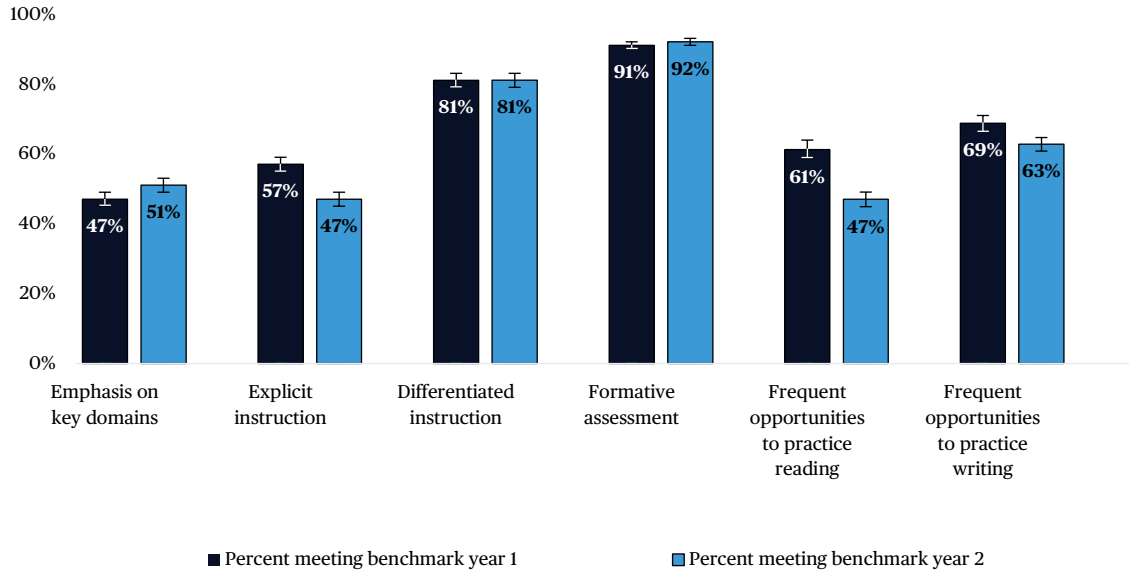
Source: SRCL 2019-20 K-12 teacher survey, Q19, Q21, Q25, Q27, Q29, Q32, Q33. The unweighted sample size is 169 teachers of grades 9–12 who responded to the survey items for all six benchmarks.

Exhibit C.21. Percentage of pre-K teachers meeting each benchmark for features of comprehensive literacy instruction, 2018–19 and 2019–20



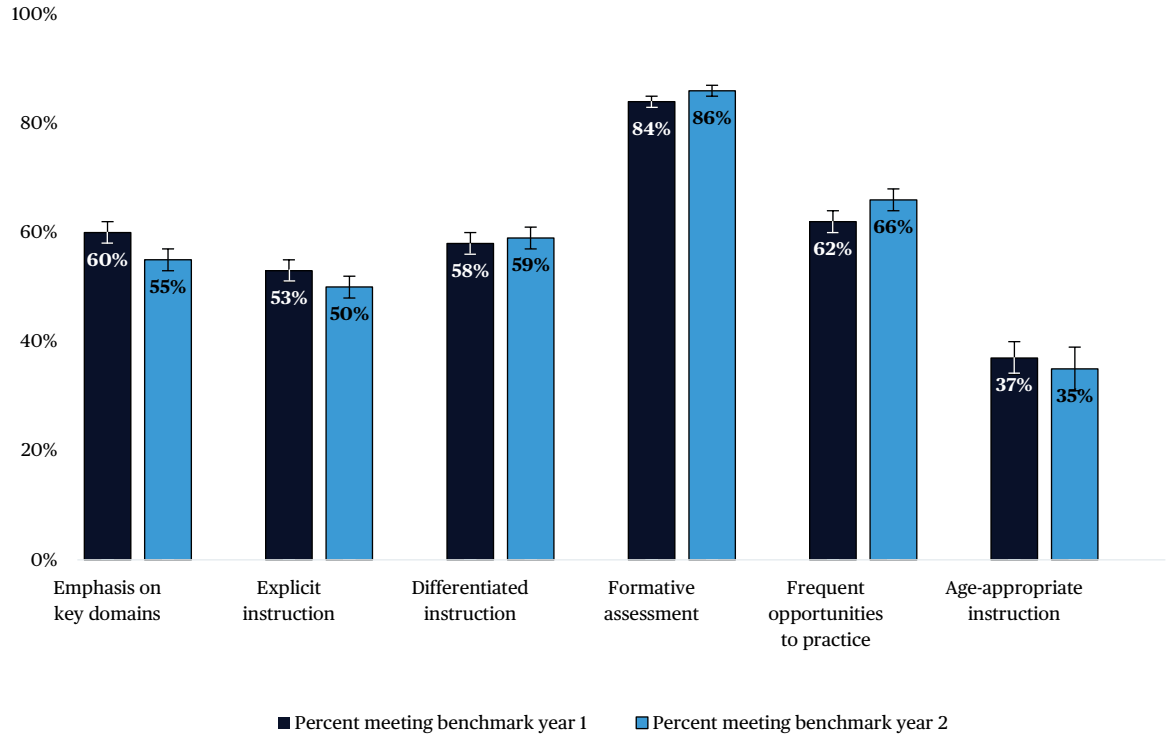
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL pre-K teacher surveys, 2018-19 and 2019-20. The unweighted sample size for all features *except* differentiated instruction is 285 pre-K teachers in 2018-19 and 332 pre-K teachers in 2019-20. The unweighted sample size for differentiated instruction is 238 pre-K teachers in 2018-19 and 284 pre-K teachers in 2019-20.

Exhibit C.22. Percentage of teachers meeting each benchmark for features of comprehensive literacy instruction in grades K–3, 2018–19 and 2019–20



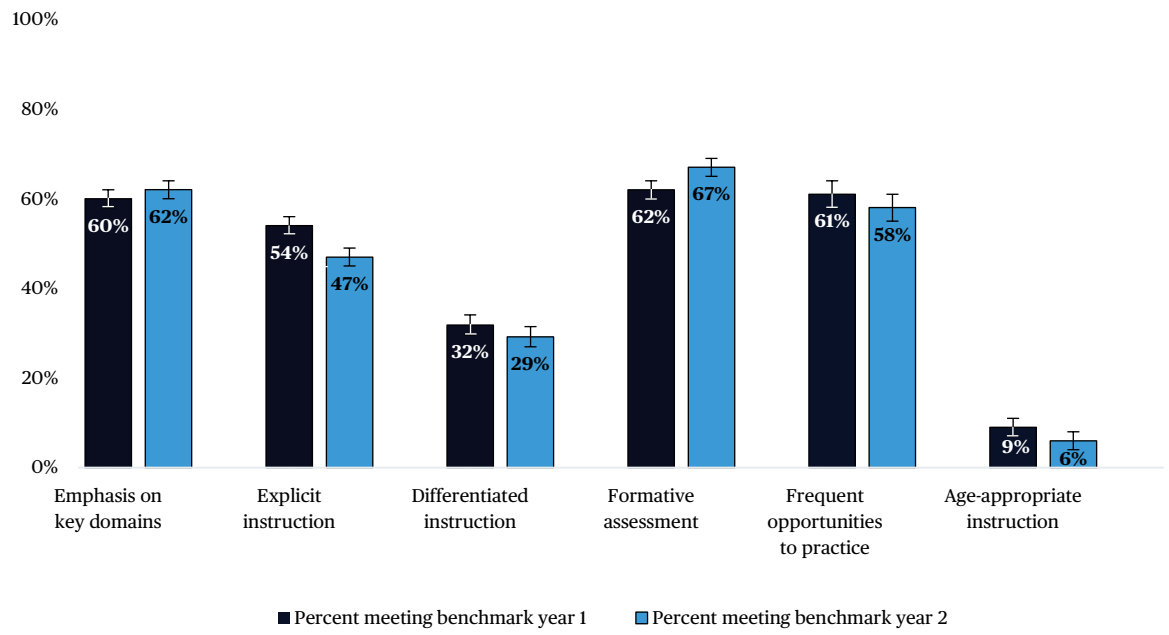
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher surveys, 2018-19 and 2019-20. The unweighted sample size ranges from 657 to 687 teachers of grades K-3 in 2018-19 and ranges from 765 to 881 teachers of grades K-3 in 2019-20.

Exhibit C.23. Percentage of teachers meeting each benchmark for features of comprehensive literacy instruction in grades 4–8, 2018–19 and 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher surveys, 2018-19 and 2019-20. The unweighted sample size for the first five features of literacy instruction ranges from 923 to 980 teachers of grades 4-8 in 2018-19 and from 1,108 to 1,171 teachers in 2019-20. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 362 teachers of grades 4-8 in 2018-19 and 392 teachers in 2019-20.

Exhibit C.24. Percentage of teachers meeting each benchmark for features of comprehensive literacy instruction in grades 9–12, 2018–19 and 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher surveys, 2018-19 and 2019-20. The unweighted sample size for the first five features of literacy instruction ranges from 385 to 452 teachers of grades 9-12 in 2018-19 and from 432 to 517 teachers in 2019-20. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 196 teachers of grades 9-12 in 2018-19 and 208 teachers in 2019-20.

C.4. Achievement Trends Analysis Results

The estimated relationships between receipt of SRCL and ELA achievement ranged from -0.071 standard deviation in New Mexico to 0.036 standard deviation in Louisiana, with a meta-analytic average of almost exactly zero. Two positive results, in Louisiana and Ohio, are significantly different from zero (Exhibit C.25).[28] A detailed discussion of the analytic approach used to identify the association between receipt of SRCL funds and ELA achievement can be found in Appendix B.8.

Similarly, the analysis found no statistically significant relationship between receipt of SRCL and ELA outcomes for students who are disadvantaged. When the sample is limited to disadvantaged students (English learners, students with disabilities, and students who are eligible for free or reduced-price lunch), the average difference in student performance remains almost exactly zero (Exhibit C.25). The estimated differences ranged from -0.075 standard deviation in New Mexico to 0.041 standard deviation in Louisiana; none are significantly distinguishable from zero.

[28] These two statistically significant results should be interpreted with caution due to the possibility of inflated Type I error given that statistical tests were conducted for 11 states.

Exhibit C.25. Differences in student ELA score trends between SRCL schools and comparable non-SRCL schools, by state

State	All Students				Students Who Are Disadvantaged			
	Difference (Standard Deviation Units)	Standard Error	p-Value	Number of Student-Year Observations [Schools]	Difference (Standard Deviation Units)	Standard Error	p-Value	Number of Student-Year Observations [Schools]
Georgia	-0.007	0.003	.230	428,248 [796]	-0.004	0.011	.789	363,838 [796]
Kansas	0.005	0.021	.842	14,425[a] [382]	0.015	0.020	.594	8,587[a] [382]
Kentucky	0.013	0.024	.673	542,092 [364]	0.010	0.026	.775	369,648 [364]
Louisiana	0.036	0.003	.030*	484,878 [284]	0.041	0.009	.142	339,603 [284]
Maryland	-0.063	0.023	.222	472,416 [295]	-0.058	0.016	.174	189,465 [295]
Minnesota	0.019	0.007	.224	323,213 [175]	0.036	0.013	.228	197,340 [175]
Montana	-0.016	0.112	.910	85,817 [115]	-0.041	0.121	.792	5,855 [115]
New Mexico	-0.071	0.016	.138	200,957 [166]	-0.075	0.020	.162	183,138 [166]
North Dakota	-0.004	0.007	.647	733[a] [55]	0.004	0.020	.880	642[a] [55]
Ohio	0.021	0.001	.035*	879,131 [448]	0.018	0.003	.118	465,383 [448]
Oklahoma	-0.018	0.021	.534	189,554 [137]	-0.050	0.035	.386	143,310 [137]
All	0.000	0.006	.994	3,606,306[b] [3,217]	0.000	0.006	.980	2,157,580[b] [3,217]

* Statistically significant at $p = .05$.

[a] Sample sizes for Kansas and North Dakota are at the school-year level rather than the student-year level.

[b] Aggregate sample sizes exclude Kansas and North Dakota.

Note: ELA is English language arts.

Source: State education agency administrative data sets in 8 states, Kentucky data provided by the Kentucky Center for Statistics; aggregated standardized test scores in Kansas and North Dakota. The unweighted sample size is 3,606,306 students using state administrative data and 15,158 students using aggregated scores in Kansas and North Dakota.

APPENDIX D. ADDITIONAL ANALYSES

This appendix includes additional analyses not discussed in the body of the report. The details provide additional context for readers interested in more detail on SRCL implementation. This appendix includes information on the use of literacy programs supported by evidence (section D.1); comprehensive literacy instruction in SRCL classrooms (section D.2); district use of needs assessments (section D.3); data for continuous improvement (section D.4); funds for professional development (section D.5); professional development related to comprehensive literacy instruction (section D.6); teacher and school leader reports of the usefulness of professional development (section D.7); school leader reports of changes due to SRCL (section D.8); and the level of emphasis on key domains of reading by grade band (section D.9).

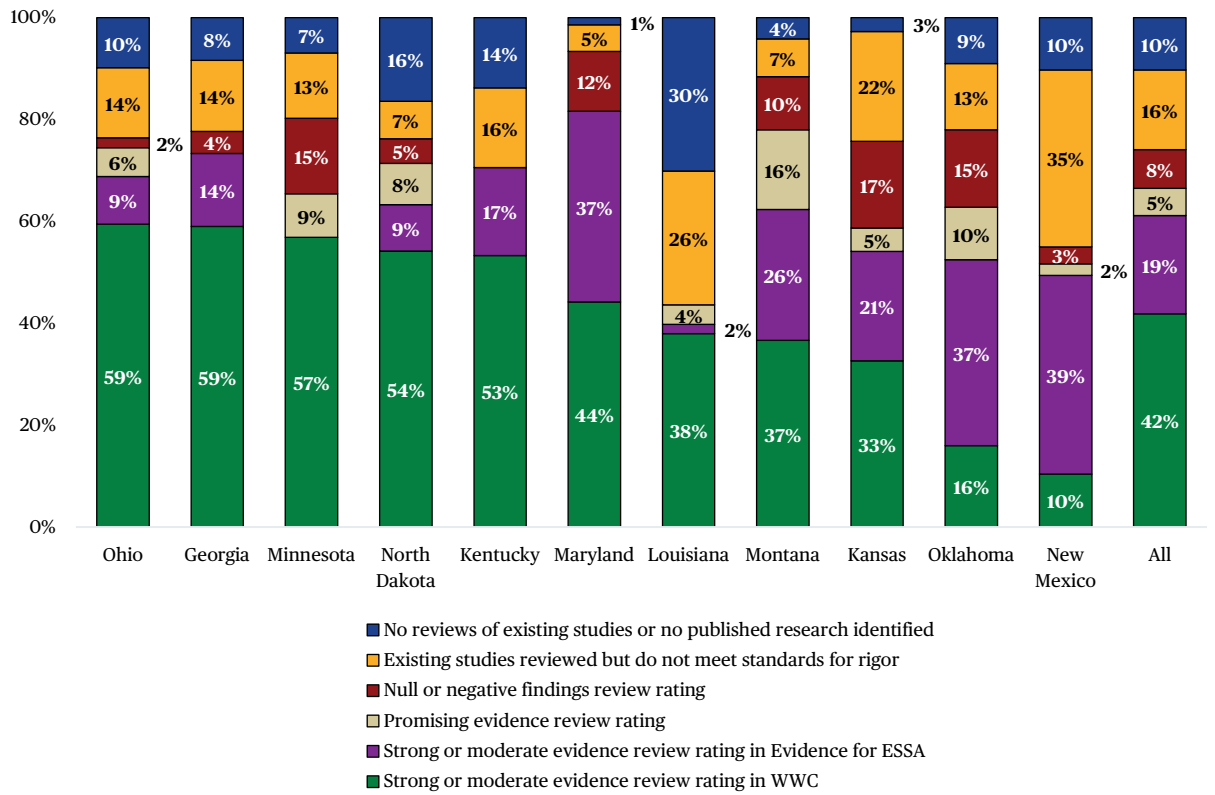
D.1. Additional Analyses on Use of Literacy Programs Supported by Evidence

This appendix section includes supplementary analyses related to use of programs supported by strong or moderate evidence in SRCL districts and schools. The section begins with a summary of principal reports of literacy programs by the highest level of evidence (D.1.1), followed by a discussion of the consistency of district purchase of programs supported by strong or moderate evidence (D.1.2). Section D.1.3 depicts the relationship between the average number of district program purchases with the percentage of districts that purchased a program supported by strong or moderate evidence.

D.1.1. Principal-Reported Use of Evidence Based Literacy Programs

In addition to gathering data on the programs districts purchased with SRCL funds (shown in Exhibits 3 and 4 in the report and appendix section C.2) and programs used by teachers (appendix section C.2), the study also gathered data from principals on the programs used in their schools with English learners, students with disabilities, and struggling readers. (This analysis was not conducted for Early Childhood Center Directors, because few listed named literacy programs.) Nearly all principals (95 percent) indicated using at least one literacy program with one or more of these student groups. Exhibit D.1 displays the percentage of principals that reported using at least one program supported by strong or moderate evidence. Programs were classified by the study team using the same methods as the programs reported by SRCL districts (see section B.2). Overall, 42 percent of SRCL school principals reported using at least one program with a strong or moderate evidence rating in the WWC. The percentage increases to 61 percent when including SRCL schools that reported using at least one program with a strong or moderate evidence rating in Evidence for ESSA. These percentages are higher than the percentage of teachers using at least one supplementary literacy program with struggling students supported by strong or moderate evidence (Exhibits C.13a and C.13b). One potential explanation for the difference is that principals reported on programs used by any teacher in the school, rather than programs used by a single teacher.

Exhibit D.1. Percentage of schools by highest evidence level of programs reported by principals, 2018–19, by state



Note: The order of data in the exhibit matches the order of labels in the legend.
This exhibit combines the categories of "no review of existing studies" and "no published research identified" from Exhibit 3 in the report. The categories of "strong or moderate evidence review rating in WWC" and "strong or moderate evidence review rating in Evidence for ESSA" were combined in Exhibits 3 and 4 in the report. This exhibit excludes the 5 percent of principals who did not report use of a literacy program.
Source: SRCL school leader survey, Q11-13, and review of WWC and Evidence for ESSA. The unweighted sample size is 331 K-12 schools in which the principal reported using at least one literacy program for English learners, students with disabilities, or struggling readers.

D.1.2. Consistency of District Purchase of Literacy Programs Supported by Strong or Moderate Evidence

Analyses presented in Exhibits C.11 and C.12 display the percentage of SRCL districts by the highest level of evidence of literacy programs purchased with SRCL funds, but these analyses do not depict whether SRCL districts consistently purchased programs supported by strong or moderate evidence. To examine this issue, a district-level analysis was performed, examining the proportion of each SRCL district's purchases that were supported by strong or moderate evidence as determined by the WWC.[29] The results, shown in Exhibit D.2, indicate that 5 percent of districts that purchased at least one

[29] Programs with strong or moderate evidence according to Evidence for ESSA, but not according to the WWC, are not counted as supported by strong or moderate evidence in this analysis.

program with SRCL funds purchased programs that were *all* supported by strong or moderate evidence. However, most of those districts purchased a single program. Among districts that purchased two or more programs with SRCL funds, only 2 percent bought *only* programs supported by strong or moderate evidence.

Exhibit D.2. Percentage of districts purchasing only programs supported by strong or moderate evidence, 2018–19, by state

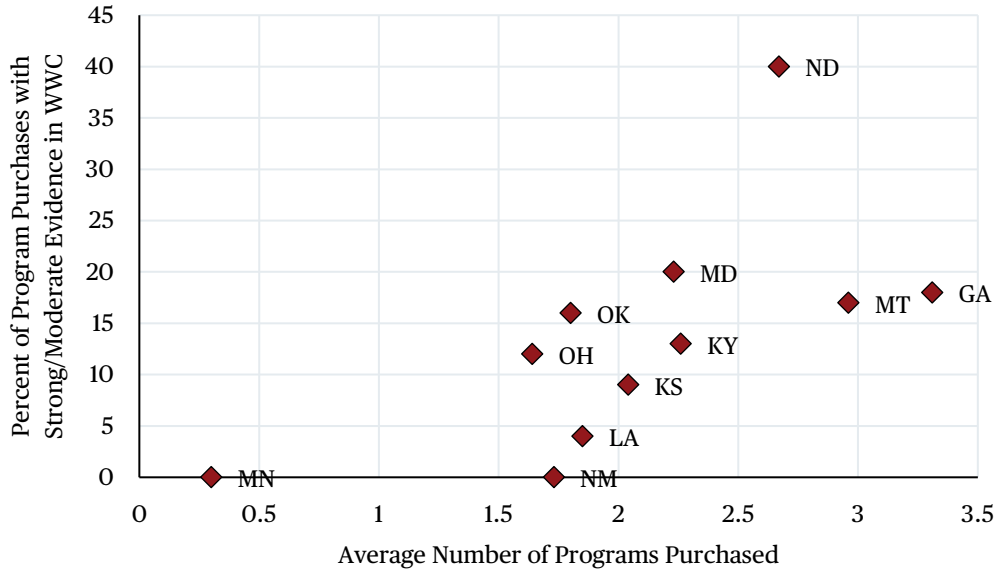
	Percentage of Districts Purchasing Programs With SRCL Funds	Percentage of Districts Purchasing One or More Strong or Moderate Programs	Percentage of Districts Purchasing Only Strong or Moderate Programs	Percentage of Districts Purchasing Two or More Programs With SRCL Funds	Percentage of Districts That Purchased Two or More Programs Purchasing Only Strong or Moderate Programs
North Dakota	100%	73%	13%	87%	15%
Montana	92%	45%	9%	71%	0%
Ohio	78%	24%	6%	49%	0%
Kentucky	84%	31%	6%	66%	4%
Maryland	77%	41%	6%	68%	0%
Oklahoma	80%	30%	5%	64%	0%
Kansas	79%	23%	5%	64%	0%
Georgia	94%	57%	3%	91%	3%
Louisiana	82%	7%	0%	55%	0%
Minnesota	26%	0%	0%	4%	0%
New Mexico	91%	0%	0%	64%	0%
All	80%	30%	5%	62%	2%

Note: Districts that did not purchase at least one literacy program with SRCL funds are excluded.
Source: SRCL district survey, Q14–15, and review of WWC and Evidence for ESSA. The unweighted sample size is 389 districts in the first column, 310 districts in the second and third columns, 389 districts in the fourth column, and 229 districts in the fifth column.

Exhibit D.2 shows that it was rare for districts that purchased multiple programs to buy only programs supported by strong or moderate evidence. Still, it is possible that districts that bought more programs with SRCL funds tended to buy a higher proportion supported by evidence. Exhibit D.3 presents a scatter plot of the percentage of program purchases in a state that were supported by strong or moderate evidence and the average number of programs purchased by districts in a state. In general, states in which districts tended to buy more programs were generally also states in which districts more consistently bought programs backed by strong or moderate evidence; the correlation between the average number of programs purchased and the percentage of programs purchased with strong or moderate evidence is 0.63. The most extreme examples are Minnesota (most districts bought no programs, and no districts bought programs with strong or moderate evidence) and North Dakota (the average district bought almost three programs, and 40 percent of those purchases were of programs

with strong or moderate evidence), but even with those two states removed, the correlation between the average number of programs purchased and the percentage of program purchases with strong or moderate evidence is 0.58 for the remaining nine states.

Exhibit D.3. Percentage of program purchases supported by strong or moderate evidence and average number of programs purchased by districts, by state



Source: SRCL district survey, Q14-15, and review of WWC and Evidence for ESSA. The unweighted sample size is 389 districts.

D.2. Comprehensive Literacy Instruction

This section includes additional analyses of issues related to comprehensive literacy instruction (CLI). Section D.2.1 compares results on the SRCL teacher survey with national results from the Early Childhood Longitudinal Survey, for items included on both surveys. Section D.2.2 explores the sensitivity of results to missing data on the sixth feature of CLI, the assignment of age-appropriate reading materials. Finally, section D.2.3 presents sensitivity analyses that adjust the thresholds necessary to meet benchmarks for emphasis on key domains and explicit instruction.

D.2.1. Comparisons of Survey Results Against National Samples

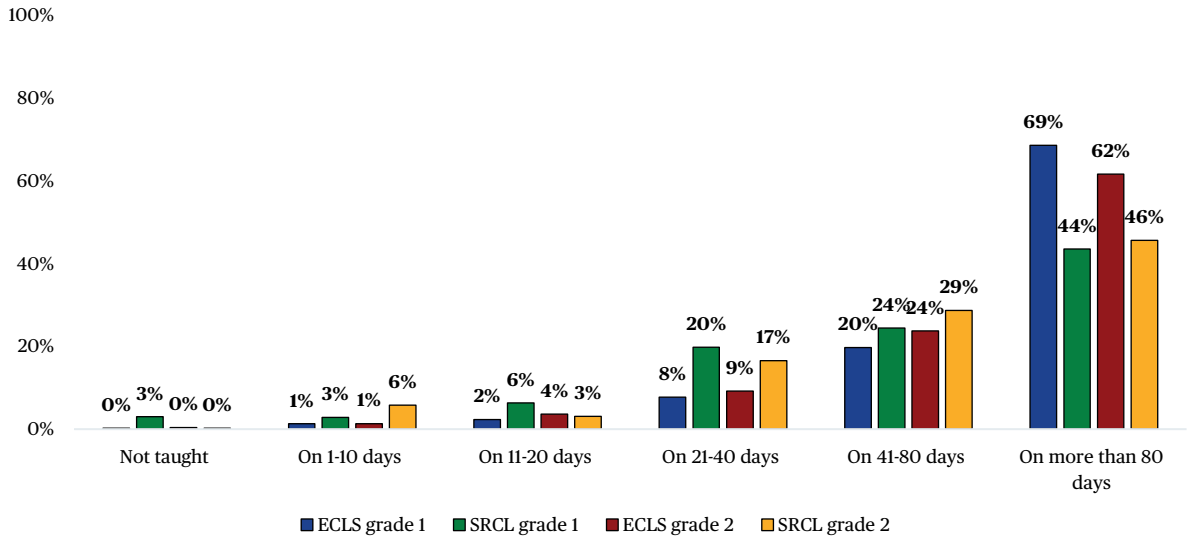
Most of the survey items used to measure teachers' implementation of comprehensive literacy instruction (CLI) were drawn from nationally representative surveys. This provides an opportunity to compare the responses of teachers in SRCL-funded schools with a national sample.

In particular, the survey includes items from teacher questionnaires used in the Early Childhood Longitudinal Studies (ECLS) Program, which followed a cohort of students that were in kindergarten in the 2010-11 school year. Students who progressed at the typical rate were in grade 1 in 2011-12 and in grade 2 in 2012-13. The sample of schools represented in the ECLS is meant to be representative of teachers of students in the relevant grades for the United States as a whole (in spring 2012 for grade 1 and spring 2013 for grade 2), whereas the sample of schools represented in SRCL is largely composed of

disadvantaged schools, since states were expected to prioritize disadvantaged schools for SRCL funding. In addition, the SRCL teacher survey was administered in the spring of 2020. Thus, the SRCL and ECLS samples differ in student composition and in the time period measured.

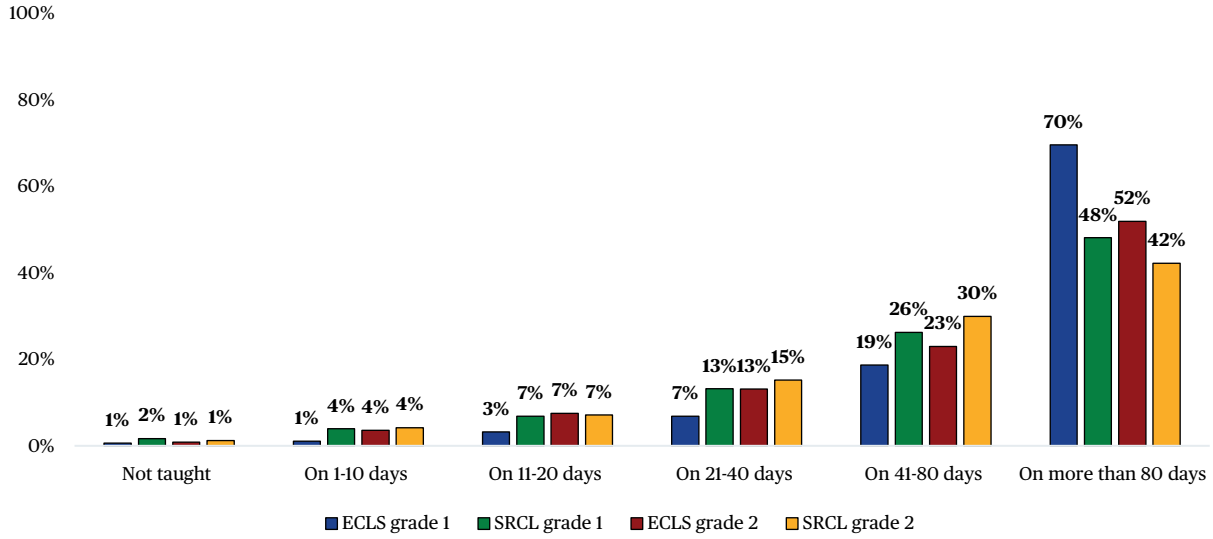
Teachers were asked, "From the first day of school until today, please indicate how many days each of the following reading/ELA [English language arts] skills and concepts has been covered in your class? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Not all skills/concepts may apply to your grade level. Check one per row." Results for the time devoted to predicting what might come next in a text are shown in Exhibit D.4 and results for the time devoted to reading irregularly spelled words are shown in Exhibit D.5. In general, the results for the nationally representative sample of teachers in the ECLS sample and for teachers in SRCL schools are similar, but the ECLS teachers were statistically significantly more likely to implement the practices measured on both surveys at least 80 days per year than the SRCL teachers.

Exhibit D.4. Frequency of *predicting what might occur next in the text* for ECLS-K teachers (2011–12, 2012–13) and SRCL teacher survey respondents (2019–20)



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K–12 teacher survey, 2019–20, Q27; Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) first-grade Spring Classroom Teacher Questionnaire, Teacher Level, QC2, ECLS-K:2011 second-grade Spring Teacher Questionnaire, Teacher Level, QC2. The unweighted sample size for SRCL teachers is 378 teachers of grade 1 and/or grade 2. The unweighted sample size for ECLS-K teachers is 13,240 teachers of grade 1 and 12,191 teachers of grade 2. All ECLS results are from public-use data.

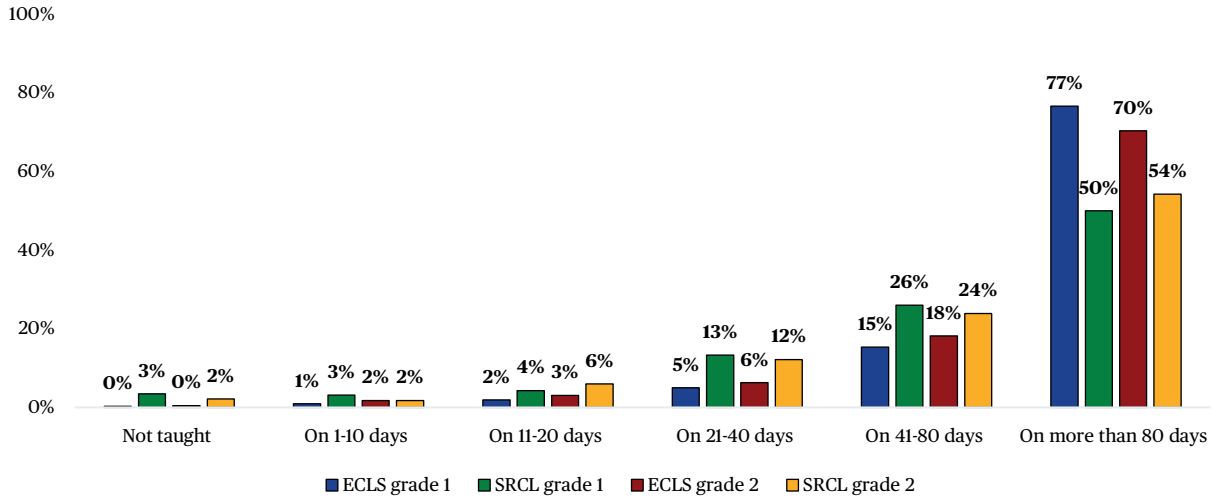
Exhibit D.5. Frequency of *reading irregularly spelled words* for ECLS-K teachers (2011-12, 2012-13) and SRCL teacher survey respondents (2019-20)



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher survey, 2019-20, Q28; ECLS-K:2011 first-grade Spring Classroom Teacher Questionnaire, Teacher Level, QC2, ECLS-K:2011 second-grade Spring Teacher Questionnaire, Teacher Level, QC2. The unweighted sample size for SRCL teachers is 379 teachers of grade 1 and/or grade 2. The unweighted sample size for ECLS-K teachers is 13,214 teachers of grade 1 and 12,109 teachers of grade 2. All ECLS results are from public-use data.

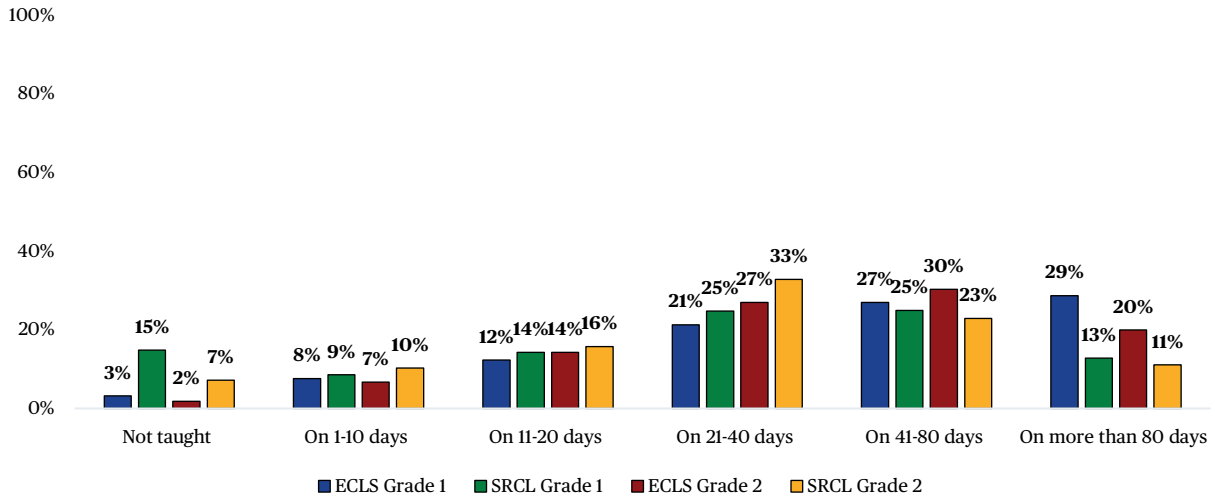
Additionally, to determine the extent to which teachers offered frequent opportunities to practice reading, teachers were asked, "From the first day of school until today, how many days have students in your class been asked to read and reread passages orally with guidance on pacing, intonation, and expression as part of reading/ELA instruction? Please include the time during which you provide direct instruction as well as the time you spend supervising students as they work. Check one." For frequent opportunities to practice writing, teachers were asked, "From the first day of school until today, how frequently have students in your class been asked to do the following types of writing? Check one per row." Results are shown in Exhibits D.6 and D.7.

Exhibit D.6. Frequency of *asking students to read and reread passages orally* for ECLS-K teachers (2011–12, 2012–13) and SRCL teacher survey respondents (2019–20)



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher survey, 2019-20, Q29; ECLS-K:2011 first-grade Spring Classroom Teacher Questionnaire, Teacher Level, QC2, ECLS-K:2011 second-grade Spring Teacher Questionnaire, Teacher Level, QC2. The unweighted sample size for SRCL teachers is 379 teachers of grade 1 and/or grade 2. The unweighted sample size for ECLS-K teachers is 13,260 grade 1 teachers and 12,141 grade 2 teachers. All ECLS results are from public-use data.

Exhibit D.7. Frequency of *asking students to write a narrative with two or more sequenced events* for ECLS-K teachers (2011–12, 2012–13) and SRCL teacher survey respondents (2019–20)



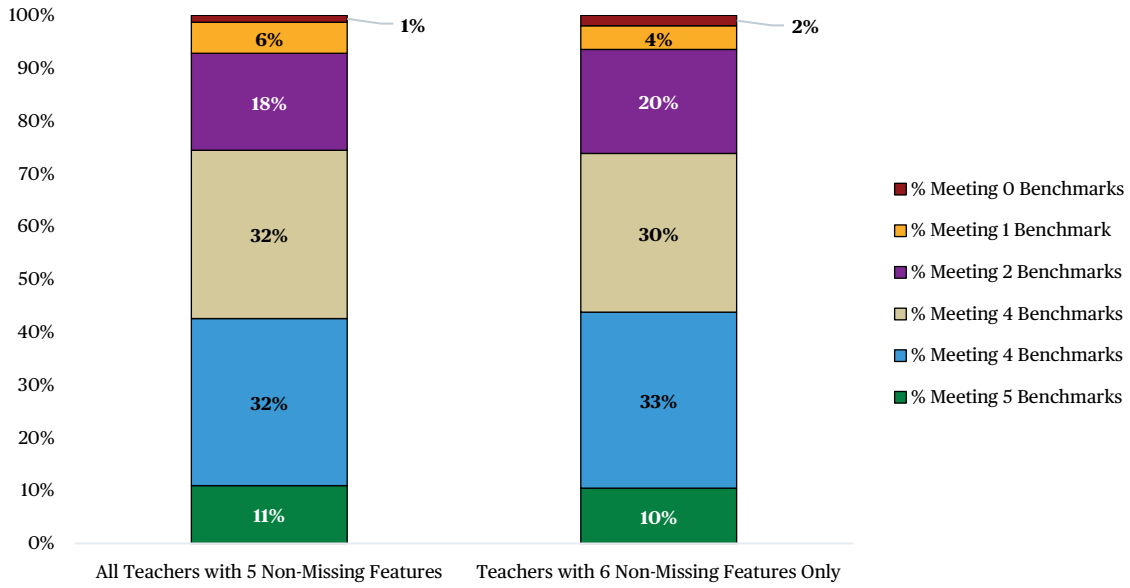
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL K-12 teacher survey, 2019-20, Q30; ECLS-K:2011 first-grade Spring Classroom Teacher Questionnaire, Teacher Level, QC2, ECLS-K:2011 second-grade Spring Teacher Questionnaire, Teacher Level, QC2.The unweighted sample size for SRCL teachers is 379 teachers of grade 1 and/or grade 2. The unweighted sample size for ECLS-K teachers is 13,236 grade 1 teachers and 12,176 grade 2 teachers. All ECLS results are from public-use data.

D.2.2. Sensitivity Analyses for Number of Benchmarks Met, Excluding Assignment of Age-Appropriate Novels

Exhibit 5 in the report presents the percentage of teachers in each grade band meeting six, five, four, three, two, one, and zero benchmarks for features of comprehensive literacy instruction. The sample for this exhibit consists of teachers with non-missing data on all six features. As shown in Exhibit B.12, many teachers in grades 4–12 do not have data on all six features, generally because they did not assign a novel with a Lexile level in the two weeks prior to taking the survey and thus do not have data on age-appropriate instruction.

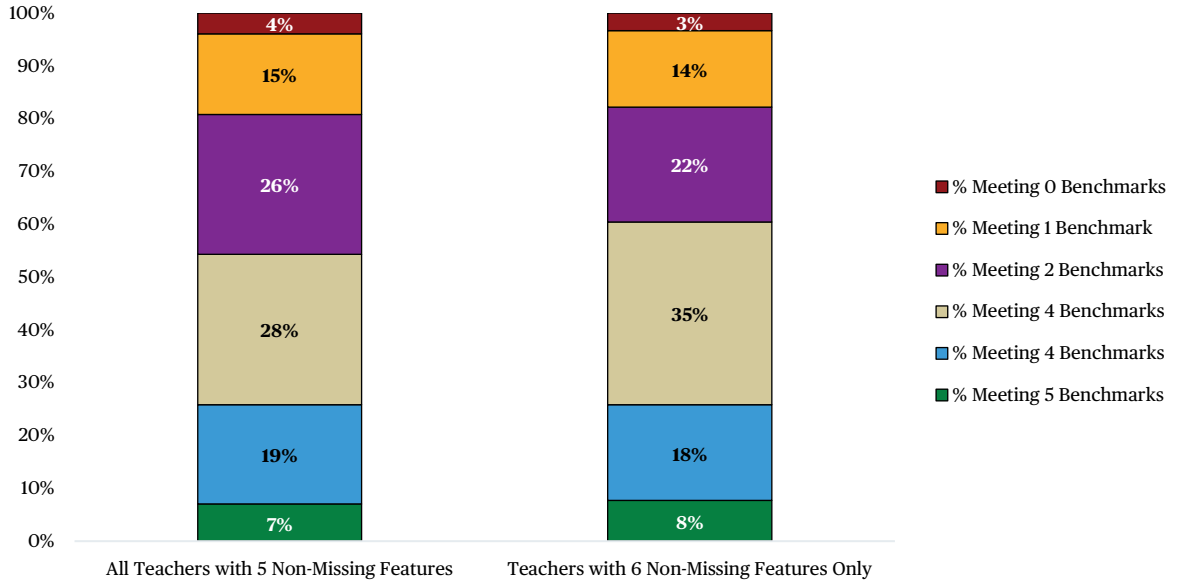
To examine whether the findings in Exhibit 5 were affected by restricting the sample to teachers with data on all six features, including age-appropriate instruction, we conducted a set of sensitivity analyses presented in Exhibits D.8-D.11. Exhibit D.8 presents the percentages of teachers in grades 4–8 in 2019-20 meeting five, four, three, two, one, and zero benchmarks for features of comprehensive literacy instruction, leaving out the benchmark for assignment of age-appropriate novels -- separately for all teachers with data on the remaining five features and for teachers with data on all six original features (including assignment of age-appropriate novels). Exhibit D.9 presents parallel results for grades 9-12. D.10 and D.11 present parallel results for 2018-19. In both grade bands in both years, the distribution of the number of benchmarks met is broadly similar regardless of whether teachers without data for assignment of age-appropriate novels are excluded from the sample.

Exhibit D.8. Percentage of teachers meeting benchmarks for 0–5 features of comprehensive literacy instruction, 2019–20, grades 4–8



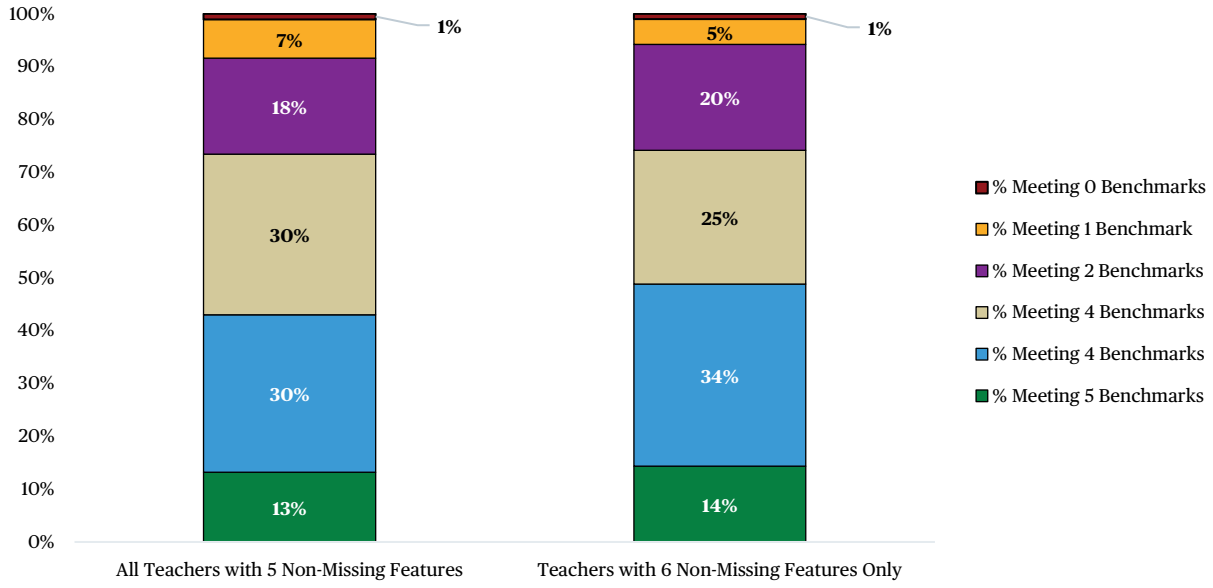
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL teacher survey, 2019–20. The unweighted sample size is 1,053 teachers with 5 non-missing features and 351 teachers with 6 non-missing features.

Exhibit D.9. Percentage of teachers meeting benchmarks for 0–5 features of comprehensive literacy instruction, 2019–20, grades 9–12



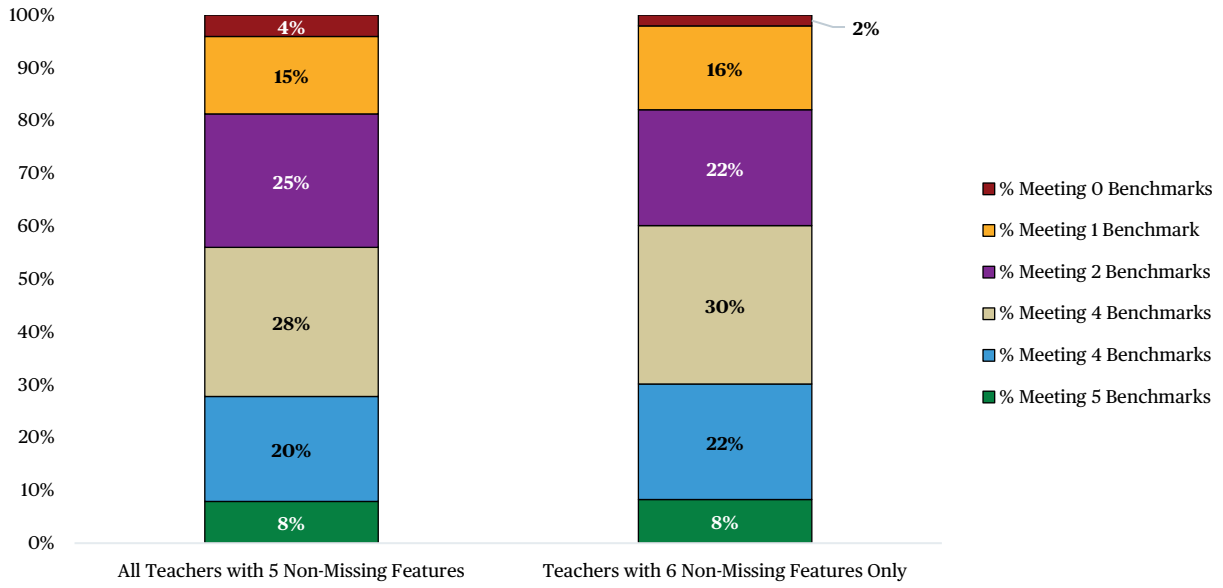
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL teacher survey, 2019–20. The unweighted sample size is 425 teachers with 5 non-missing features and 169 teachers with 6 non-missing features.

Exhibit D.10. Percentage of teachers meeting benchmarks for 0–5 features of comprehensive literacy instruction, 2018–19, grades 4–8



Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL teacher survey, 2018–19. The unweighted sample size is 877 teachers with 5 non-missing features and 327 teachers with 6 non-missing features.

Exhibit D.11. Percentage of teachers meeting benchmarks for 0–5 features of comprehensive literacy instruction, 2018–19, grades 9–12

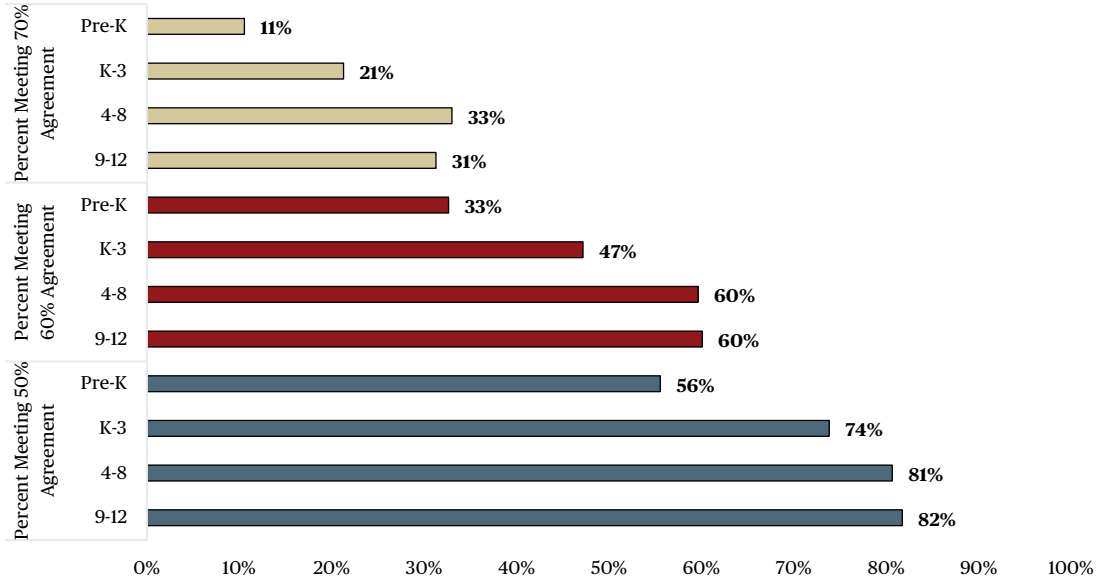


Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL teacher survey, 2018–19. The unweighted sample size is 379 teachers with 5 non-missing features and 158 teachers with 6 non-missing features.

D.2.3. Sensitivity Analyses Related to Thresholds for Expert-Scored Features

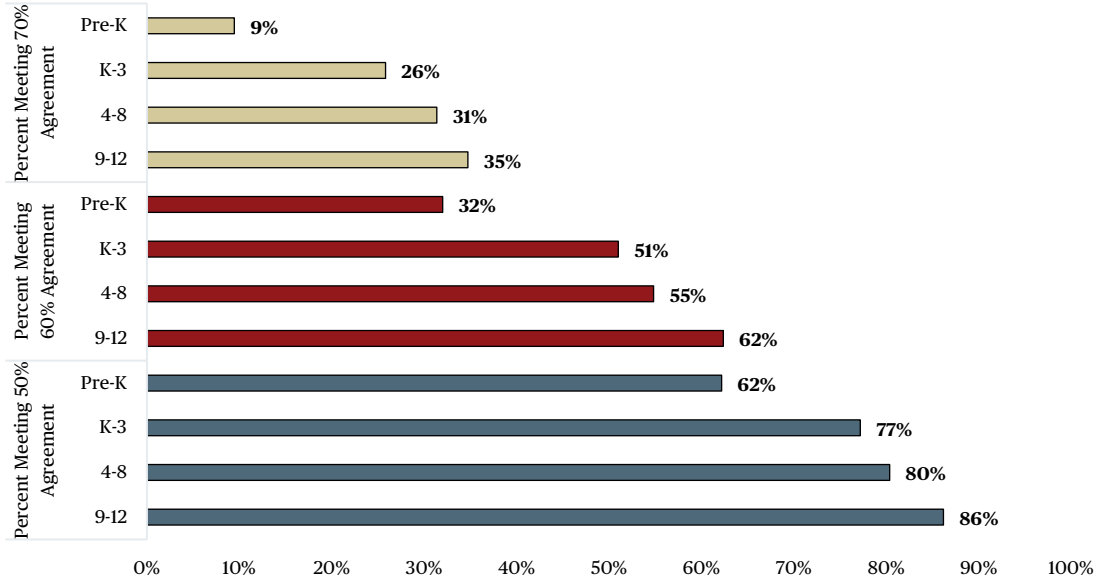
Two features of comprehensive literacy instruction, emphasis on key domains and explicit instruction, were scored based on how closely teachers' responses to survey questions matched the responses recommended by a panel of experts. Teachers were credited with meeting the benchmark for a feature if they received at least 60 percent of the maximum points available for that feature. To test the sensitivity of the results to the specific benchmark chosen, we conducted parallel analyses using thresholds of 50 and 70 percent. Exhibits D.12–D.15 below present the percentages of teachers meeting the benchmarks when the thresholds were set at 70 percent, 60 percent, and 50 percent of the possible points. As expected, fewer teachers met the benchmarks when higher numbers of points were required, and more teachers met the benchmarks when lower numbers of points were required. In general, however, the pattern across grade bands is similar for all three thresholds. For example, in general, more teachers meet the thresholds in grades 4-8 and 9-12 than Pre-K and K-3 for emphasis on key domains; while more meet the thresholds in Pre-K and K-3 for explicit instruction.

Exhibit D.12. Percentage of teachers by grade band meeting strict, standard, and lenient benchmarks for emphasis on key domains, 2018–19



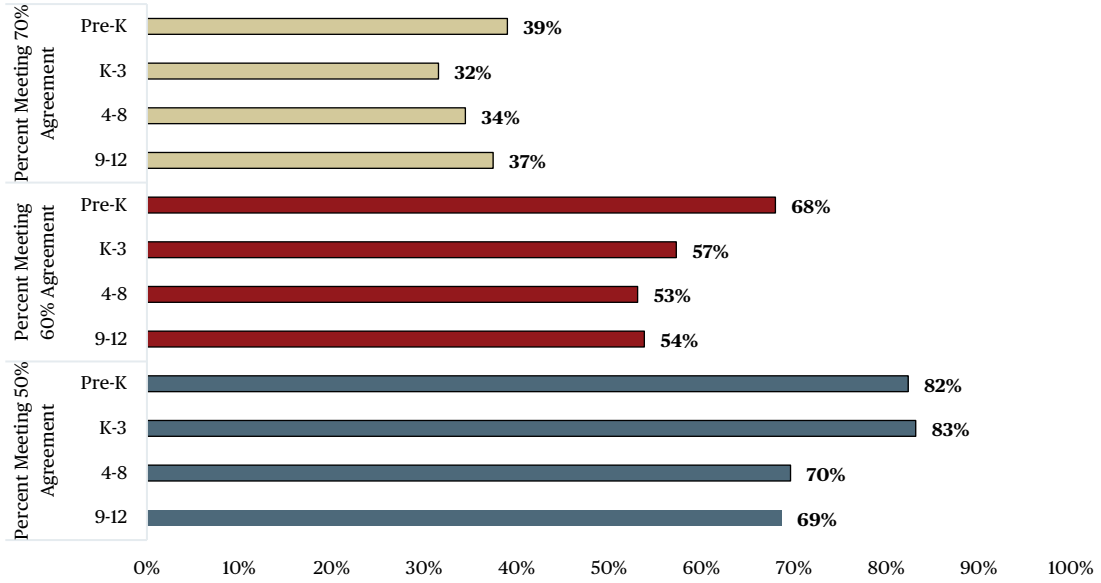
Source: SRCL teacher survey, 2018–19. The unweighted sample size is 285 pre-K teachers, 687 teachers in grades K–3, 980 teachers in grades 4–8, and 452 teachers in grades 9–12.

Exhibit D.13. Percentage of teachers by grade band meeting strict, standard, and lenient benchmarks for emphasis on key domains, 2019–20



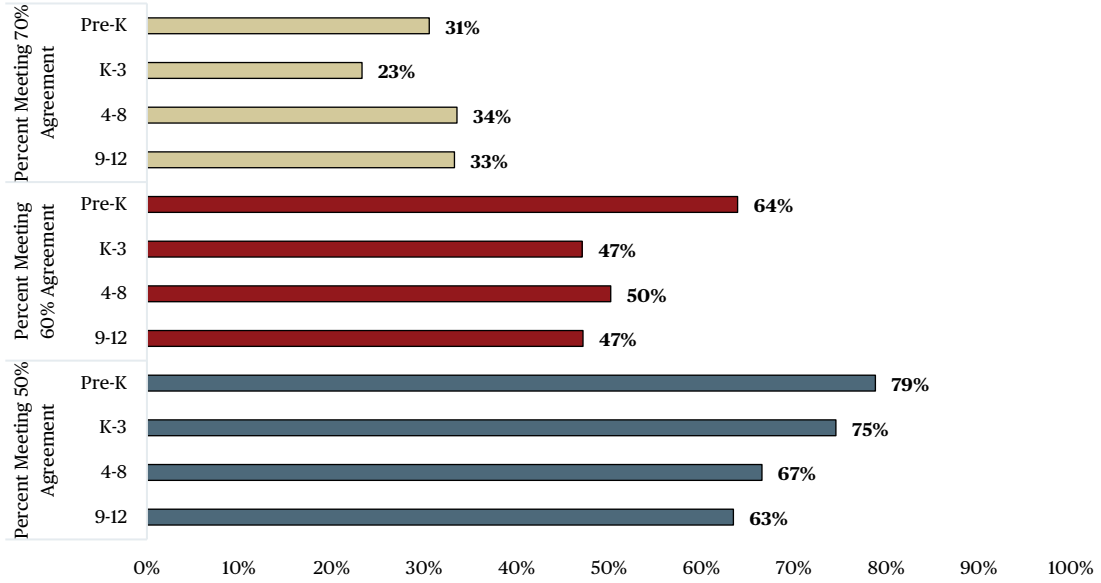
Source: SRCL teacher survey, 2019–20. The unweighted sample size is 332 pre-K teachers, 811 teachers in grades K–3, 1,171 teachers in grades 4–8, and 517 teachers in grades 9–12.

Exhibit D.14. Percentage of teachers by grade band meeting strict, standard, and lenient benchmarks for explicit instruction, 2018–19



Source: SRCL teacher survey, 2018–19. The unweighted sample size is 285 pre-K teachers, 659 teachers in grades K–3, 978 teachers in grades 4–8, and 452 teachers in grades 9–12.

Exhibit D.15. Percentage of teachers by grade band meeting strict, standard, and lenient benchmarks for explicit instruction, 2019–20

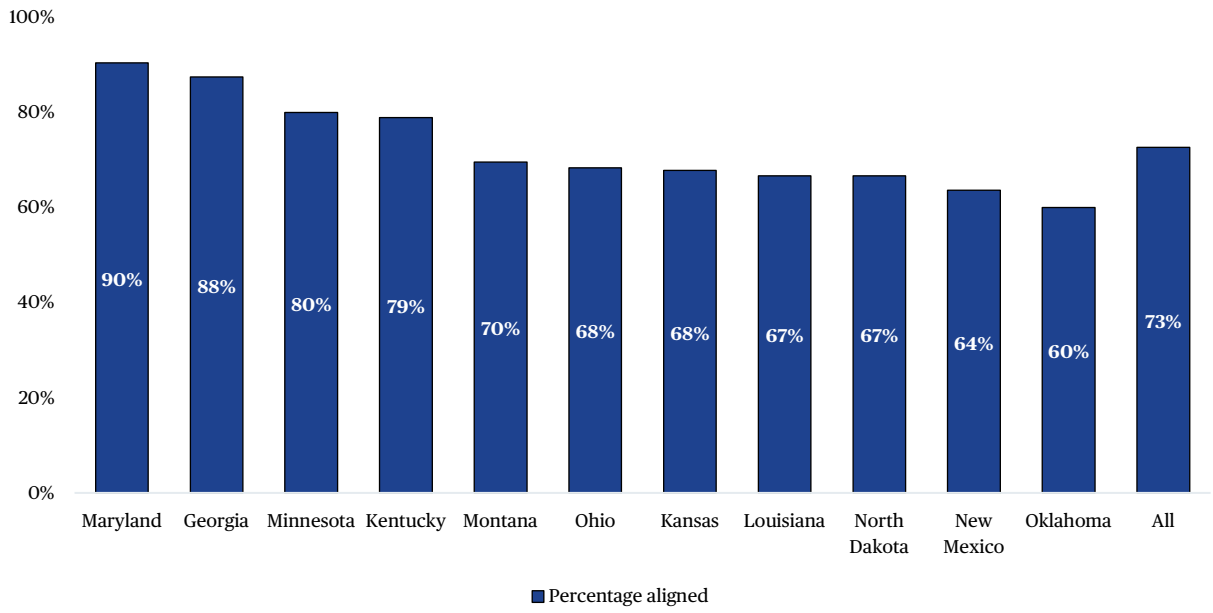


Source: SRCL teacher survey, 2019–20. The unweighted sample size is 332 pre-K teachers, 766 teachers in grades K–3, 1,169 teachers in grades 4–8, and 517 teachers in grades 9–12.

D.3. District Use of Needs Assessments for SRCL

SRCL required funded districts to align their work with state comprehensive literacy plans and local needs assessments. The study drew on data from the district survey to assess the extent to which districts' SRCL expenditures were aligned with their identified local needs. The survey asked districts to report the needs identified in their most recent needs assessment (choosing from a list provided in the survey), as well as planned SRCL expenditures. Districts were classified as having aligned their purchases with their needs if all of the top three priorities for SRCL funds were identified on their local needs assessment. If a district spent SRCL funds on fewer than three priorities, district were classified as having aligned expenditures if all of the SRCL-funded priorities were identified on the needs assessment. Exhibit D.16 reports on the percentage of districts in each state that were considered to have aligned their expenditures with their needs assessment. Overall, the majority of SRCL districts aligned their purchases with their identified needs.

Exhibit D.16. Percentage of districts using SRCL expenditures to address district need

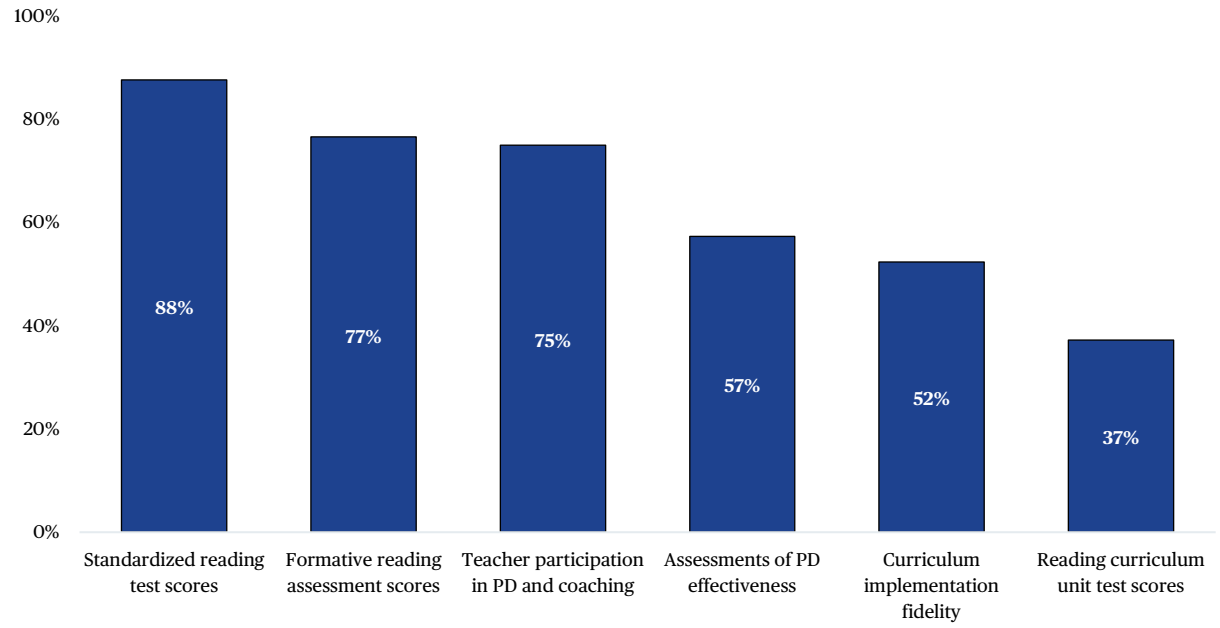


Source: SRCL district survey, Q11, Q13. The unweighted sample size is 380 districts.

D.4. District Use of Data for Continuous Improvement

States receiving SRCL funds were expected to use data to inform continuous improvement and decision making, as well as to consider applicants' plans for using data to track student outcomes when evaluating subgrant applications. To that end, the study's district survey asked districts about their use of various data types to change or improve their SRCL programs. The most commonly reported data used for program improvement included student performance data on standardized state and formative reading assessments (88 percent and 77 percent, respectively) and teacher participation in reading/ELA-related professional development or coaching (75 percent; Exhibit D.17).

Exhibit D.17. Percentage of districts using each type of data for continuous improvement

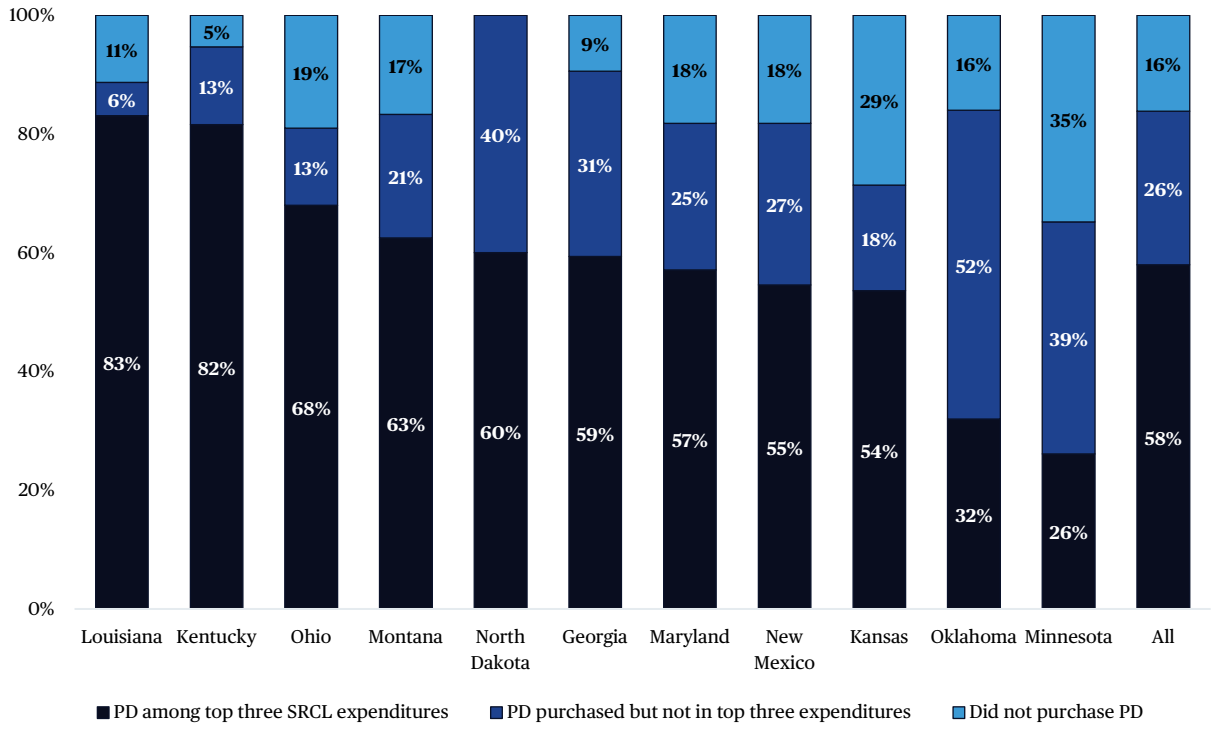


Note: PD is professional development.
Source: SRCL district survey, Q41. The unweighted sample size is 377 districts.

D.5. District Use of Funds for Professional Development

According to data from the district survey, professional development (PD) was the most frequently reported use of SRCL funds, with about 84 percent of districts purchasing PD. In all states except Oklahoma and Minnesota, professional development was among the top three SRCL expenditures in at least half of responding districts. In Louisiana and Kentucky, more than three quarters of responding districts reported that professional development was one of their three largest SRCL expenditures (Exhibit D.18).

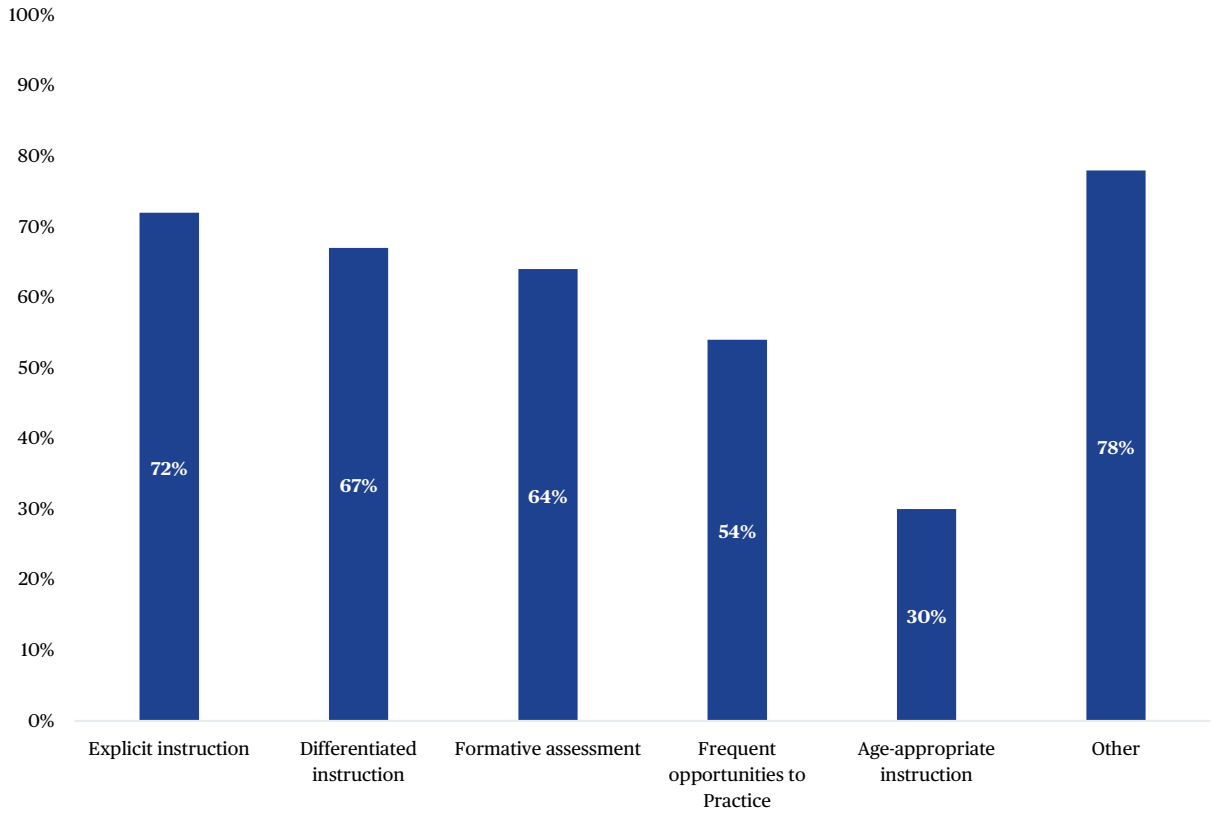
Exhibit D.18. Percentage of SRCL districts purchasing PD, by state



Note: The order of data in the exhibit matches the order of labels in the legend.
PD is professional development.
Source: SRCL district survey, 2018-19, Q13. The unweighted sample size is 388 districts.

Additionally, professional development activities covering explicit instruction (or emphasis on key domains; see the figure note for Exhibit D.19), differentiated instruction, formative assessment, and frequent opportunities to practice reading and/or writing were each provided by at least half of districts using SRCL funds, even when districts that did not fund professional development are included in the calculation. Exhibit D.19 below presents the percentage of SRCL districts that funded professional development covering each of the features of comprehensive literacy instruction, plus professional development that covered other topics not included in any of the features of comprehensive literacy instruction, such as classroom management and the use of technology.

Exhibit D.19. Percentage of SRCL districts covering features of comprehensive literacy instruction in their SRCL-funded professional development



Note: Districts can be included in more than one column if they provide professional development covering multiple features. "Explicit instruction" also includes professional development related to emphasis on key domains; "Frequent Opportunities to Practice" includes professional development related to frequent opportunities to practice reading and/or writing. Districts that did not fund professional development with SRCL money are included in the percentages but counted as not providing professional development in any of the features.
Source: SRCL district survey, 2018-19, Q31 The unweighted sample size is 389 districts.

D.6. Professional Development Related to Comprehensive Literacy Instruction

This section expands upon a report finding that participation in literacy-related professional development was associated with meeting benchmarks for comprehensive literacy instruction. Exhibit 5 in the report presents the percentage of teachers meeting from zero to six benchmarks corresponding to the features of comprehensive literacy instruction. The results show substantial variation in the number of benchmarks met. Exhibit 6 in the report provides detail about the percentage of teachers that met each of the individual benchmarks. One potential reason why some teachers were able to implement features of comprehensive literacy instruction and others were not may be teachers' varying exposure to relevant professional development. One might hypothesize that teachers who received professional development on how to implement a feature of comprehensive literacy instruction might be more likely to meet the benchmark for that feature, drawing on the knowledge gained from the professional development.

To examine the association between professional development and the implementation of comprehensive literacy instruction, the study team conducted supplemental analyses for each of the six features of comprehensive literacy instruction. For each feature, the analysis compared teachers who received professional development related to the feature with teachers who did not.

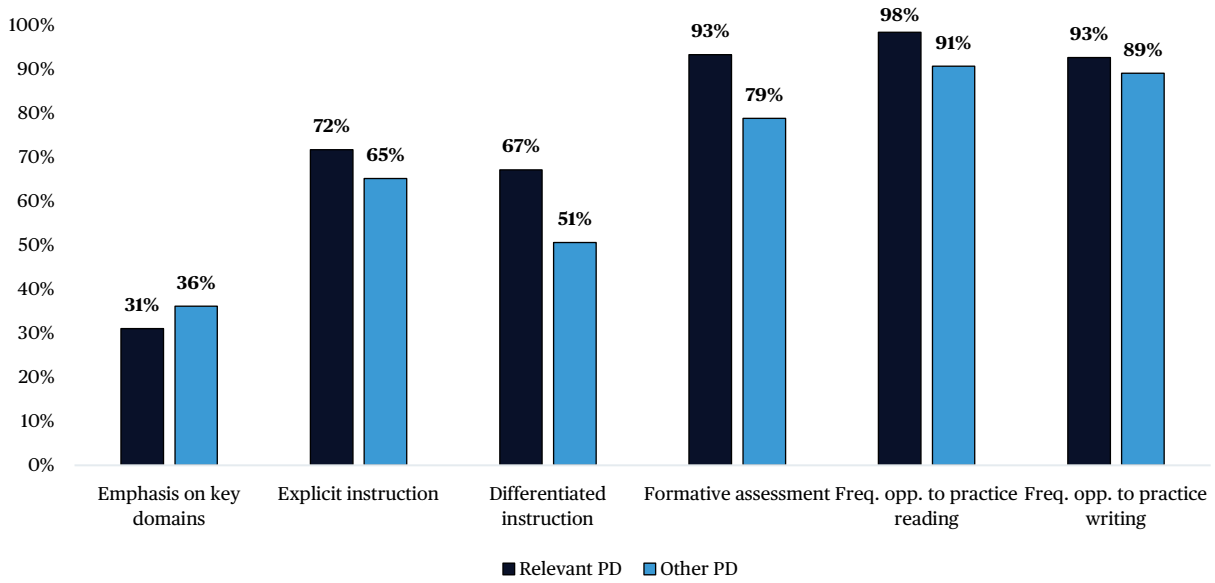
Professional development was classified as relevant based on the topics that teachers said were covered in their professional development on questions 47 and 48 of the K–12 teacher survey or questions 25 and 26 of the pre-K teacher survey. For emphasis on key domains and explicit instruction, to be classified as completing relevant professional development, a teacher needed to receive professional development covering more than half of the topics relevant to emphasis on key domains and explicit instruction in their grade band; for all other features, a teacher needed to receive professional development covering at least one topic relevant to that feature. Topics related to emphasis on key domains and explicit instruction correspond to the domains in which the expert panel (described in section B.3.1 and B.3.2) recommended the highest level of emphasis for a given grade band. The testing method outlined in section B.4 was used to test whether teachers whose professional development was relevant to a particular feature were more likely to achieve the benchmark for that feature than teachers whose professional development was not related to that feature. Exhibit D.20 below presents the topic areas defined as relevant to each feature of comprehensive literacy instruction.

Exhibit D.20. Professional development topics defined as relevant to each feature of comprehensive literacy instruction

Professional Development Topic	Feature of Comprehensive Literacy Instruction
Decoding (e.g., phonological awareness, phonics skills)	Emphasis on key domains and explicit instruction (pre-K through second grade)
Reading fluency	Emphasis on key domains and explicit instruction (first grade through third grade)
Vocabulary	Emphasis on key domains and explicit instruction (all grades)
Reading comprehension (e.g., comprehension strategies)	Emphasis on key domains and explicit instruction (second grade through eighth grade)
Oral language (listening comprehension, speaking)	Emphasis on key domains and explicit instruction (pre-K through second grade)
Writing	Emphasis on key domains and explicit instruction (third through eighth grade)
Handwriting	None
Grammar, punctuation, and capitalization (conventions of English)	Emphasis on key domains and explicit instruction (third grade)
Spelling	Emphasis on key domains and explicit instruction (second and third grade)
Critical analysis of literary text	Emphasis on key domains and explicit instruction (6th through 12th grade)
Critical analysis of informational text	Emphasis on key domains and explicit instruction (6th through 12th grade)
How to use evidence-based reading programs	None
How to choose literary text for instruction (e.g., how to select text at the right level of difficulty, with an appropriate focus)	Age-appropriate instruction
How to choose informational text for instruction (e.g., how to select text at the right level of difficulty, with an appropriate focus)	None (the age-appropriate instruction feature only measures the age appropriateness of novels)
How to integrate reading and writing instruction	Frequent opportunities to practice reading and writing
How to organize small group instruction	Differentiated instruction
How to identify reading problems	Formative assessment
How to interpret and use assessment data to guide instruction	Formative assessment
How to help struggling readers	None
How to motivate readers	None
Strategies for teaching reading/ELA to English learners (ELs)	Differentiated instruction
Strategies for teaching reading/ELA to students with disabilities (including students with individualized education programs and those with 504 plans)	Differentiated instruction

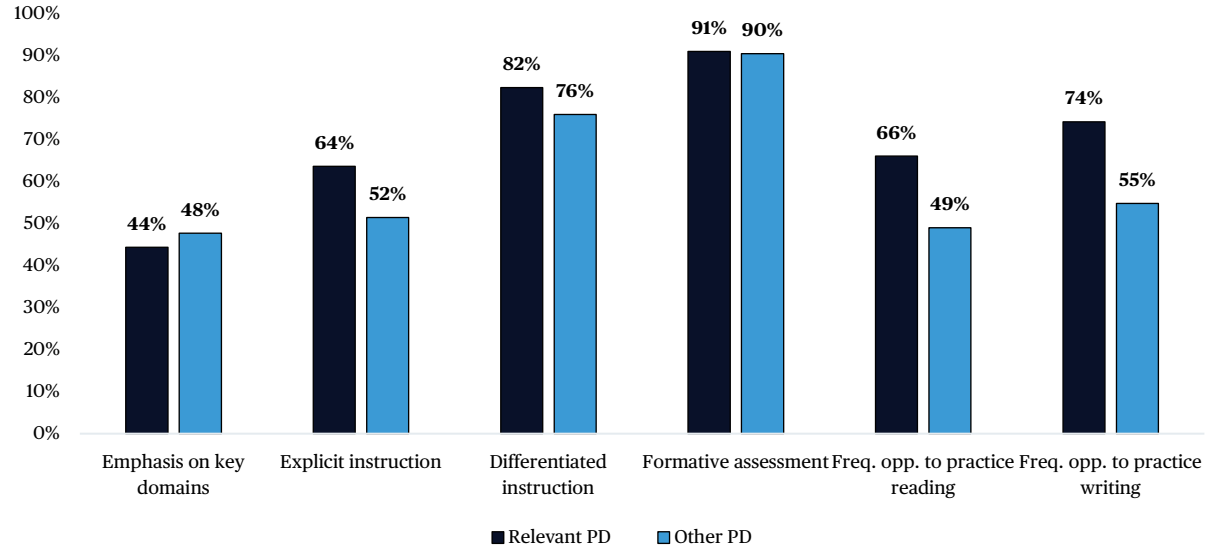
Exhibits D.21–D.28 present the survey-weighted percentage of teachers meeting the benchmarks for each feature for teachers who received professional development relevant to that feature, with separate figures by grade band and year. The percentage of teachers who received unrelated ("other") professional development is also shown. Few teachers reported receiving no professional development; teachers who received no professional development in the year that they took the survey are excluded from the sample in the figure. Teachers are included in the sample for all features for which they have outcome data in the given year.

Exhibit D.21. Percentage of teachers meeting each benchmark, by professional development received, pre-K, 2018–19



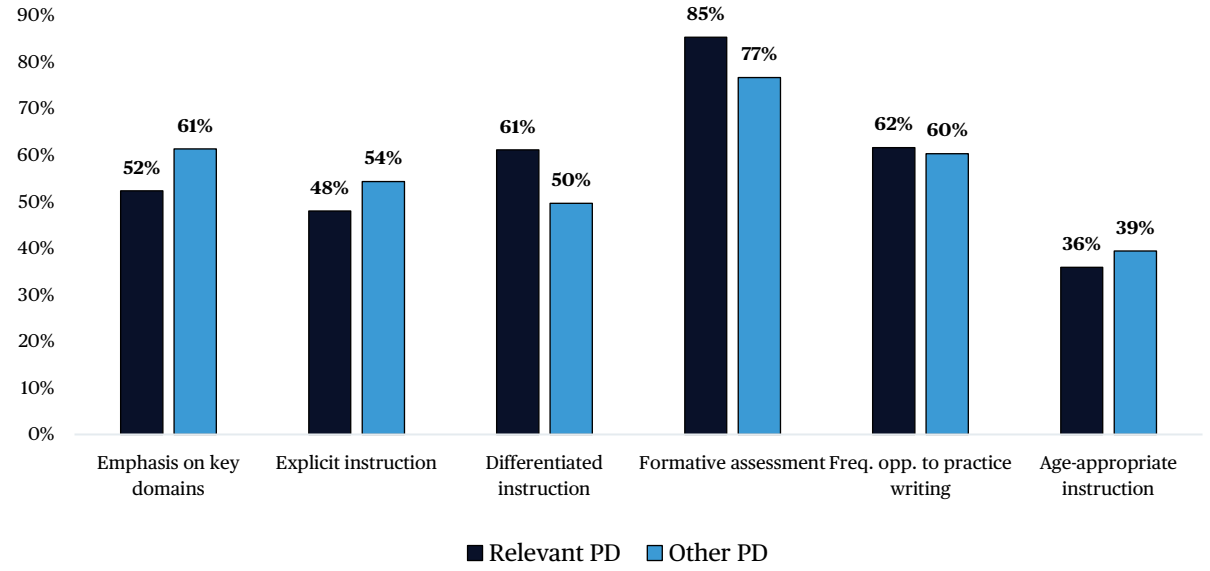
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL pre-K teacher survey, 2018-19. The unweighted sample size for all features *except* differentiated instruction is 285 pre-K teachers. The unweighted sample size for differentiated instruction is 238 pre-K teachers.

Exhibit D.22. Percentage of teachers meeting each benchmark, by professional development received, grades K–3, 2018–19



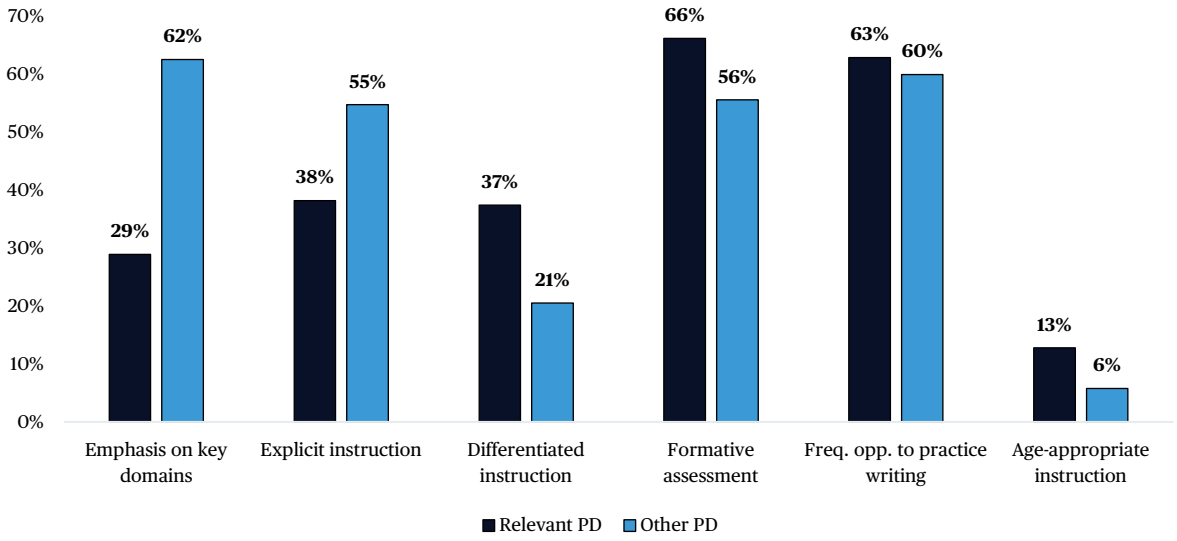
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2018–19. The unweighted sample size ranged from 657 to 687 teachers of grades K-3.

Exhibit D.23. Percentage of teachers meeting each benchmark, by professional development received, grades 4–8, 2018–19



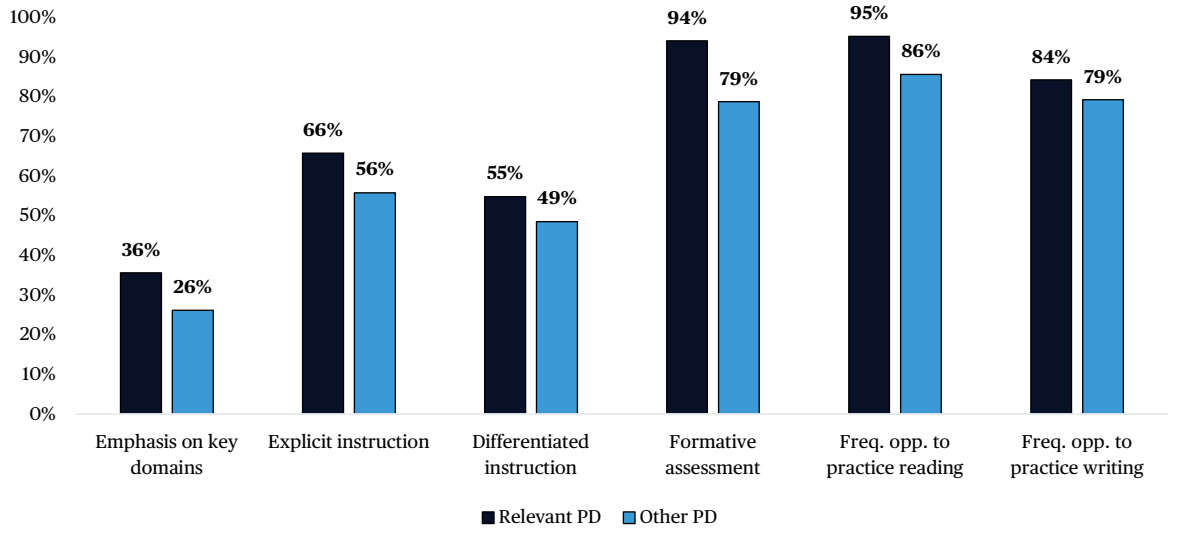
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2018–19. The unweighted sample size for the first five features of literacy instruction ranges from 923 to 980 teachers of grades 4-8. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 362 teachers.

Exhibit D.24. Percentage of teachers meeting each benchmark, by professional development received, grades 9–12, 2018–19



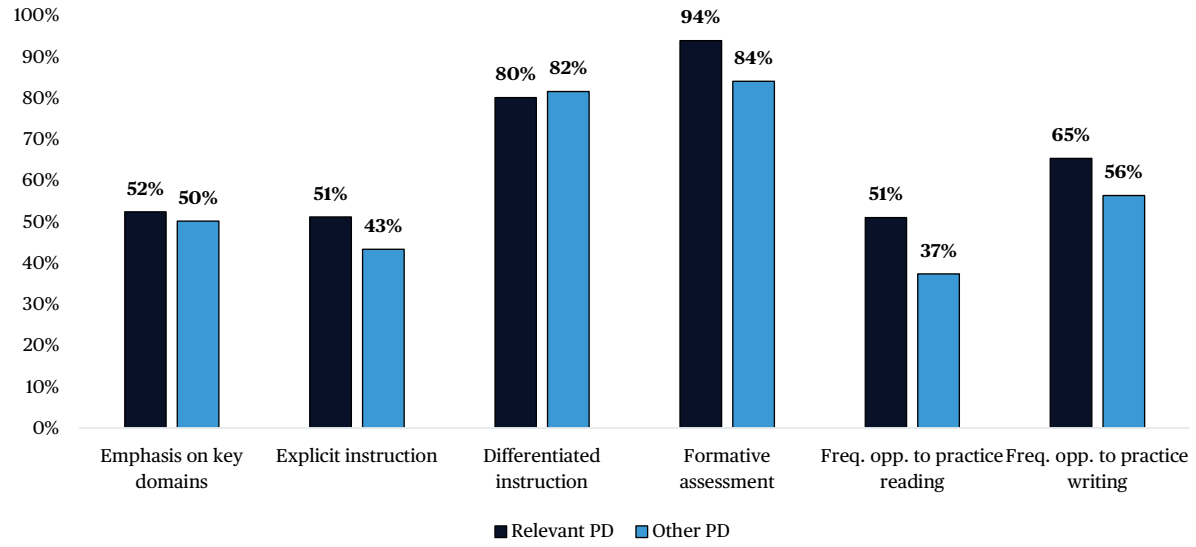
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2018–19. The unweighted sample size for the first five features of literacy instruction ranges from 385 to 452 teachers of grades 9-12. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 196 teachers.

Exhibit D.25. Percentage of teachers meeting each benchmark, by professional development received, pre-K, 2019–20



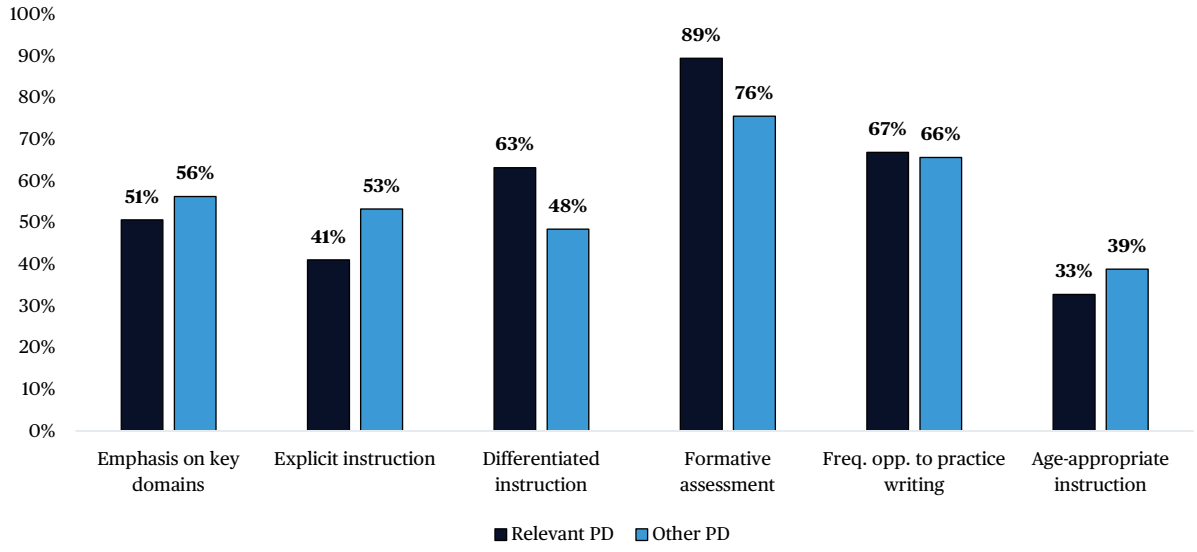
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL pre-K teacher survey, 2019–20. The unweighted sample size for all features *except* differentiated instruction is 332 pre-K teachers. The unweighted sample size for differentiated instruction is 284 pre-K teachers.

Exhibit D.26. Percentage of teachers meeting each benchmark, by professional development received, grades K–3, 2019–20



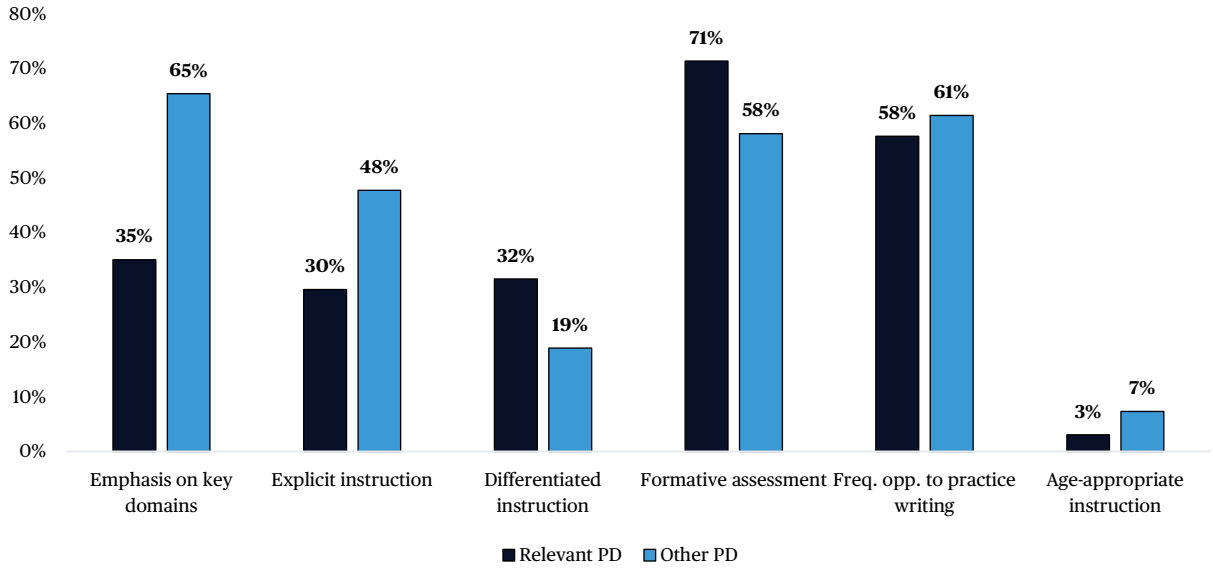
Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2019–20. The unweighted sample ranged from 765 to 881 teachers of grades K-3.

Exhibit D.27. Percentage of teachers meeting each benchmark, by professional development received, grades 4–8, 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2019–20. The unweighted sample size for the first five features of literacy instruction ranges from 1,108 to 1,171 teachers of grades 4-8. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 392 teachers.

Exhibit D.28. Percentage of teachers meeting each benchmark, by professional development received, grades 9–12, 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend. PD is professional development.
Source: SRCL K-12 teacher survey, 2019–20. The unweighted sample size for the first five features of literacy instruction ranges from 432 to 517 teachers of grades 9-12. The sample for age-appropriate instruction is restricted to teachers who assigned a novel in the previous two weeks for which a Lexile level could be assigned. The unweighted sample size for age-appropriate instruction is 208 teachers.

Exhibit D.29 below identifies the statistically significant differences, separately by year, in the fraction of teachers meeting benchmarks between teachers receiving relevant professional development and teachers receiving professional development unrelated to the feature being measured. In 12 of the tests, the percentage of teachers meeting the benchmark is statistically significantly higher for teachers who participated in relevant professional development than those who participated in other professional development; in three of the tests, the percentage of teachers meeting the benchmark is statistically significantly higher for teachers who participated in other professional development. In the remaining tests, the results were not statistically significant.

Exhibit D.29. Statistically significant differences in the percentage of teachers meeting benchmarks, by professional development relevance

Feature	Pre-K	Grades K–3	Grades 4–8	Grades 9–12
Formative assessment	Year 2: relevant PD > other PD	Year 2: relevant PD > other PD	Years 1 and 2: relevant PD > other PD	Years 1 and 2: relevant PD > other PD
Explicit instruction	–	Years 1 and 2: relevant PD > other PD	Year 2: other PD > relevant PD	Year 2: other PD > relevant PD
Emphasis on key domains	–	–	–	Year 2: other PD > relevant PD
Frequent opportunities to practice reading	Year 2: relevant PD > other PD	Years 1 and 2: relevant PD > other PD	n/a	n/a
Frequent opportunities to practice writing	–	Years 1 and 2: relevant PD > other PD	–	–
Differentiated instruction	Year 1: relevant PD > other PD	–	Years 1 and 2: relevant PD > other PD	Years 1 and 2: relevant PD > other PD
Age-appropriate instruction	n/a	n/a	–	Year 1: relevant PD > other PD

Note: PD is professional development. "X > Y" means that teachers who received professional development of type X were significantly more likely to meet the given benchmark than teachers who received professional development of type Y. "n/a" indicates that data were not available to report on the feature of comprehensive literacy instruction for the indicated grade level. "–" indicates no statistically significant differences.
Source: SRCL teacher surveys, 2018–19 and 2019–20. Sample sizes vary by feature. For features relevant to teachers in all grades, including formative assessment, explicit instruction, emphasis on key domains, frequent opportunities to practice writing, and differentiated instruction, sample sizes range from 2,203 to 2,404 teachers in 2018–19 and from 2,597 to 2,831 teachers in 2019–20. Only teachers in pre-K through grade 3 were assessed on frequent opportunities to practice reading; sample sizes for frequent opportunities to practice writing are 944 teachers in 2018–19 and 1,097 teachers in 2019–20. Only teachers in grade 4 through grade 12 who assigned a novel with a Lexile level in the two weeks prior to taking the survey were assessed on age-appropriate instruction; sample sizes for age-appropriate instruction are 558 teachers in 2018–19 and 600 teachers in 2019–20.

Exhibits D.30 and D.31 present the percentage of teachers receiving relevant professional development, other professional development, and no professional development for each feature, broken out by grade band, separately by year. The sample for each cell includes only teachers who have nonmissing outcome data for the relevant feature, so the total sample sizes may differ slightly between features within a year and grade band. For instance, teachers who did not assign a novel are excluded from the sample for age-appropriate instruction.

Exhibit D.30. Number of teachers by professional development exposure for each feature, 2018–19

Feature	Pre-K	Grades K–3	Grades 4–8	Grades 9–12
Formative assessment	Relevant PD: 52% Other PD: 44% No PD: 4%	Relevant PD: 78% Other PD: 18% No PD: 3%	Relevant PD: 77% Other PD: 18% No PD: 5%	Relevant PD: 63% Other PD: 29% No PD: 8%
Explicit instruction	Relevant PD: 71% Other PD: 25% No PD: 4%	Relevant PD: 51% Other PD: 45% No PD: 4%	Relevant PD: 28% Other PD: 67% No PD: 6%	Relevant PD: 5% Other PD: 87% No PD: 8%
Emphasis on key domains	Relevant PD: 71% Other PD: 25% No PD: 4%	Relevant PD: 50% Other PD: 47% No PD: 3%	Relevant PD: 27% Other PD: 67% No PD: 6%	Relevant PD: 5% Other PD: 87% No PD: 8%
Frequent opportunities to practice reading and/or writing	Relevant PD: 54% Other PD: 42% No PD: 4%	Relevant PD: 65% Other PD: 31% No PD: 4%	Relevant PD: 68% Other PD: 27% No PD: 5%	Relevant PD: 67% Other PD: 24% No PD: 8%
Differentiated instruction	Relevant PD: 73% Other PD: 24% No PD: 3%	Relevant PD: 74% Other PD: 22% No PD: 3%	Relevant PD: 71% Other PD: 24% No PD: 5%	Relevant PD: 64% Other PD: 28% No PD: 8%
Age-appropriate instruction	n/a	n/a	Relevant PD: 46% Other PD: 49% No PD: 5%	Relevant PD: 37% Other PD: 60% No PD: 4%

Note: PD is professional development. Only teachers with nonmissing data for a feature are included in the sample related to that feature. Percentages may not add to 100 percent due to rounding. "n/a" indicates that data were not available to report on the feature of comprehensive literacy instruction for the indicated grade level.
Source: SRCL teacher surveys, 2018–19. Sample sizes vary by feature. For features relevant to teachers in all grades, including formative assessment, explicit instruction, emphasis on key domains, frequent opportunities to practice writing, and differentiated instruction, sample sizes range from 2,203 to 2,404 teachers. Only teachers in pre-K through grade 3 were assessed on frequent opportunities to practice reading; the sample size for frequent opportunities to practice writing is 944 teachers. Only teachers in grade 4 through grade 12 who assigned a novel with a Lexile level in the two weeks prior to taking the survey were assessed on age-appropriate instruction; the sample size for age-appropriate instruction is 558 teachers.

Exhibit D.31. Number of teachers by professional development exposure for each feature, 2019–20

Feature	Pre-K	Grades K–3	Grades 4–8	Grades 9–12
Formative assessment	Relevant PD: 50% Other PD: 40% No PD: 11%	Relevant PD: 80% Other PD: 17% No PD: 3%	Relevant PD: 77% Other PD: 19% No PD: 5%	Relevant PD: 71% Other PD: 23% No PD: 6%
Explicit instruction	Relevant PD: 64% Other PD: 25% No PD: 11%	Relevant PD: 51% Other PD: 46% No PD: 3%	Relevant PD: 26% Other PD: 69% No PD: 5%	Relevant PD: 6% Other PD: 88% No PD: 6%
Emphasis on key domains	Relevant PD: 64% Other PD: 25% No PD: 11%	Relevant PD: 50% Other PD: 47% No PD: 3%	Relevant PD: 26% Other PD: 69% No PD: 5%	Relevant PD: 6% Other PD: 88% No PD: 6%
Frequent opportunities to practice reading and/or writing	Relevant PD: 57% Other PD: 33% No PD: 11%	Relevant PD: 65% Other PD: 32% No PD: 3%	Relevant PD: 71% Other PD: 25% No PD: 5%	Relevant PD: 70% Other PD: 24% No PD: 6%
Differentiated instruction	Relevant PD: 63% Other PD: 26% No PD: 11%	Relevant PD: 77% Other PD: 20% No PD: 3%	Relevant PD: 71% Other PD: 24% No PD: 5%	Relevant PD: 68% Other PD: 26% No PD: 6%
Age-appropriate instruction	n/a	n/a	Relevant PD: 46% Other PD: 49% No PD: 5%	Relevant PD: 43% Other PD: 52% No PD: 5%

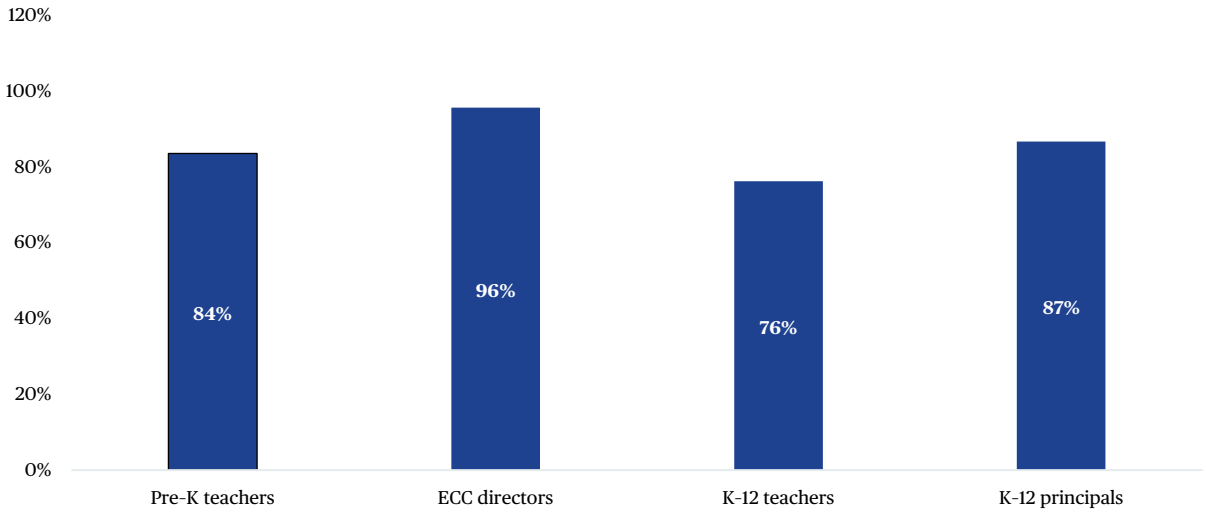
Note: PD is professional development. Only teachers with nonmissing data for a feature are included in the sample related to that feature. Percentages may not add to 100 percent due to rounding. "n/a" indicates that data were not available to report on the feature of comprehensive literacy instruction for the indicated grade level.
Source: SRCL teacher surveys, 2019–20. Sample sizes vary by feature. For features relevant to teachers in all grades, including formative assessment, explicit instruction, emphasis on key domains, frequent opportunities to practice writing, and differentiated instruction, sample sizes range from 2,597 to 2,831 teachers. Only teachers in pre-K through grade 3 were assessed on frequent opportunities to practice reading; the sample size for frequent opportunities to practice writing is 1,097 teachers. Only teachers in grade 4 through grade 12 who assigned a novel with a Lexile level in the two weeks prior to taking the survey were assessed on age-appropriate instruction; the sample size for age-appropriate instruction is 600 teachers.

D.7. Teachers' and School Leaders' Reports of the Usefulness of Professional Development

According to the teacher and school leader surveys, most teachers and leaders in SRCL schools who received professional development related to literacy instruction indicated that they perceived it to be useful. Exhibit D.32 describes the percentage of leaders of SRCL schools that participated in at least one professional development activity they described as useful, as well as the percentage of pre-K and K-12

teachers who agreed or strongly agreed that the professional development and coaching they received "better prepared [them] to teach all of the components of reading/ELA."[30]

Exhibit D.32. Percentage of SRCL teachers and school leaders describing PD as useful



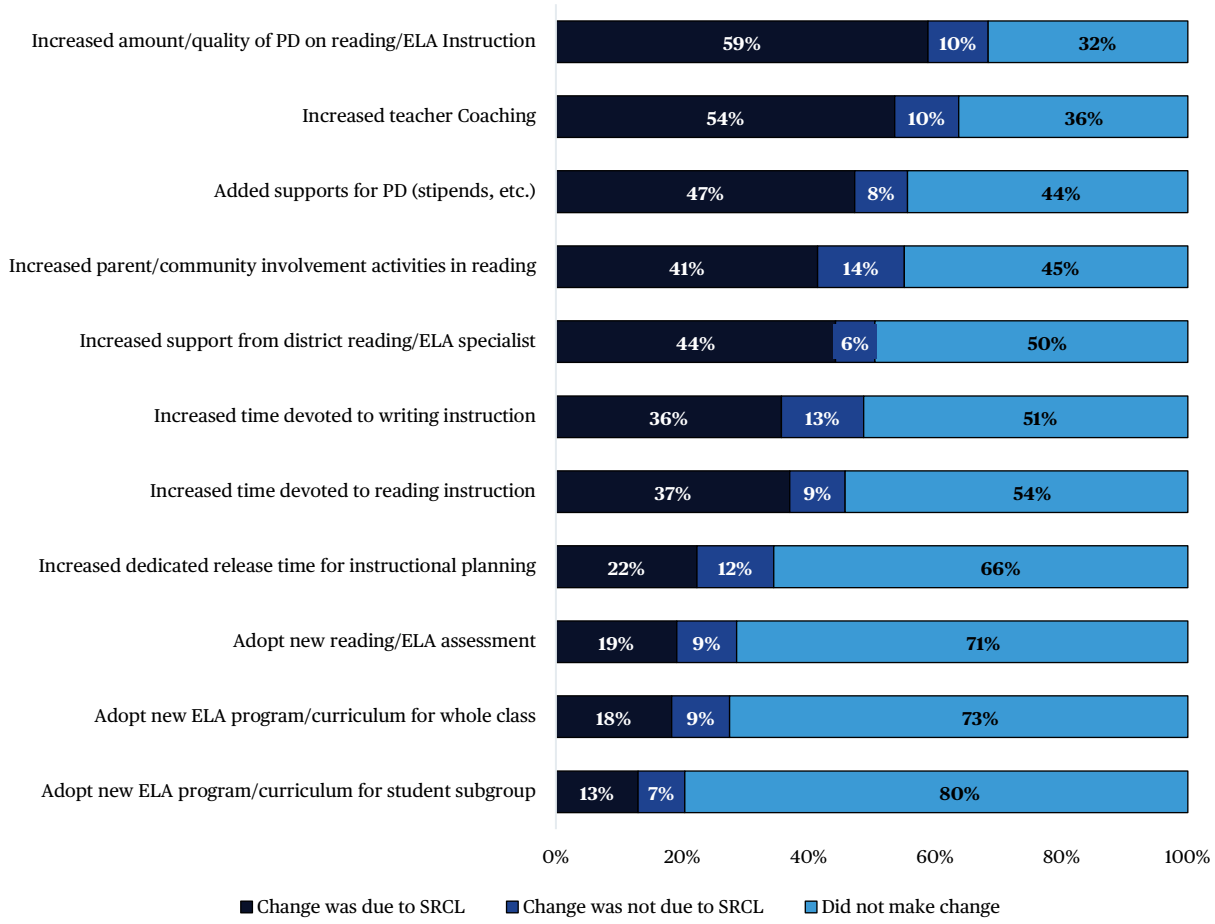
Note: PD is professional development.
Source: SRCL pre-K teacher survey, 2019-20, Q30; SRCL school leader survey, 2018-19, early childhood center director Q37; SRCL K-12 teacher survey, 2019-20, Q51; SRCL school leader survey, 2018-19, K-12 principal Q19. The unweighted sample size is 331 pre-K teachers, 93 early childhood center directors, 2,495 K-12 teachers, and 353 K-12 principals.

D.8. School Leader Reports of Changes Due to SRCL

The school leader survey asked principals in SRCL schools and directors of early childhood centers to report on changes that occurred in their schools in the last year (see Exhibits D.33 and D.34 for a complete list of these changes). For each of the changes to which the school leaders responded "yes," they were then asked whether those changes were due to their SRCL program. For each change cited, the majority of school leaders reporting that change attributed it to SRCL. The most commonly reported changes among K-12 SRCL schools were increased amount or quality of professional development relating to reading/ELA instruction (69 percent) and increased teacher coaching (64 percent) (Exhibit D.33). For early childhood centers, the most commonly reported changes were increased parent and community involvement activities (76 percent) and increased amount or quality of professional development relating to reading/ELA instruction (68 percent; Exhibit D.34).

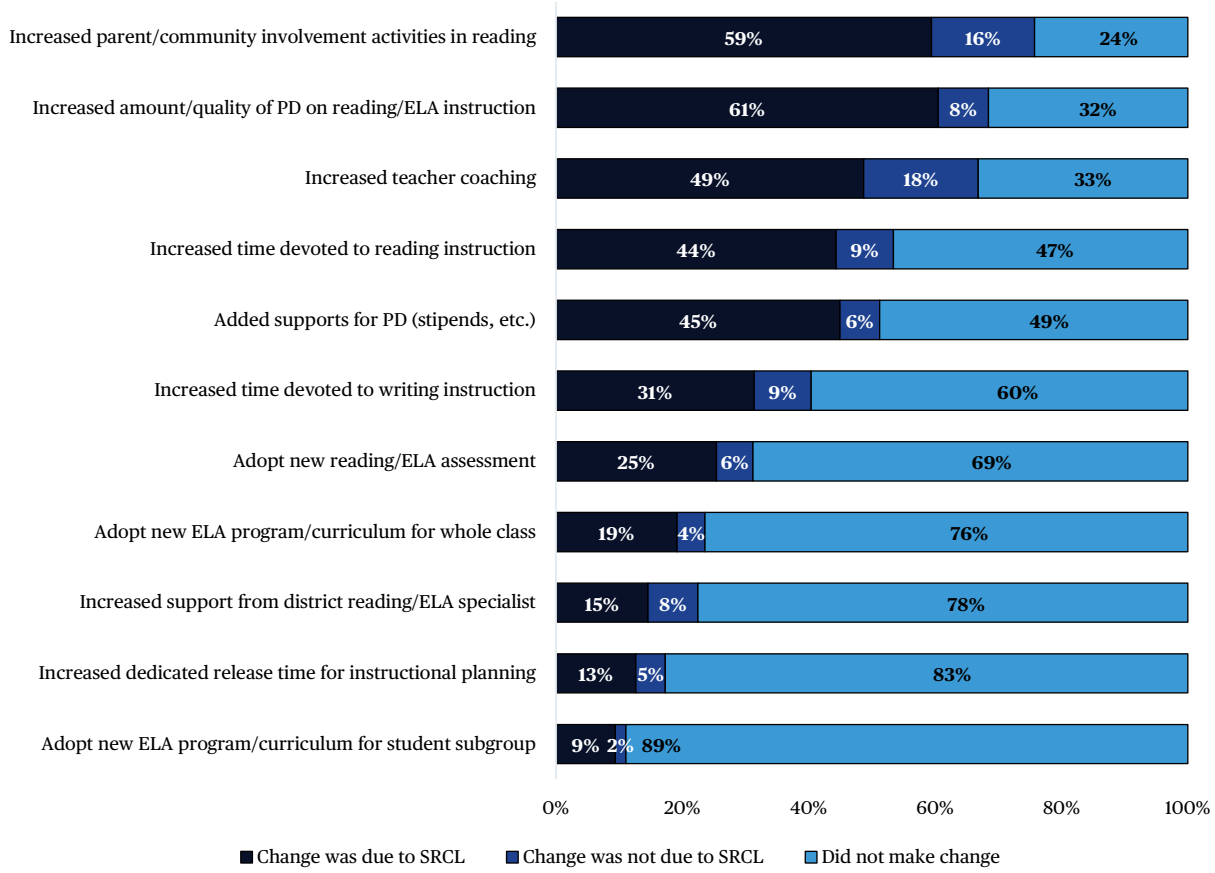
[30] A vast majority of survey respondents received some form of professional development; 88 percent of school leaders reported receiving professional development related to literacy in 2018-19, while 89 percent of pre-K teachers reported receiving professional development related to literacy in 2019-20, and between 94 percent and 97 percent of teachers in any K-12 grade band reported receiving professional development related to literacy in 2019-20.

Exhibit D.33. Percentage of school leaders reporting school changes due and not due to SRCL



Note: The order of data in the exhibit matches the order of labels in the legend. ELA is English language arts. PD is professional development.
Source: SRCL school leader survey, 2018-19, Q23. The unweighted sample size is 350 K-12 principals.

Exhibit D.34. Percentage of early childhood center directors reporting school changes due and not due to SRCL



Note: The order of data in the exhibit matches the order of labels in the legend. ELA is English language arts.
PD is professional development.
Source: SRCL school leader survey, 2018-19, early childhood center director Q43. The unweighted sample size is 93 early childhood center directors.

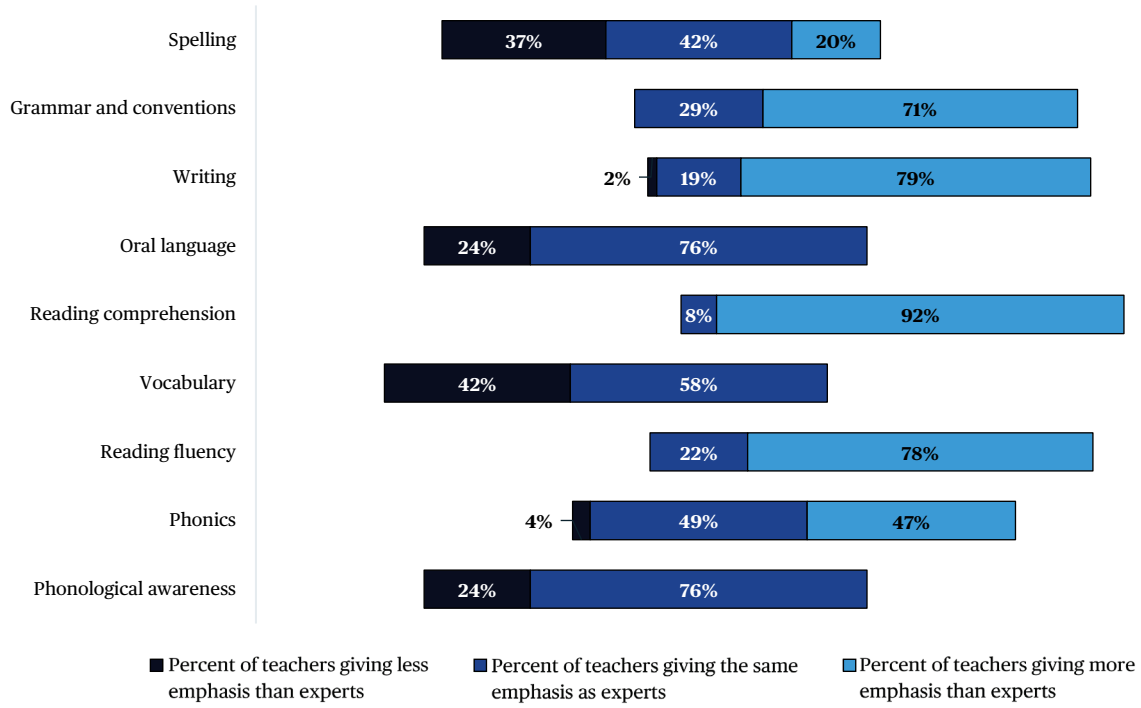
D.9. Level of Emphasis on Key Domains of Reading

As described in section B.3, the study drew on five experts in reading instruction to set the benchmark to measure teachers' emphasis on key domains of reading. The team asked the experts to indicate the emphasis that should be given to each domain for teachers teaching prekindergarten, kindergarten, and grades, 1, 2, 3, 4-5, 6-8, and 9-12, basing their judgment on their analysis of the available research evidence. In doing so, the experts observed that some domains should be given less emphasis in the earlier grades (such as analysis of informational text) while others should be given less emphasis in high school grades (decoding, for example). Thus, it is possible for a teacher not to meet the benchmark for emphasis on key domains by either giving moderate or major emphasis to a domain or strategy on which the experts placed little emphasis or for giving no or little emphasis to a domain that the experts thought should be emphasized for a given grade level.

This analysis provides information on whether teachers who did not meet the benchmark for emphasis on key domains gave less emphasis than was recommended by the experts, or more emphasis. Exhibits D.35–D.38 present the survey-weighted percentage of teachers in each grade band in 2019–20 who gave more emphasis, the same emphasis, or less emphasis than the experts for each domain of literacy. For example, as shown in Exhibit B.23, a teacher in grades 9–12 who placed moderate emphasis on reading comprehension would be classified as aligned with the experts, whereas a teacher in grades 9–12 who placed major emphasis on reading comprehension would be classified as giving more emphasis to reading comprehension than the experts, and a teacher in grades 9–12 who placed little or no emphasis on reading comprehension would be classified as giving less emphasis to reading comprehension than the experts. For some domains in some grade bands, the experts recommended either no emphasis or major emphasis; in the former case, teachers could not be classified as giving less emphasis than the experts, and in the latter case, teachers could not be classified as giving more emphasis than the experts.

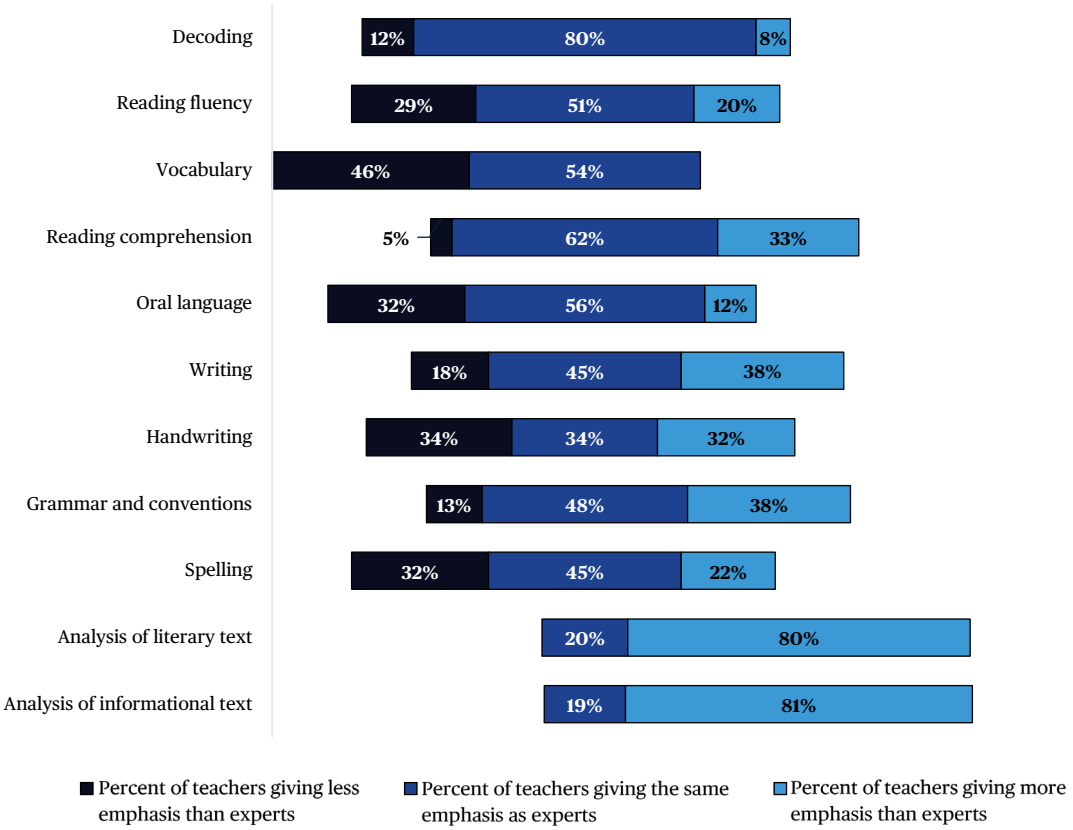
Across all grade levels, some teachers gave more or less emphasis than experts recommended to some domains of literacy instruction. For instance, in high school, the experts gave major emphasis to vocabulary, and about half of high school teacher respondents placed less emphasis on vocabulary than the experts recommended; on the other hand, the experts gave less emphasis to grammar and conventions for high school instruction, and 81 percent of high school teachers placed more emphasis on grammar and conventions than the experts recommended. High school teachers often emphasized domains that the experts emphasized for younger students, including decoding, reading comprehension, and grammar.

Exhibit D.35. Percentage of SRCL pre-K teachers giving less, the same, or more emphasis compared to expert recommendations in key domains of reading, 2019–20



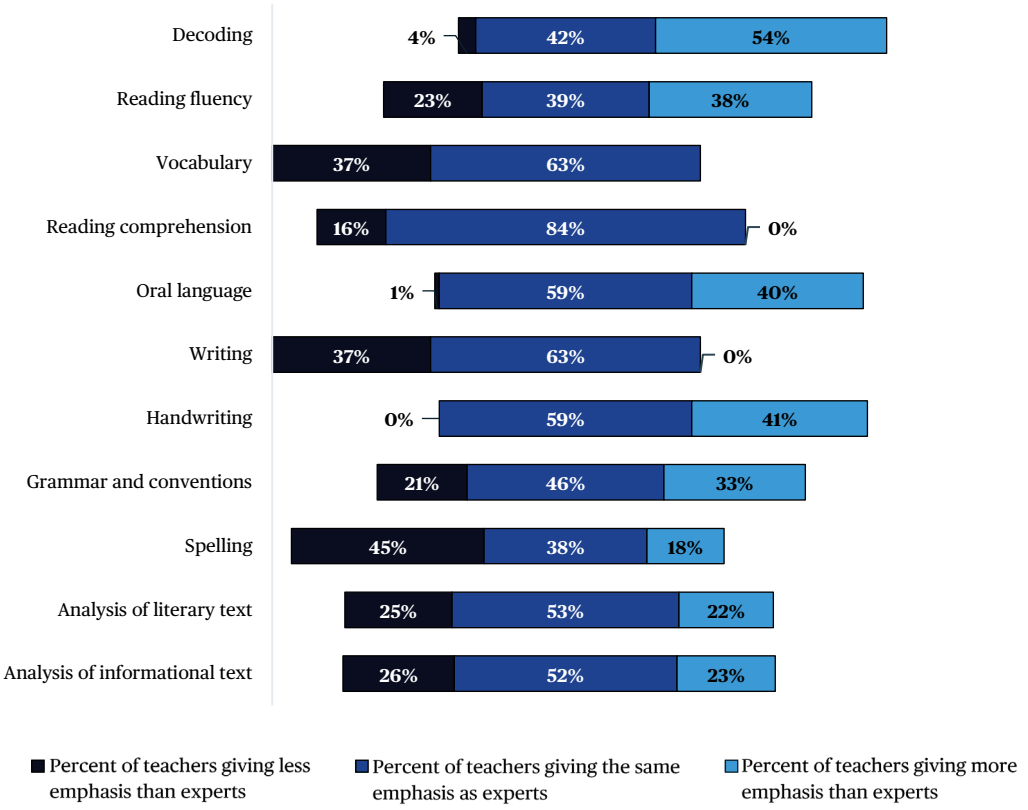
Note: The order of data in the exhibit matches the order of labels in the legend.
Source: SRCL pre-K teacher survey, 2019-20, Q25. The unweighted sample size is 332 pre-K teachers.

Exhibit D.36. Percentage of SRCL teachers in grades K–3 giving less, the same, or more emphasis compared to expert recommendations in key domains of reading, 2019–20



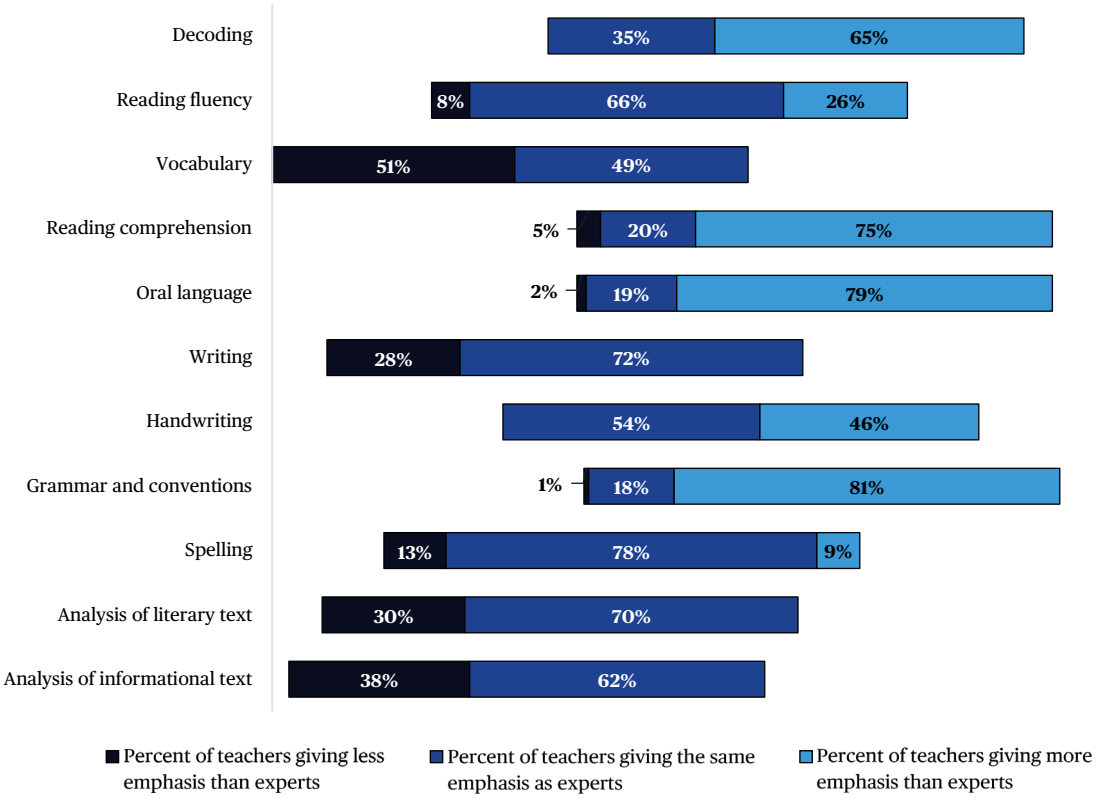
Note: The order of data in the exhibit matches the order of labels in the legend.
The experts recommended the highest possible level of emphasis on vocabulary, so it was impossible for teachers to give more emphasis than the experts in this domain. There was no domain in which the experts recommended the lowest possible level of emphasis for all grades between kindergarten and third grade.
Source: SRCL K-12 teacher survey, 2019-20 Q25. The unweighted sample size is 811 teachers of grades K-3.

Exhibit D.37. Percentage of SRCL teachers in grades 4–8 giving less, the same, or more emphasis compared to expert recommendations in key domains of reading, 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
The experts recommended the highest possible level of emphasis on vocabulary, reading comprehension, and writing, so it was impossible for teachers to give more emphasis than the experts in these domains. The experts recommended the lowest possible level of emphasis on handwriting, so it was impossible for teachers to give less emphasis than the experts in this domain.
Source: SRCL K-12 teacher survey, 2019-20, Q25. The unweighted sample size is 1,171 teachers of grades 4-8.

Exhibit D.38. Percentage of SRCL teachers in grades 9–12 giving less, the same, or more emphasis compared to expert recommendations in key domains of reading, 2019–20



Note: The order of data in the exhibit matches the order of labels in the legend.
The experts recommended the highest possible level of emphasis on vocabulary, writing, analysis of literary text, and analysis of informational text; thus, it was impossible for teachers to give more emphasis than the experts in these domains. The experts recommended the lowest possible level of emphasis on decoding and handwriting; thus, it was impossible for teachers to give less emphasis than the experts in these domains.
Source: SRCL K-12 teacher survey, 2019–20, Q25. The unweighted sample size is 517 teachers of grades 9-12.

APPENDIX E. EVIDENCE FOR FEATURES OF COMPREHENSIVE LITERACY INSTRUCTION

This appendix provides additional information on the features of comprehensive literacy instruction measured through this study and described in the body of the report. The Striving Readers Comprehensive Literacy (SRCL) Notice Inviting Applications used the Every Student Succeeds Act (ESSA) definition of comprehensive literacy instruction (Part B, Section 2221), which involves the 12 components listed in Exhibit E.1. For the purposes of the evaluation, eight of these components were collapsed into six features, defined as shown in Exhibit E.1. The evaluation did not measure the other four of the 12 components defined by ESSA (rows i through l of Exhibit E.1) because they were not readily measured based on survey data or they involved aspects of teacher practice other than instruction (e.g., collaboration).

Exhibit E.1. Components of comprehensive literacy instruction

Component of Comprehensive Literacy Instruction From Notice Inviting Applications	Study-Defined Features of Comprehensive Literacy Instruction					
	Emphasis on Key Domains	Explicit Instruction	Differentiated Instruction	Use of Formative Assessment	Frequent Opportunities to Practice Reading and Writing	Age-Appropriate Instruction
a. Includes developmentally appropriate, contextually explicit, and systematic instruction, and frequent practice, in reading and writing across content areas;		○			○	○
b. Includes age-appropriate, explicit, systematic, and intentional instruction in phonological awareness, phonic decoding, vocabulary, language structure, reading fluency, and reading comprehension;	○	○				○
c. Includes age-appropriate, explicit instruction in writing, including opportunities for children to write with clear purposes, with critical reasoning appropriate to the topic and purpose, and with specific instruction and feedback from instructional staff;	○	○				○
d. Makes available and uses diverse, high-quality print materials that reflect the reading and development levels, and interests of children;						○
e. Uses differentiated instructional approaches, including individual and small-group instruction and discussion;			○			
f. Provides opportunities for children to use language with peers and adults in order to develop language skills, including developing vocabulary;	○					
g. Includes frequent practice of reading and writing strategies;					○	

Component of Comprehensive Literacy Instruction From Notice Inviting Applications	Study-Defined Features of Comprehensive Literacy Instruction					
	Emphasis on Key Domains	Explicit Instruction	Differentiated Instruction	Use of Formative Assessment	Frequent Opportunities to Practice Reading and Writing	Age-Appropriate Instruction
h. Uses age-appropriate, valid, and reliable screening assessments, diagnostic assessments, formative assessment processes, and summative assessments to identify a child's learning needs, to inform instruction, and to monitor the child's progress and the effects of instruction;				○		
i. Uses strategies to enhance children's motivation to read and write and children's engagement in self-directed learning[a];						
j. Incorporates the principles of universal design for learning[a];						
k. Depends on teachers' collaboration in planning, instruction, and assessing a child's progress and on continuous professional learning[a];						
l. Links literacy instruction to the State's challenging academic standards, including standards relating to the ability to navigate, understand, and write about complex subject matters in print and digital formats.[a]						

[a] These components were not directly measured in the study.

Exhibits E.2–E.7 list recommendations from the What Works Clearinghouse practice guides that support each of these six features. Survey items used to measure the features are shown in Appendix B.

Exhibit E.2. Emphasis on key domains

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
K-3	Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/wwc_foundationalreading_040717.pdf	Teach students academic language skills, including the use of inferential and narrative language, and vocabulary knowledge (Foorman et al., 2016) Develop awareness of the segments of sounds in speech and how they link to letters. Teach students to decode words, analyze word parts, and write and recognize words.
K-3	Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/readingcomp_pg_092810.pdf	Teach students how to use reading comprehension strategies.
Adolescents	Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008-4027). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/docs/practiceguide/adlit_pg_082608.pdf	Provide opportunities for extended discussion of text meaning and interpretation. Provide explicit vocabulary instruction. Provide direct and explicit comprehension strategy instruction.

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
4–9	Vaughn, S., Gersten, R., Dimino, J., Taylor, M. J., Newman-Gonchar, R., Krowka, S., Kieffer, M. J., McKeown, M., Reed, D., Sanchez, M., St. Martin, K., Wexler, J., Morgan, S., Yañez, A., & Jayanthi, M. (2022). *Providing reading interventions for students in grades 4–9* (WWC 2022007). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/WWC-practice-guide-reading-intervention-full-text.pdf	Build students' decoding skills so they can read complex multisyllabic words. Provide purposeful fluency-building activities to help students read effortlessly.

Exhibit E.3. Explicit instruction

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
Adolescents	Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008-4027). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/docs/practiceguide/adlit_pg_082608.pdf	Provide explicit vocabulary instruction. Provide direct and explicit comprehension strategy instruction.
Adolescents	Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J., Olson, C. B., & Smither Wulsin, C. (2016). *Teaching secondary students to write effectively* (NCEE 2017-4002). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/508_WWCPG_SecondaryWriting_122719.pdf	Explicitly teach appropriate writing strategies using a Model-Practice-Reflect instructional cycle.

Exhibit E.4. Differentiated instruction

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
K-3	Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades: A practice guide.* (NCEE 2009-4045). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_reading_pg_021809.pdf	Provide time for differentiated reading instruction for all students based on assessments of students' current reading level.

Exhibit E.5. Use of formative assessment

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
K-3	Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades: A practice guide.* (NCEE 2009-4045). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_reading_pg_021809.pdf	Screen all students for potential reading problems at the beginning of the year and again in the middle of the year. Regularly monitor the progress of students at risk for developing reading disabilities.
Adolescents	Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J., Olson, C.B., & Smither Wulsin, C. (2016). *Teaching secondary students to write effectively* (NCEE 2017-4002). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/508_WWCPG_SecondaryWriting_122719.pdf	Use assessments of student writing to inform instruction and feedback.

Exhibit E.6. Frequent opportunities to practice reading and writing

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
K-3	Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/wwc_foundationalreading_040717.pdf	Ensure that each student reads connected text every day to support reading accuracy, fluency, and comprehension.
K-3	Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide* (NCEE 2012-4058). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/practiceguide/writing_pg_062612.pdf	Provide daily time for students to write.

Exhibit E.7. Age-appropriate instruction

Grade Band for Practice Guide	Practice Guide	Recommendations From Practice Guide
K-3	Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/readingcomp_pg_092810.pdf	Choose texts with word recognition and comprehension difficulty appropriate for the students' reading ability and the instructional activity.

APPENDIX F. INDEPENDENT EVIDENCE REVIEW

This appendix describes the findings of an independent review of 10 programs frequently purchased by school districts using Striving Readers Comprehensive Literacy (SRCL) subgrant funds. The review was conducted to determine whether any of the 10 programs were supported by strong, moderate, or promising evidence. The results of this evidence review were not incorporated into the analyses described in the report because subgrantees did not have access to this information when choosing literacy-related programs. The intent is that the information from this evidence review be added to the knowledge base regarding the effectiveness of literacy programs.

This appendix supplements the information on program evidence provided in the main report, which was based on a search of the What Works Clearinghouse (WWC) website and databases, the Evidence for Every Student Succeeds Act (ESSA) website, and the SRCL resource page. The searches of these sources revealed that studies had been conducted on 47 percent of the programs purchased with SRCL funds, but many of those studies had not yet been reviewed by the WWC. In addition, some of the reviews performed by the WWC were published 8 or more years ago, indicating that the programs may be ready for review in light of new research.

This appendix begins with a discussion of how literacy programs were chosen for a WWC evidence review (section F.1). It then outlines the standards and tools used in determining evidence ratings (section F.2). Finally, it presents the results of the evidence review (section F.3).

F.1. Method Used for Choosing Literacy Programs for a WWC Evidence Review

The study team selected 10 literacy programs for review based on a systematic process (Exhibit F.1) as follows:

- The initial pool of literacy programs consisted of all 236 programs on which districts spent SRCL funds.
- The pool of eligible programs was restricted to 165 programs that had never been reviewed by the WWC or Evidence for ESSA, did not appear on the SRCL resource page,[31] or had been reviewed by the WWC in 2013 or before (generally using Version 1.0 or 2.0 of the WWC standards).
- Programs purchased by five or fewer districts were excluded.

The last two steps reduced the pool of programs to 28.

[31] The SRCL resource page (https://dev.srclgrants.com/public/srResources) was developed by the U.S. Department of Education-funded technical assistance provider Synergy Enterprises Inc. Grantees could search the page for evidence-based practices.

Next, a team searched databases of research literature for publications relevant to the programs. The team used the EBSCOhost platform to search multiple research databases at the same time. Specific terms included in database search strings are listed in Exhibit F.2. The searches retrieved a list of references and abstracts for all relevant research articles, reports, dissertations, and conference presentations that were indexed by those databases.

Box F.1. Research Databases Searched

- Academic Search Premier
- EconLit
- Education Source conference papers
- Education Resources Information Center (ERIC)
- PsycInfo

The search team also looked for relevant studies on Literacy program developers' websites. Program developers often include on their websites summaries of studies or actual studies that support the mechanisms on which the program is based or support the actual programs. The abstracts and/or executive summaries for the studies listed on program developers' websites were added to those found during the database search.

Next, a group of abstract screeners removed the nonrelevant studies by reading each abstract and executive summary and answering three questions for that abstract:

1. Does the publication examine the program of interest?
2. Is the publication an empirical study of the program?
3. Does the study include some sort of a comparison group?

If screeners answered the questions with some combination of "Yes" and "Do not know," the study was categorized as "potentially eligible for review," and forwarded to a research librarian for full-text retrieval. Study abstracts for which screeners answered "No" to any of the three questions were screened out of the pool of eligible studies. The 10 programs selected for full evidence review were those for which there were four or more "potentially eligible studies."

Exhibit F.1. Process for selecting 10 programs ready for WWC review

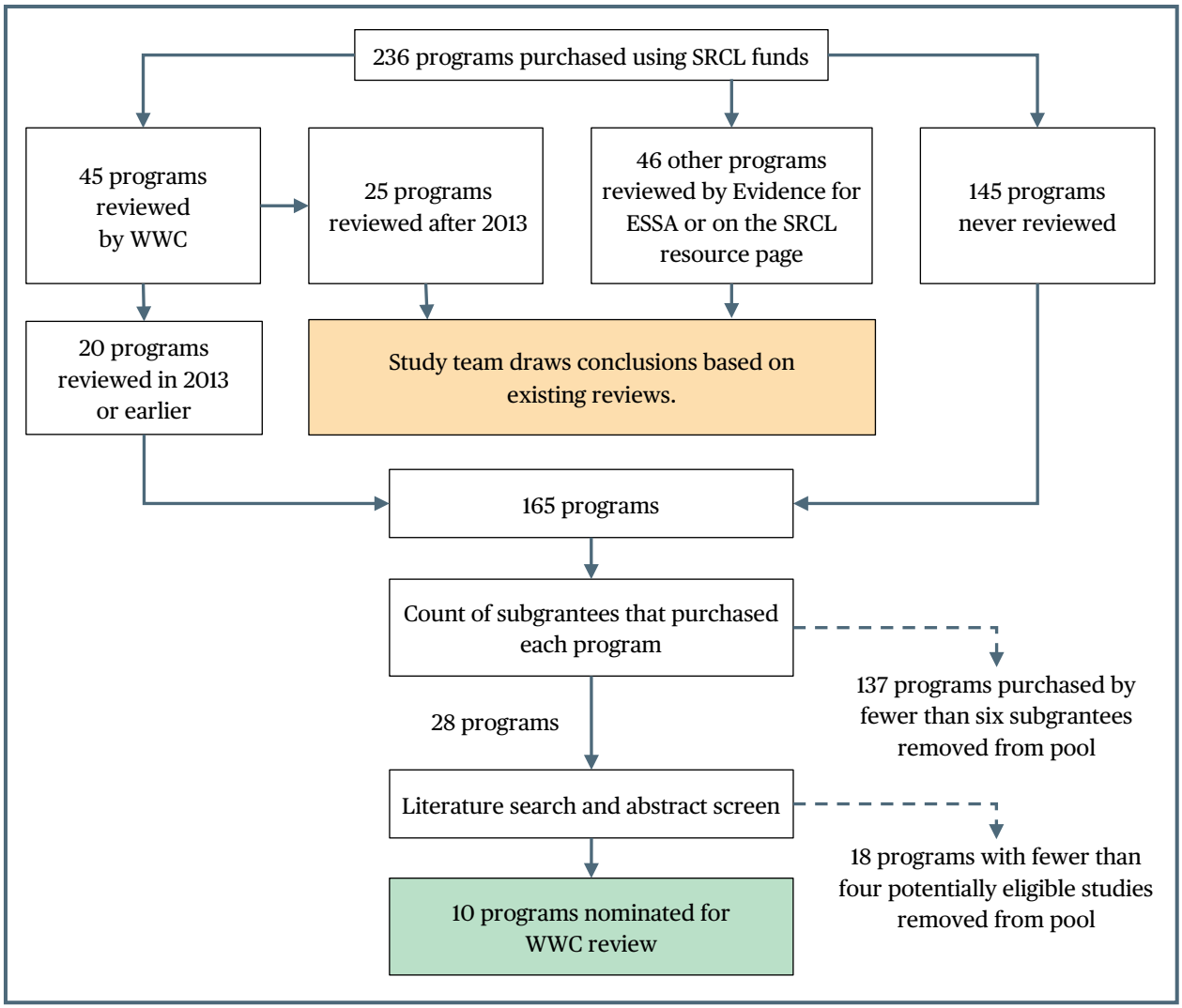


Exhibit F.2. Terms entered for database searches

Category	Search Terms
Evaluation terms	"Analys*" OR "Assess*" OR "Compar*" OR "Estimat*" OR "Evaluat*" OR "Evidence" OR "Examin*" OR "Literature review" OR "Measur*" OR "Meta analys*" OR "Metaanalys*" OR "Predict*" OR "Systematic review" OR "Varia*" OR "Impact" OR "Random*" OR "Experiment*"
Intervention terms	Name of program entered for the search
Analysis terms	"Achiev*" OR "Affect*" OR "Benefit*" OR "Declin*" OR "Decreas*" OR "Develop*" OR "Effect*" OR "Efficac*" OR "Gain" OR "Growth" OR "Higher" OR "Increas*" OR "Impact*" OR "Improv*" OR "Lower" OR "Progress" OR "Reduc*" OR "Statistic*" OR "Success*"

Category	Search Terms
Outcome terms	"Phonem* awareness" OR "Alphabet" OR "Phonolog* awareness" OR "Language" OR "Phonics" OR "Verbal development" OR "Word Recognition" OR "Lexicography" OR "Letter identification" OR "Print awareness" OR "Phonics" OR "Read*" OR "Fluency" OR Comprehen*" OR "Vocabulary" OR "Comprehen*" OR "English language arts" OR "English Language Arts" OR "ELA" OR "Literacy" OR "English" OR "Writing"
Delimiters	Subject ages (5–18); Grade levels (PK–Grade 12); Language (English); Setting (U.S. only); Publication dates (2001 to present)

Note: * denotes a wildcard, which means that any words containing the word stem to the left of the asterisk were included. All terms including the name of the program were entered into each program-specific search, with categories of terms separated by "and."

To prevent the possibility of duplication of work, the evidence review team confirmed that none of the 10 programs on the list were scheduled for review by the WWC.

Of the 10 literacy programs that were selected for the evidence review (Exhibit F.3), five had never been reviewed by the WWC. The other five had been the focus of WWC reviews performed in 2013 or earlier and were the subject of four or more potentially eligible studies since that previous review.

F.2. Review Standards and Tools

To determine the evidence ratings for the 10 selected literacy programs, the first step was to determine whether the studies of these programs met WWC standards with or without reservations. Certified WWC reviewers reviewed the studies of the 10 literacy programs according to the *WWC Procedures and Standards Handbook, Version 4.0*, using the WWC's Online Study Review Guide. For studies that met WWC standards, the study team retrieved information about the analysis sample and findings for reading-related outcomes and classified the 10 programs into the evidence rating categories as defined in section B.2.

Exhibit F.3. Programs selected for WWC review

Program Name	Program Type	Number of Subgrantees Purchasing Program	Number of Records Found in Literature Search (Potentially Eligible Studies)[a]	Previously Reviewed by WWC? (Year and Protocol)
Bookworms	Curriculum (language arts)	11	72 (4)	No
Corrective Reading	Curriculum	7	13 (6)	Yes (2007, BR) (2010, AL)
Creative Curriculum	Curriculum (preschool)	24	13 (4)	Yes (2013, ECE)
i-Ready	Curriculum	9	18 (8)	No
Lexia/Lexia Core5	Curriculum (computer based)	13	97 (41)[a]	Yes (2009, BR)
Raising A Reader	Early literacy and family engagement program	6	24 (6)	No
Read Naturally	Curriculum	12	23 (6)	Yes (2013; AL, BR)
Reading Mastery	Curriculum	10	29 (21)	Yes (2006, EL) (2010; AL, SWD) (2013, BR)
REWARDS	Curriculum	15	268 (4)	No
System 44	Intervention (computer based)	22	19 (6)	No

[a] A study was identified as "potentially eligible" if the study team could not definitively judge the study to be "ineligible" based on the abstract alone. Many of those studies first labeled "potentially eligible" were subsequently determined to be ineligible once the review team examined the full text of the study. Only studies that were eligible moved on to the full WWC review stage.
Note: AL = adolescent literacy protocol; BR = beginning reading protocol; ECE = early childhood education protocol; EL = English learners protocol; SWD = students with disabilities protocol. The publication year for previous WWC reviews and the WWC protocol used for the review are listed in parentheses.

F.3. Results of the WWC Reviews

The results of the review of the 10 programs are provided in Exhibit F.4. To provide a high-level summary of the results, the exhibit presents findings at the outcome domain level for each study that met WWC standards. Detailed findings at the individual outcome level for these studies are available upon request.

One of the 10 programs reviewed, i-Ready, was found to meet the moderate evidence definition. Of the four studies of i-Ready eligible for the review, two met WWC standards with reservations. One study– Dvorak et al. (2019a)– examined the impact of i-Ready on reading-related outcomes for 37,932 students in grades 1-5. The study showed that students participating in i-Ready outperformed comparison students in general literacy achievement by a statistically significant margin. A similar study–Dvorak et al. (2019b)–examined the impact of i-Ready on the literacy achievement of 18,809 students in grades 6-8. The impact estimate for that sample was not statistically significant. The WWC rating for three other studies of i-Ready was "does not meet standards."

Three programs– Lexia/Lexia Core5, Read Naturally, and Reading Mastery–were supported by promising evidence; all three were rated as having promising evidence in earlier reviews, and the new review did not change these ratings. The programs are still classified as having promising evidence because studies that examined the programs and met WWC standards showed statistically significant positive findings, but those findings were not based on 350 or more students.

The other six programs either had no studies that met WWC standards (Bookworms and System 44) or had studies that met WWC standards but showed no statistically significant positive or negative effects (Corrective Reading, Raising A Reader, and REWARDS).

Two previously unreviewed studies on Creative Curriculum were found, but neither of the studies met WWC standards.

Exhibit F.4. Summary of WWC reviews on selected literacy programs

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
Bookworms	Walpole et al. (2017)	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						Existing studies reviewed but do not meet standards of rigor
Corrective Reading	Torgesen et al. (2006)[a,b]	**Meets WWC standards with reservations**	Grade 5	86 students	Yes	Alphabetics	*g* = 0.13 (II=+4) Indeterminate	Null or negative findings review rating
			Grade 5	86 students	Yes	Reading fluency	*g* = 0.08 (II=+3) Indeterminate	
Creative Curriculum	Teaching Strategies (2013a)	**DNM.** QED with no baseline equivalence information						Null or negative findings review rating
	Teaching Strategies (2013b)	**DNM.** QED with no baseline equivalence information						
i-Ready	Curriculum Associates (2019)	**DNM.** QED in which baseline equivalence criterion not met						Strong or moderate evidence review rating in WWC
	Durfee et al. (2019)	**DNM.** QED in which baseline equivalence criterion not met						
i-Ready	Dvorak et al. (2019a)	**Meets WWC standards with reservations**	Grades 1-5	37,932 students	Yes	General literacy achievement	*g* = 0.11 (II=+4) *Stat sig. + effect*	

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
	Dvorak et al. (2019b)	**Meets WWC standards with reservations**	Grades 6-8	18,809 students	Yes	General literacy achievement	*g* = 0.07 (II=+3) Indeterminate	
	Torres (2019)	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						
Lexia/Lexia Core5	Macaruso et al. (2006)[a]	**DNM.** QED in which baseline equivalence criterion not met						Promising evidence review rating
	Macaruso & Walker (2008)[a]	**Meets WWC standards with reservations**	Kindergar-ten	71 students	Yes	Alphabetics	*g* = -0.138 (II=-5) Indeterminate	
		DNM: Equivalence of clusters in each group is necessary but not provided	Kindergar-ten	71 students	Yes	Compre-hension		
	Ness et al. (2013)	**Meets WWC standards without reservations**	Years 1-6 (~grades K-5/6)	37 students	No	Reading fluency	*g* = -0.35 (II=-14) Indeterminate	
						Alphabetics	*g* = 0.48 (II=+18) Indeterminate	

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
Lexia/Lexia Core5	O'Callaghan et al. (2016)	**DNM.** RCT with approach to handling for missing data not acceptable						
	Schechter et al. (2015)	**Meets WWC standards without reservations**	Grades 1–2	83 students	No	Compre-hension	*g* = 0.44 (II=+17) Indeterminate	
	Taylor (2018)	**Meets WWC standards without reservations**	Grades 1–3	132 students	No	Reading fluency	*g* = 0.69 (II=+26) *Stat sig. + effect*	
	Wilkes et al. (2020)	**DNM.** QED in which baseline equivalence criterion not met						
	Wilkes et al. (2016)	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
Raising A Reader	Anthony et al. (2014)	**Meets WWC standards with reservations**	Preschool	339 children	Yes	Compre-hension	*g* = -0.10 (II=-4) Indeterminate	Null or negative findings review rating
			Preschool	336 children	Yes	Cognition	*g* = -0.11 (II=-4) Indeterminate	
			Preschool	341 children	Yes	Alphabetics	*g* = 0.0 (II=0) Indeterminate	
		DNM. High-attrition RCT without baseline equivalence	Preschool	335 children	Yes	Language competencies		
	Chao et al. (2015)	**DNM.** QED in which baseline equivalence criterion not met						

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
Read Naturally	Arvans (2009)[a]	**DNM**. RCT with approach to handling for missing data not acceptable						Promising evidence review rating
	Berkeley et al. (2011)	**DNM.** High-attrition RCT without baseline equivalence						
	Christ & Davie (2009)[a]	**Meets WWC standards without reservations**	Grade 3	106 students	Yes	Reading fluency	*g* = 0.28 (II=+11) *Stat sig. + effect*	
			Grade 3	105 students	Yes	Compre-hension	*g* = -0.07 (II=+3) Indeterminate	
			Grade 3	105 students	Yes	Alphabetics	*g* = 0.13 (II=+5) Indeterminate	
	Tucker (2010)[a]	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						
	Wright (2006)	**DNM.** Study does not use an outcome that is sufficiently valid/reliable						

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
Reading Mastery	Green (2010)	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						Promising evidence review rating
	Stockard (2010)[a]	**Meets WWC standards without reservations**	Grade 4	52 students	No	Reading fluency	*g* = 0.70 (II=+26) *Stat. sig + effect*	
			Grade 4	52 students	No	Compre-hension	*g* = 0.52 (II=+20) Indeterminate	
	Stockard (2011a)	**DNM.** High-attrition RCT without baseline equivalence						
	Stockard (2011b)	**DNM.** High-attrition RCT without baseline equivalence						
	Stockard & Engelmann (2010)	**DNM.** Measures of effectiveness cannot be attributed solely to the program (*N* = 1 confound)						

Program	Reference	WWC Rating	Grade Level(s)	Analytic Sample (*N*)	Multisite?	Outcome Domain	Impact Estimate & WWC Characterization of Findings	Evidence Rating
REWARDS	Newman & Kundert (2012)	**Meets WWC standards without reservations**	Grade 7	469 students	Yes	Literacy achievement	*g* = 0.00 (II=0) Indeterminate	Null or negative findings review rating
	Pierson & Gowan (2014)	**DNM.** QED in which baseline equivalence criterion not met						
	Shippen et al. (2005)	**DNM.** High-attrition RCT without baseline equivalence						
System 44	Borrego (2017)	**DNM**. QED in which baseline equivalence criterion not met						Null or negative findings review rating

[a] These studies were included in the original WWC product and were reviewed again by the study team using the *WWC Standards and Procedures Handbook, Version 4.0*.
[b] The intervention report published in 2007 suggests potentially positive effects, but those analyses reflect the findings of grade 3 students. When the study was reviewed again using the *WWC Standards and Procedures Handbook, Version 4.0*, the study team concluded that the outcomes for grade 3 students did not meet standards because of high attrition and failure to establish baseline equivalence.
Note: DNM = Does not meet WWC standards; RCT = randomized controlled trial; QED = quasi-experimental design; II = improvement index. In accordance with current WWC practice, effects that exceed 0.25 standard deviations but are not statistically significant were labeled "indeterminate" rather than "substantively important."
Source: Study team's WWC reviews.

F.4. Changes in the Strength of Evidence for the 10 Programs

The results of these evidence reviews were not incorporated into the findings presented in the main report (Exhibits F.2–F.4). However, the new reviews would have made little change in the overall distribution of programs by level of evidence.

- Ratings for six programs were unchanged by the new reviews: Bookworms, Creative Curriculum, Lexia/Lexia Core5, Read Naturally, Reading Mastery, and REWARDS (Exhibit F.5).
- The rating for one program (i-Ready) changed from the "No review of existing studies" category to the "Strong or moderate evidence review rating in WWC" category (but with rounding it did not change the percentage of programs in the strong/moderate evidence review rating category in Exhibit 3 in the report, which remained 12 percent).
- The rating for the Raising A Reader program changed from "No review of existing studies" to "Null or negative findings review rating." The evidence review team found that the one study that met WWC standards with reservations showed null effects on literacy outcomes.
- System 44 was initially rated as showing null or negative effects. Our review suggests that no studies of System 44 met WWC standards.
- Corrective Reading was initially rated as having promising evidence. Based on the new review, the study team now classifies it into the "Null or negative findings review rating" category. The WWC's previous review of Torgesen et al. (2006), performed by the Beginning Reading topic area team, focused on the findings for grade 3 students, and excluded students in grade 5 because such students fell within the Adolescent Literacy topic area. Our rereview of this study using the *WWC Standards and Procedures Handbook, Version 4.0*, indicates that the findings for grade 3 students did not meet WWC standards, while the findings for grade 5 students did meet standards. The grade 5 findings showed no statistically significant effects.

Exhibit F.5. Changes to evidence ratings for programs selected for independent evidence review

Program Name	Evidence Rating Used in Report	Updated Evidence Rating
Bookworms	Existing studies reviewed but do not meet standards of rigor	Existing studies reviewed but do not meet standards of rigor
Corrective Reading	Promising evidence review rating	Null or negative findings review rating
Creative Curriculum	Null or negative findings review rating	Null or negative findings review rating
i-Ready	No review of existing studies	Strong or moderate evidence review rating in WWC
Lexia/Lexia Core5	Promising evidence review rating	Promising evidence review rating
Raising A Reader	No review of existing studies	Null or negative findings review rating
Read Naturally	Promising evidence review rating	Promising evidence review rating
Reading Mastery	Promising evidence review rating	Promising evidence review rating
REWARDS	Null or negative findings review rating	Null or negative findings review rating
System 44	Null or negative findings review rating	Existing studies reviewed but do not meet standards of rigor

Source: Study team's WWC reviews.

REFERENCES

Anthony, J. L., Williams, J. M., Zhang, Z., Landry, S. H., & Dunkelberger, M. J. (2014). Experimental evaluation of the value added by Raising A Reader and supplemental parent training in shared reading. *Early Education and Development, 25*(4), 493–514. https://eric.ed.gov/?id=EJ1030899

Arvans, R. (2009). *Improving reading fluency and comprehension in elementary students using Read Naturally* [Doctoral dissertation]. Western Michigan University.

Berkeley, S., Mastropieri, M. A., & Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *Journal of Learning Disabilities, 44*(1), 18–32.

Borrego, J. D. (2017). *English as a Second Language achievement and online reading programs* [Doctoral dissertation]. Dallas Baptist University.

Chall, J. S. (1983). *Stages of' reading development*. McGraw Hill.

Chao, S. L., Mattocks, G., Birden, A., & Manarino-Leggett, P. (2015). The impact of the Raising A Reader program on family literacy practices and receptive vocabulary of children in pre-kindergarten. *Early Childhood Education Journal, 43*(5), 427–434. https://eric.ed.gov/?id=EJ1071432

Christ, T. J., & Davie, J. (2009). *Empirical evaluation of Read Naturally effects: A randomized control trial* (Unpublished journal article). University of Minnesota, Minneapolis.

Curriculum Associates. (2019). *i-Ready efficacy: Research on i-Ready instruction program impact* (Report No. RR 2019-55). https://www.curriculumassociates.com/-/media/1a95af7827d74adcb87ccc7a53a7af22.ashx

Durfee, M., Call, T., Throndsen, J., & Nielsen, D. (2019). *Early intervention reading software program report.* The Utah State Board of Education. https://www.schools.utah.gov/policy/_policy_/_utahlegislativereports_/_2019_/2019NovemberEarlyInteractiveReadingSoftware.pdf

Dvorak, R. N., Randel, B., & Swain, M. (2019a). *An impact evaluation of mathematics and reading i-Ready instruction for elementary grades: Final report* (No. 062). Human Resources Research Organization.

Dvorak, R. N., Randel, B., & Swain, M. (2019b). *An impact evaluation of reading i-Ready instruction for middle school grades: Final report* (No. 074). Human Resources Research Organization.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First impact study final report* (NCEE 2009-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Green, A. K. (2010). Comparing the efficacy of SRA Reading Mastery and guided reading on reading achievement in struggling readers. *Dissertation Abstracts International, 71*(11A), 3969.

Griffith, D., & Duffett, A. (2018). *Reading and writing instruction in America's schools*. Thomas B. Fordham Institute.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65.

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29*(2), 162–172. https://eric.ed.gov/?id=EJ733029

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*(3), 266–287.

Ness, M., Couperus, J., & Willey, M. (2013). A comparison study of the effectiveness of the Lexia Reading Programme. *Kairaranga, 14*(1), 16–24. https://eric.ed.gov/?id=EJ1011403

Newman, D. L., & Kundert, D. K. (2012). *Report of intent to treat estimates of program impacts on student achievement: Gates-MacGinitie Reading Tests*. Striving Readers Project. Evaluation Consortium.

O'Callaghan, P., McIvor, A., McVeigh, C., & Rushe, T. (2016). A randomized controlled trial of an early-intervention, computer-based literacy program to boost phonological skills in 4- to 6-year-old children. *British Journal of Educational Psychology, 86*(4), 546–558. https://eric.ed.gov/?id=EJ1120013

Pierson, M. R., & Gowan, D. (2014). Decoding multisyllabic words: Structural analysis in reading by groups. *Journal of Gender and Power, 2*(2), 99–117. http://hdl.handle.net/10593/12677

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*(1), 24–46.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal, 48*, 965–995.

Polikoff, M.S., Zhou, N., & Campbell, S.E. (2015). Methodological choices in the content analysis of textbooks for measuring alignment with standards. *Educational Measurement Issues and Practice, 34*(3), 10-17.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008a). Bright Beginnings and Creative Curriculum: Vanderbilt University. In *Effects of preschool curriculum programs on school readiness* (pp. 41–54, Appendix C, and Appendix D). U.S. Department of Education, Institute of Education Sciences, National Center for Education Research.

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008b). Creative Curriculum: University of North Carolina at Charlotte. In *Effects of preschool curriculum programs on school readiness* (pp. 55–64, Appendix C, and Appendix D). U.S. Department of Education, Institute of Education Sciences, National Center for Education Research.

Reardon, S. F., Shear, B., R., and Castellano, K.E. (2017). Using Heteroskedastic Ordered Probit Models to Recover Moments of Continuous Test Score Distributions from Coarsened Data. *Journal of Educational and Behavioral Statistics, 42*(1), 3–45.

Shanahan, T., & Duffett, A. (2013). *Common core in the schools: A first look at reading assignments*. Thomas B. Fordham Institute.

Schechter, R., Macaruso, P., Kazakoff, E. R., & Brook, E. (2015). Exploration of a blended learning approach to reading instruction for low SES students in early elementary grades. *Computers in the Schools, 32*(3–4), 183–200.

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. (2005). A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education, 26*(3), 175–182.

Stockard, J. (2010). *Fourth graders' growth in reading comprehension and fluency: A pretest-posttest randomized control study comparing Reading Mastery and Scott Foresman Basal Reading Program* (Technical Report 2010-3). National Institute for Direct Instruction. https://www.nifdi.org/research/recent-research/technical-reports/287-2010-3-fourth-graders-growth-in-reading-comprehension-and-fluency-a-pretest-posttest-randomized-control-study-comparing-reading-mastery-and-scott-foresman-basal-reading-program/file

Stockard, J. (2011a). Increasing reading skills in rural areas: An analysis of three school districts. *Journal of Research in Rural Education, 26*(8).

Stockard, J. (2011b). Direct instruction and first grade reading achievement: The role of technical support and time of implementation. *Journal of Direct Instruction, 11*(1), 31–50.

Stockard, J., & Engelmann, K. (2010). The development of early academic success: The impact of Direct Instruction's Reading Mastery. *Journal of Behavior Assessment & Intervention in Children, 1*(1), 2–24.

Taylor, M. (2018). *Using the Lexia reading program to increase NWEA MAP reading scores in grades 1 to 3.* https://files.eric.ed.gov/fulltext/ED581492.pdf

Teaching Strategies. (2013a). *Comparing child outcomes: Examining the impact of The Creative Curriculum® for preschool.* https://teachingstrategies.com/wp-content/uploads/2017/03/TS-Comparing-Child-Outcomes-2013.pdf

Teaching Strategies. (2013b). *Effectiveness study: The Creative Curriculum® for preschool.* https://teachingstrategies.com/wp-content/uploads/2017/03/TS-Curriculum-Effectiveness-Study-2013.pdf

Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., Stancavage, F., Durno, D., Javorsky, R., & Haan, C. (2006). *National assessment of Title I interim report: Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for striving readers*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED491144

Torres, R. A. (2019). *The effect of the i-Ready Reading Program on student scores on the Northwest Evaluation Association (NWEA®) Measures of Academic Progress (MAP) Reading Assessment* (Publication No. 1167) [Doctoral dissertation, Cleveland State University]. https://eric.ed.gov/?id=ED599290

Tucker, C. N. (2010). *Response to intervention: Increasing fluency, rate, and accuracy for students at risk for reading failure* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN. ProQuest Dissertations and Theses.

U. S. Department of Education. (2016). *Non-regulatory guidance: Using evidence to strengthen education investments.* https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf

Walpole, S., Amendum, S., Pasquarella, A., Strong, J. Z., & McKenna, M. C. (2017). The promise of a literacy reform effort in the upper elementary grades. *The Elementary School Journal, 118*(2), 257–280. https://eric.ed.gov/?id=EJ1162546

What Works Clearinghouse. (2020). *What Works Clearinghouse procedures handbook*, Version 4.1. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf

Wilkes, S., Kazakoff, E. R., Prescott, J. E., Bundschuh, K., Hook, P. E., Wolf, R., Hurwitz, L. B., & Macaruso, P. (2020). Measuring the impact of a blended learning model on early literacy growth. *Journal of Computer Assisted Learning, 36*(5), 595–609. https://eric.ed.gov/?id=EJ1267201

Wilkes, S., Macaruso, P., Kazakoff, E., & Albert, J. (2016). Exploration of a blended learning approach to reading instruction in second grade. In *Proceedings of EdMedia 2016—World Conference on Educational Media and Technology* (pp. 797–802). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/173040/

Wright, S. A. (2006). *The effects of Read Naturally on students' oral reading fluency and reading comprehension* [Unpublished master's thesis]. California State University-San Marcos, San Marcos, CA.