

# Scaling Students' Success with STARI: Expanding Eligibility, Support, and Spread

## Table of Contents

|   |           |
|---|-----------|
| <b>Significance</b>   | <b>1</b>  |
| STARI: A Highly Promising Strategy                          | 2         |
| Contribution of Proposed Project                            | 6         |
| <b>Strategy to Scale</b>                                    | <b>6</b>  |
| Barriers to Scale   | 7         |
| Management Plan   | 12        |
| Capacity to Scale   | 15        |
| Dissemination   | 18        |
| <b>Project Design</b>                                       | <b>19</b> |
| Conceptual Framework  | 19        |
| Goals, Objectives and Outcomes                              | 21        |
| Needs of the Target Population                              | 22        |
| <b>Project Evaluation</b>                                   | <b>24</b> |
| Meeting WWC Standards Without Reservations                  | 24        |
| Components, Mediators, Outcomes and Thresholds for Fidelity | 31        |
| Guidance about Effective Replication Strategies             | 35        |
| Feedback about Performance and Progress                     | 36        |

## **Scaling Students' Success with STARI: Expanding Eligibility, Support, and Spread**

The SERP Institute, in collaboration with MDRC, proposes an Expansion Grant to scale the Strategic Adolescent Reading Intervention (STARI), a program designed to accelerate the reading growth of students in grades 6-8 who read two or more years below grade level. Over the course of four years, STARI will reach 15,000 struggling readers in 80 schools in four states. STARI addresses multiple barriers that prevent middle school students from accessing grade-level texts, including slow and inaccurate decoding, poor command of sentence structure, weak vocabulary, and limited background knowledge (Catts et al., 2012). STARI's defining characteristic is the attention to student motivation and engagement woven into the program's design—factors that are critical contributors to learning in the adolescent years (National Academies, 2018). STARI is supported by strong evidence (**Absolute Priority 1**) and is a field-initiated intervention (**Absolute Priority 2**). In addition, because STARI serves the highest-need students in the country, accelerating their reading growth in the wake of the pandemic, the current proposal also meets **Competitive Priorities 1 and 2**.

### **Significance**

In 2019, according to the National Assessment of Educational Progress (NAEP), only 34% of U.S. eighth graders could read at a proficient level. By middle school, these students have experienced repeated failure to meet grade-level expectations. As the gap expands between students' reading skills and the complexity of grade-level texts, so does their risk for dropping out of school (Kamil, 2003; Biancarosa & Snow, 2004). The cost to students who do not complete high school can be measured in lower wages, higher unemployment and rates of incarceration, and poorer health, outcomes that ripple through society and future generations (Burrus & Roberts, 2012; Reim, 2014). The data also reveal significant racial and socioeconomic

disparities. In 2019, only 15% of Black students, 22% of Hispanic students, and 20% of students living in poverty scored proficient on NAEP, as compared to 42% of White students and 46% of students not eligible for free and reduced-price meals. Because STARI is designed with specific attention to these high-need populations—including culturally responsive practices and highly engaging texts—this proposal addresses **Competitive Preference Priority 1 - Promoting Equity in Student Access to Educational Resources and Opportunities.**

Furthermore, the pandemic and resulting disrupted schooling have both depressed overall reading scores and increased inequities. In the fall of 2021, in schools serving predominantly students of color, students were five months behind the already low pre-pandemic levels of achievement in reading (Dorn et al, 2021). Thus, the pandemic has widened the achievement gap between White students and students of color by approximately a third, such that students in schools serving mostly students of color are now a full 12 months behind students in majority-White schools (Dorn et al, 2021). Recovery from the pandemic requires that students have the opportunity to make up lost ground, but that they do so without missing out on grade-level content. Unlike other reading interventions, STARI *integrates* instruction in foundational skills with instruction in deep comprehension and grade-level skills. Therefore, STARI is exceptionally well positioned to **address the impacts of COVID-19 on students, educators and faculty (Competitive Preference Priority 2).**

### **STARI: A Highly Promising Strategy**

STARI was developed through a partnership among SERP, the Boston Public Schools, Wheelock College, and Harvard University in response to a problem of practice posed by Boston teachers and administrators: that students in middle and high school did not have sufficient reading skills to comprehend their textbooks (██████████ et al., 2013). This proposal therefore

meets **Absolute Priority 2 - Field-Initiated Innovations**. After an extensive iterative development process, STARI was made publicly available in 2014; all curriculum materials are freely downloadable, which facilitates its scalability (printed versions are available for purchase). Thousands of registrants download STARI each year, from all 50 states and Washington, DC.

STARI is designed to accelerate adolescents' reading comprehension, to improve word-level skills and reading fluency, and to address gaps in background knowledge (Denton et al., 2015; Geva & Farnia, 2012). Skilled comprehension involves developing a personal, critical stance on what is read (Hemphill et al., 2019; Newell et al., 2011), and thus requires that students be engaged in reading. Yet disengagement and lack of efficacy are characteristic of adolescents with reading difficulties (Guthrie et al., 2012). STARI differs from other interventions in the extent to which it incorporates features aimed at addressing engagement and efficacy. These features include: (i) engaging, relevant themes and texts with culturally familiar content (Ivey & Johnston, 2013; Tatum, 2008); (ii) integration of basic skills with cognitively challenging content (Hemphill et al., 2019); (iii) frequent opportunities for peer collaboration (Fuchs et al., 2011); (iv) texts matched to students' reading levels, so students experience success and build confidence (Fulmer & Tulis, 2013; O'Connor et al., 2002); and (v) structured discussion of cognitively challenging content aligned to 21<sup>st</sup>-century standards (Allen et al., 2011).

STARI is intended to be used for a full class period (45 minutes minimum), at least three days a week, in addition to the regular English class (See Appendix J.2). STARI is organized into three series for grades 6-8 (See Appendix J.3-8 for scope and sample materials). Each series includes three thematic units, motivated by an essential question such as "What makes a family?" and includes teacher lesson plans, student workbooks, and leveled fluency passages.

Research-based key components address both foundational and higher-level literacy

skills. In daily *fluency practice*, students engage with a partner in repeated readings of short texts. Leveled passages build background knowledge needed to comprehend core unit texts and also provide targeted practice with decoding and morphological patterns taught in mini-lessons. Each passage includes a “mini debate” for partners to contrast personal stances on the topic. Because expanded exposure to text is a key goal, students engage in daily *partner reading or guided reading with novels and nonfiction texts*. For *guided* reading, teachers are provided with questions to build students’ reasoning skills and scaffold discussion of unit novels. In *partner* reading, workbooks guide students to read, stop and discuss, and record answers to text-based questions with a partner. Each STARI unit also contains at least one *debate* that raises an engaging text-based question (e.g., “In *Locomotion*, which characters are most like family to Lonnie?”) and requires work in teams to make a claim and defend that claim using textual evidence. Finally, each unit contains *mini-lessons in comprehension strategies and decoding*. *Comprehension strategy* instruction is based on the Reciprocal Teaching (RT) model (Palincsar & Brown, 1984). *Decoding* instruction focuses on skills such as finding base words and syllable division rules. Students frequently practice comprehension and decoding strategies in context with fluency passages and unit novels and nonfiction.

**Promise for Improving Student Outcomes.** STARI was evaluated in a randomized trial in eight schools in urban and suburban districts in Massachusetts (■■■■ et al., 2017). The study focused on students in grades 6-8 (n=402) who read two or more years below grade level. The sample was mostly low-income students of color. Students were randomly assigned to STARI or to another intervention program. Their reading skills were measured using the RISE (renamed the ReadBasix), a computer-based assessment developed by ETS (O’Reilly et al., 2012).

**The RfU study found that STARI students made meaningful gains on multiple**

**reading skills, while control students made little or no progress on those skills.** Students assigned to STARI outperformed control students on subtests of word recognition ( $d = .20$ ), morphological awareness ( $d = .18$ ), and efficiency of basic reading comprehension ( $d = .21$ ) (Kim et al., 2017). An instrumental variables analysis demonstrated greater impacts for students who completed more of the curriculum. These findings are meaningful because most adolescent reading interventions produce little or no impact (Scammacca et al., 2013), and adolescents who are significantly below grade level are the least likely to be successful in existing reading interventions (Kemple et al., 2008; Kim et al., 2011). The findings are also notable because a majority of students in the control group received another reading intervention. RfU study results met WWC standards without reservations (ref. no. 755352265). Therefore, because this study (a) meets WWC standards without reservations, (b) includes multiple statistically significant and positive effects on relevant outcomes, (c) includes no statistically significant and negative effects, and (d) is based on a sample of 402 students from eight schools in four urban and suburban districts, this proposal meets **Absolute Priority 1 - Strong Evidence**.

In 2017, SERP was awarded an Education and Innovation in Research (EIR) mid-phase grant to support and evaluate implementation of STARI in several high-needs school districts, as well as build capacity to scale STARI. The intention was to evaluate STARI in New York City, Washington, DC, Baltimore, MD, and Jackson, MS during the 2019-20 and 2020-21 school years. The project has been impacted by COVID-19, as collecting student assessment data in 2020 and 2021 proved impossible. Despite pandemic-related obstacles, SERP and MDRC extended implementation into a third study year, 2021-22 (in Jackson and New York only due to budget constraints). Although data collection and analysis are not yet complete, preliminary results are promising, with positive effects that are larger in magnitude than in the RfU study ( $d$

= .27 -  $d = .41$ ), across all six subcomponents of the ReadBasix. Resources from the mid-phase grant also provided the opportunity to develop new program components that will make it possible to serve a broader population of students (including those reading below a 3rd grade level), over a much broader geographic area (due to an online professional learning series that can be accessed from anywhere), laying the groundwork for the proposed expansion study.

### **Contribution of Proposed Project**

Successfully addressing adolescent literacy development for students who are behind grade level expectations is both challenging and critically important. STARI's demonstrated success in the initial RCT involving four districts in a single state is therefore noteworthy. For this reason, we applied for (and were awarded) a mid-phase EIR grant intended to test STARI with a much larger sample of students in more diverse geographic settings. However, the constraints imposed by COVID significantly impeded our ability to conduct that study. Now in its final implementation months, the extension year provides only a similar sized sample to the original study, with data from two additional districts in different states. The expansion study will reveal the extent to which impacts can be replicated across a larger, more geographically diverse sample (15,000 students in 80 schools in four states), at the same time that it will answer questions about STARI's effectiveness for particular subgroups. The proposed study will also address questions about what resources are needed to support schools' adoption and sustained implementation of STARI, in order to inform national scaling. Most crucially, this project will offer STARI to students, teachers, and schools who desperately need promising interventions to support struggling adolescent readers, a population that has increased due to the pandemic.

### **Strategy to Scale**

The proposed project's goal is to scale STARI to meet the national need for interventions

to support struggling adolescent readers. The EIR mid-phase project allowed SERP to identify a number of barriers to scale—including barriers to *spread, depth, sustainability, and shift in reform ownership* (Coburn, 2003)—and to develop solutions. This project will allow us to *implement and evaluate* these solutions across a large and geographically diverse sample.

## **Barriers to Scale**

**Barrier 1: Teacher Knowledge and Professional Learning.** To date, the primary barrier to scaling STARI has been the cost and feasibility of professional learning (PL). While freely downloadable STARI curriculum materials are a primary lever to improve student outcomes, they must be paired with structured teacher learning opportunities to support high-quality implementation (Amendum & Fitzgerald, 2013; Biancarosa et al., 2010; Garet et al., 2008; Sailors & Price, 2010). Research shows that middle school generalist classroom teachers—those who typically implement STARI—have limited understanding of reading pedagogy and often are unfamiliar with high-impact literacy practices (Spear-Swerling & Cheeseman, 2012) and with high-leverage teacher-student interactions (Hamre et al., 2012). Moreover, deep and lasting instructional change requires more than just following lesson plans—it requires changing teachers’ *beliefs and underlying pedagogical principles* (Coburn, 2003). Previous evaluations of STARI, including the EIR mid-phase evaluation, have addressed teacher learning via a multi-day, in-person summer institute, regular visits from a SERP-trained STARI coach, and monthly coach-facilitated professional learning community (PLC) meetings. Outside of grant-funded evaluations of STARI, SERP has hosted STARI summer institutes in 2016-2019 and offered local institutes at cost. However, only 11 districts have underwritten the relatively high cost of a local STARI summer training institute, and only 25 have paid to send representatives to national STARI institutes hosted by SERP. Even then, such one-shot PL



opportunities have demonstrated relatively weak effects on teachers' instructional practice or student outcomes (Yoon et al., 2007). Many districts have inquired about ongoing support from SERP, but both the cost to the district and SERP's limited capacity to provide ongoing support across the country have prevented scaling of this model of STARI PL.

As part of the mid-phase EIR project, SERP has developed a 17-session online PL series (See Appendix J.9 for overview). Recent research has shown that online PL may result in positive shifts in teacher practice and student learning similar to those achieved by in-person teacher workshops, with more flexibility and at a lower cost (Hill et al., 2013; Snell et al., 2019). The online STARI PL incorporates key, research-based features of effective online PL, including the use of classroom video to provide models of effective practice, and facilitated discussions through accompanying PLCs, including reflection on teachers' own practice (Beisiegel et al., 2018; Delaco et al., 2022). The online PL series provides rich content that can be used in routine district PL structures: in-service summer workshops and ongoing PLCs during the year. Although the series is designed to be compatible with self-paced, independent use, research demonstrates that the most meaningful PL is collaborative (e.g., Gibbons & Cobb, 2017). Therefore, participating schools and districts are encouraged to implement the PL in person, led by a local facilitator (e.g., a district literacy coach or intervention coordinator). An accompanying comprehensive facilitator's guide includes suggested language for introducing and debriefing the instructional videos and the hands-on activities (See Appendix J.10 for sample page). Teachers can access any sessions they may miss or want to revisit at their convenience.

Thus, the online PL series addresses a number of major barriers to scale. The asynchronous format facilitates *spread* by enabling access to the expertise of STARI developers even in small or rural schools where in-person support has been prohibitively difficult and/or

costly in the past. This format also facilitates *sustainability* because anyone can be trained at any time—valuable flexibility for districts with high teacher turnover. The series facilitates *depth* of reform by directly addressing teachers’ beliefs and underlying pedagogical principles (Coburn, 2003) through ongoing, hands-on PL. And finally, the locally facilitated model allows a *shift in reform ownership*, away from researchers and program developers to local instructional leaders. The proposed study will allow us to evaluate and iteratively improve this more affordable, scalable model of professional learning.

**Barrier 2: Limited Target Student Population.** STARI was originally designed for students who read, at minimum, at a 3rd-grade level. However, in many U.S. districts, **a significant proportion of middle school students read below a 3rd grade level.** For example, in District of Columbia Public Schools (DCPS), a partner in the EIR mid-phase project, even *prior to the pandemic*, approximately 10% of middle school students read below the 3rd-grade level, a proportion that has grown due to pandemic-related disrupted learning (Boots, 2021). To better serve *all* students in need of literacy intervention in middle school, DCPS leaders asked SERP to develop a phonics supplement to STARI. SERP developed a computer-based component—Phonics Boost—that can be seamlessly integrated into STARI instructional routines (See Appendix J.11 for scope and sequence and sample activities). Accessed via a web application on any internet-connected device, Phonics Boost activities include decoding and encoding practice, using both real and nonsense words (Castles et al., 2018), brief tutorials in new letter-sound patterns, and occasional practice with sight words (Castles et al., 2018; Joseph & Schisler, 2009). Students begin at an individualized starting point, and spend as much time as needed practicing each skill. Teachers have access to real-time data to monitor progress.

Phonics instruction in the absence of comprehension instruction has not been shown to

improve the reading comprehension of older students (Joseph & Schisler, 2009; National Reading Panel, 2000). By integrating Phonics Boost into STARI rather than offering it as a standalone intervention, students experience this instruction in the context of engaging, age-appropriate materials. Critically, they also receive instruction in grade-appropriate comprehension skills, so that they do not continue to fall further behind.

Often, schools are unable to support multiple reading interventions, and prefer an intervention that can reach their full population of struggling readers. Therefore, this field-initiated addition to the original STARI program contributes to STARI's *spread*, improving alignment with market needs by serving a broader population of students. Phonics Boost may also promote *depth*, as teachers develop their understanding of foundational literacy skills alongside their students. The proposed study will allow us to demonstrate STARI's effectiveness with this expanded population of target students.

### **Barrier 3: Limited Information About For Whom and Under What Conditions**

**STARI is Effective.** Students who struggle with reading often belong to subgroups with specific needs. For example, DCPS' Department of Specialized Instruction has expressed interest in using STARI with special education students, and New York City's Division of Multilingual Learners has inquired about STARI for English learners. However, STARI evaluations to date have not had a large enough sample to conduct subgroup analyses. The proposed study's large and purposely diverse sample will allow subgroup analyses to better support recommendations regarding these populations, thus increasing STARI's *spread* to subpopulations of students.

**Barrier 4: Sustainability and Shift in Reform Ownership.** Since 2017, more than forty districts and charter networks have purchased STARI materials and PL from SERP, and many more have launched STARI using only the downloadable curriculum materials. Some districts

have expressed that they are accustomed to receiving substantial support from for-profit intervention vendors (e.g., weekly meetings and monthly visits from a representative). As a small nonprofit organization providing the curriculum as an open educational resource, SERP is unable to provide this level of support outside the context of a grant-funded research study, potentially driving districts away from STARI, and toward more costly, less effective interventions.

Therefore, as a part of this project, SERP intends to rigorously evaluate two models of support (*enhanced* vs. *standard*) for schools launching STARI, to determine the most sustainable and cost-effective option. The *enhanced* support version will include direct support from SERP's literacy specialist including individualized guidance for scheduling STARI, ordering and distributing materials, scheduling and supporting PL, monitoring student progress, and evaluating STARI teachers and holding them accountable to high levels of implementation. Although this level of support would undoubtedly help schools launch, implement, and sustain STARI, and may be comparable to that provided by some for-profit intervention vendors, it would increase the cost of the program. Therefore, SERP also proposes to test a version of *standard* implementation support that may be offered to schools at a marginal additional cost. This package would include monthly *tips and reminders* to complete STARI PL pushed directly to teachers' phones, and access to *monthly webinars* and an *online, searchable STARI discussion forum*, both of which would allow stakeholders to have their questions answered by STARI experts and/or peers who may have encountered and found solutions to similar challenges. We will examine whether differences occur in implementation fidelity and student outcomes between schools receiving these two support models, and whether support levels are associated with schools' decision about whether to continue to implement STARI after the first year. Comparing these two conditions will improve SERP's strategy to serve and advise future STARI users.

## **Management Plan**

**Management.** As the prime organization, SERP will oversee the entire project, communicating regularly with the independent evaluator (MDRC) and engaging consultants and districts (**See Appendix J.1 for a comprehensive management plan and timeline**). SERP will lead school recruitment and delivery of STARI materials and PL to these schools. MDRC will lead the impact evaluation. SERP and MDRC are productively working together as part of the EIR mid-phase project, and management of the proposed project will reflect prior successful collaborations. Virtual meetings between SERP and MDRC will be held twice per month in Years 1-4, and will cover every aspect of the project, including relationships with districts and schools, recruitment, planning for training, reports from specific school sites, financial management, and product design. The bimonthly schedule will ensure close attention to markers of progress and a quick response to reported challenges.

**Recruitment.** SERP plans to recruit 80 schools across two school years: Cohort 1 schools will be recruited by Spring 2023 (with participation beginning in Fall 2023) and Cohort 2 schools will be recruited by Spring 2024 (with participation beginning in Fall 2024). We will begin recruitment at the state level both to aid in recruiting a large student sample, and to echo a common adoption process for national scaling. In Cohort 1, senior state department of education leaders in Mississippi and New Mexico are enthusiastic about STARI and have agreed to assist with recruiting districts to participate. SERP has an ongoing relationship with Mississippi, as the State Superintendent is on SERP's Board, and one district (Jackson Public Schools) participated in the EIR mid-phase project. New Mexico leaders learned about STARI through the US Department of Education's 2021 Striving Readers Comprehensive Literacy (SRCL) and Comprehensive Literacy State Development (CLSD) National Convening; SERP was then asked

to present at New Mexico’s 2021 statewide convening. These two states were deliberately selected to inform national scaling because they represent very different high-need populations. Both states include urban and suburban settings, which overlap with the settings from the RfU study, as well as rural settings that offer opportunities to scale to currently underserved populations. During the 2023-24 school year, we will recruit two additional states. To facilitate national scaling, we will target different regions of the country than those served in Cohort 1, or the RfU or EIR mid-phase studies, seeking a state from the Midwest and one from the West or Northwest. Districts in Wisconsin, Minnesota, Ohio, Illinois, Utah, Colorado and California have purchased STARI materials and PL, making these likely target sites. We chose to wait to recruit Cohort 2 states due to the impracticality of asking stakeholders to learn about and commit to this project more than two years in advance.

In the recruitment process, state-level leaders will identify *districts* that might be a good fit for participation, and facilitate connections; Mississippi has 151 school districts with 213 schools serving 6th-8th graders, and New Mexico has 89 districts with 192 schools serving 6th-8th grades. Once districts have been recruited, as in the mid-phase study, district-level leaders will facilitate connections with *schools*; SERP will then meet with schools to confirm eligibility and interest, which will be formalized by having each school sign an agreement that outlines roles and responsibilities. To participate, schools will agree to randomization and data collection procedures and to implement STARI in at least two grade levels (6<sup>th</sup> and 7<sup>th</sup> grade; 8th grade will be optional). Table 1 summarizes the implementation plan. In each cohort, schools will be randomly assigned to either begin implementing STARI in the coming school year with an enhanced level of support (“immediate implementation” or II group) or to begin STARI implementation two years later (“delayed implementation” or DI group). Schools in the DI group

will continue to implement their “business as usual” (BAU) reading instruction in their first two years of study participation; in their third year, they will receive STARI, and be randomly assigned to either “enhanced” or “standard” levels of support from SERP. This design will allow us to not only explore STARI’s effect compared to BAU instruction, but also to explore whether schools can achieve similarly strong implementation and student outcomes with less support.

**Table 1. Timeline for STARI implementation by Research Group and School Cohort**

|  | SY<br>23-24 | SY<br>24-25 | SY<br>25-26 | SY<br>26-27 |
|--|-------------|-------------|-------------|-------------|
| <b>School cohorts and research groups</b>              |             |             |             |             |
| <b>School Cohort 1: 40 schools</b>                     |             |             |             |             |
| 20 immediate implementation schools (enhanced support) | Year 1      | Year 2      |             |             |
| 20 delayed implementation schools                      | BAU         | BAU         |             |             |
| 10 schools with enhanced support                       |             |             | Year 1      |             |
| 10 schools with standard support                       |             |             | Year 1      |             |
| <b>School Cohort 2: 40 schools</b>                     |             |             |             |             |
| 20 immediate implementation schools (enhanced support) |             | Year 1      | Year 2      |             |
| 20 delayed implementation schools                      |             | BAU         | BAU         |             |
| 10 schools with enhanced support                       |             |             |             | Year 1      |
| 10 schools with standard support                       |             |             |             | Year 1      |

Notes: Year X = STARI Implementation year (first or second). BAU = Business as usual reading instruction.

**Study Sample.** Schools implementing STARI will be encouraged to offer it to all 6th-8th graders who were not proficient on the ELA state test the prior spring (about 70% of students enrolled in study schools). Unlike prior studies, students who are very far below grade level, students with IEPs, and ELLs will not be excluded. We expect about 20% of students in the study sample to have an IEP, 15% to be ELLs, and 40% to have especially low reading skills (“below basic” on the state ELA test). School administrators will identify teachers for STARI training to mimic routine conditions. The number of teachers depends on the number of STARI sections per teacher, but we anticipate roughly 240 teachers *per cohort* (120 II; 120 DI).

**Local Support.** Effective implementation and evaluation of STARI requires partnership with district-level employees who have familiarity and influence with district- and school-level

systems. In each participating district, SERP will hire a **Local Project Support Associate (LPSA)** to: assist schools with scheduling STARI into the school day and identifying students to participate in the intervention; coordinate ordering materials with SERP and ensure their prompt delivery to the correct classrooms, and collaborate with the research team to schedule FoI observations. One lesson learned from the EIR mid-phase grant is that responsiveness to local conditions is critical to allow smooth integration of an intervention into normal educational practices. Therefore, the exact role and time allocation for each LPSA will be determined in partnership with the district. SERP has successfully used this adaptive approach to staffing district on-the-ground support for a previous study for the IES Academic Language Intervention Project implementing WordGen Elementary in districts in five states.

The project director, assistant project director, and literacy specialist will work intensively with LPSAs to provide logistical support for implementation, and to debrief areas of strength and required improvements; we expect that meetings would be weekly during the launch period, then bimonthly for the rest of the project period. The project director, assistant project director, and literacy specialist will join the LPSA's meetings with school and district administrators as needed. The project director and/or literacy specialist will also make site visits at least twice per year to speak with coaches, administrators, and teachers in person.

### **Capacity to Scale**

SERP is a nonprofit, 501(c)(3) organization founded in 2003 to harness the expertise of researchers, practitioners and designers for purposes of solving problems of practice identified by school district partners. SERP's Design Studio provides expertise on all aspects of product design and production for school settings. The school district partnerships that SERP establishes create the contexts for better understanding the complexity of problems, the needs of educators



and students, and the constraints on solution design. SERP leaders bring expertise in research, practice, design, collaborative processes, and communications, and are noteworthy for the ability to cross boundaries and institutional cultures. Since its founding, SERP has successfully managed over \$45M in awards and developed more than a dozen practice-driven, research-based products available to the public free of charge (see [serpinstitute.org](http://serpinstitute.org)).

**SERP Personnel.** [REDACTED] **EdD**, will serve as the project's Principal Investigator (PI) and Project Director. She will manage and coordinate all aspects of the project, co-author reports, and disseminate findings. [REDACTED] is currently Project Director and co-PI for the EIR mid-phase grant, supervising coaches, developing PL opportunities for teachers, and working directly with district- and school-level personnel to ensure implementation in compliance with research standards. Previously, [REDACTED] collaborated on the evaluation of STARI impacts in the RfU study, conducted independent research into teacher implementation of STARI during and after that grant period, and disseminated results in peer-reviewed journals.

[REDACTED], **Ph.D.**, project co-PI, is founding Executive Director of the SERP Institute. She has served as PI on numerous federal grants totaling over \$40 million, including the IES RfU grant that resulted in the development of STARI and the EIR mid-phase grant now in its final year. [REDACTED] will assist with recruitment and monitor partnerships with states, districts, and schools, regularly attend meetings and coordinate with project staff to advise and provide support on all aspects of the project, and assist with dissemination activities.

**Other SERP personnel** will contribute to the success of the project. [REDACTED], **Assistant Project Director**, brings several years of experience working closely with STARI, making curriculum revisions, managing materials, conducting observations and interviews, assisting with data, and providing support to districts and teachers. A **Literacy Specialist** will be

hired to support implementation, coordinate with LPSAs, train observers, monitor implementation, and develop additional resources, as needed. [REDACTED], **Video Producer**, has won awards for his work on education videos and has produced all videos to date for SERP's online PL. [REDACTED], **Creative Director**, brings a practice-informed design lens to his work designing all SERP materials and maintaining all SERP websites. [REDACTED], **Assistant Director**, has experience managing several federal grants, assisting with design decisions, developing and monitoring web content, and disseminating materials.

**Independent Evaluator.** MDRC, the independent evaluator for the proposed project, is a nonprofit, nonpartisan education and social policy research organization with a 40-year history of conducting large-scale demonstrations and evaluations. MDRC's staff of over 250 includes highly competent researchers with expertise in all aspects of qualitative and quantitative evaluation (study design, data collection, impact and implementation analysis, cost studies). MDRC has in-house fiscal and legal teams and a robust technology infrastructure with all the necessary systems in place to secure, safely store, and manage student data. [REDACTED], **Edd, Senior Research Associate at MDRC**, will lead the impact evaluation, oversee the collection and preparation of student records, direct the impact analysis, coauthor reports, and disseminate findings. [REDACTED] is currently leading the mid-phase EIR independent evaluation of STARI (with SERP) and has led the impact analysis for three other federally funded random assignment studies focused on reading interventions.

SERP and MDRC have extensive combined expertise in the areas necessary for this project: (i) a deep knowledge of STARI and reading interventions; (ii) expertise in all aspects of impact and implementation evaluation and cost analysis; (iii) substantial prior experience with large-scale evaluations of reading interventions; (iv) experience with the preparation of products

aimed at disseminating research findings to different audiences; and (v) project management.

**Consultants.** [REDACTED], **Edd**, served as Co-PI on the RfU and EIR projects, leading the development of the STARI curriculum and fidelity scale, and training coaches.

[REDACTED] will co-develop approaches for assessing teacher knowledge and instructional quality, advise on implementation fidelity, and co-author reports. [REDACTED], **Edd**, an expert on literacy and experimental design, and [REDACTED], **PhD**, an expert on measurement of teacher-student interactions, will advise on development and validation of a teacher knowledge measure.

## **Dissemination**

**Disseminating the Product.** We will disseminate program information and findings through email and social media campaigns, advertising in education publications, and via other nonprofit organizations with broad reach. SERP email campaigns reach more than 40,000 contacts and experience high open and click through rates (~15-40% and 5-40%, respectively). Publications such as EdWeek provide multiple marketing possibilities—from advertisements to briefs and papers—that can be targeted to specific audiences during specific times when districts make decisions about program adoptions for the following year. In addition, we will use search engine optimization (SEO) techniques to drive traffic to the STARI website. Google Ads available to SERP free of charge (up to \$10,000/mo) will help target specific, high-leverage keywords to drive traffic to the site (e.g., searches for reading interventions for adolescents).

SERP staff will attend national literacy conferences (e.g., International Literacy Association) and local conferences (e.g., Mississippi's Making Connections Conference). We have budgeted for SERP's literacy specialist, accompanied by a local practitioner, to attend conferences in each partner state in years 2-5 of the proposed project, including setting up tables at which participants can see the product, watch videos of its use, and talk with staff. We will

also use SERP's ties to other nonprofit organizations (for example, Instructional Partners, Teaching Matters, Learning Ally, and the Pennsylvania Teacher Technical Assistance Network), who have broad national networks, and who benefit from pointing districts to effective solutions.

**Disseminating the Findings.** Information on the effectiveness of the STARI program when implemented with the online PL model, with Phonics Boost, and for diverse student subgroups will be shared as soon as results become available. Specific findings will be incorporated into all materials and website texts that describe and provide access to the program ([serpinstitute.org/stari](http://serpinstitute.org/stari)). We will leverage academic and policy *conferences* as a way to disseminate interim and summative findings to different stakeholders (e.g., American Educational Research Association, Council of Chief State School Officers). SERP will also host a webinar about the study's findings, lessons learned, and implications for adolescent literacy programming, which will be advertised through SERP and MDRC networks.

We will prepare written products geared both at a general audience and at more specific audiences. Accessible products for practitioners and policymakers will include a short brief describing the intervention and summarizing the study's findings, with links to provide access to STARI materials and implementation support. Products for researchers and policymakers will include a peer-reviewed journal article summarizing evaluation findings. Written research products will be posted on ERIC, and will be available for free on MDRC and SERP websites, which are visited by more than 50,000 people each month, and circulated via SERP/MDRC social media networks, which reach thousands of stakeholders.

## **Project Design**

### **Conceptual Framework**

The conceptual framework for the proposed project appears in Figure 1, and is enlarged

in Appendix G. Underlying this framework is the assumption that learning to read fluently and comprehend effectively in the middle grades requires commitment and disciplined attention on the part of a student. While STARI materials and practices are key inputs, successful outcomes will be driven by mediators: students’ self-efficacy and social and cognitive engagement. We assume further that the effective implementation of the program depends on teachers’ capacity to enact effective STARI instructional practices, and that the required teacher capacity is supported by professional learning resources and activities.

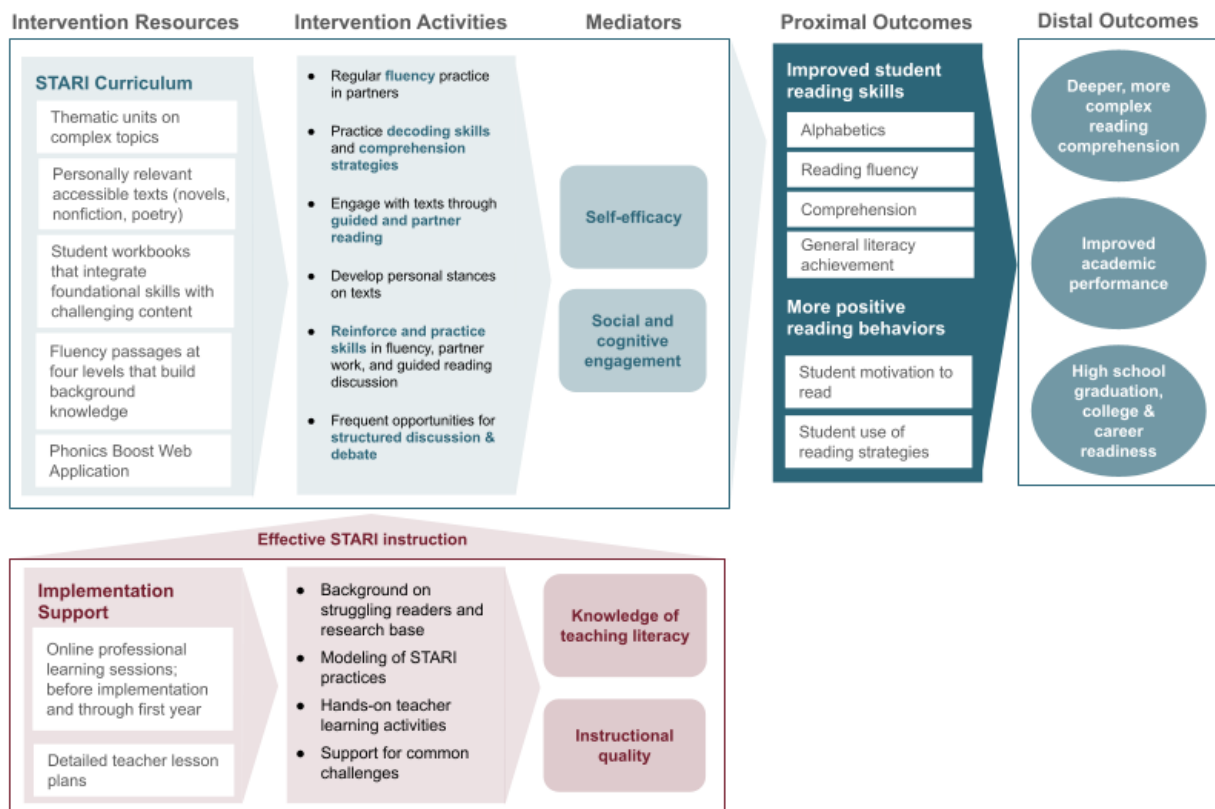


Figure 1. Conceptual Framework

The change process begins with *STARI intervention resources and activities* that are attentive to supporting student motivation. *Effective STARI instruction* is supported by the online PL series through which teachers develop an expanded knowledge base on struggling readers, understand the rationale behind key instructional practices, and become familiar with

how they play out in the classroom. However, even a well-designed and taught intervention will only be effective if students do the work required to make progress. In this framework, it is *because* materials attend to accessibility for students at multiple skill levels that students experience the reading success that leads to self-efficacy; it is *because* the activities are designed for collaboration and discussion that students become socially engaged in developing reading skills, and it is *because* the topics are complex and provide rich opportunities for discussion and debate that students become cognitively engaged. These emergent responses to the STARI resources and activities are the mediators of the proximal, and eventually distal, outcomes.

### Goals, Objectives and Outcomes

|   |   |
|---|---|
| <b>Goal 1: To improve the reading skills—including both foundational skills and deep comprehension—of middle school students who struggle with reading, resulting in improved academic performance, progress toward closing the gap between students’ current performance and grade-level expectations, and ultimately, high school graduation and college and career readiness</b> |   |
| Objectives  | Outputs and Outcomes  |
| 1.1 Recruit 80 schools and randomize them to II or DI condition   | 1.1 List of school assignments with signed participation agreements   |
| 1.2 Implement STARI with ~15,000 high-need students total across 4 years  | 1.2 Class rosters from study schools  |
| 1.3 Demonstrate effects of STARI on: reading behaviors, <i>proximal</i> reading skills, progress toward grade-level expectations, <i>leading indicators of distal</i> academic outcomes   | 1.3 Impact of STARI (compared with BAU) on: student reading surveys, ReadBasix scores, state test scores, course pass rates   |
| <b>Goal 2: To build teacher capacity for effective literacy instruction and to measure the impact of STARI professional learning on teachers’ knowledge and instructional practice</b>  |   |
| Objectives  | Outputs and Outcomes  |
| 2.1 Recruit ~120 STARI teachers per group   | 2.1 List of participating teachers  |
| 2.2 Increase teachers’ knowledge for literacy instruction   | 2.2 Register 100% of teachers for PL; at least 80% of teachers complete at least 80% of PL. Develop and validate a measure of teachers’ knowledge for literacy instruction; At least 80% of teachers complete assessments of knowledge; Growth is demonstrated on average |
| 2.3 Teachers implement quality literacy   | 2.3 Annual observations of practice meet  |

|   |  |
|---|--|
| instruction in STARI classrooms   | thresholds for fidelity and quality of instruction   |
| <b>Goal 3: Determine for whom and under what conditions STARI is most effective</b>   |  |
| Objectives  | Outputs and Outcomes   |
| 3.1 Recruit a sufficiently large and diverse sample to analyze subgroups of interest  | 3.1 Sample will include: 35% students who read on a 1st or 2nd grade level; 10% English learners; 15% students with IEPs |
| 3.2 Demonstrate subgroup effects of STARI on all measures   | 3.3 Impacts on student reading surveys; ReadBasix; state test scores; course pass rates                                  |
| <b>Goal 4: To create an affordable, sustainable package of supports for schools that choose to implement STARI outside of a grant-funded research study</b> |  |
| Objectives  | Outcomes and Outputs   |
| 4.1 Provide all DI schools access to STARI materials and PL in their third year   | 4.1 Register DI teachers for STARI PL series; Ship materials to DI schools   |
| 4.2 Randomly assign DI schools to enhanced vs. standard support; Provide enhanced or standard support to schools  | 4.2 List of school assignments; LPSA activity logs; Records of engagement with online materials                          |
| 4.3 Determine relationship between support condition and likelihood of sustaining STARI   | 4.3 List of whether or not schools implement STARI in the year following DI  |
| 4.4 Demonstrate cost-effectiveness of STARI as compared to BAU  | 4.4 Net cost per student of STARI; Cost-effectiveness ratio  |
| <b>Goal 5: To disseminate results and expand the use of STARI nationwide</b>  |  |
| Objectives  | Outcomes and Outputs   |
| 5.1 Implement online marketing strategies (e.g., SEO, Google Ads, Edweek online ads)  | 5.1 Increase in web traffic and downloads  |
| 5.2 Develop and distribute dissemination materials to promote STARI   | 5.2 Increase in PL registrations and STARI bundle orders   |
| 5.3 Disseminate results from evaluation and implementation studies  | 5.3 List of conferences and papers   |
| 5.4 Disseminate STARI to practitioners  | 5.4 List of conferences or webinars/meetings   |

**Needs of the Target Population**

The target population for STARI and for the study will include middle school students who are not proficient on their state’s spring ELA test. Just as in the RfU and EIR mid-phase studies, the students who read below grade level in the proposed partner states are predominantly low-income students of color. The COVID pandemic has disproportionately impacted this target

population, exacerbating existing inequities and increasing the urgency of intervention.

**Table 2. Characteristics of Middle Schools in Cohort 1 States**

| <b>School Characteristics</b>                            | <b>Mississippi</b> | <b>New Mexico</b> |
|--|--------------------|-------------------|
| Title I eligibility (%)                                  | 81.7               | 94.3              |
| Free-reduced price lunch (%)                             | 80.0               | 75.1              |
| Racial-ethnic composition (%)                            |                    |                   |
| Black  | 47.0               | 1.6               |
| Hispanic   | 2.7                | 50.7              |
| White  | 46.3               | 23.5              |
| Asian  | 0.5                | 1.1               |
| Other  | 1.6                | 13.2              |
| Location (%)   |                    |                   |
| Urban or Suburban  | 15.0               | 41.1              |
| Town   | 20.7               | 19.8              |
| Rural  | 64.3               | 39.1              |
| Median enrollment per grade (# students)                 |                    |                   |
| Grade 6  | 86                 | 54                |
| Grade 7  | 89                 | 53                |
| Charter (%)  | 0.5                | 27.1              |
| 6 <sup>th</sup> graders not proficient on ELA state test | 73%                | 69%               |
| 8 <sup>th</sup> graders not proficient on NAEP Reading   | 75%                | 77%               |
| English Language Learners (%)                            | 2%                 | 16%               |
| Students with IEP (%)                                    | 12%                | 19%               |
| Number of schools serving grades 6 and 7                 | 213                | 192               |

Source: Common Core of Data (CCD), state department of education websites, and NAEP 2019.

STARI has demonstrated effectiveness in meeting the needs of this population. STARI provides intervention resources with elements that improve students’ reading skills while also attending to their self-efficacy and engagement with reading, including texts that are culturally relevant (Ivey & Johnston, 2013; Moje et al., 2008, Tatum, 2008) and accessible to below-grade level readers (Ehri et al., 2007; Fulmer & Tulis, 2013; O’Connor et al., 2002), but that are age-appropriate and have characteristics of cognitive challenge (Kamil et al., 2008). In contrast to other reading interventions, which often are not engaging or age appropriate and/or focus on isolated skills practice at the expense of higher level comprehension skills (Compton et al., 2014). STARI’s tightly integrated program elements and its focus on motivation and self-efficacy



make STARI likely to produce better outcomes than current practice.

### **Project Evaluation**

MDRC will conduct an independent evaluation of STARI in the expansion sites. The evaluation will include: an **impact study** that will evaluate STARI’s effect on student outcomes (Project Goal 1) and meet What Works Clearinghouse (WWC) standards *without* reservations (WWC, 2020); an **implementation study** that will assess the fidelity with which STARI is implemented and explore the mediating processes leading to student gains (Project Goal 2); a **service contrast study** to understand how STARI differs from other Tier 2 reading interventions used in the study schools; a **replication study** that will generate lessons for STARI’s delivery and scale-up (Project Goals 3-4); and a **cost and cost-effectiveness** analysis (Project Goal 4).

### **Meeting WWC Standards Without Reservations**

**Cluster (school-level) random assignment research design.** STARI’s effects on student outcomes will be evaluated using a two-stage school-level random assignment research design. In the first stage, schools will be randomly assigned to begin STARI implementation immediately (“immediate implementation” or II group, 40 schools) or to begin STARI implementation two years later (“delayed implementation” or DI group, 40 schools). To improve the precision of estimated effects, random assignment will be blocked by district, or by groups of schools with similar reading proficiency rates (in rural areas and small towns).

A school-level random assignment design is proposed because school-level STARI implementation—in which all eligible students get STARI as a Tier 2 intervention—is more feasible for schools to implement than student-level random assignment, and better reflects how STARI would be implemented in a non-study setting. An important lesson from the RfU and EIR mid-phase studies is that student-level random assignment is often infeasible or extremely

complicated, for example because schools often use a scheduling model whereby all students in a cohort take the same classes together. Also, districts typically require that Tier 2 support be provided to all students who need it, requiring schools to support two literacy interventions (one for control students), increasing the strain on instructional leaders.

In the proposed project, STARI's effects on student outcomes, compared to "business as usual" (BAU) instruction, will be evaluated in the first two years of implementation (Year 1 and Year 2 in Table 1). The study sample will include students in study schools who, in Year 1 or Year 2: (1) are enrolled in 6th or 7th grade; and (2) were not proficient on their state's ELA state test at baseline. "Baseline" will be defined as the spring before a student's first opportunity to receive STARI (i.e., spring of 5th grade for students who enter the study sample as 6th graders; spring of 6th grade for students who enter the study sample as 7th graders). This sample definition will be applied consistently across all II and DI schools.

**Reliable measures of student outcomes.** The impact study will examine STARI's effect on the student outcomes in the conceptual framework (Figure 1). This includes four (proximal) domains prioritized in the WWC protocol for adolescent literacy interventions (WWC, 2016): alphabetics, reading fluency, comprehension, and general literacy achievement. Students' reading behaviors (proximal) and their course performance (distal and a leading indicator of high school graduation) will also be examined. To maximize cross-study learning, the outcome measures will be the same as prior STARI studies (see Table 3 for a summary of measures). The reliability of all measures exceeds the WWC threshold of 0.60.

Like the RfU and mid-phase EIR studies, **alphabetics, fluency, and comprehension** will be assessed using the ReadBasix, a computer-administered assessment developed by ETS (Sabatini et al., 2015). The ReadBasix includes subtests for the reading skills that pose the most

difficulty for struggling adolescent readers: word recognition/decoding, morphological awareness, reading efficiency, vocabulary, sentence processing, and reading comprehension. ReadBasix scores are predictive of state ELA scores (O'Reilly et al., 2012) and subtest reliability is high ( $\alpha=0.77-0.93$ ). The test takes 45 minutes to 1 hour to complete. Where available, we will also explore using students' scores on district formative reading tests (from administrative records) as a supplemental, policy-relevant measure of comprehension (z-scored for pooling).

**Table 3. Research Questions and Measures for the Impact Study**

| Research Questions                          | Student Outcomes  | Data Source                             |
|---|---|---|
| What is the effect of STARI on alphabetics? | Word recognition score (50 items, $\alpha=0.91$ ) <sup>a</sup><br>Morphological awareness (32 items, $\alpha=0.90$ ) <sup>a</sup><br>*Alphabetics composite (average of two subtests)                                   | ReadBasix                               |
| ... on reading fluency?                     | *Efficiency of reading (36 items, $\alpha=0.93$ ) <sup>a</sup>  | ReadBasix                               |
| ... on comprehension?                       | Vocabulary (38 items, $\alpha=0.86$ ) <sup>a</sup><br>*Reading comprehension (22 items, $\alpha=0.77$ ) <sup>a</sup><br>Sentence processing (26 items, $\alpha=0.84$ ) <sup>a</sup><br>District formative reading tests | ReadBasix assessment<br>Student records |
| ... on general literacy achievement?        | *ELA state tests score (z-score)<br>ELA state test proficiency  | Student records                         |
| ... on course performance in core subjects? | Overall GPA in core subjects<br>Courses passed in core subjects   | Student records                         |
| ... on students' reading behaviors?         | Reading motivation (14 items) <sup>b</sup><br>Use of effective reading strategies (4 items) <sup>c</sup>  | Student survey                          |

Notes: \*Confirmatory outcomes. <sup>a</sup> Sabatini et al. (2015); <sup>b</sup> Wigfield & Guthrie (1997); <sup>c</sup> Somers *et al.* (2010). Domains shaded in grey are prioritized by the WWC protocol for adolescent literacy interventions (WWC, 2016).

Like the mid-phase EIR study, students' **general literacy achievement** will be assessed using students' scores on ELA state tests. State tests are as reliable as commercial tests and are more widely used and policy relevant. The content of ELA state tests also substantially overlaps across states, with a consistently heavy focus on deep, complex reading comprehension, both literary and informational. To pool across assessments, test scores will be z-scored by school year, grade, and assessment (May et al., 2009).

Like the mid-phase EIR study, students' **reading behaviors** will be assessed using an online student survey, administered after the ReadBasix, that takes about 5-10 minutes to complete. Intrinsic reading motivation will be measured using a subscale from the Motivations for Reading Questionnaire (MRQ; Wigfield & Guthrie, 1997) which is appropriate for middle school students reading at an elementary school level and has a reliability of 0.70 (Davis et al., 2018; Troyer et al., 2019). Students' use of effective reading strategies will be measured using a scale adapted from the Enhanced Reading Opportunities study (Somers et al., 2010), which has reliability above 0.75. Finally, **course performance** will be measured using course marks and courses passed in core content areas (ELA, math, science, social studies).

The evaluation will also measure the **baseline characteristics and reading achievement** of students, using administrative records. This information will be used to determine STARI eligibility, describe the sample, assess baseline equivalence, and as covariates in the analysis. Demographic data will include students' gender, race/ethnicity, grade level, and eligibility for special programs (e.g., free or reduced-price meals, ELL programs, IEP). Baseline equivalence will be assessed using students' baseline state ELA scores (z-scored for pooling), which is an acceptable baseline measure for adolescent literacy intervention (WWC, 2016).

The ReadBasix and survey will be administered in the spring of Years 1 and 2 (2024 and 2025 for Cohort 1 schools; 2025 and 2026 for Cohort 2). Administrative records (state tests, course performance) will be requested annually to measure students' outcomes while in middle school and at baseline. To reduce burden on districts, these records will be obtained from state departments of education. If a state cannot provide data, records will be obtained from districts. Parents and guardians will have an opportunity to opt their child out of data collection

(ReadBasix, student survey, and the release of records with identifiers to allow linking across data sources). Schools will agree to a passive consent process as a condition of participation.

Given this plan, the impact study will be able to rigorously assess the effect of **one year of STARI** compared to BAU instruction (6th & 7th graders in Year 1; 6th graders in Year 2), which is prioritized by the WWC. The study will also explore (non-experimentally) the effect of **two years of STARI** (7th graders in Year 2) and whether effects are **sustained** one year after the end of the intervention (8th graders in Year 2).

**Limiting the threat of bias caused by joiners and low attrition.** To meet WWC standards without reservations, the confirmatory (primary) impact analysis will focus on STARI's effect in the *first implementation year* (SY 23-24 for Cohort 1 schools and SY 24-25 for Cohort 2). In prior studies, STARI implementation was strong in the first year, so effects are expected even in Year 1. Importantly, focusing on Year 1 will limit the threat of bias caused by joiners. Because the study schools will serve grades 6-8, 6<sup>th</sup> grade students will be enrolling for the first time in the fall *after* random assignment, so they will be “joiners” by WWC standards. Because STARI is a low-profile class and not a whole-school intervention, it is not expected to change parents' enrollment decisions (or student composition), especially in the first year of implementation when parents are less aware of STARI. Thus, to meet WWC standards without reservations, the confirmatory student sample will be limited to *early joiners*, defined as 6<sup>th</sup> and 7<sup>th</sup> grade students enrolled in the study schools in the first 6 weeks of Year 1.

Focusing on Year 1 will also limit the threat of bias caused by school and student nonresponse. Schools are expected to engage most actively in the data collection in Year 1, when many of the school staff consulted during school recruitment will still be present (whereas

leadership turnover in later years could lead to lower engagement). In the first year, we expect about 76.5% of eligible students to complete the ReadBasix and survey, based on reasonable assumptions about passive consent (90%) and attendance on testing days (85%). Given this level of attrition (23.5%), the study will meet WWC standards without reservations if the differential attrition between II and DI schools is less than 9.4%, which the RfU and EIR mid-phase studies indicate is achievable. Offering STARI to the DI schools later in the project, and using a passive parent consent process, will also help minimize differential nonresponse at the school and student level. For records-based outcomes (e.g., ELA state tests), the study will also meet WWC standards without reservations because records will be available for all schools and students.

**Accounting for clustering, missing data, and multiple testing.** To account for the cluster-level study design, the analysis of effects will be based on a two-level model, with students nested in schools. (The classroom level can be omitted because middle school students have multiple teachers, and nearly identical results are obtained whether or not the classroom level is omitted; see Zhu et al., 2012). The following model will be used:

$$\text{Level 1 (students): } Y_{ij} = \alpha_j + \delta ELA_{-1ij} + \theta G7_{ij} + \sum_M \omega_m D_{mij} + \varepsilon_{ij}$$

$$\text{Level 2 (schools): } \alpha_j = \alpha + \beta T_j + \sum_K \lambda_k B_{kj} + \mu_j$$

In this model,  $Y_{ij}$  is the 6<sup>th</sup> grade outcome of interest for student  $i$  enrolled at school  $j$  in the analysis sample.  $T_j$  is an indicator for whether a student's school was assigned to the II group (=1) or to the DI group (=0), and  $\beta$  is an intent-to-treat (ITT) estimate of the impact of STARI on outcome  $Y$ , whose statistical significance will be assessed at a 5 percent significance level using a two-tailed t-test. The model also includes covariates to improve the precision of estimated

effects. At the school level, the covariates are a set of  $K$  random assignment block indicators ( $\mathbf{B}_{kj}$ ). At the student level,  $ELA_{.ij}$  is a (z-scored) measure of a student's score on their baseline state ELA test,  $G7_{ij}$  is an indicator for grade (=0 for 6<sup>th</sup> and =1 for 7<sup>th</sup>), and  $\mathbf{D}_{mi}$  is a set of  $M$  baseline demographic characteristics (eligibility for free/reduced price lunch, race/ethnicity, gender, ELL status, and whether the student has an IEP).

The analysis sample will be limited to students in the study sample who have outcomes data. By definition, all of these students will have a baseline ELA state test score because it is used to define the sample. Demographic covariates ( $\mathbf{D}_{mi}$ ) may be missing for a small number of students and will be imputed using the indicator variable method, an approach that is considered acceptable by the WWC. (Missing data will be imputed with a constant, and indicators of missingness for each characteristic will be added to the model.)

To reduce false positives from multiple testing, one confirmatory outcome has been pre-specified for each of the four domains prioritized by the WWC (see Table 3). Conclusions about STARI's effectiveness will be based on its effect on these four outcomes in Year 1 for the confirmatory sample. All other findings will be exploratory and will only be used to contextualize the confirmatory findings. Based on WWC guidelines, p-values will not be adjusted for multiple testing because there is only one confirmatory outcome per domain.

The study will be well powered to detect STARI's effect on student outcomes (Project Goal 1) of the size found in prior research. The confirmatory sample will include 80 schools and about 7,040 students (44 students per school per grade). After accounting for student nonresponse, the minimum detectable effect size (MDES), at the 5 percent level, is expected to be 0.13 for effects on alphabetics, fluency, and comprehension, and 0.08 for ELA state test scores. Effects of this size are meaningful because they amount to 30-48% of the annual growth

made by middle school students (Hill et al., 2008). Importantly, the study will also be well powered to explore for whom and under what conditions STARI is effective (Goal 3). It will be possible to detect effects for key student subgroups: students with the lowest reading skills (MDES=0.09-0.14), students with an IEP (0.12-0.16), and ELLs (0.14-0.18). It will also be possible to estimate and compare the effect of one year of STARI for 6th graders in Year 1 and Year 2 (MDES by year = 0.10-0.14), in order to determine whether impacts are larger in the second year of STARI implementation. These MDES are based on parameter assumptions from the RfU study, the EIR mid-phase study, and other adolescent literacy studies.

The confirmatory analysis will focus on the “intent-to-treat” effect of STARI, or the difference in outcomes between eligible students enrolled in II schools and DI schools in Year 1. However, some eligible students in the II schools (5-10%) may *not* receive STARI for reasons including scheduling challenges or discretionary judgments about STARI’s suitability for that student. Thus, the effect of STARI on students *who received it* (complier average causal effect or CACE) will also be examined, using a two-stage least squares (2SLS) model with one binary endogenous variable (student received STARI vs. did not) and one binary instrument (II vs. DI school). CACE estimates will meet WWC standards without reservations because the required assumptions will be met (i.e., the exclusion restriction and instrument strength).

### **Components, Mediators, Outcomes and Thresholds for Fidelity**

**Implementation fidelity.** The evaluation will examine whether the STARI resources and activities in the conceptual framework (Figure 1) are implemented as intended by II schools in their first two years of implementation, and by DI schools in their first year. Our framework explicitly states that increased teacher knowledge will lead to improved instruction, which will



improve student outcomes. Furthermore, the proposed project explicitly intends to evaluate whether the scalable online PL model can achieve similar results to the cost-intensive in-person coaching used in prior studies of STARI. Therefore, it is critically important to measure the link between PL and student outcomes (i.e., instructional practice). For this reason, we have planned and budgeted to observe two class periods per year in each STARI teacher’s classroom. Fidelity will be measured using an adapted version of the framework for the mid-phase EIR study, which includes indicators and pre-specified benchmarks for acceptable implementation. **Fidelity to professional learning** (“intervention resources”) will include the amount of training that teachers receive, using metadata from the online PL series. **Fidelity to the intervention** (“intervention activities”) will be measured during annual classroom observations; the STARI fidelity tool (used in the RfU study) will be used to rate teachers’ delivery of the curriculum and their use of recommended instructional practices. The analysis will report fidelity scores and the proportion of schools with acceptable implementation, by implementation year.

These observations will also allow for a **continuous improvement process** that echoes the one used in the mid-phase grant, which resulted in a number of improvements to STARI’s usability. For example, we observed that teachers struggled with pacing: in response, we added timestamps to each lesson activity, to support teachers’ time management during lessons. We expect to make comparable adaptations to both products and processes in the proposed study, particularly those that are newer (such as the online PL series and the Phonics Boost app) or those that are developed as part of this project (such as the package of implementation supports).

**Classroom and student mediators.** The evaluation will also explore the mediating processes hypothesized to improve student outcomes in the conceptual framework. For *student mediators*, the quality of **student engagement** will be rated during classroom observations using

the STARI fidelity tool. Students' **reading self-efficacy** will be measured using an established scale (Wigfield & Guthrie, 1997) included in the student survey discussed earlier. The number of hours of STARI received by students (**dosage**) will be measured using daily attendance rates.

The evaluation will also measure the two *teacher mediators* in the conceptual framework (and Project Goal 2). During STARI classroom observations, **instructional quality** will be rated using the “accountable talk” subscale of the Instructional Quality Assessment (IQA; Junker et al., 2004) which focuses on the quality of discussion. Gains in STARI teachers' **knowledge for teaching literacy** will also be assessed. In the mid-phase EIR study, teacher knowledge was measured using an abbreviated version of Spear-Swerling and Cheesman's (2012) Teacher Knowledge Survey. However, because this survey is designed for K-6 teachers, it places insufficient emphasis on teaching higher-level comprehension. Furthermore, it does not attend to teachers' knowledge of teacher-student interactions, which is essential to effective STARI implementation. Therefore, the SERP team, advised by [REDACTED], [REDACTED], and [REDACTED], will develop an instrument to measure teachers' content and pedagogical content knowledge on all components of literacy relevant to middle- and upper-grade readers, including those below grade level (e.g., decoding, fluency, basic and deep comprehension, teacher-student interactions and literacy analysis). Data from the large study sample will allow this instrument to be validated, and linked with instructional practice and with average student outcomes—an important contribution to the field, as no measure of knowledge for literacy instruction (e.g., Carlisle et al, 2011; Phelps et al, 2014; Spear-Swerling and Cheesman, 2012) has yet been conclusively linked with instructional quality or student growth. This instrument will be administered at baseline (before teachers begin PL) and again each spring to measure knowledge gains. For the evaluation, the analysis will report mediator levels/gains by year and

examine the association between mediator values and effects on students at the block level.

**Table 4. Research Questions and Measures for the Implementation Study**

| Key Research Questions  | Measures  | Data Sources   |
|---|---|--|
| <b>Implementation Fidelity</b>  |   |  |
| Were the STARI resources and activities implemented with fidelity?                | <u>Professional learning fidelity:</u><br>- % teachers completing STARI PL series<br><u>Intervention fidelity (Cohens' K=0.84):</u><br>- Fluency work, guided reading, & partner work delivered as intended (18 items) <sup>a</sup><br>- Teacher uses recommended instructional strategies (9 items) <sup>a</sup> | Metadata from online professional learning series; structured classroom observations using a fidelity tool |
| <b>Mediators</b>  |   |  |
| How many hours of STARI did students receive?                                     | - % eligible students who received STARI<br>- Average hours of STARI received   | Student records  |
| What was the level of student engagement with STARI? Their reading self-efficacy? | - Student responsiveness (6 items, Cohens' K=0.84) <sup>a</sup><br>- Reading self-efficacy (3 items) <sup>b</sup>   | Structured classroom observations  |
| Did STARI teachers experience gains in their instructional knowledge?             | - Knowledge for teaching literacy (scale to be developed; see earlier discussion)   | STARI teacher survey (baseline; each spring)   |
| Was instructional quality high in STARI classrooms?                               | - Instructional quality (IQA, "accountable talk" subscale, 6 items) <sup>c</sup>  | Structured classroom observations  |

Notes: <sup>a</sup>Kim *et al.* (2016); <sup>b</sup>Wigfield & Guthrie (1997); <sup>c</sup>Junker et al. (2004). Implementation fidelity and mediators will be measured annually in II schools and in DI schools after they begin implementation.

***How STARI differs from other Tier 2 reading interventions.*** To further interpret STARI's implementation and effects, the evaluation will compare STARI to the other Tier 2 BAU reading interventions used by DI schools in the confirmatory year (Year 1). Intervention data will be obtained through a principal survey, and interventions will be reviewed and coded based on whether they include STARI features. Classroom observations of the interventions will be conducted in DI schools in Year 1 using the STARI fidelity tool (to determine whether aspects of STARI instruction are present) and the IQA (to describe general instructional practice differences between STARI and other interventions). Reading intervention teachers in DI schools will also complete the teacher knowledge assessment to enable comparisons across research groups. At the student level, student records will be used to examine the proportion of eligible

students in DI schools enrolled in a Tier 2 reading class.

### **Guidance about Effective Replication Strategies**

The evaluation will include a replication study that will generate insights about the strategies, conditions, and settings that lead to stronger implementation and bigger effects. To explore the *drivers of implementation fidelity and impacts*, we will examine whether STARI intervention fidelity is associated with: STARI teacher characteristics (years of experience, highest degree, certification) from the teacher survey; teacher baseline knowledge for reading instruction; professional learning fidelity; and school characteristics (state, rural/urban). The evaluation will also examine whether higher fidelity scores are associated with larger impacts on student outcomes (at the block level). To understand the *implementation experiences of school staff*, annual semi-structured phone interviews will be conducted with STARI teachers, LPSAs, and district literacy leaders to identify facilitators and barriers to fidelity, and adaptations to the intended implementation of STARI (and reasons for these adaptations). MDRC will also conduct annual interviews with SERP to document the types of challenges encountered by schools and strategies used by SERP to support implementation. To explore *for whom and in what contexts STARI is more effective*, the study will examine whether STARI's effects differ by baseline reading level ("below basic" vs. "basic" on state tests), by whether students have an IEP, by whether they are ELLs, and by grade (6<sup>th</sup> vs. 7<sup>th</sup>). STARI's effects will also be examined by implementation year (Year 1 vs. Year 2), by state, and by setting (rural/urban). We will also explore (non-experimentally) the effect of two years of STARI, and whether effects are sustained one year after the end of the intervention. To *test different implementation support strategies*, the study will compare DI schools randomized to enhanced vs. standard levels of STARI support

in their first year of implementation (Project Goal 4). The analysis will compare the proportion of schools in each group that continue to use STARI post-grant (Fall 2026 for Cohort 1; Fall 2027 for Cohort 2). DI schools will also be compared based on their implementation fidelity, mediator levels, and, in Cohort 1, student ELA state test scores and course performance.

A **cost study** will be conducted using the ingredients approach (Levin & McEwan, 2001) to inform resource allocation decisions (Project Goal 4). The impact study will collect data on costs for the necessary ingredients to implement STARI *and* the BAU reading interventions in DI schools. Cost data will come from key informant interviews; budget documents; teacher salary schedules; the NCES Fiscal Survey; and the **Cost Out** tool. The average per student cost of STARI will be estimated, overall and by ingredient type, implementation year, level of support (enhanced vs. standard), and by state. Start-up and ongoing costs will be presented separately and by the constituent that bears them (SERP, districts, schools). The analysis will also estimate the *net* per student cost of STARI, relative to the average per student cost of the BAU interventions in DI schools. STARI's cost-effectiveness will be calculated by dividing its net per student cost by its estimated (complier) effect size on student outcomes.

### **Feedback about Performance and Progress**

The evaluation plan includes annual analyses of implementation and student data, which will allow for periodic feedback and continuous improvement. MDRC will share and discuss findings as they are known, and also prepare annual internal memos summarizing implementation and interim impact findings and progress toward objectives. In support of Project Goal 5, the evaluation plan also includes reports of findings to practitioners and policymakers, a two-page project synopsis, and conference presentations.

## References

- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science (New York, N.Y.)*, 333(6045), 1034–1037. <https://doi.org/10.1126/science.1207998>
- Amendum, S. J., & Fitzgerald, J. (2013). Does structure of content delivery or degree of professional development support matter for student reading growth in high poverty settings? *Journal of Literacy Research*, 45(4), 465-502.
- Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal*, 111(1), 7-34.
- Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy*. Retrieved from Carnegie Corporation of New York.
- Beisiegel, M., Mitchell, R., & Hill, H. C. (2018). The design of video-based professional development: An exploratory experiment intended to identify effective features. *Journal of Teacher Education*, 69(1), 69-89.
- Boots, J (2021). *Unfinished Learning in DC: How the pandemic impacted student growth during 2020-21*. EmpowerK12.  
<https://research.empowerk12.org/unfinished-learning-in-dc/index.html>
- Burrus, J., & Roberts, R. D. (2012). Dropping out of high school: Prevalence, risk factors, and remediation strategies. *R&D Connections*, 18, Princeton, NJ: Educational Testing Service.

- Carlisle, J. F., Cortina, K. S., & Katz, L. A. (2011). First-grade teachers' response to three models of professional development in reading. *Reading and Writing Quarterly, 27*, 212-238.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5-51.  
doi:10.1177/1529100618772271
- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology 104*(1), 166-181.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3-12.
- Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of “quick fix” interventions for children with reading disability? *Scientific Studies of Reading, 18*(1), 55-73.  
<https://doi.org/10.1080/10888438.2013.836200>
- Davis, M. H., Tonks, S. M., Hock, M., Wang, W., & Rodriguez, A. (2018). A review of reading motivation scales. *Reading Psychology, 39*(2), 121-187.
- Delaco, R., Samuelson, C., Grifenhagen, J., Davis, D. S., & Kosanovich, M. (2022). Using insights from teachers to inform online professional development in early literacy instruction. *Literacy Research and Instruction, 61*(1), 84-111.  
<https://doi.org/10.1080/19388071.2021.1921889>
- Denton, C., Wolters, C. A., York, M. J., Swanson, E., Kulesz, P. A., & Francis, D. J. (2015). Adolescents' use of reading comprehension strategies: Differences related to reading proficiency, grade level, and gender. *Learning and Individual Differences, 37*(1), 81-95.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2021). COVID-19 and education: The lingering effects of unfinished learning. *McKinsey & Company*.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*(2), 414-448.

<https://doi.org/10.2307/30069443>

Fuchs, D., Fuchs, L. S., Thompson, A., Svenson, E., Yen, L., Al Otaiba, S., Yang, N., McMaster, K. N., Prentice, K., Kazdan, S., & Saenz, L. (2011). Peer-assisted learning strategies in reading: Extensions for kindergarten, first grade, and high school. *Remedial and Special Education*, *22*(1), 15-21.

Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction*, *27*(1), 11-20.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., Szejnberg, L., & Silverberg, M. (2008). *The impact of two professional development interventions on early reading instruction and achievement*, NCEE 2008-4030. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.



Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1.

*Reading and Writing, 25*(8), 1819-1845.

Gibbons, L.K., & Cobb, P. (2017). Focusing on Teacher Learning Opportunities to Identify Potentially Productive Coaching Activities. *Journal of Teacher Education, 68*(4), 411-425.

Guthrie, J. T., Wigfield, A., & You, W. (2012). *Instructional contexts for engagement and achievement in reading*. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601-634). New York: Springer.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., Howes, C., LaParo, K., & Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88-123. <https://doi.org/10.3102/0002831211434596>



Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172-177.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher, 42*(9), 476-487.

Ivey, G., Johnston, P. H. (2013). Engagement with young adult literature: Outcomes and processes. *Reading Research Quarterly, 48*(3), 255-275.

Joseph, L. M., & Schisler, R. (2009). Should adolescents go back to the basics?: A review of teaching word reading skills to middle and high school students. *Remedial and Special Education, 30*(3), 131-147. doi:10.1177/0741932508315646

Junker, B., Matsumura, L., Crosson, A., Wolf, M., Levison, A., Weisberg, Y., & Resnick, L. (2004, 01/01). Overview of the instructional quality assessment.

Kamil, M.L. (2003). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., and Torgesen, J. (2008). Improving adolescent literacy: Effective classroom and intervention practices: A Practice Guide (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc>

Kemple, J. J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., & Drummond, K. (2008). The enhanced reading opportunities study: Early impact and implementation findings. National Center for Education Evaluation and Regional Assistance.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Levin, H. M., & McEwan, P. J. (2001). *Cost effectiveness analysis: Methods and applications*. Thousand Oaks, CA: Sage Publications.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Moje, E. B., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review*, 78(1), 107-154.

National Assessment of Educational Progress (NAEP). (2019). *The Nation's Report Card: Results from the 2019 mathematics and reading assessments*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. Washington, DC: The National Academies Press.

National Reading Panel, Q. N. R. P. (2000). *National reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction : Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health. Washington, D.C.

- Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of the literature. *Reading Research Quarterly, 46*(3), 273-304.
- O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology, 94*(3), 474-485.
- O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a rti framework. *Reading Psychology, 33*(1-2), 162-189. doi:10.1080/02702711.2012.631865
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.
- Phelps, G., Weren, B., Croft, A., & Gitomer, D. (2014). Developing content knowledge for teaching assessments for the measures of effective teaching study. *ETS Research Report Series, 2014*(2), 1-92. <https://doi.org/10.1002/ets2.12031>
- Reim, M. C. (2014). *Barriers to high school completion create barriers to economic mobility* (CPI Discussion Paper No. 17). Washington, DC: Center for Policy Innovation, The Heritage Foundation.
- Sabatini, J., Bruce, K., Steinberg, J., & Weeks, J. (2015). SARA Reading Components Tests, RISE Forms: Technical Adequacy and Test Design, 2nd Edition. ETS Research Report, 13-32.
- Sailors, M., & Price, L. R. (2010). Professional development that supports the teaching of cognitive reading strategy instruction. *The Elementary School Journal, 110*(3), 301–322.

Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, K. K. (2013). A meta-analysis of interventions for struggling readers in grades 4-12: 1980-2011. *Journal of Learning Disabilities*, 0022219413504995. doi:10.1177/0022219413504995

Snell, E. K., Hindman, A. H., & Wasik, B. A. (2019). A review of research on technology-mediated language and literacy professional development models. *Journal of early childhood teacher education*, 40(3), 205-220.  
<https://doi.org/10.1080/10901027.2018.1539794>

Somers, M.-A., Corrin, W., & Sepanik, S. (2010). *The Enhanced Reading Opportunities Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Spear-Swerling, L., & Cheesman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading & Writing: An Interdisciplinary Journal*, 25(7), 1691–1723.

Tatum, A.W. (2008). Toward a more anatomically complete model of literacy instruction: A focus on African American male adolescents and texts. *Harvard Educational Review*, 78(1), 155–180.

[REDACTED]

[REDACTED]

[REDACTED]

What Works Clearinghouse. (2016). *WWC Evidence Review Protocol Adolescent Literacy Interventions, Version 3.0*. Washington, DC: What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education.

- What Works Clearinghouse. (2020). *What Works Clearinghouse Standards Handbook, Version 4.1*. Washington, DC: What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education.
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*(3), 420.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (REL2007–No. 033). Washington, DC: Regional Educational Laboratory Southwest, Institute of Education Sciences, U.S. Department of Education.