Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

An Evaluation of the Strategic Adolescent Reading Intervention (STARI) in Two Urban School Districts During the

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Somers, Marie-Andree	MDRC	https://orcid.org/0000-0002-8079-

Publication/Completion Date—(if *In Press,* enter year accepted or completed) June 2023

Check type of content being submitted and complete one of the following in the box below:

- ☐ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Evaluation report

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Office of Elementary and S through **[Grant number]** U411B1700 to **Institution]** the Strategic E. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Office of Elementary and S or the U.S. Department of Education.

An Evaluation of the Strategic Adolescent Reading Intervention (STARI) in Two Urban School Districts During the COVID-19 Pandemic

Prepared by:

Marie-Andrée Somers, MDRC

June 2023

ACKNOWLEDGEMENTS

This study was supported by an Education Innovation and Research (EIR) grant from the US Department of Education.

Table of Contents

OVERVIEW

The learning losses caused by the COVID-19 pandemic make this an important time to identify strategies that can help students accelerate their learning. The Strategic Adolescent Reading Intervention (STARI) is a year-long supplemental reading intervention that is intentionally designed to address the motivational barriers of middle school students who are reading below grade level. STARI is intended to be used for a full class period (45 minutes minimum), at least three days a week, in addition to their regular English Language Arts class. STARI has a Tier 1 (strong) ESSA evidence rating based on a student-level randomized experiment conducted in SY 2013-14 in a northeastern state, which found that STARI had statistically significant positive effects on students' reading skills, including their word recognition, efficiency of basic reading comprehension, and morphological awareness.

To explore the replicability of these results, the effect of STARI was evaluated in 11 middle schools in two urban school districts in SY 2021-22, a school year when instruction was still periodically disrupted by the COVID-19 pandemic. Eligible students were randomly assigned to the STARI class (n=183) or to a business-as-usual elective or reading class at their school (n=215). Students' English Language Arts (ELA) state test scores and their course marks at the end of the program year were obtained from school districts, and students' reading skills and reading motivation and self-efficacy were measured using a reading assessment and survey administered in Spring 2022. In one of the two study districts, the findings are challenging to interpret due to low response rates related to the COVID-19 pandemic; in the other study district, response rates were higher and more balanced across research groups. In that district, students in the STARI group had higher ELA state test scores than students in the control group and the difference between research groups is statistically significant at the 10 percent level (n=154,

effect size = 0.32, p-value=0.071). Students in the STARI group also had consistently higher scores on subtests of reading skills than students in the control group (*n*=80, effect size = 0.15 to 0.32). Although not statistically significant, the magnitude of these effects mirrors the findings from the prior evaluation of STARI, suggesting that STARI is a promising strategy for helping struggling adolescent readers catch up across different settings, including a time period when instruction was still challenging because of the pandemic. An upcoming large-scale evaluation of STARI will continue to build evidence on its implementation and effectiveness in additional settings and for different subgroups of students.

INTRODUCTION

Students' success in middle school has profound implications for their future. Attendance, grades, and test scores during the middle school years all predict students' odds of graduating from high school (Balfanz, 2009; Balfanz, Herzog, & MacIver, 2007; Kieffer and Marinell, 2012). Students who struggle with reading are at especially high risk of dropping out of school (Kamil, 2003; Biancarosa & Snow, 2004). By middle school, students are expected to read texts with complex words, academic sentence structures, and sophisticated content. Yet multiple barriers limit access to grade level texts for struggling middle school readers, including slow and inaccurate decoding, weak vocabulary, poor command of sentence structure and insufficient background knowledge (Catts *et al.*, 2012).

Helping middle school students who are struggling with reading is particularly challenging, however, due to the changes in motivation and engagement that occur during this period of child development (Eccles, 1999; Juvonen *et al.*, 2004). In early adolescence, students start to experience dramatic changes cognitively, physically, socially, and emotionally. Middle school students are more self-conscious, and therefore, they experience more anxiety when participating in activities in which they are less skilled, which can hinder their learning (Eccles, 1999; Eccles & Wigfield, 2000). This means that to be effective, reading interventions for struggling middle school students need to be both engaging and relevant, while also building students' self-confidence in their reading skills.

Developed by the Strategic Education Research Partnership (SERP) Institute, Harvard University, and Wheelock College, the Strategic Adolescent Reading Intervention (STARI) is a year-long intervention that is designed to address the motivational barriers faced by struggling middle school readers. STARI tackles gaps in students' basic reading skills and background

knowledge, but unlike other interventions, it seamlessly integrates those skills with complex comprehension tasks, and it gives a central role to student motivation. Students receive the STARI intervention in addition to their regular English Language Arts (ELA) class, during an elective period or an intervention period, and it can be used as a Tier 2 (targeted) intervention. All STARI curriculum materials are freely downloadable and printed versions are available for purchase. Thousands of registrants download STARI each year, from all 50 states and Washington, DC.

In 2017, SERP was awarded an Education and Innovation in Research (EIR) mid-phase grant from the US Department of Education to scale up and evaluate STARI in several high-needs school districts. The intention was to implement STARI in four urban school districts for three school years (SY 2018-19, SY 2019-20, and SY 2020-21) and to evaluate STARI's effect on student outcomes in Year 2 and Year 3. The project, however, was deeply impacted by the COVID-19 pandemic, starting with school closures in March 2020, followed by a year of virtual and hybrid learning in SY 2020-21. These circumstances made it challenging to implement STARI as intended and created significant barriers to collecting student data for the evaluation. Due to these exceptional circumstances, SERP decided to continue to support STARI implementation for a fourth year in two of the four study districts, in SY 2021-22, when in-person instruction had resumed in both districts.

This report presents the findings from the impact evaluation conducted in SY 2021-22, an unprecedented time when struggling adolescent readers had fallen even further behind because of the instructional disruptions caused by the COVID-19 pandemic. In 2022, only 31% of U.S. eighth graders could read at a proficient level, down from 34% prior to the pandemic (National Assessment of Educational Progress, 2022). Racial and socioeconomic disparities also persisted:

in 2022, only 16% of Black students, 21% of Hispanic students, and 19% of students living in poverty scored proficient on NAEP, as compared to 38% of White students and 41% of students not eligible for free and reduced-price meals. The deep learning losses caused by the COVID-19 pandemic make this an especially important time to evaluate the effectiveness of STARI, a program that is intentionally designed to help middle school students accelerate their learning.

Although the COVID-19 pandemic provides a policy relevant context for the study, it also created several challenges that resulted in the sample size being smaller than originally planned. In the first instance, the pandemic altered the study timeline, so the evaluation includes one cohort of students instead of two cohorts, and two districts instead of four districts. The pandemic also made it more difficult to recruit schools and students into the study, because parents and school staff faced competing demands on their time. The resulting small sample size of the study reduces its ability to statistically detect STARI's effect on student outcomes.

Fortunately, this study is not a standalone study – it is one of three rigorous evaluations of STARI – so its findings can be interpreted not solely based on their statistical significance, but also based on their alignment with the results of other evaluations. The first evaluation of STARI, which was conducted in four school districts in Massachusetts in SY 2013-14, found statistically significant positive effects on students' reading skills. An upcoming evaluation, which will begin in SY 2023-24, will examine STARI's effectiveness in a more geographically diverse sample of schools that will also include rural schools, and with a larger sample of students that will make it possible to explore STARI's effect for student subgroups. Thus, the pattern of findings from this report—when interpreted alongside the findings from these other evaluations—can still contribute to building evidence about STARI's effectiveness across different settings.

STARI: A Promising Strategy

STARI was developed in response to a practical problem faced by Boston teachers and administrators: middle and high school students do not have sufficient reading skills to comprehend their textbooks (Donovan *et al.*, 2013). Comprehension involves developing a personal and critical position on what has been read (Hemphill *et al.*, 2019; Newell *et al.*, 2011), which in turn requires that students be engaged in what they are reading. Yet adolescents with reading difficulties are typically experiencing high levels of disengagement and a lack of self-efficacy (Guthrie *et al.*, 2012).

STARI is intended to be used for a full class period (45 minutes minimum), at least three days a week for the entire school year, in addition to their regular English Language Arts (ELA) class. Because many middle school teachers are generalist classroom teachers with limited experience with reading pedagogy, STARI includes teacher lesson plans with detailed guidance on the implementation of classroom activities and student progress monitoring. These materials make it possible for STARI to be taught successfully by any teacher who has a growth mindset, including ELA teachers, special education teachers, and even paraprofessionals.

STARI is organized into three year-long series for grades 6-8. Each STARI series is divided into three thematic units, each of which includes lessons for approximately 40 instructional days (120 days per year across all three units).[1] Each unit is motivated by an essential question such as "What makes a family?" and is framed around a core novel, as well as thematically related nonfiction texts.

[1] If offered for 45 minutes daily, this represents about 8 weeks of instruction per STARI unit, or 24 weeks of instruction across all three units in the year-long series.

The activities included in each STARI unit are designed to build students' basic reading skills—decoding, morphological analysis, and fluency—as well as their vocabulary and reading comprehension:

- Fluency, which is the ability to read (or decode) words at a conversational rate with ease and accuracy and expression, is an essential building block for reading comprehension. In daily *fluency practice*, students engage with a partner in repeated readings of short texts (leveled fluency passages). These fluency passages build the background knowledge needed to comprehend core unit texts and provide targeted practice with decoding (i.e., how to apply knowledge of letter-sound relationships and letter patterns to correctly pronounce written words) and morphological analysis (i.e., analyzing the smallest units of meaning, like root words, prefixes, and suffixes, to understand the meaning of a word). Each fluency passage includes a "mini debate" for partners to contrast personal stances on the topic.
- Students also engage in daily *partner reading or guided reading with novels and nonfiction texts*. For guided reading, all students read the same age-appropriate novel, and teachers are provided with questions to build students' reasoning skills and scaffold discussion. 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*.[2] Students frequently practice these strategies in context with fluency passages and unit novels and nonfiction.

A 45-minute STARI lesson typically contains two of the three major components of STARI: fluency practice, guided reading, and/or partner reading. Some lessons also include decoding and comprehension mini lessons. Debates happen at the end of each unit. Each activity in the lesson plan includes a timestamp so that teachers can adapt the sequence of activities to fit class periods of different lengths.

STARI's design differs from that of other interventions in the extent to which it incorporates features – related to both topic selection and instructional practices – that prior research suggests can effectively increase student motivation and self-efficacy (Guthrie *et al.*, 2012):

- *Engaging, relevant themes and texts with culturally familiar content*: STARI focuses on timely topics such as bullying, which were chosen for their relevance to students' lives. This provides opportunities for both perspective taking and critical thinking skills that are important for deep comprehension and that are used in unit debates that students find highly motivating (Ivey & Johnston, 2013; Tatum, 2008). For example, in one unit, students read texts with different representations of young people's first jobs.
- *Texts matched to students' reading levels:* STARI aims to build self-efficacy through the matching of unit texts to students' reading levels, so that students can experience success and build their confidence (Fulmer & Tulis, 2013; O'Connor *et al.*, 2002). Many

[2]Comprehension strategy instruction is based on the Reciprocal Teaching (RT) model (Palincsar & Brown, 1984), while decoding instruction focuses on skills such as finding base words and syllable division rules.

adolescents with reading difficulties view themselves as less competent (Wolters *et al*. 2014), so STARI strategically selects novels with Lexile levels that make them accessible to below-grade level readers, but which still contain elements of cognitive complexity (e.g., complex characters and ambiguous plot points) that address grade-level skills and standards. Fluency passages are written at four levels of difficulty so that students can be matched to the level that will promote the greatest progress.

- *Structured discussion of cognitively challenging content:* STARI students engage in daily discussion of text. Lesson plans and student workbooks provide complex, engaging questions aligned with unit themes. In addition, every unit includes at least one whole-class debate.
- *Integration of basic skills work with challenging content*: Typical interventions teach component skills in isolation, while STARI embeds decoding practice into cognitively complex and engaging activities (Hemphill *et al.*, 2019). For example, students are taught strategies for chunking words such as 'underdog' and 'backstabber' in the bullying unit.
- *Frequent opportunities for peer collaboration*: STARI incorporates frequent opportunities for peer collaboration, a practice found to be effective in past research (Fuchs *et al.*, 2011). STARI students work daily with partners to practice reading fluently and to compare and contrast perspectives on highly topical readings. Students also work in teams during debate activities, voicing their own perspectives to peers.

A small-scale randomized trial of STARI conducted in four school districts in Massachusetts in SY 2013-14 found statistically positive effects on students' foundational reading skills (Kim *et al.*, 2017). The eight schools in the study qualified for Title I status, with

73% of students eligible for free or reduced-price lunch. Professional learning for STARI teachers included a three-day summer institute, regular in-class guidance from project literacy coaches, district-based professional learning communities (PLCs), and three statewide network meetings. The study focused on students in grades 6-8 scoring below proficient on the state ELA assessment, placing them in the lowest third of test-takers statewide (*n*=402). Eligible students were randomly assigned to STARI or their school's own intervention program. Students' reading skills were measured using the RISE (now renamed the ReadBasix), a computer-based assessment developed by ETS (O'Reilly *et al.*, 2012). Students assigned to STARI outperformed control students by a statistically significant amount on subtests of word recognition and decoding (effect size=0.20), morphological awareness (effect size=0.18), and efficiency of basic reading comprehension *(*effect size=0.21).[3] 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 (Kim *et al.*, 2011). The findings are also notable because 70% of students in the control group received another reading intervention. STARI has received a Tier 1 (Strong) ESSA evidence rating based on the findings from this study.

Scaling and Evaluating STARI

In 2017, SERP was awarded an Education and Innovation in Research (EIR) mid-phase grant from the US Department of Education, to scale-up and support STARI implementation in four high-needs school districts and to conduct a rigorous evaluation of its effect in these new settings. As part of the grant-funded implementation of STARI, SERP planned to provide the following curriculum and professional learning resources to participating schools and teachers:

[3] See Appendix A for a description of each ReadBasix subtest.

- *STARI curriculum materials*: STARI materials, including detailed teacher lesson plans, leveled fluency passages, and student workbooks, are available for free online. However, as part of the EIR grant, each STARI teacher received printed copies of the teacher lesson plans and student workbooks, as well as the unit novels and nonfiction texts which would normally need to be purchased.
- *Teacher professional learning*: Effective implementation of STARI depends on teachers' capacity to enact STARI instructional practices. Many middle school teachers, however, do not have specialized training in how to provide reading instruction to struggling students, and therefore a key component of the implementation plan for the EIR grant – much like the prior study in Massachusetts – was to provide a set of robust professional learning supports for STARI teachers. This included district-based multi-day summer training institutes for new STARI teachers;[4] regular coaching throughout the school year from SERP-hired STARI coaches;[5] and monthly professional learning community (PLC) activities. The PLC activities included an online guidance session on a selected topic each month, paired with an in-person reflection session on that same topic that were held

[4]The summer institutes were held in Summer 2018 and 2019 and covered topics such as: the research base on struggling adolescent readers and skilled reading; STARI's approach to helping struggling readers; and how to implement the STARI curriculum elements and associated activities (fluency routines, guided reading, partner work, and debate). The training was supplemented by videos of classroom practices. In Summer 2018, the institute was 3 days and was taught by SERP staff and experienced STARI teachers from non-study districts. In Summer 2019, the institute was 2 days and the SERP-hired STARI coaches also assisted with the teaching.

[5] SERP hired seven coaches to provide support to STARI teachers during the first three years of implementation in the four study districts. Coaches were expected to provide at least 21 coaching sessions to each STARI teacher per school year. SERP-hired coaches received training. This included a summer training institute prior to coaches' first year of coaching, as well as regular cross-district check-ins with SERP to discuss the challenges faced by teachers. The training for SERP coaches happened in Summer 2018 in Washington DC and lasted 4 days. In Summer 2019, before the pandemic, SERP also invited districts' literacy coaches to a training to build capacity for sustaining STARI implementation post-grant.

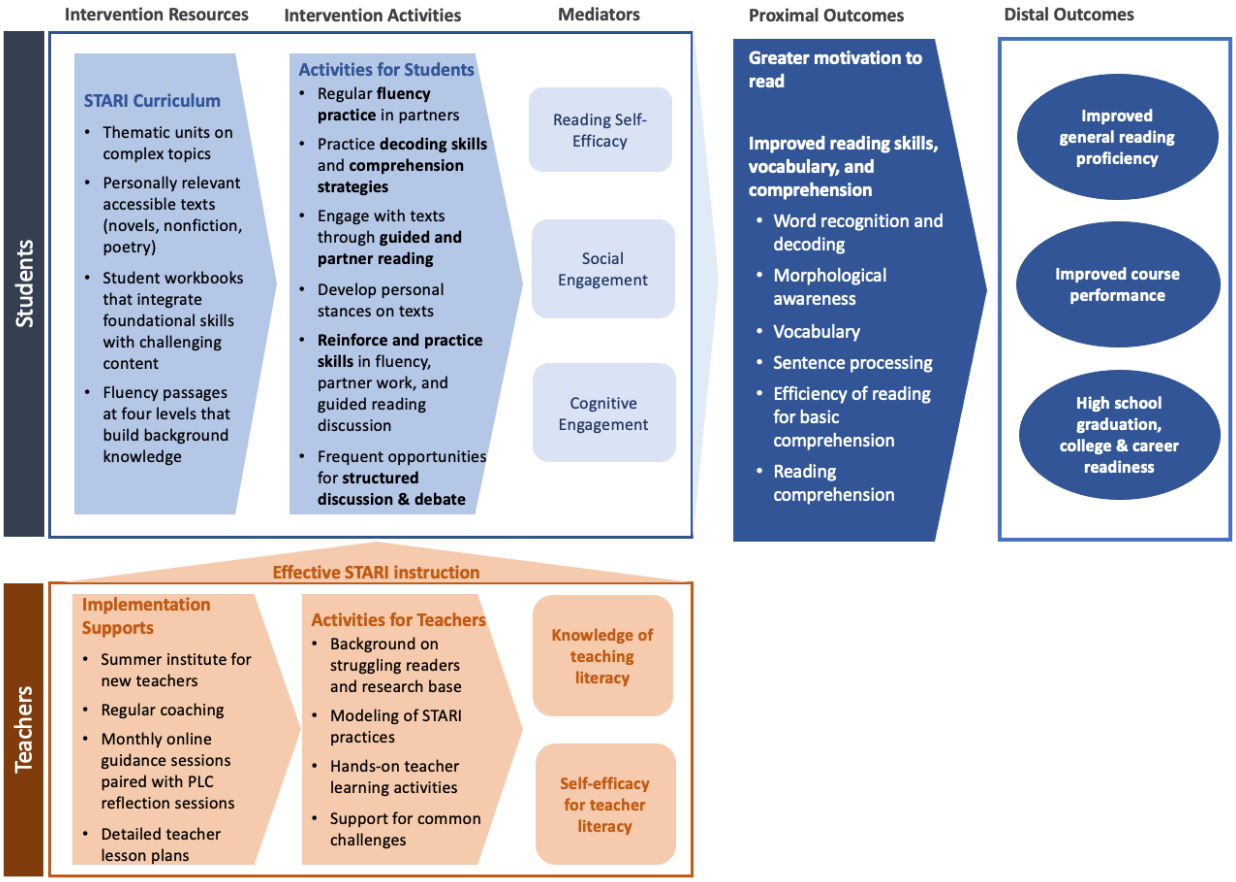
at the district level and moderated by coaches.[6] Coaching and PLCs were primarily targeted at teachers who were in their first two years of STARI implementation, with lighter-touch supports provided to more experienced STARI teachers as needed.

Exhibit 1 shows the logic model for the grant-funded implementation and evaluation of STARI. The change process begins with STARI curriculum resources and activities that are attentive to enhancing student motivation. Effective STARI instruction is supported by professional learning resources for STARI teachers (summer institute, professional learning series, and coaching) 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. In turn, effective STARI instruction is hypothesized to help students feel more positively about their reading self-efficacy and their motivation to read, and to help them accelerate the development of their reading skills, vocabulary and reading comprehension. In turn, students' general reading proficiency will also improve and they will perform better in their classes, which in the longer term will help students complete high school and prepare them for college and their careers.

The original plan for the EIR-funded project had been to implement STARI in four school districts for three school years (SY 2018-19, SY 2019-20, and SY 2020-21). The intention was to evaluate STARI's effect on student outcomes after teachers had had the chance to learn and become comfortable implementing the program. Thus, Year 1 (SY 2018-19) was a "training year" in which no collection of student data was planned, and the impact evaluation was intended to focus on STARI's effect in Year 2 (SY 2019-20) and Year 3 (SY 2020-21). In the first

[6]The monthly topics addressed key STARI practices and obstacles that teachers may encounter. To accommodate the fact that some teachers had longer travel times and/or extended school days, the PLC meetings in some districts were a mix of in-person meetings and moderated virtual meetings.

Exhibit 1: STARI Logic Model



implementation year (SY 2018-19; Year 1), which predates the pandemic, STARI was implemented in the target number of districts as intended based on the original implementation plan.[7] In subsequent school years (Year 2 and Year 3), however, the COVID-19 pandemic affected the availability of student data for the impact study[8] and disrupted the implementation of STARI classes and the professional supports for teachers:

- At the onset of the pandemic in SY 2019-20 (Year 2), the study schools were in their second year of STARI implementation. When schools closed in March 2020, many

[7] See Appendix H for an overview of implementation fidelity during this school year.
[8] State testing was suspended in Spring 2020 in all study districts. In Spring 2021, state testing had resumed at a typical level in one district; administration had resumed but participation rates were lower in the other study districts.

STARI teachers in the study schools had been implementing STARI and receiving professional learning supports for 1.5 years. After schools closed, STARI coaches continued to check in with STARI teachers where possible, but STARI classes stopped. To prepare for possible continued virtual instruction, SERP staff worked rapidly to move all STARI curriculum materials online, including fillable PDF versions of student workbooks, short video versions of decoding and comprehension mini lessons, along with guidance for teachers about how to use the resources and how to prioritize synchronous instructional time. SERP also worked quickly to move an abbreviated version of their teacher professional learning series online; using an Open EdX platform, they created a series of videos and interactive activities for teachers to learn the principles and practices of STARI.

- In SY 2020-21 (Year 3), the study districts attempted to offer STARI to students, but instruction was fully virtual for most of the year in two districts and hybrid in the other two districts. SERP and three full-time STARI coaches continued to support STARI teachers to the best of their ability, and mostly remotely due to restrictions on entering school buildings. However, several contextual factors outside of SERP's control made STARI implementation challenging, including constraints on the availability of tablets and at-home internet access for students, technical difficulties with virtual instruction, and the fact that teachers had to prioritize the teaching of core subject areas and/or social and emotional learning, as opposed to supplemental interventions like STARI.

Given these exceptional circumstances, SERP decided to continue to support STARI's implementation for an additional unplanned year in two of four study districts (Year 4, SY 2021-22). This additional year made it possible to provide reading intervention to help students catch

up at a time when they needed it the most, and to continue to build evidence on STARI's effect across different settings.

Study Goals and Research Questions

This study, which examines the effect of STARI on student outcomes in SY 2021-22, is guided by the following research questions:

- What is STARI's effect on students' reading self-efficacy and their reading motivation at the end of the program year?
- What is STARI's effect on students' reading skills, vocabulary, and comprehension at the end of the program year?
- What is STARI's effect on students' general reading proficiency (as measured by scores on English Language Arts state tests) and on their ELA course grades at the end of the program year?

The effect of STARI was evaluated using a student-level random assignment research design. Students were randomly assigned to STARI or to "business as usual" classes, which could include other reading intervention classes already offered by their school. To provide further context for interpreting the impact findings, the study also examined the extent to which students in the STARI group were enrolled and present in the STARI class to receive the intervention, and whether students in the control group were enrolled in another reading intervention and/or receiving other reading supports outside of their regular classes.

The setting for the study is 11 middle schools in two urban school districts during the 2021-22 school year, a period when schools were still recovering from the COVID-19 pandemic. The evaluation's timing represents a policy relevant context for examining not only the replicability of STARI's effects, but also an opportunity to build evidence on strategies to help

struggling adolescent readers catch up. As described earlier, STARI's design integrates instruction in basic reading skills (decoding, fluency, and morphological analysis) with deep comprehension of challenging texts, so it is well positioned to address learning gaps and help struggling adolescent readers make up lost ground without missing out on grade-level content.

The following section describes the study design in more detail, including data sources, measures, and the analytic approach used to evaluate STARI's effects. This is followed by a presentation of the study's findings for each research question. The report concludes by taking stock on what the current body of research on STARI suggests in terms of its effectiveness across different settings, and it discusses lessons and avenues of inquiry for future research.

METHODS[9]

Study Setting and Intervention Delivery

At the beginning of the project in 2018, four urban school districts in the Eastern US were recruited to implement STARI and to participate in the evaluation. These districts were chosen because they enrolled a high proportion of middle school students reading below grade level, and because they were willing to facilitate STARI implementation in the study schools and to participate in the evaluation. In each district, participating schools were chosen based on their capacity and willingness to implement STARI and to participate in the study.[10] A total of 30 middle schools were recruited into the study—10 schools in District 1, 10 schools in District 2, 6

[9] An analysis plan was pre-registered at https://sreereg.icpsr.umich.edu/sreereg/ (see Appendix G) and a restricted use file for the study will be archived at the Inter-university Consortium for Political and Social Research (ICPSR) https://www.icpsr.umich.edu/web/pages/.

[10] The school eligibility criteria were as follows: (1) the school enrolls students in middle school grades (6-8); (2) the school has 1-2 teachers available to be trained to teach STARI; (2) the school can fit STARI into its schedule in addition to the ELA class and ideally for the same amount of time as a regular class, and (3) the school has more eligible students than can be served (oversubscription) and agrees to use a random assignment process to select which students will be offered the STARI class.

schools in District 3, and 4 schools in District 4. These schools implemented STARI in SY 2018-19 (Year 1) and in SY 2019-20 (Year 2) until schools closed in March due to the pandemic.

As explained earlier, the impact evaluation was originally supposed to be conducted in all four study districts in SY 2019-20 (Year 2) and SY 2020-21 (Year 3), however there were significant barriers to STARI implementation and data collection during these years due to the COVID-19 pandemic. Therefore, it was decided to offer STARI to schools for an additional school year (SY 2021-22; Year 4) so that more struggling readers could be served and to make it possible to evaluate STARI's effect during an especially policy-relevant time period. Given remaining project resources, the implementation and evaluation of STARI could only be extended in two of the four school districts. One district (District 3) could not continue to participate in the study in Year 4 because research activities had been suspended districtwide due to the pandemic. Of the remaining three districts, it was decided to conduct the impact study in District 1 and District 2, the two districts with the largest number of study schools, to maximize the sample size. These two districts are geographically diverse—District 1 is in the Southeast and District 2 is in the Northeast—and they implemented STARI with varying levels of fidelity in Year 1, so they are representative of different implementation experiences (see Appendix H).

Of the original 20 study schools in District 1 and 2, a total of 11 schools (6 schools in District 1 and 5 schools in District 2) agreed to implement STARI in Year 4 and to remain in the study.[11] Schools that declined to continue to participate cited challenges related to the pandemic, rather than dissatisfaction with STARI.

[11] In Year 1, ten schools in each district had implemented STARI. In District 1, 6 of these 10 schools are included in the Year 4 impact study (2 schools closed, 1 school did not implement STARI, and 1 school is not included due to limited information about STARI implementation and student enrollment in STARI.). In District 2, 5 of 10 schools are included in the Year 4 impact study (5 schools did not implement STARI.)

Intervention delivery in Year 4. Exhibit 2 provides an overview of the implementation of STARI in each study district in SY 2021-22, described further below.

Grade levels and instructional time. All six schools in District 1 offered STARI classes in grade 6 only. In District 2, two schools offered STARI in grade 6 only, one school in grade 7 only, one school in grade 6 and 7, and one school in all three grade levels (6, 7, 8). In District 1, STARI classes began at the start of the school year. In contrast, in District 2, students were diagnostically assessed at the start of the school year by the district to determine their need for intervention, and supplemental Tier 2 interventions like STARI did not begin until later in the fall after the assessment was completed. Given their respective scheduling models, STARI was offered for an average of 50 minutes per day in District 1, and 45 minutes per day in District 2.[12]

Exhibit 2. STARI Implementation in the Study Schools, SY 2021-22

Implementation characteristics	District 1	District 2
Number of schools	6	5
STARI grade levels	6	6-8
Start of STARI classes	Start of school year	End of October
Average STARI minutes per day	50	45
Percent of new STARI teachers	1/6 (17%)	6/11 (55%)
Number and percent of teachers that completed…		
Less than one STARI unit	3 (50%)	1 (9%)
Unit 1 only	2 (33%)	10 (91%)
Units 1 and 2	1 (17%)	0 (0%)
Units 1, 2 and 3	0 (0%)	0 (0%)

Staffing. In Year 4, 11 teachers taught STARI classes across the two study districts: 6 teachers in District 1 (one teacher per school) and 11 teachers in District 2 (1-3 teachers per school). STARI teachers were reading interventionists, English Language Arts teachers, or special education teachers, and in District 1 some STARI teachers were also paraprofessionals. In District 1, 1 of 6 STARI teachers (17%) was implementing the intervention for the first time,

[12] District 2 uses an alternating AB block schedule; STARI was offered 100 minutes a day every two days. District 1 uses a bell schedule with 45-minute periods; STARI was offered daily.

while in District 2 more than half of STARI teachers were new to the intervention (6 of 11 teachers, or 55%).

Teacher supports. In Year 4, two full-time STARI coaches (one in each district) supported teachers' implementation of the intervention. These coaches, both of whom had also served as STARI coaches in Years 1-3 of the study, trained new STARI teachers one-on-one and they were expected to meet with these new teachers weekly to provide coaching. The more experienced STARI teachers had already been trained at an in-person summer institute before the COVID-19 pandemic; coaches were expected to meet with these teachers every other week. In District 1, formal group PLC meetings for STARI teachers were held regularly and were well attended; in District 2, travel time for teachers made it more challenging to organize group PLCs, so the content of the PLCs was covered in individual coaching sessions instead, and inter-school visits for STARI teachers were organized by the STARI coach as a way for teachers to meet. This level of communication and support is typical of what would have been provided before the pandemic, through with slightly more virtual as opposed to in-person communication.

Pacing. Although in-person classroom instruction had resumed in both study districts in Year 4, STARI classes were disrupted by individual and cohort-level quarantines and occasional school closures arising from high COVID caseloads, which may have affected teachers' progress through the curriculum. Recall that each year long STARI series includes 3 units. In District 1, 3 teachers (50%) partially completed Unit 1, 2 teachers (33%) completed Unit 1 only, and 1 teacher (17%) also completed Unit 2. In District 2, 1 teacher (9%) partially completed Unit 1, 9 teachers (82%) completed Unit 1 only, and 1 teacher (9%) partially completed Unit 2; no teachers completed Unit 2.[13]

[13] Based on data from coach observations and an end-of year teacher survey administered by SERP.

The implementation of STARI in Year 4 differs in several ways from the earlier study in Massachusetts. On the one hand, the level and types of support provided to teachers (training, coaching, and PLCs) were similar across the two studies. On the other hand, in the current study some teachers were teaching STARI for the first time, whereas in the Massachusetts study all participating teachers were in their second year or beyond of teaching STARI. Teachers in that study progressed further into the STARI curriculum—all teachers completed Unit 2 and more than half (53%) completed Unit 3, which means that STARI students in the Massachusetts study received more of the STARI curriculum than in the current study.

Target Population

Students in the study schools were eligible for STARI if they were enrolled in a grade level in which STARI sections were offered at their school, and if they were reading more than one year below grade level. In District 1, whether a student was reading below grade level was determined based on their ELA state test scores from the prior spring (Spring 2021). In District 2, state tests were not consistently administered in Spring 2021, so eligibility was determined based on ELA state test scores from the prior spring if available, and otherwise eligibility was determined based on students' Fall 2021 test scores on the formative reading assessment used by the district for placement in interventions. (As noted earlier, STARI classes in District 2 began *after* this fall assessment was administered.)

In addition, students were not eligible for STARI if their reading skills were at a basic or beginning level that would prevent them from fully engaging in STARI's content. For this reason, the following students were not included in the eligible population: (1) students reading below a third-grade level; (2) students in substantially separate special education classes; (3)

students who were Level 1 or 2 English learners;[14] and (4) students whose special education plan required an intensive, rules-based phonics intervention.

In total, there were 398 eligible students in the study sample across schools (166 students in District 1 and 232 students in District 2). In District 1, 99% of eligible students were Black; all students were from families with low income; and 4% had an individualized education plan (IEP). In District 2, 61% of eligible students are Hispanic and 30% are Black; 83% of students were from families with low income; 28% had an IEP; and 9% were English language learners.[15]

Eligible students in the current study are notably different from the students who participated in the earlier evaluation of STARI in Massachusetts, both demographically and in terms of their educational needs. In the latter study, most students were white (50%) or Hispanic (24%), and 73% were from a family with low income. Students in the Massachusetts study were eligible if they were not proficient on the ELA state assessment (i.e., approximately below grade level), whereas students in the current study had to be at least one year below grade level and were, therefore, further behind on average.[16] On the other hand, compared to the present study, the earlier Massachusetts study had a higher proportion of English language learners (16%) and a higher percentage of students with an IEP (33%). This variation in the characteristics of students across the two studies is useful for building evidence about STARI's effectiveness for different student populations.

[14] Level 1 learners are beginners who are in the early phase of learning English and have a small English vocabulary and little/no ability to use English in academic settings. Level 2 learners can use basic English in routine academic activities and communicate simply in English about familiar topics, but do not understand all the details and need substantial support to understand instruction.
[15] Appendix D provides additional information on the characteristics of students. The characteristics described above are based on students in the state test sample.
[16] The ReadBasix assessment was administered in both studies, but the scores of students in the control group cannot be compared across studies due to changes in scaling.

Research Design

The effect of STARI was evaluated using a student-level random assignment research design. More students were eligible for STARI than could be served, so lotteries were used as a fair way to determine which eligible students would be offered STARI for one school year (STARI group, *n*=183) and which students would not be offered STARI (control group, *n*=215). The outcomes of students in the STARI group and the control group were compared to evaluate the effect of STARI. Random assignment was blocked by school and by grade level.[17]

Students in the control group could be scheduled into any of the other "business as usual" classes offered at their school, including another reading intervention, another academic class, or a non-academic elective. Schools are often required to provide a Tier 2 reading intervention to students who need it, which means that STARI's effects in this study are being compared to the effects of other reading intervention classes available to students in the control group. In the earlier study of STARI in Massachusetts, for instance, 70% of students in the control group were enrolled in another reading or literacy class.

Data Sources and Measures

STARI's effect was examined on the student outcomes that the intervention aims to improve—reading self-efficacy, reading motivation, basic reading skills, vocabulary, comprehension, general reading proficiency, and course performance. These outcomes were measured at the end of the program year (Spring 2022) for students in both research groups (Exhibit 3).

[17]There are 14 blocks in the study sample (6 blocks in District 1, and 8 blocks in District 2). Each school was given a randomly sorted (and rank ordered) list of students for each grade level and instructed to schedule the first *x* students from the list into the STARI class(es), where *x* is the number of available STARI spots given the number of sections offered at the school. An average of 13 students per school and grade were assigned to the STARI group (10 students in District 1 and 15 students in District 2). This is aligned with SERP's recommendation that STARI sections should be about 14 students.

Exhibit 3: Student Outcome Measures

Research Questions	Student Outcomes (Measured in Spring 2022)	Data Source
Proximal Outcomes		
What is the effect of STARI on reading skills, vocabulary, and comprehension?	• Word recognition and decoding (50 items, α=0.91)[a] • Morphological awareness (32 items, α=0.90)[a] • Vocabulary (38 items, α=0.86)[a] • Sentence processing (26 items, α=0.84)[a] • Efficiency of reading for basic comprehension (36 items, α=0.93)[a] • Reading comprehension (22 items, α=0.77)[a]	ReadBasix
... on reading self-efficacy and motivation to read?	• Reading self-efficacy (2 items, α=0.56)[b] • Intrinsic reading motivation (4 items, α=0.57)[b]	Student survey
Distal Outcomes		
What is the effect of STARI on general reading proficiency?	• English Language Arts state test score (z-score)	Student records
... on course performance in English Language Arts?	• Course marks in English Language Arts (0-100%)	Student records

Notes: [a] Sabatini *et al.* (2015); [b]Wigfield & Guthrie (1997). α = Cronbach's alpha.

The data sources for the study include an adaptive reading assessment (ReadBasix) and a survey administered in Spring 2022 to students in both research groups, as well as individual student records provided by the study districts. Parents in both districts were asked for their active consent to administer the ReadBasix and survey to their child; in District 2, parents were also separately asked for their permission to access their child's records.[18] The study team worked with schools to distribute and collect parent consent forms, which were provided to parents in the languages identified by schools. To encourage the return of consent forms, parents were provided a paper consent form but also offered the option of filling out an electronic form. The return of consent forms was also incentivized using individual gift cards for students who

[18] A separate consent question was included in the consent form for the ReadBasix/survey and student records, allowing parents to consent to their child's participation in one type of data collection but not the other.

returned a form.[19] The consent process occurred in Fall/Winter 2021 after the random assignment of students. (Given school calendars, it was not possible to conduct the parent consent process before random assignment or before the start of STARI classes.)

Mediating and proximal outcomes. Students' reading skills, vocabulary, and reading comprehension were assessed using the ReadBasix, a computer-administered adaptive 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, so it is appropriate for students who are reading far below grade level. ReadBasix scores are predictive of state ELA scores (O'Reilly *et al.*, 2012) and subtest reliability is high (α=0.77-0.93). The ReadBasix was also used to measure students' reading skills in the prior evaluation of STARI in Massachusetts.[20] The test takes 45 minutes to 1 hour to complete.[21]

Students' reading self-efficacy and motivation to read were measured using a short (5-10 minute) online student survey that was administered after the ReadBasix. The two constructs were assessed using subscales from the Motivations for Reading Questionnaire (MRQ; Wigfield & Guthrie, 1997), which is appropriate for middle school students reading at an elementary school level. To minimize survey length and burden, the study focuses on intrinsic motivation (as opposed to extrinsic motivation) because STARI is hypothesized to primarily affect the former. The reliability of the survey scales is moderate (α=0.56-0.57).[22]

---

[19] Students received a small gift card for returning a completed form, regardless of whether or not their parent agreed or declined to the student's participation in the study.
[20] The assessment was called the RISE at the time of the prior study of STARI.
[21] The Read Basix is not overly aligned with STARI and is considered an independent measure by the What Works Clearinghouse: https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC-SRP50-508.pdf.
[22] In prior studies, the reliability of these scales was 0.70 or above (Davis *et al*., 2018; Troyer *et al.*, 2019). However, to keep the survey short, only a subset of items from the MRQ scales were used, which resulted in lower reliability.

The ReadBasix and survey were administered to students whose parents or guardians had actively consented to their participation in this data collection. Students were also asked to assent prior to taking the test. The ReadBasix and survey were administered in Spring 2022, as late in the school year as districts would allow. In District 1, administration occurred in April, prior to state testing; in District 2, administration happened in June, after state testing. The study team worked with each school to schedule a time to administer the ReadBasix and survey to students and to schedule a make-up session. On average, students spent 48 minutes taking the ReadBasix and 4 minutes taking the survey. All students completed a sufficient number of items to receive a score on each subtest.[23]

General reading proficiency and course performance. Students' general reading proficiency and their course performance in English Language Arts (ELA) were measured using student records provided by the study districts. District 1 provided student records data for all students in the study sample; District 2 provided records for students whose parents or guardians had actively consented to the release of their child's records. Student records, including state tests, were collected by the study districts as part of routine educational conditions that were the same for students in the STARI group and the control group.

General reading proficiency was measured using students' scores on ELA state tests, which are as reliable as commercial tests, more widely used, and policy relevant. To pool across assessments, test scores were z-scored by district and by grade based on the mean and standard deviation for students in the state test sample (May *et al.*, 2009).[24]

[23] On average, students in the STARI group spent 2 more minutes taking the ReadBasix than students in the control group. As a sensitivity analysis, the effect of STARI was estimated controlling for test-taking time as a covariate; the results from this analysis (Appendix A) are similar to those from the main analyses presented in this report.

[24] As recommended by the What Works Clearinghouse, scores were standardized using the within-group mean and standard deviation (What Works Clearinghouse, 2022).

Students' performance in their ELA courses was measured using their course marks in ELA classes, which are on a percentage scale (0-100%). Data were also collected on students' course performance in other core content areas (social studies, math, and science) to examine whether the reading skills gained by students also helped them in their other classes. The analysis of course performance in other core subjects is exploratory because STARI is not necessarily expected to have an effect in other core subjects in the intervention year.

Student characteristics and baseline reading achievement. Students' characteristics and outcomes prior to random assignment were measured using student records provided by the study districts. These measures were used to describe the sample and to assess whether students in the STARI and control group had similar reading achievement at baseline. For students in District 1, reading achievement at baseline was measured using the ELA state test score from the prior spring (Spring 2021). In District 2, state tests were not consistently administered in Spring 2021 due to the COVID-19 pandemic, so baseline reading achievement was measured using students' ELA state test scores from the prior spring if available, and their score on the district's formative reading test from Fall 2021 otherwise.[25] To allow for pooling, baseline test scores were standardized (z-scored) by district, assessment, and grade, based on the mean and standard deviation of students in the study sample.

Receipt of STARI. To benefit from STARI, students must be in class to receive the curriculum. Two factors may have affected the extent to which students were present to receive STARI instruction in SY 2021-22. First, students assigned to STARI may not have been *enrolled* in the STARI class, due to scheduling conflicts or other factors. As previously noted, SY 2021-22 was the first year of fully in-person instruction in the study districts, and the transition back to

[25] In District 2, the ELA state test score from the prior spring is used for 28% of students in the state test sample, and the district reading test score is used for 72% of students.

a more typical school schedule may have resulted in problems with scheduling students into STARI. To explore this possibility, student enrollment in the STARI class was measured using scheduling information provided directly by the study schools and confirmed using course enrollment data provided by the districts. Second, students enrolled in the STARI class may not have been consistently *present* at school to receive the curriculum, due to reticence by families to send their children to school in periods of high COVID caseloads and transmission. Student records data were used to examine average attendance rates and chronic absenteeism rates for students in the study during the intervention year.

Service contrast. An important factor affecting the magnitude of STARI's effects on students is the difference in the amount of the reading supports received by students in the STARI group compared to students in the control group. As noted earlier, providing Tier 2 interventions to all students who need it was encouraged in District 1 and District 2, so students in the control group had access to other literacy interventions or classes during the school day. Course data provided by the study districts was used to examine whether students in the control group were taking classes flagged as reading interventions. In addition, students in the control group may have received supplemental reading supports outside of their regular school classes. To examine this possibility, in the student survey, students were asked whether they received help with their reading from an adult who is not their teacher, and the number of times they received that help across the school year.

Response Rates and Baseline Characteristics

The effect of STARI was evaluated for students who have data on the outcome of interest at the end of the program year and who also have a baseline reading score. The latter condition

makes it possible to confirm that students in the STARI group and the control group started off with similar reading achievement levels before the intervention.[26]

The most important factor affecting whether students have data is parent consent. Contacting and communicating with parents continued to be challenging in SY 2021-22 due to the COVID-19 pandemic, resulting in low consent form return rates. Only 64% of parents in District 1 returned a consent form, and in District 2 only 40% of parents returned a form.[27] Other factors affecting response rates are that students had to be present on the days of testing or make-up sessions, and they had to assent to taking the Read Basix and survey (see Appendix C for further information on the causes of non-response).

Exhibit 4 shows the proportion of students included in the analysis sample for each outcome domain (state test sample, course performance sample, ReadBasix sample, and survey sample). In District 1, almost all (93%) of students in the study sample are included in the state test sample (n=154) and the course sample (n=156), because parent consent was not required for the release of student records. However, less than half (48%) of students are included in the ReadBasix sample (n=80) and the survey sample (n=81), primarily due to low consent form return rates. In District 2, where parent consent was required for all data sources including student records, only 20% of students in the study sample are included in the state test and course samples (n=46), and 10% of students are included in the ReadBasix and survey samples (n=24).

[26] Most students with program year data also had a baseline measure of reading achievement, so this additional requirement did not significantly reduce the analysis samples. See Appendix C.

[27] The return rates were higher in District 1 because project staff were able to enter schools in person to help with the consent process and gather forms, whereas in District 2 there were still pandemic-related restrictions in place that limited the extent to which non-district staff could enter schools. Among parents who returned forms, the majority of parents said yes to their child's participation in the study (89% in District 1; 72-76% in District 2 depending on the data source.)

In contrast, in the earlier evaluation of STARI in Massachusetts, 83% of students in the study sample were included in the analysis sample.

For some outcomes, the response rate also differs across research groups, with higher response rates for students in the STARI group than the control group. This is likely because STARI students were a more easily accessible group – STARI teachers could remind students to return their consent forms, whereas control students were enrolled across several different classes. In District 2, for example, response rates for the state test sample and the course samples are 10 percentage points higher for students in the STARI group than for the control group. Based on What Works Clearinghouse standards (WWC, 2022), attrition is low for the state test sample and the course sample in District 1;[28] all other samples have high attrition.[29]

The baseline reading scores of students in the analysis samples were examined to assess whether high attrition rates affected the comparability of the two research groups (Exhibit 5). Students' baseline reading achievement is a key predictor of the outcomes of interest, so it is a useful metric by which to assess the similarity of students in the STARI and control group.[30] As explained earlier, reading achievement at baseline was measured using a combination of ELA state test scores and district formative reading test scores pre-dating random assignment, which were standardized (z-scored) to allow for pooling. In District 1, the difference between STARI and control group students' baseline reading scores does not exceed an effect size of 0.25 for any

[28] In District 1, there are differences in the percentage of students who have student records data across the two groups—8 percent *fewer* students in the STARI group are included the state test sample compared to the control group. However, it is highly unlikely that STARI increased student mobility, so this difference is probably due to chance.

[29] This is based on the optimistic attrition boundary, which applies to our study because the intervention is supplemental and targeted and because it is unlikely to influence dropout.

[30] The correlation between baseline reading achievement and the ReadBasix subtests ranges from 0.23 to 0.42 in District 1, and 0.23 to 0.59 in District 2. The correlation between baseline reading achievement and ELA state test scores at follow-up is 0.33 in District 1 and 0.58 in District 2. See Appendix F for correlations between measures.

Exhibit 4. Response Rates by Student Outcome

	Study Sample			STARI Group			Control Group			Difference	
Analysis Sample	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Adjusted Response Rate	STARI - Control	P-Value
District 1											
State test	166	154	92.8%	60	53	88.3%	106	101	96.9%	-8.6%	0.098 *
Course	166	156	94.0%	60	54	90.0%	106	102	97.6%	-7.6%	0.120
ReadBasix	166	80	48.2%	60	34	56.7%	106	46	43.6%	13.0%	0.111
Student survey	166	81	48.8%	60	34	56.7%	106	47	45.1%	11.6%	0.160
District 2											
State test	232	46	19.8%	123	31	25.2%	109	15	15.7%	9.5%	0.085 *
Course	232	46	19.8%	123	31	25.2%	109	15	15.7%	9.5%	0.085 *
ReadBasix	232	24	10.3%	123	15	12.2%	109	9	8.8%	3.4%	0.415
Student survey	232	24	10.3%	123	15	12.2%	109	9	8.8%	3.4%	0.415

Notes: N = Number of students (randomized or in the sample). For the control group, the response rate shown in the table is not necessarily equal to the response rate directly calculated using the number of students randomized and in the analysis sample, because the response rate for the control group is regression-adjusted for the blocking of random assignment. The differential response rate between STARI and control students is also regression-adjusted. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

analysis sample.[31] In District 2, however, only the state test and course samples meet this standard; for all other samples, the standardized difference in baseline reading scores is larger than 0.25. (See Appendix D for additional baseline student characteristics by research group for each analysis sample.)

Exhibit 5: Reading achievement scores at baseline, by analysis sample

	STARI Group			Control Group			Difference		
Analysis Sample	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1									
State test	53	-0.09	0.97	101	-0.02	1.02	-0.07	-0.07	0.702
Course	54	-0.12	0.98	102	-0.02	1.02	-0.10	-0.10	0.571
ReadBasix	34	-0.16	1.03	46	-0.18	1.07	0.02	0.02	0.942
Student survey	34	-0.16	1.03	47	-0.19	1.08	0.02	0.02	0.921
District 2									
State test	31	0.01	0.95	15	0.13	0.97	-0.12	-0.13	0.711
Course	31	0.01	0.95	15	0.13	0.97	-0.12	-0.13	0.711
ReadBasix	15	0.05	1.03	9	-0.21	0.81	0.26	0.27	0.570
Student survey	15	0.05	1.03	9	-0.21	0.81	0.26	0.27	0.570

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. Baseline reading scores are standardized (z-scored) by district and by assessment and by grade, based on the mean and standard deviation of students in the study sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Taken together, this means that the strength of the evidence (internal validity) of the impact findings presented in this report varies across student outcomes and districts (Exhibit 6):

- The findings for District 1 generally provide the most rigorous evidence of STARI's effects. For ELA state tests and course performance, the level of evidence for District 1 is *strong*—there is low sample attrition, and STARI and control group students had similar levels of reading achievement at baseline. Thus, the difference between STARI and

[31] A benchmark effect size of 0.25 is used by the What Works Clearinghouse to assess baseline equivalence. The effect size is calculated using the within-group standard deviation, as defined in the What Works Clearinghouse standards (2022).

control students is unlikely to be confounded with other factors, and the estimated effects of STARI for these outcomes and samples have high internal validity. For reading skills and reading self-efficacy and motivation, the level of evidence for District 1 is *moderately strong*—there are higher levels of attrition, but students in both research groups had similar level of reading achievement at baseline, which means that bias due to unobserved differences between the two research groups is likely to be small. Sensitivity analyses also suggest that the estimated effect of STARI on ReadBasix subtests and survey outcomes in District 1 – even though they are based on a smaller subset of students – are representative of STARI's effect on these outcomes for the full study sample.[32]

- In contrast, the findings for District 2 provide moderately strong evidence at best. For reading skills and reading self-efficacy and motivation, the level of evidence for District 2 is *limited*, because response rates are low and students in the STARI and control group differ with respect to their reading achievement at baseline. This means that unobserved differences between the STARI and control group are more likely, and by extension, that the estimated effect of STARI should be interpreted with caution because there is a higher risk of bias due to confounding factors.

Because response rates and the strength of the evidence varies across districts, the findings in this report are presented by district, rather than pooled across districts.[33]

[32] In District 1, the effect of STARI on ELA state test scores *for the subset of students in the ReadBasix sample* is similar in magnitude to its effect on ELA state tests for the full state test sample. This suggests that STARI's estimated effect on ReadBasix subtests - even though they are based on a smaller subsample of students - are likely to be generalizable to the larger study sample. See Appendix C for further information.

[33] For reporting purposes, the estimated effect of STARI *for both districts pooled together* is provided in Appendix F. Pooled results are also presented in other relevant appendices.

In addition to internal validity, it is also important to consider the reliability (precision) with which STARI's effects can be estimated. Depending on the outcome, the analysis samples range from 80 to 156 students for District 1, and 24 to 46 students for District 2. At a 10 percent level, the minimum detectable effect size (MDES) for District 1, which has the largest sample, is 0.43 for ELA state test scores and 0.36 for ELA course performance, and it ranges from 0.57 to 0.62 for the ReadBasix subtests and reading motivation and self-efficacy.[34] These MDES are larger than the effect sizes found in the prior study of STARI conducted in Massachusetts, suggesting that the current study may not be sufficiently powered to statistically detect effects of the size that one would expect STARI to produce, due to its smaller than expected sample sizes. However, when interpreted alongside the findings from the earlier study in Massachusetts, the pattern of findings from this study can still provide important information about the consistency of STARI's effect across different contexts.

Exhibit 6. Levels of Evidence (Internal Validity)

Outcome domains	Pooled sample	District 1	District 2
General reading proficiency (ELA state tests)	Strong	Strong	Moderate
Course performance	Strong	Strong	Moderate
Reading skills, vocabulary, and comprehension	Moderate	Moderate	Limited
Reading self-efficacy and motivation	Moderate	Moderate	Limited

Notes: Attrition ratings are based on What Works Clearinghouse (2022). ELA= English Language Arts.
Strong = Low differential attrition (optimistic boundary) and the STARI-control difference in baseline reading achievement is less than 0.25 standard deviation; expected to meet WWC standards *without* reservations.
Moderate = High differential attrition, and the STARI-control difference in baseline reading achievement is less than 0.25 standard deviation; expected to meet WWC standards *with* reservations.
Limited = High differential attrition, and the STARI-control difference in baseline reading achievement is greater than 0.25 standard deviation; may not meet WWC standards.

[34] These are the actual MDES based on the standard errors for estimated effects on each outcome. See Appendix F for the MDES by outcome and sample.

Analytic Approach

The impact of STARI was estimated by comparing the outcomes of students in the STARI group and the control group. To do so, the following ordinary least squares (OLS) linear regression model was fitted to the relevant analysis sample:

$$Y_{ij} = \beta T_i + \sum_K \lambda_k B_{ki} + \delta READ_{-1i} + \sum_M \omega_m D_{mi} + \varepsilon_i$$

In this model, $\boldsymbol{Y_i}$ is the outcome of interest for student *i* in the analysis sample. $\boldsymbol{T_i}$ is an indicator for whether a student was randomly assigned to the STARI group (=1) or the control group (=0). **β** is an intent-to-treat (ITT) estimate of the effect of being assigned to the STARI group on outcome *Y*, or the difference in outcome Y for students assigned to the STARI group and students assigned to the control group. The model also includes covariates to improve the precision of estimated effects and to adjust for baseline differences. $\boldsymbol{B_{kj}}$ is a set of *K* random assignment block indicators to account of the blocking or random assignment; $\boldsymbol{READ_{-1i}}$ is a (z-scored) measure of a student's baseline reading achievement score; and $\boldsymbol{D_{mi}}$ is a set of M baseline demographic characteristics, including indicators for whether a student is eligible for free/reduced price lunch, race/ethnicity (Black, Hispanic, other), gender, English Language Learner status, and whether the student has an individualized education plan. The statistical significance of estimated effects was assessed using a two-tailed t-test.

As noted earlier, the impact analyses are limited to students who have data on a given outcome. Students' demographic covariates ($\boldsymbol{D_{mi}}$) are missing for a very small number of students and were imputed using the indicator variable approach. Missing data on these characteristics were imputed with a constant, and indicators of missingness for each characteristic were added to the statistical model. Students' baseline reading score ($\boldsymbol{READ_{-1i}}$) was not imputed because all students in the analysis samples have data on this measure.

FINDINGS

Reading classes and other literacy supports

Receipt of STARI. An important factor affecting the magnitude of STARI's effects on students is the extent to which STARI students were enrolled in the STARI class and how often they were at school in person to receive the intervention. Overall, the findings indicate that most students in the STARI group were present to receive instruction:

- In each district, almost all students in the STARI group (91%) were enrolled in the STARI class. The causes of non-enrollment included scheduling conflicts[35] and/or the discretionary judgment of school staff (teachers, counselors, and principals) about STARI's suitability for some students.
- Student attendance rates in SY 2021-22 were similar to pre-pandemic school years, indicating that students were at the school in person to receive instruction. The average attendance rate for students in the study sample was 91%. This is similar to the attendance rate in the study schools in SY 2018-19, suggesting that attendance was not atypically low in SY 2021-22 due to COVID-related absences.[36]

Service contrast. Another important factor affecting the magnitude of STARI's effect is the extent to which students in the control group were enrolled in other reading or literacy classes. In District 1, about a third of students in the control group (34%) were enrolled in

[35] In some instances, for example, students could not be scheduled into STARI because they had also been identified for a supplemental math intervention. In other instances, parents objected to their child missing out on fun electives or classes, or to their child being pulled out of band to take STARI when the parent had already paid to rent the instrument.

[36] In District 1, the average daily attendance rate for students in the course sample was 90%, and 27% of students were chronically absent, defined as having an average daily attendance rate of less than 90%. The average chronic absenteeism rate was similar, 26%, for the study schools in SY 2018-19 before the pandemic. In District 2, the average daily attendance rate for students in the course sample was 92%, and 30% of students were chronically absent. The average chronic absenteeism rate for the study schools before the pandemic, in SY 2018-19, was also 30%.

another reading class or intervention (Exhibit 7).[37] In District 2, almost half of control students (47%) were enrolled in another reading class. These proportions are much lower than in the earlier evaluation of STARI in Massachusetts (70%), possibly due to a delayed return to pre-pandemic levels of supplemental reading instruction in SY 2021-22, or to between-study differences in district policies related to providing supplemental reading intervention.

Students in the control group as well as the STARI group also received tutoring and extra help with their reading outside of their regular classes (Exhibit 8). In both districts, more than half of control group students reported receiving extra help.[38] In District 1, 56% of students in the control group received additional reading support, but a high percentage of students in the STARI group (65%) also received extra support. Conversely, in District 2, control group students were more likely to receive extra support than students in the STARI group—52% of students in the control group students reported receiving extra help with their reading, compared to 33% of students in STARI group. Students in the control group also received this support more frequently—36 times during the school year, compared to 3 times for the STARI group, a statistically significant difference. Of the students in the control group in District 2 receiving these extra supports, 80% were also enrolled in an alternative reading class, so these supports appear to be a supplement rather than a substitute for a reading class.[39] At first glance, these findings suggest that the service contrast in reading supports between students in the STARI group and the control group may have not have been as strong in District 2 as in District 1,

[37] This was determined using course records data, by flagging students enrolled in a class with "reading" in the course name.

[38] These findings are based on students in the survey sample. These students had lower baseline reading achievement scores than students who did not take the survey (see Appendix C), and they were more likely to be enrolled in another reading class than students in the course sample – 43% in District 1 and 60% in District 2 (see Appendix E). This suggests that the average level of reading supports received by these students may be higher than for the full study sample.

[39] In contrast, in District 1, 35% of students receiving extra supports were also enrolled in a reading class.

however these findings should be interpreted with caution because the sample for District 2 is small (n=9 students in the control group).

On balance, the amount of additional reading intervention received by STARI students in this study – relative to students in the control group – is comparable to the earlier Massachusetts study. On the one hand, STARI students in the Massachusetts study received more reading intervention than STARI students in the current study, because their teachers made it further into the STARI curriculum. (As noted earlier, none of the STARI teachers in the current study progressed to Unit 3, whereas about half of teachers in the Massachusetts study completed Unit 3.) On the other hand, students in the control group in the Massachusetts study also received relatively more reading intervention—70% of control group students in the Massachusetts study were enrolled in an alternative reading class, compared to 34% of control group students in District 1. Considered together, these two factors suggest that the service contrast between STARI and control students was similar across studies.

Exhibit 7: Reading Classes and Other Reading Supports

Intervention or Service Receipt Measure	STARI Group			Control Group			Difference	
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	P-Value
District 1								
Reading classes								
Enrolled in STARI (%)	54	91%	29%	102	0%	0%	91%	<0.001 ***
Enrolled in STARI or another reading class (%)	54	91%	29%	102	34%	47%	57%	<0.001 ***
Reading support from an adult who is not a teacher								
Received help[a] (%)	34	65%	49%	47	56%	50%	9%	0.438
Number of times received help during school year[b]	34	26.8	47.7	47	21.4	32.7	5.3	0.567
District 2								
Reading classes								
Enrolled in STARI (%)	31	90%	30%	15	0%	0%	90%	<0.001 ***
Enrolled in STARI or another reading class (%)	31	94%	25%	15	47%	52%	46%	<0.001 ***
Reading support from an adult who is not a teacher								
Received help[a] (%)	15	33%	49%	9	52%	53%	-19%	0.406
Number of times received help during school year[b]	15	3.3	10.2	9	36.0	57.9	-32.7	0.047 **

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. Enrollment in STARI and reading classes is reported for students in the course sample; the receipt of other reading supports is reported for students in the student survey sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]*Other than your teachers, did an adult help you with your reading this school year? It can be someone at your school, outside of school, or at an after-school program.* [b]*How LONG did you get this help with your reading? How OFTEN did you get this help with your reading?*

Effects on Student Outcomes

STARI is designed to improve students' reading skills and their reading self-efficacy and motivation, which are in turn hypothesized to improve students' general reading proficiency and their course performance. The effect of STARI on these outcomes was examined for each district separately, by comparing the outcomes of students in the STARI group and the control group at the end of the intervention year in Spring 2022. As discussed previously, a small proportion of students in the STARI group (9%) did not enroll in the STARI class. The findings presented in this section represent the estimated intent-to-treat effect of being *assigned* to the STARI class, as opposed to the effect of being enrolled in the class.[40]

As noted earlier, the sample size for most analyses is too small to reliably detect effects of the magnitude found in the prior study in Massachusetts. However, the pattern of results from this evaluation – when interpreted alongside the larger body evidence about STARI – can still provide useful information about STARI's effectiveness in different settings. For this reason, the findings in this section are discussed not solely based on statistical significance, but also based on whether the magnitude and pattern of effects is consistent with the prior evaluation of STARI in Massachusetts.

Because the internal validity of the findings is stronger for District 1 than for District 2, the results are presented separately for each district. Differences (or similarities) in the pattern of effects across the two districts should be interpreted with caution, because they could be due to

[40] The effect of enrolling in STARI can be estimated by dividing the effect of being assigned STARI by the proportion of students who enrolled in the class (Bloom, 1984). Because STARI enrollment rates are high (91%), the estimated effect of enrolling in STARI is similar in size to the effect of being assigned to STARI. The p-value for the two types of effect estimates is the same, because the standard error of the impact estimate must also be scaled up by the enrollment rate.

differences in the internal validity of the impact estimates across districts, as opposed to true variation in STARI's effect across districts.

General Reading Proficiency and Course Performance

STARI's longer-term aim is to improve students' odds of graduating from high school and to prepare them for postsecondary education and their careers. Accordingly, STARI's effects were examined on two leading predictors of high school graduation and college and career readiness (Exhibit 8)—students' general reading proficiency (as measured by ELA state test scores) and their performance in their classes (as measured by their ELA course grades). Recall that the findings for District 1 (n=154) provide strong evidence of effectiveness for these two outcomes, and the findings for District 2 (n=46) provide moderately strong evidence (Exhibit 6).

District 1. In District 1, students in the STARI group outperformed students in the control group on ELA state tests and this difference is statistically significant at the 10 percent level (effect size = 0.32, p-value = 0.071).[41] Similarly, the proportion of students proficient on the ELA state test was 6.5 percentage points higher in the STARI group than in the control group (p-value=0.084).[42] As a reference point, the reading scores of middle school students typically increase by an effect size of 0.27 annually (Hill *et al.*, 2008). This means that STARI's estimated effect (0.32) represents a little more than an *extra year* of gains beyond typical growth – or double the expected growth of middle school students in reading. These gains are policy relevant given the deep learning losses experienced by students during the COVID-19 pandemic.[43] The estimated effect of STARI on state tests in the present study is also larger than in the

[41] As noted earlier, these findings represent the effect of being assigned to the STARI group; for students who were *enrolled*, the effect size is even larger – 0.35. This is equal to the "intent to treat" effect of being assigned to STARI (0.32) divided by the proportion of students who enrolled in STARI (91%).
[42] See Appendix F for estimated effects on ELA state test proficiency rates.
[43] Estimated effects on ELA state tests for the subsample of students in District 1 who are also in the ReadBasix sample are similar in magnitude (effect size =0.39). See Appendix C.

Massachusetts study; in that study, STARI and control students had similar ELA state test scores and the difference between groups was not statistically significant.

Exhibit 8: Estimated Effect of STARI on Students' ELA State Test Scores and ELA Course Grades

	STARI Group			Control Group			Difference		
Outcome	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1									
ELA state test score (z-score)[a]	53	0.21	1.11	101	-0.10	0.92	0.31	0.32	0.071 *
ELA course grades (0-100%)	54	72.4	8.4	102	70.2	8.5	2.2	0.26	0.075 *
District 2									
ELA state test score (z-score)	31	-0.01	0.97	15	0.24	1.02	-0.25	-0.26	0.234
ELA course grades (0-100%)	31	81.3	9.4	15	80.1	13.7	1.2	0.11	0.714

Notes: ELA = English Language Arts. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the state test sample and the course sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]The estimated STARI-control difference on ELA state test scores is statistically larger in District 1 than in District 2 (p-value=0.038).

In District 1, students in the STARI group also outperformed students in the control group in their ELA classes. Students in the STARI group had an average ELA course mark of 72%, compared to an average course mark of 70% for students in the control group, a difference of two percentage points that is statistically significant at the 10 percent level (effect size = 0.26, p-value=0.075). STARI's effect on students' performance in other core subjects was also examined. The estimated effect of STARI on math and science course grades was smaller in size and not statistically significant, but there were positive and statistically significant effects on students' marks in their social studies courses, possibly due to the amount of informational

reading required in this subject area.[44] (STARI's effects on course performance were not examined in the earlier Massachusetts study.)

District 2. In District 2, STARI's estimated effect on ELA state test scores and ELA course performance does not follow a consistent pattern, perhaps due to the smaller sample size. The estimated effect of STARI on ELA course grades is positive in direction but not statistically significant (effect size=0.11, p-value=0.714).[45] Conversely, the estimated effect of STARI on ELA state test scores is in the opposite direction and not statistically significant (effect size=-0.26, p-value=0.234). This inconsistent pattern of results is not due to a lack of correlation between the two outcomes (which is 0.46 in District 2). It is also interesting to note that ELA course marks were 80-81 percent for both research groups in District 2, which is higher than expected given the target population of students. This may be due to relaxed grading standards during the pandemic, but it could also be a further indication that the results from District 2 need to be interpreted with caution because of its small sample.

Reading Skills, Vocabulary, and Comprehension

STARI is designed to help students develop the reading skills and vocabulary that are needed for deep comprehension of complex texts. Accordingly, STARI's effect on the following ReadBasix subtest scores was examined at the end of the program in Spring 2022: word recognition and decoding, morphological awareness, vocabulary, sentence processing, efficiency of reading for basic comprehension, and reading comprehension (Exhibit 9). Recall that the findings for District 1 (*n*=80) provide moderately strong evidence for these outcomes, while the

[44] See Appendix F for the estimated effect of STARI on these supplemental outcomes.
[45] The estimated effect of STARI on ELA state test scores is statistically larger in District 1 than in District 2.

findings for District 2 (*n*=24) provide limited evidence and should be interpreted more cautiously.

Exhibit 9: Estimated Effect of STARI on Students' Reading Skills, Vocabulary, and Comprehension

	STARI Group			Control Group			Difference		
Outcome	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
District 1[a]									
Word recognition and decoding	34	232.5	13.3	46	228.4	13.3	4.1	0.32	0.196
Morphological awareness	34	236.8	10.4	46	233.8	10.4	3.1	0.31	0.176
Vocabulary	34	233.1	7.0	46	232.0	7.0	1.1	0.15	0.504
Sentence processing	34	241.6	12.2	46	239.3	12.2	2.3	0.20	0.424
Efficiency of reading for comprehension	34	233.1	7.0	46	232.0	7.0	1.1	0.15	0.504
Reading comprehension	34	233.1	7.0	46	232.0	7.0	1.1	0.15	0.504
District 2									
Word recognition and decoding	15	238.7	11.0	9	234.0	11.0	4.7	0.33	0.393
Morphological awareness	15	241.3	5.9	9	243.4	5.9	-2.1	-0.28	0.616
Vocabulary	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.098 *
Sentence processing	15	240.9	5.4	9	240.2	5.4	0.6	0.07	0.894
Efficiency of reading for comprehension	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.098 *
Reading comprehension	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.098 *

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the ReadBasix sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. [a]The estimated STARI-control differences in this table do not differ by a statistically significant amount across school districts.

District 1. In District 1, students in the STARI group had consistently higher ReadBasix scores than students in the control group at the end of the program year, with effect sizes ranging

from 0.15 to 0.32. Estimated effects are not statistically significant due to the small number of students who took the ReadBasix assessment. However, the magnitude of estimated effects is similar in size to the effects observed in the Massachusetts study. In that study, effects ranged 0.08 to 0.21, with statistically significant effects on word recognition (effect size = 0.20), efficiency of basic reading (effect size = 0.18), and morphological awareness (effect size = 0.21).

The magnitude of effects in this study, and in the prior study of STARI, is also policy relevant. In this study, for example, STARI's estimated effect on reading comprehension (effect size = 0.15) represents a 12-percentage point increase in the proportion of students who are comprehending text at grade level. As noted earlier, typical annual reading achievement growth in middle school is 0.27 (Hill *et al.*, 2008), so the estimated effect of STARI on comprehension represents about an extra half year of gains, which is a meaningfully sized gain in the context of helping students catch up.[46]

District 2. In District 2, there is more variation in the estimated effect of STARI across ReadBasix subtests because the sample size is small (*n*=24). Effect sizes range from an estimated effect of -0.28 on morphological awareness (not statistically significant) to an effect of +0.68 on vocabulary that is statistically significant at the 10 percent level. As noted earlier, the effect sizes for District 2 should be interpreted with caution.

Reading Self-Efficacy and Motivation

STARI is intentionally designed to improve students' reading achievement by improving their reading self-efficacy and their motivation to read. The estimated effect of STARI on these two outcomes was examined at the end of the program in Spring 2022 (Exhibit 10). Recall that

[46] See Appendix F for the effect of STARI on the proportion of students who performed at grade level on the ReadBasix subtests.

the findings for District 1 provide moderately strong evidence of effectiveness (*n*=81), and the findings for District 2 provide limited evidence (*n*=24).

District 1. The findings suggest that, as intended, STARI helped students in District 1 feel more confident about their reading skills and more enthused about reading. Students in the STARI group had more positive feelings about their reading self-efficacy than students in the control group, and this difference is statistically significant at the 5 percent level (effect size=0.61, p-value=0.015). Students in the STARI group were also more highly motivated to read than students in the control group, but the estimated effect is not statistically significant (effect size= 0.23, p-value=0.348). Estimated effects on reading self-efficacy and motivation were not examined in the prior evaluation of STARI in Massachusetts.

District 2. The general pattern of effects on these outcomes is positive in direction in District 2 but not statistically significant. Effect sizes range from 0.47 to 0.58.

Exhibit 10: **Estimated Effect of STARI on Reading Self-Efficacy and Motivation**

	STARI Group			Control Group			Difference		
Outcome	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1[a]									
Reading self-efficacy (1-4)	34	3.2	0.7	47	2.8	0.8	0.5	0.61	0.015 **
Reading motivation (1-4)	34	3.0	0.6	47	2.9	0.7	0.1	0.23	0.348
District 2									
Reading self-efficacy (1-4)	15	3.1	0.6	9	2.8	0.7	0.3	0.47	0.479
Reading motivation (1-4)	15	3.2	0.4	9	2.8	1.0	0.4	0.58	0.411

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the survey sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. [a]The estimated STARI-control differences in this table do not differ by a statistically significant amount across school districts.

DISCUSSION

The findings from this study provide plausible evidence that STARI helped accelerate students' reading achievement in SY 2021-22, a school year when instruction and student learning was still affected by the COVID-19 pandemic. In District 1, which provides the strongest and most internally valid evidence, students in the STARI group outperformed students in the control group on ELA state tests (effect size=0.32) and in their ELA classes (effect size=0.26), and these differences are statistically significant at the 10 percent level. These effects represent an extra year's worth of reading gains – double the typical growth made by middle school students – suggesting that STARI accelerated students' learning and helped them catch up. STARI students also had statistically higher self-reported reading self-efficacy than students in the control group, as intended by the program's design. Due to small samples, the estimated effects of STARI on students' reading motivation, reading skills, vocabulary and comprehension are not statistically significant, but the direction of effects on these outcomes was consistently positive in direction.

When considered together with the findings from the earlier evaluation of STARI in Massachusetts, there is suggestive and growing evidence that STARI is an effective strategy for helping struggling adolescent readers catch up, across different geographical settings (6 districts in 3 states across the two studies) as well as temporal settings (pre-pandemic and during the pandemic). The growing body of evidence also suggests that STARI can be effective across different implementation circumstances, including when it is delivered by paraprofessionals and when the instructional environment is unstable and there is slower progress through the STARI curriculum. STARI also appears to be more effective than the supplemental reading classes currently being offered in school districts, because in both studies a substantial proportion of

students in the control group were enrolled in another reading class. The consistency and size of STARI's effects are especially noteworthy because RCTs of educational interventions typically have effects of about 0.05-0.07 on average on standardized reading outcomes in grades 6-8 (Kraft, 2020).

As discussed earlier, the COVID-19 pandemic profoundly altered the plans and scope for the current study, which introduced several limitations. The first is that the student sample size is smaller than originally intended—and smaller than in the earlier evaluation of STARI in Massachusetts. One factor affecting the sample size is that the study only includes two school districts (instead of four as originally planned) and only one cohort of students (instead of two cohorts). In addition, the return rate for parent consent forms was low because contacting and communicating with parents remained challenging during the pandemic. STARI students and their families were more easily reachable than control group students, which also led to a difference in consent rates between the two research groups for several outcomes and samples. As a result, some of the study findings (including all results from District 2) need to be interpreted cautiously. A related limitation, also related to the small sample size, is that it was not possible to examine STARI's effect for student subgroups—English language learners, students with an IEP, and students who are reading especially below grade level.

Despite these limitations, the study's findings still contribute to the evidence base on STARI's effectiveness across different contexts, and the project also yielded several positive outputs in terms of STARI's reach and its design. Importantly, SERP was able to meet its targets with respect to the number of teachers trained and students served (see Appendix I). SERP was also able to use the time during school closures and virtual learning to develop several new tools to address implementation barriers and improve STARI's scalability. For example, SERP learned

that attendance at trainings and PLC sessions was lower in school districts where travel and commuting were more complicated. In response, SERP developed a 17-session online professional learning series that STARI teachers can complete at their own pace, or that teachers can view together as a group with coordination from a district facilitator. SERP also learned from districts that they wanted to use STARI with students reading below a third-grade level. These students are not typically eligible for STARI because they lack the decoding skills needed to successfully access STARI content. In response, SERP developed STARI Boost, a computer-based phonics supplement to STARI that can be integrated into STARI instructional routines. Accessed via a web application, Boost activities include decoding and encoding practice, brief tutorials in new letter-sound patterns, and practice with sight words. Importantly, Boost provides individualized instruction—students begin at their own individual starting point and spend as much time as necessary practicing each skill. Teachers also have access to data to monitor student progress. Boost is intended to be used by students reading below a third-grade level during the first 10 minutes of the STARI fluency routine.

These implementation and curricular enhancements will make it possible to expand STARI's reach to an even more diverse group of districts and students. The online format of the STARI professional learning series will make it possible to bring STARI to rural districts and to shift the delivery and ownership of teacher training away from SERP and into the hands of local instructional leaders. The addition of Boost will make it possible for districts to also offer STARI to students who are reading below a third-grade level, so that they can build their phonics skills at the same time as their reading comprehension.

The implementation and impact of STARI – including these new enhancements – will be evaluated as part of an EIR Expansion grant awarded to SERP, which began in January 2023.

STARI will be implemented in 80 middle schools across several states. To encourage more sustainable implementation, STARI teachers will be trained and supported by local coaches and district staff. The contribution of this new project will be to examine whether STARI's effects can be replicated in a more geographically diverse set of districts, including rural districts, and to evaluate STARI's effects for an expanded target population of students that also includes students reading below a third-grade level, students with an IEP and English language learners. STARI's effects will be evaluated using a school-level random assignment research design, where schools will be randomly assigned to begin STARI implementation immediately or to begin implementation two years later. The study will also examine what resources are needed to support adoption and sustained implementation of STARI, to inform national scaling and to continue to build evidence on how best to support struggling middle school readers.

REFERENCES

Allensworth, E. M., & Easton, J. Q. (2007). What Matters for Staying On-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year. Research Report. *Consortium on Chicago School Research*. Retrieved from: https://eric.ed.gov/?id=ED498350.

Balfanz, R. (2009). *Putting Middle Grades Students on the Path to Graduation: A Policy and Practice Brief.* Westerville, OH: National Middle School Association.

Balfanz, R., Herzog, L., & MacIver, D.J.. (2007). Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions." *Educational Psychologist* 42, 4: 223-235.

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.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, *8*, 225–246.

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.

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.

Deke, J. (2014). Using the Linear Probability Model to Estimate Impacts on Binary Outcomes in Randomized Controlled Trials. *Evaluation Technical Assistance Update for OAH & ACYF Teenage Pregnancy Prevention Grantees, Brief 6*. Washington, DC: U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Office of Adolescent Health.

Donovan, M. S., Snow, C. & Daro, P. (2013). The SERP Approach to Problem-Solving Research, Development, and Implementation. In Fishman, B., Penuel, W. R., Allen, A., & Cheng, B. H. (Eds.), Design-based implementation research: Theories, methods, and exemplars. *National Society for the Study of Education Yearbook*, 112(2), 400-425.

Eccles, J.S., & Wigfield, A.. (2000). "Schooling's Influences on Motivation and Achievement." Chapter 5 in S. Danziger and J. Waldfogel (eds.), *Securing the Future: Investing in Children from Birth to College.* New York: Russell Sage Foundation.

Eccles, J.S. (1999). "The Development of Children Ages 6 to 14." *Future of Children* 9, 2: 30-44.

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.

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.

Hemphill, L., Kim, J., & Troyer, M. (2019). Improving struggling readers' literacy skills through talk about text. In V. Grøver, P. Uccelli, M. Rowe, & E. Lieven (Eds.), *Learning through language: Towards an educationally informed theory of language learning* (pp. 123-134). Cambridge, UK: Cambridge University Press.

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.

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

Juvonen, J., Vi-Nhuan, L., Kaganoff, T., Augustine, C., & Constant, L.. (2004). *Focus on the Wonder Years: Challenges Facing the American Middle School.* Santa Monica, CA: RAND Education.

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

Kieffer, M.J., & Marinell, W.H. (2012). *Navigating the Middle Grades: Evidence from New York City.* New York: Research Alliance for NYC Schools.

Kim, J. S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a mixed-method literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program? Results from a randomized controlled trial of READ 180 enterprise. *Educational Evaluation and Policy Analysis, 33*(2), 183-201. doi:10.3102/0162373711399148

Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly, 52*(3), 357-382. doi:10.1002/rrq.171

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

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.

National Assessment of Educational Progress (NAEP). (2022). *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. Retrieved from: https://www.nationsreportcard.gov/reading/nation/achievement/?grade=8

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.

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.

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* 48(4), 369-390. doi:10.1177/0022219413504995

Schochet, P. Z. (2013). Statistical power for school-based RCTs with binary outcomes. *Journal of Research on Educational Effectiveness*, *6*(3), 263-294. https://doi.org/10.1080/19345747.2012.725803

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.

Troyer, M., Kim, J. S., Hale, E., Wantchekon, K. A., & Armstrong, C. (2019). Relations among intrinsic and extrinsic reading motivation, reading amount, and comprehension: A conceptual replication. *Reading and Writing*, *32*(5), 1197-1218.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report

is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/Handbooks.

What Works Clearinghouse. (2020). *Review Protocol for Practice Guide for Assisting Students Struggling with Reading: Intervention in Grades 4-9, Version 4.0*. Washington, DC: What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education.

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.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology, 89*(3), 420.

Wolters, C. A., Denton, C. A., York, M. J., & Francis, D. J. (2014). Adolescents' motivation for reading: Group differences and relation to standardized achievement. *Reading and Writing 27*(3), 503-533.

APPENDIX A. READBASIX SUBTESTS AND ADMINISTRATION

Exhibit A.1 provides an overview of the reading skills measured by each ReadBasix subtest and its reliability. **Exhibit A.2** provides information on the timing of the test administration, and the amount of time that students took to complete the assessment, by research group. The key findings are that:

- Students in District 1 took the test in April (110 days into the calendar year on average). Students in District 2 took the test in June (166 days into the calendar year on average). Students in the STARI and control group took the test at the same time of year.
- On average, students in District 1 had more time to take the test than students in District 2, because class periods are longer in District 1. In District 1, test-taking time was 52 minutes on average. Students in the STARI group spent 4 extra minutes taking the test than students in the control group, and this difference is statistically significant at the 10 percent level. In District 2, average test-taking time was 35 minutes. Students in the control group spent an extra 3 minutes taking the test than students in the STARI group (this difference is not statistically significant).
- Students had sufficient time to complete almost all items for each subtest; in both districts, at least 90% of items were completed for any given subtest. (Subtest scores can still be derived even if students do not complete every item, but scores are more reliably estimated if more items are completed.) The proportion of items completed is not statistically different between students in the STARI and control group.

To test the sensitivity of the findings to differences in the amount of time that students spent taking the test, STARI's effect on the ReadBasix subtests was estimated controlling for the

number of minutes of test-taking time for the subtest of interest and the proportion of test items completed for that subtest (**Exhibit A.3**). The key findings are that:

- In District 1, the range of effect sizes when controlling for the test-taking variables (0.14 to 0.28) is similar to the range of effects observed in the main analysis (0.15 to 0.31). Estimated effects are not statistically significant.
- In District 2, the findings are more sensitive to controlling for the test-taking variables because the sample is much smaller (*n*=24). Estimated effects ranged from -0.45 to +0.68, compared to -0.26 to +0.68 for the main analysis. The sensitivity of the results to model specification in District 2 is further reason to interpret estimated effects on reading skills in this district with caution.

Exhibit A.1. Overview of ReadBasix Subtests

Subtest	Description
Word recognition and decoding (50 items; α=.91)	This subtest measures the ability to "get words off the page" accurately and efficiently. Sight word recognition is essential to fluent reading, while decoding is a skill that all readers use when they encounter novel words ("glutamate" and "uxorious"), proper names ("Bolivia" and "Myanmar") and product names ("Jenuvia" and "Kleenex"). Students are asked to identify whether the stimulus is a word, a decodable non-word, or a pseudohomophone.
Morphological awareness (32 items; α=.90)	This subtest measures the ability to understand that many words – especially academic words – are made up of several meaningful parts, as in civilization (civil + ize + ation), and to use those parts to aid in word recognition, sentence comprehension, and learning the meaning of new words. Students select which of three morphologically related words fits the syntax in a given sentence.
Efficiency of reading for basic comprehension (36 items; α=.93)	This subtest measures the ability to read text accurately and at an appropriate rate for comprehension. In the ReadBasix, both reading rate and comprehension are assessed through 36 comprehension items presented in a maze format. Students have 3 minutes to read each of 3 nonfiction passages and select appropriate words to fit sentence and passage context.
Sentence processing (26 items; α=.84)	This subtest measures the ability to comprehend sentences of varying levels of syntactic complexity, from very easy short sentences ("The book is on the table.") to more difficult sentences that use complex syntactic structures and are often found in textbooks and other kinds of academic writing ("The first ten amendments to the United States Constitution, also known as the Bill of Rights, are a series of statements that protect individual rights, though their exact meaning has been the object of much interpretation."). Students select the most appropriate word to complete sentences of increasing length and complexity.
Vocabulary (38 items; α=.86)	This subtest measures the ability to understand the meaning of individual words and their relationships to topical knowledge. Students must select a synonym or word that is topically associated with the target word.
Reading comprehension (22 items; α=.77)	This subtest measures the ability to understand the meaning of individual words and their relationships to topical knowledge. It includes a set of traditional multiple-choice items focused on the same three nonfiction passages that students read in the reading efficiency subtest.

Exhibit A.2. Timing and Duration of the ReadBasix Assessment

Testing outcome	STARI Group			Control Group			Difference	
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	P-Value
District 1								
# days into calendar year	34	110.1	3.7	46	110.2	3.7	-0.1	0.885
Total assessment minutes	34	52.6	12.5	46	48.5	10.7	4.1	0.089 *
Percent of items completed								
Word recognition	34	95.8	12.5	46	93.8	18.4	2.0	0.614
Morphological awareness	34	99.8	0.7	46	96.6	14.6	3.2	0.238
Vocabulary	34	96.9	10.1	46	97.2	9.3	-0.4	0.868
Sentence processing	34	96.3	8.3	46	93.4	15.4	2.8	0.360
Efficiency of reading	34	90.6	13.9	46	90.9	19.8	-0.3	0.944
Reading comprehension	34	98.5	6.2	46	98.6	11.2	-0.1	0.979
District 2								
# days into calendar year	15	165.1	2.0	9	165.8	2.9	-0.8	0.122
Total assessment minutes	15	34.1	9.8	9	37.4	11.8	-3.3	0.481
Percent of items completed								
Word recognition	15	100.0	0.0	9	100.0	0.0	0.0	n/a
Morphological awareness	15	96.4	9.6	9	100.3	0.0	-3.9	0.294
Vocabulary	15	100.0	0.0	9	100.0	0.0	0.0	n/a
Sentence processing	15	95.9	7.3	9	94.1	10.5	1.8	0.643
Efficiency of reading	15	91.6	22.4	9	89.3	10.8	2.3	0.786
Reading comprehension	15	98.9	4.1	9	97.6	6.3	1.3	0.593
Pooled								
# days into calendar year	49	126.9	25.8	55	127.2	21.4	-0.2	0.483
Total assessment minutes	49	46.9	14.5	55	44.5	12.4	2.4	0.262
Percent of items completed								
Word recognition	49	97.1	10.5	55	95.5	17.0	1.5	0.619
Morphological awareness	49	98.8	5.4	55	97.2	13.5	1.6	0.488
Vocabulary	49	97.8	8.5	55	98.1	8.6	-0.3	0.869
Sentence processing	49	96.2	7.9	55	93.6	14.7	2.6	0.309
Efficiency of reading	49	90.9	16.7	55	90.6	18.7	0.3	0.934
Reading comprehension	49	98.7	5.6	55	98.4	10.5	0.3	0.892

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted mean. The findings in this table are for students in the ReadBasix sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit A.3. Estimated Effect of STARI on Students' Reading Skills, Adjusted for Test Duration and Progress

	STARI Group			Control Group			Difference		
Outcome	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1[a]									
Word recognition and decoding	34	232.5	13.3	46	229.9	13.3	2.6	0.20	0.419
Morphological awareness	34	236.8	10.4	46	234.1	10.4	2.7	0.28	0.233
Vocabulary	34	233.1	7.0	46	231.9	7.0	1.2	0.17	0.463
Sentence processing	34	241.6	12.2	46	239.9	12.2	1.7	0.14	0.544
Efficiency of reading for comprehension	34	233.1	7.0	46	231.9	7.0	1.2	0.17	0.463
Reading comprehension	34	233.1	7.0	46	231.9	7.0	1.2	0.17	0.463
District 2									
Word recognition and decoding	15	238.7	11.0	9	233.9	11.0	4.7	0.33	0.416
Morphological awareness	15	241.3	5.9	9	244.8	5.9	-3.5	-0.45	0.411
Vocabulary	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.117
Sentence processing	15	240.9	5.4	9	235.4	5.4	5.5	0.58	0.295
Efficiency of reading for comprehension	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.117
Reading comprehension	15	237.5	5.2	9	234.0	5.2	3.6	0.68	0.117
Pooled									
Word recognition and decoding	49	234.4	13.6	55	231.7	13.6	2.7	0.20	0.339
Morphological awareness	49	238.2	10.7	55	236.5	10.7	1.7	0.17	0.399
Vocabulary	49	234.5	6.8	55	232.9	6.8	1.5	0.22	0.264
Sentence processing	49	241.4	11.4	55	239.3	11.4	2.0	0.18	0.390
Efficiency of reading for comprehension	49	234.5	6.8	55	232.9	6.8	1.5	0.22	0.264
Reading comprehension	49	234.5	6.8	55	232.9	6.8	1.5	0.22	0.264

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the ReadBasix sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment, for students' baseline ELA/reading test scores and other characteristics, as well as the total minutes of test-taking time for the relevant subtest and the percentage of items on that subtest completed by the student. Effect sizes are calculated by dividing

the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. [a]The estimated STARI-control differences in this table do not differ by a statistically significant amount across school districts.

APPENDIX B. STUDENT SURVEY SCALES

Three measures were created from the student survey: (1) reading self-efficacy, (2) intrinsic reading motivation, and (3) the amount of help students received with their reading. **Exhibit B.1** summarizes the items in the student survey that were used to create each measure. The measures were derived as follows:

- **Reading self-efficacy** and **intrinsic reading motivation** were created by recoding students' responses to a numeric scale (1=very different from me; 2=a little different from me; 3=a little like me; 4= a lot like me) and taking the average of a student's recoded responses across the items included in the measure. The reliability of the reading scales is moderate, ranging from 0.56 to 0.57 (see **Exhibit B.2**). To keep the survey short, only a subset of items from the original scales were used, which resulted in lower reliability.
- The **number of times a student received help with reading** was created by recoding the duration and frequency with which a student received extra help with their reading:
  - *For how LONG did you get this help with reading?* The response categories are "one month" (coded as 4 weeks), "a couple of months" (coded as 8 weeks), "one semester or term" (coded as 18 weeks), "most of the year" (coded as 27 weeks), and "all year" (coded as 36 weeks).
  - *How OFTEN did you get this help with reading?* The response categories are "less than once a month" (coded as 0.1 times a week), "once a month" (coded as .25 times a week), "every other week" (coded as .5 times a week), "once a week" (coded as 1 time a week), "twice a week" (coded as 2 times per week), "3-4 times a week" (coded as 3.5 times per week), and "every day" (coded as 5 times a week).

The number of times that each student received reading help was then calculated by multiplying their recoded response to the duration item by their recoded response to the frequency item. For example, a student who reports that they receive support for one semester every other week would receive a value of 9 (=18 weeks * 0.5 times per week).

Exhibit B.1. Items Included in Student Survey Scales

Reading Self-Efficacy (Source: MRQ; 4-point agreement scale)
I know that I will do well in reading next year I am a good reader
Intrinsic Reading Motivation (Source: MRQ; 4-point agreement scale)
If the teacher discusses something interesting I might read more about it I enjoy a long, involved story or fiction book I have favorite topics that I like to read about I make pictures in my mind when I read
Number of reading help sessions (Source: ERO)
Other than your teachers, did an adult help you one-on-one with your reading this school year, like a tutor? It can be someone at your school, outside of school, or at an after-school program. If "YES" to this question: • For how LONG did you get this help with your reading? (one month or less, a couple of months, one semester or term, more than one semester or term) • How OFTEN did you get this help with your reading? (less than once a month, once a month, every other week, once a week, twice a week, 3-4 times a week, every day)

Notes: MRQ = Motivations for Reading Questionnaire (MRQ; Wigfield & Guthrie, 1997). The scale is: very different from me; a little different from me; a little like me, a lot like me.
ERO = Enhanced Reading Opportunities Study (Somers *et al.*, 2010)

Exhibit B.2. Reliability of Reading Behavior Scales (Cronbach's Alpha)

Survey Scale	District 1	District 2	Pooled Districts
Reading self-efficacy (2 items)			
Cronbach's Alpha	0.61	0.44	0.56
Number of students	81	24	105
Intrinsic reading motivation (4 items)			
Cronbach's Alpha	0.51	0.71	0.57
Number of students	81	24	105

Notes:. The results in this table are based on students in the survey sample.

APPENDIX C. ADDITIONAL INFORMATION ON RESPONSE RATES

The study sample for this evaluation includes 398 eligible students.[47] **Exhibit C.1** shows the proportion of students in the study sample who are included in each analysis sample (state test, course, Read Basix, and survey), for both districts pooled together. **Exhibit C.2** provides additional information on the creation of the state test and course samples, and the proportion of students who were excluded from these analysis samples for different reasons. **Exhibit C.3** provides similar information, but for the creation of the ReadBasix and survey samples.

This appendix also provides information on the generalizability of the Read Basix/survey findings, by comparing the characteristics and outcomes of students in District 1 who are included *versus* excluded from the ReadBasix sample:

- **Exhibit C.4** looks at the baseline characteristics of students in the ReadBasix sample in District 1, compared to the characteristics of students who are excluded from this sample but who are included in the state test sample. Students in the ReadBasix sample had lower baseline reading test scores on average than students excluded from this sample.
- **Exhibit C.5** shows the estimated effect of STARI on ELA state test scores and course performance *for the subset of students in the ReadBasix sample*. The estimated effect of STARI on ELA state test scores for students in the ReadBasix sample is 0.39, which is similar in magnitude to its effect on ELA state tests for the full state test sample (effect size=0.32). This suggests that the estimated effect of STARI on ReadBasix subtests in District 1 would have been similar in magnitude had it been possible to assess all students

[47] In District 2, an additional 151 students were eligible for STARI and randomly assigned to the STARI class. However, these students are excluded from the study sample because their random assignment blocks (6 blocks in total) did not include any STARI students and/or any control students whose parents consented to their participation in the study.

in the study sample. In other words, the general pattern of effects on reading skills based on the ReadBasix sample can likely be generalized to the full study sample.

Exhibit C.1. Response Rates by Student Outcome, Pooled Sample

	Study Sample			STARI Group			Control Group			Difference	
Analysis Sample	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Adjusted Response Rate	STARI - Control	P-Value
State test	398	200	50.3%	183	84	45.9%	215	116	43.8%	2.1%	0.589
Course	398	202	50.8%	183	85	46.4%	215	117	43.9%	2.5%	0.513
ReadBasix	398	104	26.1%	183	49	26.8%	215	55	19.4%	7.3%	0.077 *
Student survey	398	105	26.4%	183	49	26.8%	215	56	20.0%	6.7%	0.105
District reading test	398	63	15.8%	183	23	12.6%	215	40	14.3%	-1.8%	0.497

Notes: N = Number of students (randomized or in the sample). For the control group, the response rate shown in the table is not necessarily equal to the response rate directly calculated using the number of students randomized and in the analysis sample, because the response rate for the control group is regression-adjusted for the blocking of random assignment. The differential response rate between STARI and control students is also regression-adjusted. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit C.2. Causes of Attrition and Creation of the State Test and Course Performance Analysis Samples

	Study Sample			STARI Group			Control Group			Difference	
Analysis Sample	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Adjusted Response Rate	STARI - Control	P-Value
District 1											
Student records are available	166	166	100.0%	60	60	100.0%	106	106	100.0%	0.0%	n/a
Student has baseline reading score	166	166	100.0%	60	60	100.0%	106	106	100.0%	0.0%	n/a
Student is in state test sample	166	154	92.8%	60	53	88.3%	106	101	96.9%	-8.6%	0.098 *
Student is in course sample	166	156	94.0%	60	54	90.0%	106	102	97.6%	-7.6%	0.120
District 2											
Parent returned consent form	232	92	39.7%	123	57	46.3%	109	35	34.6%	11.7%	0.075 *
Parent consent for records	232	70	30.2%	123	45	36.6%	109	25	24.6%	12.0%	0.054 *
Student has baseline reading score	232	46	19.8%	123	31	25.2%	109	15	15.7%	9.5%	0.085 *
Student is in state test sample	232	46	19.8%	123	31	25.2%	109	15	15.7%	9.5%	0.085 *
Student is in course sample	232	46	19.8%	123	31	25.2%	109	15	15.7%	9.5%	0.085 *
Pooled											
Parent consent for records[a]	398	236	59.3%	183	105	57.4%	215	131	50.3%	7.1%	0.054 *
Student has baseline reading score	398	212	53.3%	183	91	49.7%	215	121	44.1%	5.6%	0.086 *
Student is in state test sample	398	200	50.3%	183	84	45.9%	215	116	43.8%	2.1%	0.589
Student is in course sample	398	202	50.8%	183	85	46.4%	215	117	43.9%	2.5%	0.513

Notes: N = Number of students (randomized or in the sample). The state test sample is defined as students who have a baseline reading score and an ELA state test scores in spring 2022; the course sample is defined as students who have a baseline reading score and an ELA course mark in spring 2022. For the control group, the response rate shown in the table is not necessarily equal to the response rate directly calculated using the number of students randomized and in the analysis sample, because the response rate for the control group is regression-adjusted for the blocking of random assignment. The differential response rate between STARI and control students is also regression-adjusted. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Assumes a rate of 100% for District 1 where a parent consent process was not required for student records.

Exhibit C.3. Causes of Attrition and Creation of the ReadBasix and Student Survey Analysis Samples

	Study Sample			STARI Group			Control Group			Difference	
Analysis Sample	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Adjusted Response Rate	STARI - Control	P-Value
District 1											
Parent Returned Consent Form	166	106	63.9%	60	43	71.7%	106	63	61.5%	10.1%	0.166
Parent Consent for ReadBasix & Survey	166	94	56.6%	60	38	63.3%	106	56	54.8%	8.6%	0.286
Student assent for Read Basix & Survey	166	83	50.0%	60	34	56.7%	106	49	46.7%	10.0%	0.226
Student has baseline reading score	166	83	50.0%	60	34	56.7%	106	49	46.7%	10.0%	0.226
Student is in ReadBasix sample	166	80	48.2%	60	34	56.7%	106	46	43.6%	13.0%	0.111
Student is in survey sample	166	81	48.8%	60	34	56.7%	106	47	45.1%	11.6%	0.160
District 2											
Parent Returned Consent Form	232	92	39.7%	123	57	46.3%	109	35	34.6%	11.7%	0.075 *
Parent Consent for ReadBasix & Survey	232	67	28.9%	123	41	33.3%	109	26	25.3%	8.0%	0.192
Student assent for Read Basix & Survey	232	48	20.7%	123	29	23.6%	109	19	17.9%	5.7%	0.284
Student has a baseline reading score	232	29	12.5%	123	19	15.4%	109	10	10.6%	4.8%	0.281
Student is in ReadBasix sample	232	24	10.3%	123	15	12.2%	109	9	8.8%	3.4%	0.415
Student is in survey sample	232	24	10.3%	123	15	12.2%	109	9	8.8%	3.4%	0.415
Pooled											
Parent Returned Consent Form	398	198	49.7%	183	100	54.6%	215	98	43.6%	11.1%	0.024 **
Parent Consent for ReadBasix & Survey	398	161	40.5%	183	79	43.2%	215	82	34.9%	8.3%	0.092 *
Student assent for Read Basix & Survey	398	131	32.9%	183	63	34.4%	215	68	27.0%	7.5%	0.105

Analysis Sample	Study Sample			STARI Group			Control Group			Difference	
	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Response Rate	N Rando-mized	N Sample	Adjusted Response Rate	STARI - Control	P-Value
Student has baseline reading score	398	112	28.1%	183	53	29.0%	215	59	22.0%	6.9%	0.105
Student is in ReadBasix sample	398	104	26.1%	183	49	26.8%	215	55	19.4%	7.3%	0.077 *
Student is in survey sample	398	105	26.4%	183	49	26.8%	215	56	20.0%	6.7%	0.105

Notes: N = Number of students (randomized or in the sample). The ReadBasix sample is defined as students who have a baseline reading score and who took to ReadBasix in spring 2022; the survey sample is defined as students who have a baseline reading score and who completed the student survey in spring 2022. For the control group, the response rate shown in the table is not necessarily equal to the response rate directly calculated using the number of students randomized and in the analysis sample, because the response rate for the control group is regression-adjusted for the blocking of random assignment. The differential response rate between STARI and control students is also regression-adjusted. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit C.4. Baseline Reading Achievement and Characteristics of Students in the ReadBasix Sample, Compared to Students in the State Test Sample but not in the ReadBasix Sample, District 1

	ReadBasix Sample			In State Test Sample but not ReadBasix Sample			Difference		
Baseline outcome or characteristic	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	79	-0.13	1.05	75	0.13	0.94	-0.26	-0.26	0.104
Educational flags (prior school year)									
English language learner (%)	77	0%	0%	75	0%	0%	0%	0.00	n/a
Individualized education plan (%)	77	3%	16%	75	5%	23%	-3%	-0.14	0.392
Free/reduced price lunch (%)	77	100%	0%	75	100%	0%	0%	0.00	n/a
Racial/ethnic group (%)									
Black	79	99%	11%	75	99%	12%	0%	0.01	0.971
Hispanic	79	0%	0%	75	1%	12%	-1%	-0.17	0.319
Other race	79	1%	11%	75	0%	0%	1%	0.16	0.319
Gender (%)									
Female	79	48%	50%	75	32%	47%	16%	0.33	0.042 **
Male	79	52%	50%	75	68%	47%	-16%	-0.33	0.042 **
Age and grade (program year)									
Age	79	11.9	0.8	75	12.1	0.8	-0.1	-0.13	0.422
Grade 6 (%)	79	100%	0%	75	100%	0%	0%	0.00	n/a
Grade 7 (%)	79	0%	0%	75	0%	0%	0%	0.00	n/a
Grade 8 (%)	79	0%	0%	75	0%	0%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit C.5. Estimated Effect of STARI on ELA State Test Scores and ELA Course Performance, Students in the ReadBasix Sample, District 1

	STARI Group			Control Group			Difference		
Outcome	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
ELA state test score (z-score)	34	0.24	1.13	45	-0.16	0.92	0.40	0.39	0.103
ELA course grade (0-100%)	34	73.9	8.6	46	72.0	8.7	1.8	0.21	0.295

Notes: ELA = English Language Arts. N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the ReadBasix sample. The means the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

APPENDIX D. STUDENT CHARACTERISTICS

Exhibits D.1 to D.12 provide additional information on the baseline characteristics of students in each analysis sample (state test sample, course sample, Read Basix sample, and student survey sample), by study district and for both districts pooled together.

Exhibit D.1. Baseline Reading Achievement and Characteristics of Students in the State Test Sample, District 1

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	53	-0.09	0.97	101	-0.02	1.02	-0.1	-0.07	0.702
Educational flags (prior school year)									
English language learner (%)	51	0%	0%	101	0%	0%	0%	0.00	n/a
Individualized education plan (%)	51	0%	0%	101	7%	24%	-7%	-0.36	0.022 **
Free/reduced price lunch (%)	51	100%	0%	101	100%	0%	0%	0.00	n/a
Racial/ethnic group (%)									
Black	53	98%	14%	101	99%	10%	-1%	-0.06	0.798
Hispanic	53	0%	0%	101	1%	10%	-1%	-0.18	0.317
Other race	53	2%	14%	101	0%	0%	2%	0.27	0.319
Gender (%)									
Female	53	47%	50%	101	39%	48%	8%	0.16	0.373
Male	53	53%	50%	101	61%	48%	-8%	-0.16	0.373
Age and grade (program year)									
Age	53	12.0	0.8	101	12.0	0.8	0.0	-0.02	0.916
Grade 6 (%)	53	100%	0%	101	100%	0%	0%	0.00	n/a
Grade 7 (%)	53	0%	0%	101	0%	0%	0%	0.00	n/a
Grade 8 (%)	53	0%	0%	101	0%	0%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.2. Baseline Reading Achievement and Characteristics of Students in the State Test Sample, District 2

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	31	0.01	0.95	15	0.13	0.97	-0.12	-0.13	0.711
Educational flags (prior school year)									
English language learner (%)	31	6%	25%	15	12%	35%	-5%	-0.19	0.576
Individualized education plan (%)	31	26%	44%	15	36%	49%	-10%	-0.22	0.498
Free/reduced price lunch (%)	31	90%	30%	15	73%	49%	18%	0.48	0.172
Racial/ethnic group (%)									
Black	31	35%	49%	15	20%	41%	15%	0.32	0.161
Hispanic	31	55%	51%	15	75%	46%	-20%	-0.41	0.052 *
Other race	31	10%	30%	15	5%	26%	5%	0.18	0.524
Gender (%)									
Female	31	58%	50%	15	59%	51%	-1%	-0.02	0.942
Male	31	42%	50%	15	41%	51%	1%	0.02	0.942
Age and grade (program year)									
Age	31	11.7	0.9	15	11.9	0.9	-0.2	-0.22	0.240
Grade 6 (%)	31	58%	50%	15	58%	51%	0%	0.00	n/a
Grade 7 (%)	31	35%	49%	15	35%	49%	0%	0.00	n/a
Grade 8 (%)	31	6%	25%	15	6%	26%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.3. Baseline Reading Achievement and Characteristics of Students in the State Test Sample, Pooled Districts

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	84	-0.05	0.96	116	0.03	1.01	-0.08	-0.08	0.606
Educational flags (prior school year)									
English language learner (%)	82	2%	16%	116	4%	13%	-1%	-0.09	0.546
Individualized education plan (%)	82	10%	30%	116	17%	29%	-8%	-0.26	0.057 *
Free/reduced price lunch (%)	82	96%	19%	116	92%	20%	4%	0.21	0.151
Racial/ethnic group (%)									
Black	84	75%	44%	116	72%	32%	3%	0.08	0.349
Hispanic	84	20%	40%	116	26%	31%	-6%	-0.16	0.027 **
Other race	84	5%	21%	116	2%	9%	3%	0.18	0.240
Gender (%)									
Female	84	51%	50%	116	45%	49%	6%	0.12	0.460
Male	84	49%	50%	116	55%	49%	-6%	-0.12	0.460
Age and grade (program year)									
Age	84	11.9	0.8	116	12.0	0.8	-0.1	-0.07	0.636
Grade 6 (%)	84	85%	36%	116	85%	22%	0%	0.00	n/a
Grade 7 (%)	84	13%	34%	116	13%	20%	0%	0.00	n/a
Grade 8 (%)	84	2%	15%	116	2%	9%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.4. Baseline Reading Achievement and Characteristics of Students in the Course Sample, District 1

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	54	-0.12	0.98	102	-0.02	1.02	-0.10	-0.10	0.571
Educational flags (prior school year)									
English language learner (%)	52	0%	0%	102	0%	0%	0%	0.00	n/a
Individualized education plan (%)	52	0%	0%	102	7%	24%	-7%	-0.35	0.022 **
Free/reduced price lunch (%)	52	100%	0%	102	100%	0%	0%	0.00	n/a
Racial/ethnic group (%)									
Black	54	98%	14%	102	99%	10%	-1%	-0.05	0.809
Hispanic	54	0%	0%	102	1%	10%	-1%	-0.18	n/a
Other race	54	2%	14%	102	0%	0%	2%	0.26	0.319
Gender (%)									
Female	54	46%	50%	102	40%	49%	7%	0.13	0.447
Male	54	54%	50%	102	60%	49%	-7%	-0.13	0.447
Age and grade (program year)									
Age	54	12.0	0.8	102	12.0	0.8	0.0	-0.02	0.890
Grade 6 (%)	54	100%	0%	102	100%	0%	0%	0.00	n/a
Grade 7 (%)	54	0%	0%	102	0%	0%	0%	0.00	n/a
Grade 8 (%)	54	0%	0%	102	0%	0%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.5. Baseline Reading Achievement and Characteristics of Students in the Course Sample, District 2

Baseline outcome or characteristic	STARI Group			Control Group			Difference		
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	31	0.01	0.95	15	0.13	0.97	-0.12	-0.13	0.711
Educational flags (prior school year)									
English language learner (%)	31	6%	25%	15	12%	35%	-5%	-0.19	0.576
Individualized education plan (%)	31	26%	44%	15	36%	49%	-10%	-0.22	0.498
Free/reduced price lunch (%)	31	90%	30%	15	73%	49%	18%	0.48	0.172
Racial/ethnic group (%)									
Black	31	35%	49%	15	20%	41%	15%	0.32	0.161
Hispanic	31	55%	51%	15	75%	46%	-20%	-0.41	0.052 *
Other race	31	10%	30%	15	5%	26%	5%	0.18	0.524
Gender (%)									
Female	31	58%	50%	15	59%	51%	-1%	-0.02	0.942
Male	31	42%	50%	15	41%	51%	1%	0.02	0.942
Age and grade (program year)									
Age	31	11.7	0.9	15	11.9	0.9	-0.2	-0.22	0.240
Grade 6 (%)	31	58%	50%	15	58%	51%	0%	0.00	n/a
Grade 7 (%)	31	35%	49%	15	35%	49%	0%	0.00	n/a
Grade 8 (%)	31	6%	25%	15	6%	26%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.6. Baseline Reading Achievement and Characteristics of Students in the Course Sample, Pooled Districts

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	85	-0.07	0.97	117	0.03	1.01	-0.10	-0.10	0.498
Educational flags (prior school year)									
English language learner (%)	83	2%	15%	117	4%	13%	-1%	-0.09	0.546
Individualized education plan (%)	83	10%	30%	117	17%	29%	-8%	-0.26	0.057 *
Free/reduced price lunch (%)	83	96%	19%	117	92%	20%	4%	0.21	0.151
Racial/ethnic group (%)									
Black	85	75%	43%	117	72%	32%	3%	0.08	0.340
Hispanic	85	20%	40%	117	26%	30%	-6%	-0.16	0.027 **
Other race	85	5%	21%	117	2%	9%	3%	0.18	0.241
Gender (%)									
Female	85	51%	50%	117	46%	49%	5%	0.10	0.531
Male	85	49%	50%	117	54%	49%	-5%	-0.10	0.531
Age and grade (program year)									
Age	85	11.9	0.8	117	12.0	0.8	-0.1	-0.07	0.616
Grade 6 (%)	85	85%	36%	117	85%	22%	0%	0.00	n/a
Grade 7 (%)	85	13%	34%	117	13%	20%	0%	0.00	n/a
Grade 8 (%)	85	2%	15%	117	2%	9%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.7. Baseline Reading Achievement and Characteristics of Students in the ReadBasix Sample, District 1

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	34	-0.16	1.03	46	-0.18	1.07	0.02	0.02	0.942
Educational flags (prior school year)									
English language learner (%)	32	0%	0%	46	0%	0%	0%	0.00	n/a
Individualized education plan (%)	32	0%	0%	46	5%	21%	-5%	-0.34	0.162
Free/reduced price lunch (%)	32	100%	0%	46	100%	0%	0%	0.00	n/a
Racial/ethnic group (%)									
Black	34	97%	17%	46	100%	0%	-3%	-0.30	0.320
Hispanic	34	0%	0%	46	0%	0%	0%	0.00	n/a
Other race	34	3%	17%	46	0%	0%	3%	0.30	0.320
Gender (%)									
Female	34	53%	51%	46	45%	50%	8%	0.17	0.488
Male	34	47%	51%	46	55%	50%	-8%	-0.17	0.488
Age and grade (program year)									
Age	34	12.0	0.8	46	12.1	0.9	-0.1	-0.10	0.689
Grade 6 (%)	34	100%	0%	46	100%	0%	0%	0.00	n/a
Grade 7 (%)	34	0%	0%	46	0%	0%	0%	0.00	n/a
Grade 8 (%)	34	0%	0%	46	0%	0%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.8. Baseline Reading Achievement and Characteristics of Students in the ReadBasix Sample, District 2

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	15	0.05	1.03	9	-0.21	0.81	0.26	0.27	0.570
Educational flags (prior school year)									
English language learner (%)	15	0%	0%	9	10%	33%	-10%	-0.48	0.355
Individualized education plan (%)	15	7%	26%	9	38%	50%	-31%	-0.84	0.064 *
Free/reduced price lunch (%)	15	93%	26%	9	59%	53%	34%	0.90	0.110
Racial/ethnic group (%)									
Black	15	33%	49%	9	22%	44%	12%	0.25	0.372
Hispanic	15	47%	52%	9	69%	50%	-22%	-0.44	0.091 *
Other race	15	20%	41%	9	9%	33%	11%	0.28	0.497
Gender (%)									
Female	15	67%	49%	9	54%	53%	13%	0.25	0.582
Male	15	33%	49%	9	46%	53%	-13%	-0.25	0.582
Age and grade (program year)									
Age	15	11.5	1.0	9	11.9	1.1	-0.4	-0.40	0.099 *
Grade 6 (%)	15	80%	41%	9	80%	44%	0%	0.00	n/a
Grade 7 (%)	15	13%	35%	9	13%	33%	0%	0.00	n/a
Grade 8 (%)	15	7%	26%	9	7%	33%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.9. Baseline Reading Achievement and Characteristics of Students in the ReadBasix Sample, Pooled Districts

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	49	-0.10	1.02	55	-0.17	1.03	0.07	0.07	0.739
Educational flags (prior school year)									
English language learner (%)	47	0%	0%	55	2%	13%	-2%	-0.23	0.316
Individualized education plan (%)	47	2%	15%	55	14%	29%	-11%	-0.49	0.020 **
Free/reduced price lunch (%)	47	98%	15%	55	90%	26%	8%	0.37	0.103
Racial/ethnic group (%)									
Black	49	78%	42%	55	77%	34%	0%	0.00	0.979
Hispanic	49	14%	35%	55	19%	31%	-5%	-0.15	0.092 *
Other race	49	8%	28%	55	3%	13%	5%	0.23	0.236
Gender (%)									
Female	49	57%	50%	55	48%	50%	9%	0.19	0.377
Male	49	43%	50%	55	52%	50%	-9%	-0.19	0.377
Age and grade (program year)									
Age	49	11.8	0.9	55	12.0	0.9	-0.2	-0.17	0.357
Grade 6 (%)	49	94%	24%	55	94%	19%	0%	0.00	n/a
Grade 7 (%)	49	4%	20%	55	4%	13%	0%	0.00	n/a
Grade 8 (%)	49	2%	14%	55	2%	13%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.10. Baseline Reading Achievement and Characteristics of Students in the Survey Sample, District 1

Baseline outcome or characteristic	STARI Group			Control Group			Difference		
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	34	-0.16	1.03	47	-0.19	1.08	0.02	0.02	0.921
Educational flags (prior school year)									
English language learner (%)	32	0%	0%	47	0%	0%	0%	0.00	n/a
Individualized education plan (%)	32	0%	0%	47	5%	20%	-5%	-0.33	0.162
Free/reduced price lunch (%)	32	100%	0%	47	100%	0%	0%	0.00	n/a
Racial/ethnic group (%)									
Black	34	97%	17%	47	100%	0%	-3%	-0.29	0.320
Hispanic	34	0%	0%	47	0%	0%	0%	0.00	n/a
Other race	34	3%	17%	47	0%	0%	3%	0.29	0.320
Gender (%)									
Female	34	53%	51%	47	46%	50%	7%	0.13	0.574
Male	34	47%	51%	47	54%	50%	-7%	-0.13	0.574
Age and grade (program year)									
Age	34	12.0	0.8	47	12.0	0.9	0.0	-0.06	0.815
Grade 6 (%)	34	100%	0%	47	100%	0%	0%	0.00	n/a
Grade 7 (%)	34	0%	0%	47	0%	0%	0%	0.00	n/a
Grade 8 (%)	34	0%	0%	47	0%	0%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.11. Baseline Reading Achievement and Characteristics of Students in the Survey Sample, District 2

Baseline outcome or characteristic	STARI Group			Control Group			Difference		
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Baseline reading achievement (z-score)	15	0.05	1.03	9	-0.21	0.81	0.26	0.27	0.570
Educational flags (prior school year)									
English language learner (%)	15	0%	0%	9	10%	33%	-10%	-0.48	0.355
Individualized education plan (%)	15	7%	26%	9	38%	50%	-31%	-0.84	0.064 *
Free/reduced price lunch (%)	15	93%	26%	9	59%	53%	34%	0.90	0.110
Racial/ethnic group (%)									
Black	15	33%	49%	9	22%	44%	12%	0.25	0.372
Hispanic	15	47%	52%	9	69%	50%	-22%	-0.44	0.091 *
Other race	15	20%	41%	9	9%	33%	11%	0.28	0.497
Gender (%)									
Female	15	67%	49%	9	54%	53%	13%	0.25	0.582
Male	15	33%	49%	9	46%	53%	-13%	-0.25	0.582
Age and grade (program year)									
Age	15	11.5	1.0	9	11.9	1.1	-0.4	-0.40	0.099 *
Grade 6 (%)	15	80%	41%	9	80%	44%	0%	0.00	n/a
Grade 7 (%)	15	13%	35%	9	13%	33%	0%	0.00	n/a
Grade 8 (%)	15	7%	26%	9	7%	33%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit D.12. Baseline Reading Achievement and Characteristics of Students in the Survey Sample, Pooled Districts

	STARI Group			Control Group			Difference		
Baseline outcome or characteristic	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
Baseline reading achievement (z-score)	49	-0.10	1.02	56	-0.18	1.03	0.08	0.07	0.724
Educational flags (prior school year)									
English language learner (%)	47	0%	0%	56	2%	13%	-2%	-0.22	0.315
Individualized education plan (%)	47	2%	15%	56	13%	29%	-11%	-0.48	0.020 **
Free/reduced price lunch (%)	47	98%	15%	56	90%	26%	8%	0.37	0.103
Racial/ethnic group (%)									
Black	49	78%	42%	56	77%	33%	0%	0.00	0.979
Hispanic	49	14%	35%	56	19%	31%	-5%	-0.15	0.091 *
Other race	49	8%	28%	56	3%	13%	5%	0.23	0.235
Gender (%)									
Female	49	57%	50%	56	49%	50%	8%	0.16	0.444
Male	49	43%	50%	56	51%	50%	-8%	-0.16	0.444
Age and grade (program year)									
Age	49	11.8	0.9	56	12.0	0.9	-0.1	-0.14	0.446
Grade 6 (%)	49	94%	24%	56	94%	19%	0%	0.00	n/a
Grade 7 (%)	49	4%	20%	56	4%	13%	0%	0.00	n/a
Grade 8 (%)	49	2%	14%	56	2%	13%	0%	0.00	n/a

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

APPENDIX E . ADDITIONAL FINDINGS ON RECEIPT OF STARI AND OTHER READING SUPPORTS

Exhibit E.1 provides information on enrollment in reading classes and other supports received by students in each research group, for both districts pooled together.

Exhibit E.2 shows enrollment in reading classes for students in the survey sample in District 1 (i.e., the same sample of students for whom information is also available on the reading supports they received outside of regular classes). In the survey sample, 43% of students in the control group were enrolled in an alternative reading class, which is higher than the proportion for students in the larger course sample (34%). This is consistent with the finding that students in the survey sample had lower reading scores at baseline (see Appendix C). This suggests that the extra reading supports reported by students in the survey sample may overestimate the proportion of students in the full study sample who received such supports outside of their regular classes.

Exhibit E.1: Reading Classes and Other Reading Supports, Pooled Districts

Intervention or Service Receipt Measure	STARI Group			Control Group			Difference	
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	P-Value
Reading classes								
Enrolled in STARI (%)	85	91%	29%	117	0%	0%	90%	<0.001 ***
Enrolled in STARI or another reading class (%)	85	92%	28%	117	37%	48%	55%	<0.001 ***
Reading support from an adult who is not a teacher								
Received help[a] (%)	49	55%	50%	56	53%	50%	2%	0.810
Number of times received help during school year[b]	49	19.6	41.4	56	22.8	37.9	-3.2	0.694

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. Enrollment in STARI and reading classes is reported for students in the course sample; the receipt of other reading supports is reported for students in the student survey sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]*Other than your teachers, did an adult help you with your reading this school year? It can be someone at your school, outside of school, or at an after school program?* [b]*How LONG did you get this help with your reading? How OFTEN did you get this help with your reading?*

Exhibit E.2: Reading Classes and Other Reading Supports, Survey Sample

Intervention or Service Receipt Measure	STARI Group			Control Group			Difference	
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	P-Value
District 1								
Reading classes								
Enrolled in STARI (%)	34	94%	24%	47	0%	0%	94%	<0.001 ***
Enrolled in STARI or another reading class (%)	34	94%	24%	47	43%	49%	51%	<0.001 ***
Reading support from an adult who is not a teacher								
Received help[a] (%)	34	65%	49%	47	56%	50%	9%	0.438
Number of times received help during school year[b]	34	26.8	47.7	47	21.4	32.7	5.3	0.567
District 2								
Reading classes								
Enrolled in STARI (%)	15	93%	26%	9	0%	0%	93%	<0.001 ***
Enrolled in STARI or another reading class (%)	15	93%	26%	9	60%	50%	33%	0.060 *
Reading support from an adult who is not a teacher								
Received help[a] (%)	15	33%	49%	9	52%	53%	-19%	0.406
Number of times received help during school year[b]	15	3.3	10.2	9	36.0	57.9	-32.7	0.047 **
Pooled								
Reading classes								
Enrolled in STARI (%)	49	94%	24%	56	1%	0%	93%	<0.001 ***
Enrolled in STARI or another reading class (%)	49	94%	24%	56	47%	50%	47%	<0.001 ***
Reading support from an adult who is not a teacher								
Received help[a] (%)	49	55%	50%	56	53%	50%	2%	0.810
Number of times received help during school year[b]	49	19.6	41.4	56	22.8	37.9	-3.2	0.694

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. All findings in this table are based on students in the survey sample; The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]*Other than your teachers, did an adult help you with your reading this school year? It can be someone at your school, outside of school, or at an after school program?* [b]*How LONG did you get this help with your reading? How OFTEN did you get this help with your reading?*

APPENDIX F. ADDITIONAL IMPACT FINDINGS, MINIMUM DETECTABLE EFFECTS, AND CORRELATIONS

Exhibit F.1 provides estimates of the effect of STARI for both districts together, for the student outcomes discussed in the report. **Exhibits F.2** and **F.3** examine the estimated effect of STARI on supplemental outcomes related to students' performance on state tests, their course performance, and their reading skills.[48] **Exhibit F.4** provides information on the standard error and the minimum detectable effect size for each student outcome. **Exhibit F.5** shows the correlation between baseline reading scores and each of the main student outcome measures; **Exhibits F.6** and **F.7** show the correlation between each of the ReadBasix subtests, and between the two survey outcomes.

[48] The estimated effect of STARI on binary outcomes was estimated using a linear probability model (LPM). Many RCTs in education use an LPM to estimate program effects on binary outcomes, although logistic regression is also occasionally used. There are convincing arguments in favor of using LPMs in the context of RCTs, including: (1) LPMs' relative simplicity, (2) LPMs' easy interpretation, (3) the LPM estimator is unbiased, (4) the LPM standard error estimator is correct, and (5) logit models with covariates tend to reduce precision (Deke, 2014; Schochet, 2013). Heteroskedasticity-robust standard errors were used to account for the fact that using an LPM with a binary outcome can lead to heteroskedasticity.

Exhibit F.1. Estimated Effect of STARI on Student Outcomes, Pooled Districts

	STARI Group			Control Group			Difference		
Outcome	**N**	**Mean**	**SD**	**N**	**Adj. Mean**	**SD**	**STARI - Control**	**Effect size**	**P-Value**
ELA state tests and courses									
ELA state test score (z-score)	84	0.13	1.06	116	0.0	0.93	0.17	0.17	0.235
ELA course grade (0-100%)	85	75.6	9.7	117	73.8	9.8	1.8	0.18	0.127
Reading skills, vocabulary, and comprehension									
Word recognition and decoding	49	234.4	13.5	55	230.5	13.6	3.9	0.29	0.166
Morphological awareness	49	238.2	9.0	55	236.2	10.7	2.0	0.20	0.319
Vocabulary	49	234.5	6.8	55	233.1	6.8	1.4	0.21	0.301
Sentence processing	49	241.4	10.8	55	239.1	11.4	2.2	0.20	0.369
Efficiency of reading for comprehension	49	234.5	6.8	55	233.1	6.8	1.4	0.21	0.301
Reading comprehension	49	234.5	6.8	55	233.1	6.8	1.4	0.21	0.301
Reading self-efficacy and motivation									
Reading self-efficacy (1-4)	49	3.20	0.7	56	2.8	0.8	0.4	0.50	0.033 **
Reading motivation (1-4)	49	3.1	0.6	56	2.9	0.7	0.2	0.27	0.233

Notes: ELA = English Language Arts. N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. Findings for ELA state tests are based on students in the state test sample; findings for ELA course grades are based on students in the course sample; findings for reading skills, vocabulary and comprehension are based on students in the ReadBasix sample; and findings for reading self-efficacy and motivation are based on students in the survey sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Exhibit F.2. Estimated effect of STARI on Additional ELA State Test and Course Performance Outcomes

Outcome	STARI Group			Control Group			Difference		
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1[a]									
ELA state test outcomes									
ELA state test (% proficient)	53	5.7	23.3	101	-0.9	0.00	6.5	0.48	0.084 *
ELA and other course outcomes									
ELA courses passed (%)	54	59.3	49.6	102	58.5	50.2	0.8	0.02	0.920
Social studies course grade (0-100%)	54	73.2	9.4	102	70.3	8.2	2.9	0.34	0.016 **
Social studies courses passed (0-100%)	54	64.8	48.2	102	45.4	48.8	19.4	0.40	0.010 **
Science course grade (0-100%)	54	74.3	12.2	102	73.2	12.7	1.1	0.09	0.323
Science courses passed (0-100%)	54	55.6	50.2	102	53.1	50.1	2.4	0.05	0.745
Math course grade (0-100%)	54	70.7	7.0	102	69.5	8.0	1.2	0.16	0.305
Math courses passed (0-100%)	54	51.9	50.4	102	55.1	50.0	-3.2	-0.06	0.691
District 2									
ELA state test outcomes									
ELA state test (% proficient)	31	45.2	50.6	15	60.8	51.6	-15.6	-0.31	0.215
ELA and other course outcomes									
ELA courses passed (%)	31	96.8	18.0	15	93.6	13.8	3.1	0.19	0.610
Social studies course grade (0-100%)	22	77.8	10.6	8	80.0	12.6	-2.2	-0.20	0.672
Social studies courses passed (0-100%)	22	95.5	21.3	8	100.4	0.0	-4.9	-0.27	0.619
Science course grade (0-100%)	31	79.2	10.3	15	73.7	14.4	5.5	0.47	0.141
Science courses passed (0-100%)	31	92.5	25.4	15	96.7	13.8	-4.3	-0.19	0.574
Math course grade (0-100%)	31	77.0	12.4	15	76.2	15.6	0.8	0.06	0.839
Math courses passed (0-100%)	31	96.8	18.0	15	87.3	22.2	9.5	0.49	0.187

Outcome	STARI Group			Control Group			Difference		
	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
Pooled									
ELA state test outcomes									
ELA state test (% proficient)	84	20.2	40.4	116	19.3	23.92	0.9	0.03	0.819
ELA and other course outcomes									
ELA courses passed (%)	85	72.9	44.7	117	71.8	49.2	1.1	0.02	0.855
Social studies course grade (0-100%)	76	74.5	9.9	110	72.4	9.1	2.1	0.23	0.081 *
Social studies courses passed (0-100%)	76	73.7	44.3	110	57.7	49.7	16.0	0.34	0.015 **
Science course grade (0-100%)	85	76.1	11.7	117	74.3	12.9	1.8	0.15	0.115
Science courses passed (0-100%)	85	69.0	46.2	117	68.4	48.8	0.7	0.01	0.913
Math course grade (0-100%)	85	73.0	9.8	117	71.9	9.6	1.1	0.11	0.389
Math courses passed (0-100%)	85	68.2	46.8	117	66.7	49.4	1.6	0.03	0.813

Notes: ELA = English Language Arts. N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the state test sample and the course sample. Course pass rates are the number of courses passed as a proportion of the number of courses attempted. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. [a]The estimated STARI-control difference on the following outcomes is statistically larger in District 1 than in District 2: ELA state test proficiency (p-value=0.089) and social studies courses passed (p-value=0.048).

Exhibit F.3. Estimated Effect of STARI on the Percentage of Students at Grade Level on Reading Skills, Vocabulary and Comprehension

	STARI Group			Control Group			Difference		
Outcome	N	Mean	SD	N	Adj. Mean	SD	STARI - Control	Effect size	P-Value
District 1[a]									
Word recognition and decoding	34	47.1	49.8	46	42.9	49.8	4.1	0.08	0.762
Morphological awareness	34	82.4	46.5	46	71.0	46.5	11.4	0.26	0.235
Vocabulary	34	76.5	47.4	46	67.7	47.4	8.7	0.19	0.448
Sentence processing	34	73.5	47.4	46	64.8	47.4	8.7	0.19	0.434
Efficiency of reading for comprehension	34	73.5	50.6	46	58.0	50.6	15.5	0.32	0.203
Reading comprehension	34	67.6	50.5	46	56.1	50.5	11.5	0.23	0.347
District 2									
Word recognition and decoding	15	66.7	50.0	9	64.1	50.0	2.6	0.05	0.883
Morphological awareness	15	80.0	0.0	9	99.0	0.0	-19.0	-0.58	0.352
Vocabulary	15	93.3	50.0	9	73.4	50.0	20.0	0.55	0.168
Sentence processing	15	73.3	44.1	9	76.5	44.1	-3.2	-0.07	0.900
Efficiency of reading for comprehension	15	53.3	50.0	9	69.1	50.0	-15.8	-0.31	0.499
Reading comprehension	15	53.3	52.7	9	72.9	52.7	-19.6	-0.38	0.271
Pooled									
Word recognition and decoding	49	53.1	50.3	55	48.8	50.3	4.2	0.08	0.720
Morphological awareness	49	81.6	44.0	55	77.2	44.0	4.4	0.11	0.624
Vocabulary	49	81.6	47.4	55	70.8	47.4	10.8	0.25	0.279
Sentence processing	49	73.5	46.6	55	68.0	46.6	5.4	0.12	0.583
Efficiency of reading for comprehension	49	67.3	50.4	55	57.9	50.4	9.5	0.19	0.389
Reading comprehension	49	63.3	50.5	55	57.3	50.5	6.0	0.12	0.583

Notes: N = Number of students. SD = Standard Deviation. Adj. Mean = Adjusted Mean. The results in this table are based on students in the ReadBasix sample. The means for the control group, as well as the difference between the STARI and control group, are regression-adjusted for the blocking of random assignment and for students' baseline ELA/reading test scores and other characteristics. Effect sizes are calculated by dividing the estimated difference between the STARI group and the control group by the pooled within-group standard deviation of the measure for students in the analysis sample. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. [a]The estimated STARI-control differences in this table do not differ by a statistically significant amount across school districts.

Exhibit F.4. Minimum Detectable Effect Size (MDES) and Standard Errors

Domain: Outcome	District 1				District 2				Pooled			
	N	SE	MDE	MDES	N	SE	MDE	MDES	N	SE	MDE	MDES
General reading proficiency												
ELA state test score (z-score)	154	0.17	0.43	0.43	46	0.21	0.53	0.54	200	0.14	0.35	0.36
ELA state test proficiency (%)	154	3.8	9.4	0.7	46	12.4	31.4	0.62	200	4.0	9.9	0.31
ELA course performance												
ELA course grade (%)	156	1.2	3.1	0.36	46	3.2	8.2	0.75	202	1.2	2.9	0.30
ELA courses passed (%)	156	7.9	19.6	0.39	46	6.1	15.6	0.93	202	6.3	15.7	0.33
Other course performance												
Social studies course grade (0-100%)	156	1.2	3.0	0.35	30	5.1	13.3	1.19	186	1.2	3.0	0.32
Math course grade (0-100%)	156	1.2	2.9	0.38	46	3.8	9.8	0.72	202	1.2	3.1	0.32
Science course grade (0-100%)	156	1.1	2.8	0.22	46	3.7	9.3	0.79	202	1.1	2.8	0.23
Reading skills, vocabulary, and comprehension												
Word recognition and decoding	80	3.1	7.9	0.61	24	5.3	14.2	1.00	104	2.8	7.0	0.51
Morphological awareness	80	2.2	5.6	0.58	24	4.1	11.0	1.43	104	2.0	5.0	0.50
Vocabulary	80	1.6	4.0	0.57	24	2.0	5.3	1.01	104	1.4	3.4	0.50
Sentence processing score	80	2.9	7.2	0.62	24	4.7	12.6	1.33	104	2.5	6.2	0.56
Efficiency of basic reading comp.	80	2.7	6.8	0.59	24	8.0	21.3	1.26	104	2.6	6.6	0.51
Reading comprehension score	80	2.0	5.1	0.60	24	4.4	11.7	1.01	104	1.9	4.8	0.52
Reading self-efficacy and motivation												
Reading self-efficacy (1-4)	81	0.2	0.5	0.62	24	0.4	1.1	1.70	105	0.2	0.4	0.58
Reading motivation (1-4)	81	0.2	0.4	0.62	24	0.5	1.2	1.82	105	0.1	0.4	0.56

Notes: N = Number of students. SD = Standard Deviation. SE=Standard error. MDE=Minimum detectable effect. MDES = Minimum detectable effect size. The MDE is based on a 5 percent statistical significance level and 80 percent power, using the actual standard error of the estimated difference between STARI and control group students. The MDES is the MDE divided by the pooled within-group standard deviation of the measure for students in the analysis sample.

Exhibit F.5. Correlation between Baseline Reading Scores and Student Outcomes

	District 1		District 2		Pooled	
	N	Corr.	N	Corr.	N	Corr.
General reading proficiency						
ELA state test score (z-score)	154	0.33	46	0.58	200	0.38
ELA course performance						
ELA course grade (%)	156	0.32	46	0.31	202	0.30
Reading skills, vocabulary, and comprehension						
Word recognition and decoding	80	0.27	24	0.38	104	0.29
Morphological awareness	80	0.40	24	0.44	104	0.40
Vocabulary	80	0.42	24	0.36	104	0.41
Sentence processing score	80	0.23	24	0.02	104	0.20
Efficiency of basic reading comprehension	80	0.29	24	0.23	104	0.27
Reading comprehension score	80	0.24	24	0.59	104	0.33
Reading self-efficacy and motivation						
Reading self-efficacy (1-4)	81	0.14	24	-0.05	105	0.11
Reading motivation (1-4)	81	0.11	24	-0.01	105	0.08

Notes: N = Number of students. Corr. = Correlation. Reading achievement at baseline was measured using a combination of ELA state test scores and district formative reading test scores pre-dating random assignment, which were standardized (z-scored) by district and assessment and grade level to allow for pooling.

Exhibit F.6. Correlation between ReadBasix Subtests, ReadBasix Sample

	WR	MA	VO	SP	ER	RC
District 1 (n=80 students)						
Word recognition and decoding (WR)	1.00					
Morphological awareness (MA)	0.57	1.00				
Vocabulary (VO)	0.34	0.64	1.00			
Sentence processing (SP)	0.36	0.53	0.42	1.00		
Efficiency of reading for comprehension (ER)	0.32	0.51	0.32	0.41	1.00	
Reading comprehension (RC)	0.17	0.29	0.20	0.20	0.48	1.00
District 2 (n=24 students)						
Word recognition and decoding (WR)	1.00					
Morphological awareness (MA)	0.56	1.00				
Vocabulary (VO)	0.65	0.38	1.00			
Sentence processing (SP)	0.14	0.51	0.03	1.00		
Efficiency of reading for comprehension (ER)	0.15	0.46	0.20	0.52	1.00	
Reading comprehension (RC)	0.26	0.67	0.19	0.50	0.58	1.00
Pooled (n=104 students)						
Word recognition and decoding (WR)	1.00					
Morphological awareness (MA)	0.60	1.00				
Vocabulary (VO)	0.44	0.64	1.00			
Sentence processing (SP)	0.31	0.51	0.36	1.00		
Efficiency of reading for comprehension (ER)	0.29	0.50	0.31	0.43	1.00	
Reading comprehension (RC)	0.21	0.37	0.20	0.27	0.52	1.00

Notes: The correlations in this table are based on students in the ReadBasix sample.

Exhibit F.7. Correlation between Reading Self-Efficacy and Reading Motivation, Survey Sample

	District 1	District 2	Pooled
Correlation	0.40	0.62	0.39
Number of Students	81	24	105

Notes: The correlations in this table are based on students in the student survey sample.

APPENDIX G. DEVIATIONS FROM PRE-REGISTERED ANALYSIS PLAN

Our analysis plan was pre-registered in the Registry of Efficacy and Effectiveness Studies (Study #1789, https://sreereg.icpsr.umich.edu/sreereg/). Due to unanticipated limitations in data availability due to the COVID-19 pandemic, the following changes were made to the analytic strategy:

- Pooled analysis: The pre-analysis plan specified that the analysis would be based on students in both study districts pooled together. However, because the strength of the evidence (internal validity and sample size) varies across districts, the findings in this report are presented separately by district, and the pooled findings are included in the appendices. The sample size for District 1 is much larger than for District 2, so the pooled findings are similar to the findings for District 1.
- Baseline reading achievement measure: The analysis plan pre-specified that the baseline measure of reading achievement would be students' ELA state test score in spring 2021, before random assignment. This plan was followed for District 1. However, in District 2, many students did not have a baseline ELA state test score because state tests were not administered consistently in that school year due to the pandemic. For students in District 2 without a state test score, their baseline reading achievement was measured using their score on the district formative reading test from Fall 2021 (which pre-dates random assignment, because STARI classes in District 2 did not begin until later in the fall). Baseline reading scores were z-scored by district, by assessment type, and by grade level to allow for pooling.
- Impact model covariates: When estimating impacts on course performance (e.g., ELA course grades), the analysis plan pre-specified that in addition to controlling for student

characteristics and students' prior ELA test scores, the model would also include a measure of that course outcome from the prior school year (SY 2020-21). However, this covariate was not included because instruction during most of SY 2020-21 was virtual or hybrid, and therefore students' grades are not considered sufficiently reliable or valid to use as a baseline covariate.

- Student outcome domains: At the time the analysis plan was registered, the What Works Clearinghouse reviewed middle school reading interventions using a protocol developed for adolescent literacy interventions (WWC, 2016). Since then, the What Works Clearinghouse has developed a review protocol for interventions in grades 4-9 that are specifically targeted at struggling readers (WWC, 2020).[49] The name of the key student outcome domains differs across protocols. This report uses the names from the most recently developed protocol for struggling readers. For example, in this report "general reading proficiency" is used instead of "general literacy achievement" for the domain represented by ELA state test scores. However, in the analysis plan the latter term had been used.
- Confirmatory outcome measures: To reduce the risk associated with multiple hypothesis testing and false positives (Type I error), the analysis plan had pre-specified one confirmatory outcome for each of the student outcomes domains in the WWC adolescent literacy review protocol (WWC, 2016): ELA state test scores (general literacy achievement domain); ELA course grades (course performance domain); word recognition subtest scores (alphabetics domain); efficiency of reading subtest scores (reading fluency domain); and scores on a composite measure of comprehension, created

[49] The evaluation of STARI in Massachusetts has been reviewed under both protocols.

by averaging a student's scores on the sentence processing, vocabulary, and reading comprehension subtests (comprehension domain). Because estimated effects on the ReadBasix subtests are not statistically significant, STARI's effects on the composite measure of comprehension are not presented in this report because there is no need to account for false positives. Also, as noted in the previous bullet, in this report student outcome domains are referred to by their name in the recently developed WWC protocol for reading interventions for struggling readers (WWC, 2020).

- General reading proficiency confirmatory measure: For the general reading proficiency (literacy achievement) domain, the analysis plan had pre-specified that the confirmatory measure would be students' score on the Spring 2022 formative reading assessments used by the school district to determine Tier 2 intervention placement. However, in SY 2021-22, only a small number of students took these formative tests due to pandemic-related limitations on student testing (e.g., in District 1, only 3 of 6 schools administered these tests). For these reasons, the measure of general reading proficiency used in this report is students' score on the state ELA test in Spring 2022.
- Students in Cohort 1: As noted in this report, the first year of implementation (Year 1, SY 2018-19) was intended to be a "training year" and was not intended to be part of the impact study. However, to help familiarize schools with the random assignment process, lotteries of eligible students were conducted in summer/fall 2018. Because Year 2 and Year 3 (the original impact study years) were affected by the pandemic, in the pre-registered analysis plan it was noted that the study team would explore the viability of including Cohort 1 students in the impact analysis. To do so, the study team obtained course records for SY 2018-19 to determine whether Cohort 1 students were correctly

scheduled into STARI classes based on the lottery results that year. After reviewing these data, it was decided to *not* include Cohort 1 in the analysis for two reasons. First, the STARI class is not called "STARI" in school schedules and course records, so determining whether students were enrolled in the STARI class based on their course records requires knowing not only the *name* of the STARI class, but also the *period* in which it is taught and/or the section (because the other reading interventions offered in the study schools can have the same name as the STARI class). Information about the STARI course name and period/section were not available for all schools in Year 1 because, as noted earlier, the study team had not intended to include this cohort in the impact analysis. Second, even when STARI course information was available or could be inferred, schools do not appear to have followed the lottery results (all or almost all eligible students received STARI). In SY 2018-19, the study team did not monitor whether students were correctly scheduled into STARI classes or intervene if the lottery results were not being followed.

A restricted use file for this study will be archived at the Inter-university Consortium for Political and Social Research (ICPSR) https://www.icpsr.umich.edu/web/pages/.

APPENDIX H. IMPLEMENTATION FINDINGS IN YEAR 1

This appendix presents the findings from an analysis of STARI's implementation in the 2018-19 school year, which was the first year of implementation ("Year 1") in the study schools. The evaluation was originally intended to also include an analysis of implementation in the second and third years as well, but this was not possible due to the COVID-19 pandemic. Similarly, a detailed implementation analysis was not conducted in the fourth year of implementation (Year 4; 2021-22 school year), in order to ensure that sufficient resources would be available for evaluating STARI's effects in that additional (unplanned) study year.

Analytic Approach

*Implementation fidelity***.** The implementation study examined the fidelity with which three core STARI components (see Exhibit 1) were implemented in Year 1: (1) teacher training and professional development, (2) coach training and professional development, and (3) the STARI curriculum materials. Fidelity was measured based on 7 indicators representing the intended delivery of these three components. The data sources included training attendance logs, coaching logs, coach meeting notes, and STARI material delivery trackers. Each STARI teacher or coach (depending on the component) received a score of 0-1 on each indicator, and scores were then averaged by component and used to determine whether STARI was implemented with fidelity based on pre-determined thresholds agreed upon with SERP.

The sample for the fidelity analysis includes schools from the four school districts that participated in the study in Year 1. Across all districts, the sample includes 51 teachers that

implemented the STARI curriculum[50] across 30 schools[51] as well as the 7 STARI coaches that supported the teachers. Fidelity was examined by district, on average across all four Year 1 districts, and on average across the two school districts that went on to participate in the Year 4 impact study (District 1 and 2).

STARI coaching. To supplement the findings from the fidelity analysis, the Year 1 implementation study also explored the types of coaching received by teachers to support their STARI instruction, using information from coach logs completed by STARI coaches, as well as interviews with STARI coaches.

Implementation Fidelity

This section presents the Year 1 fidelity findings by STARI component. Where relevant, findings from the coach interviews are used to shed insights on factors that may have effected fidelity in Year 1.

Component 1: Teacher Training and Professional Development (PD). Across the four study districts, the average score for this component in Year 1 was 0.50 (out of 1), with 15.7% of STARI teachers receiving an adequate amount of training (Exhibit H.1). The highest overall score was in District 1 (0.587) and lowest in District 2 (0.441). Fidelity to this component was measured using four indicators: summer training attendance, participation in online guidance sessions, attendance at professional learning communities (PLCs), and the number of coaching sessions received.[52]

[50] There were in fact 59 STARI teachers in Year 1. However, 8 of these teachers were replacements for STARI teachers that left during the year. These "teacher pairs" (the teacher that left and their replacement) were counted as the same teacher for the purposes of measuring implementation fidelity across the school year, hence there are 51 full year "teachers" in the teacher sample.
[51] 10 schools in District 1, 10 schools in District 2, 6 schools in District 3 and 4 schools in District 4.
[52] The number of coaching sessions is based on the coach logs. These logs allow coaches to track their contacts with teachers, including contacts focused on logistics. Only coaching sessions that were

Exhibit H.1. Teacher Professional Development and Training: Fidelity Scores for Year 1

District	Number of Teachers	Average Score by Indicator and Across Indicators (0-1)					% of teachers meeting adequate fidelity threshold (score >=0.83)
		Summer training attendance[a]	Online guidance session partici-pation[b]	In-person PLC attendance[c]	Coaching sessions received[d]	Average across all indicators	
***District 1**	16	0.749	0.644	0.603	0.351	0.587	25.0%
***District 2**	19	0.772	0.208	0.278	0.505	0.441	0.0%
District 3	10	0.399	0.365	0.231	0.866	0.465	20.0%
District 4	6	0.443	0.442	0.495	0.720	0.525	33.3%
All	**51**	**0.653**	**0.403**	**0.396**	**0.553**	**0.501**	**15.7%**
Year 4 districts	**35**	**0.761**	**0.408**	**0.426**	**0.435**	**0.508**	**11.4%**

Notes: *Districts in the Year 4 impact study.

[a]Teacher scores for this indicator are based on teacher attendance at a 3 day summer institute training, where a score of 0 (low) was assigned to teachers who did not attend any training, a score of .33 (low-moderate) was assigned to teachers who attended 1 day at institute or attended shorter makeup training (½-1 day), a score of .66 (moderate-high) was assigned to teachers who attended 2 days at institute, and a score of 1 (high) was assigned to teachers who attended 3 days at institute.

[b]Teacher scores for this indicator are based on completed online guidance sessions that occurred throughout the year, where a score of 0 (low) was assigned to teachers who did not take any sessions, a score of .33 (low-moderate) was assigned to teachers who completed 1-2 sessions, a score of .66 (moderate-high) was assigned to teachers who completed 3-4 sessions, and a score of 1 (high) was assigned to teachers who completed 5-6 sessions (out of a total of 6).

[c]Teacher scores for this indicator are based on attendance at 7 PLCs that occurred throughout the school year, where a score of 0 (low) was assigned to teachers who did not attend any PLCs, a score of .33 (low-moderate) was assigned to teachers who attended 2-3 PLCs, a score of .66 (moderate-high) was assigned to teachers who attended 4-5 PLCs, and a score of 1 (high) was assigned to teachers who attended 6-7 PLCs.

[d] Teacher scores for this indicator are based on the number of coaching sessions that each coach logged for each individual teacher, where a score of 0 (low) was assigned to teachers who received <7 coaching sessions, a score of .33 (low-moderate) was assigned to teachers who received 7 - 13 sessions, a score of .66 (moderate-high) was assigned to teachers who received 14 – 20 sessions, and a score of 1 (high) was assigned to teachers who received >21 sessions.

meaningfully focused on instruction were counted for the fidelity indicator. (Not counted are contacts where the mode of coaching was phone or email and where the focus of the contact centered on logistics or scheduling.) About 67% of coach contacts met the inclusion criterion for being counted as a coaching session.

There was variation across the study districts with respect to the types of PD that teachers received. In District 1, attendance at the summer institute and school-year trainings (PLCs, online guidance sessions) was quite high (> 0.5), but teachers received fewer coaching sessions (< 0.35). However, the latter result could be due to underreporting of coaching sessions by coaches in that district; in an end-of-year survey of STARI teachers conducted by SERP, teachers reported receiving coaching 1-2 times every two weeks on average, as expected. In District 2, 3 and 4, teachers received a substantial amount of coaching (score > 0.5), but their participation in online guidance sessions and PLCs was lower (scores < 0.5), particularly in District 2 (score = 0.21).

Interviews with STARI coaches provided additional insight into these findings. With respect to PLC attendance, coaches noted that teachers had to travel to different school locations to attend the meetings, which was especially relevant in District 2. Coaches also noted that many STARI teachers had conflicting responsibilities that made it challenging for them to consistently attend PLCs after school. Sometimes teachers were not able to get to PLC meetings on time, which also created pacing challenges for the coaches in their delivery of PLCs. Finally, coaches noted that the scope and sequence of the PLCs and online guidance sessions were not always well aligned with teachers' specific needs with respect to the curriculum, which made it more challenging for coaches to use the PLCs to address specific issues that teachers were facing.

Component 2: Coach Training and Professional Development (PD). On average across the four Year 1 study districts, 100% of STARI coaches received adequate training (Exhibit H.2). Fidelity for this component was measured using 2 indicators: attendance at the coach summer training institute, and coaches' participation in regular check-ins with SERP.

Exhibit H.2. Coach Professional Development and Training: Fidelity Scores for Year 1

District	Number of Coaches	Average Score by Indicator and Across Indicators (0-1)			% of coaches meeting adequate fidelity threshold (score >=0.83)
		Coach summer training attendance[a]	Participation in regular check-ins with SERP[b]	Average across all indicators	
***District 1**	2	1.000	1.000	1.000	100%
***District 2**	2	1.000	1.000	1.000	100%
District 3	2	1.000	1.000	1.000	100%
District 4	1	1.000	1.000	1.000	100%
All	**7**	**1.000**	**1.000**	**1.000**	**1.000**
Year 4 districts	**4**	**1.000**	**1.000**	**1.000**	**1.000**

Notes: *Districts in the Year 4 impact study.

[a]Coach scores for this indicator are based on coach attendance at a 4 day summer institute training, where a score of 0 (low) was assigned to coaches who did not attend any training, a score of .33 (low-moderate) was assigned to coaches who attended 1-2 days at institute or attended a shorter makeup training, a score of .66 (moderate-high) was assigned to coaches who attended 3 days at institute, and a score of 1 (high) was assigned to coaches who attended 4 days at institute.

[b]Coach scores for this indicator are based on attendance at check-in calls with SERP that occurred throughout the year, where a score of 0 (low) was assigned to coaches who attended less than 25% of calls, a score of .33 (low-moderate) was assigned to coaches who attended 26-50% of calls, a score of .66 (moderate-high) was assigned coaches who attended 51-75% of calls, and a score of 1 (high) was assigned to coaches who attended more than 75% of calls.

Component 3: STARI curriculum materials. Across all four study districts, 88% of teachers received their STARI materials in an adequate time frame in Year 1 (Exhibit H.3). In the two districts that went on to participate in the Year 4 impact study, all teachers (100%) received their materials on time. Fidelity for this component was measured using one indicator: whether teachers received the STARI materials (lesson plans, student workbooks, reading library) before or during the first week of STARI classes.

Exhibit H.3. Curriculum Materials: Fidelity Scores for SY 2018-19

District	Number of teachers with delivery date information	Materials delivery (average score, 0-1)[a]	% of teachers meeting adequate fidelity threshold (score=1)
***District 1**	16 (of 16)	1.000	100%
***District 2**	19 (of 19)	1.000	100%

District	Number of teachers with delivery date information	Materials delivery (average score, 0-1)[a]	% of teachers meeting adequate fidelity threshold (score=1)
District 3	4 (of 10)	0.250	25%
District 4	3 (of 6)	0.333	33%
All	**42 (of 51)**	**0.881**	**88%**
Year 4 districts	**35 (of 35)**	**1.000**	**100%**

Notes: *Districts in the Year 4 impact study. These findings are based on only 42 teachers (out of 51) because data on delivery dates were unavailable for some teachers in Districts 3 and 4.

[a]Teacher scores for this indicator are based on whether teachers received materials on time, where a score of 0 (low) was assigned to teachers who received the materials more than 2 weeks after the start of STARI classes, a score of .5 (moderate) was assigned to teachers who received the materials the second week of STARI classes, and a score of 1 (high) was assigned to teachers who received materials within the first week of classes.

STARI Coaching

As a supplement to the fidelity analysis, the Year 1 implementation study also used data from coach logs to better understand the coaching that teachers received to support their STARI instruction. Exhibits H.4 and H.5 summarize the topics covered during these coaching sessions and the strategies used by coaches to support teachers. Where relevant, findings from the coach interviews are used to provide additional context on teachers' experience implementing STARI.

Across the four Year 1 study districts, coaches met an average of about 9 times with each teacher, or 5 times per teacher in the two districts that participated in the Year 4 impact study. Some coaches filled out their coaching logs retrospectively rather than in "real time", which may have contributed to fewer reported coaching sessions (coaches may not have been able to recall all their coaching sessions).

Focal Areas of the Coaching. Across the Year 1 study districts, the most common topics covered during the coaching sessions were fluency routines, partner work, and classroom management (Exhibit H.4). The coach interviews confirm these results and provide further context for why teachers needed more support from coaches in these areas:

Exhibit H.4. Focus of the Coaching Sessions in Year 1

Focus of Coaching	*District 1	*District 2	District 3	District 4	All districts	Year 4 districts
Fluency Routines	49%	41%	34%	39%	39%	43%
Partner Work	20%	26%	30%	12%	25%	24%
Classroom Management	18%	22%	24%	13%	21%	21%
Guided Reading	18%	14%	26%	12%	19%	15%
Discussion	9%	14%	15%	6%	13%	12%
Word Study	11%	11%	8%	16%	10%	11%
Reading Comprehension Strategies	9%	6%	10%	12%	9%	7%
# of coach interactions	65	138	197	69	469	203
Average # of coach contacts per teacher	4.06	7.26	19.70	11.50	9.20	5.08

Notes: *Districts in the Year 4 impact study. The percentages do not sum to 100% across topics, because a coaching session can cover more than one topic.

- Fluency and partner work: Understanding how to do the fluency routines was an area where coaches reported that teachers needed a lot of support. This may have been partly due to STARI teachers' backgrounds and expertise: coaches noted that just being an ELA teacher (as opposed to a reading specialist) does not necessarily equip someone with the content knowledge needed to teach reading skills like fluency. This may be especially true for middle school ELA teachers, whose expertise is more geared towards helping students conduct deep text analysis as opposed to teaching the fundamentals of reading. The fact that teachers struggled with how to teach fluency routines impeded their ability to set up students successfully for partner work, since the fluency routines are intended to happen with students working in pairs. One coach noted that they often saw teachers teaching fluency routines to the whole group instead of having students work in pairs or small groups.
- Classroom management: In District 2 and 3 especially, classroom management was another area where coaches noted that teachers needed extra help. Coaches discussed

how classroom management was particularly challenging for newer teachers and teachers who were not used to teaching large classes. While this is not unique to STARI, it may have implications for identifying which teachers are better equipped to teach STARI successfully.

Coaching Strategies used with STARI Teachers. Across all Year 1 study districts, the most common strategies used by coaches were observations, feedback on observations, and co-planning (Exhibit H.5). The least common strategies were classroom videos, and modeling or demonstrating teaching. These findings are generally consistent across districts. In interviews, coaches reported that because there is no formal coaching model for STARI, they relied on their own background and training or personal coaching philosophies when deciding what types of supports to offer teachers. Despite this flexibility to decide what to focus on, the specific strategies used by coaches were still consistent across districts.

Exhibit H.5. Types of Coaching Sessions in Year 1

Types of Coaching	*District 1	*District 2	District 3	District 4	All districts	Year 4 districts
Observation	34%	62%	63%	54%	58%	53%
Feedback on observation	22%	57%	47%	33%	45%	46%
Co-planning	43%	32%	56%	36%	44%	35%
Co-teaching	6%	29%	16%	0%	16%	22%
Student Work	12%	18%	7%	1%	10%	16%
Modeling	3%	10%	5%	4%	6%	8%
Demonstration Teaching	3%	10%	5%	3%	6%	8%
Classroom Video	2%	0%	1%	0%	0%	0%
# of coach interactions	65	138	197	69	469	203
Average # of coach contacts per teacher	4.06	7.26	19.70	11.50	9.20	5.80

Notes: *Districts in the Year 4 impact study. The percentages do not sum to 100% across coaching strategies, because a coach can use different strategies during the same session.

APPENDIX I. SCALE-UP GOALS

Exhibit I.1 provides information on the five scale-up goals for this grant, which were set at the start of the project and achieved by the end of the project.

Exhibit I.2. Scale-Up Goals by the End of the Grant

Scale-up Goal	Minimum Threshold	Actual number	Goal met?
84 teachers are trained to implement STARI	Same as goal	93	Yes
District literacy coaches are trained as STARI coaches	At least 75% of literacy coaches are trained to support STARI	92%	Yes
3,444 students are served	Same as goal	3,945	Yes
STARI readiness checklist is developed	Same as goal	N/A	Yes
STARI video for administrators is developed	Same as goal	N/A	Yes

Notes: Information about the number of teachers trained and the number of students served comes from SERP annual reports to the US Department of Education. Information about the STARI readiness checklist and administrator video are from communications with SERP.

Please note that the threshold for the first goal was adjusted due to the COVID-19 pandemic. The minimum threshold for this goal had originally been defined based on the number of teachers who were *adequately trained* (i.e., teachers participating in the trainings and professional development at a high level). However, due to changes in the project plan caused by the pandemic, data on training participation were not collected across all project years. Therefore, the minimum threshold for this goal was revised and is based on the number of teachers trained (not just "adequately trained"). The minimum threshold was revised from 63 "adequately trained" teachers to 84 teachers trained.

APPENDIX J. COST PER STUDENT

This appendix examines the cost per student of STARI during the project implementation period. It is important to note that this does not represent the cost of implementing STARI in a non-study setting, due to multiple factors including the fact that the implementation plan was altered by the COVID-19 pandemic.

Reporting Periods

Data on total expenditures and the number of students served during the grant were obtained from SERP's annual grant reports to the US Department of Education for each project year. The project reporting periods start in October and end in September of the following year, and therefore each project year spans two implementation years:

- The first project year (October 2017 to September 2018) was a planning year that included recruiting districts and schools and hiring and training STARI coaches. The first project year also included summer and fall activities for **Implementation Year 1**, like the summer training for teachers, the printing of materials, and the purchase of reading libraries in fall 2018.
- The second project year (October 2018 to September 2019) included coaching and other supports for STARI teachers during **Implementation Year 1**, as well as summer and fall activities for **Implementation Year 2** (summer training for teachers and district coaches, and STARI materials for Year 2).
- The third project year (October 2019 to September 2020) included coaching and other supports for STARI teachers during **Implementation Year 2**, which was disrupted by school closures in March 2020 due to the pandemic. The third project year also included

activities in preparation for virtual learning in **Implementation Year 3**, like converting STARI's materials to PDF and creating an online PLC for teachers.

- The fourth project year (October 2020 to September 2021) included coaching and other supports for STARI teachers during **Implementation Year 3**. As noted earlier, fewer schools implemented STARI in Year 3 and instruction was virtual or hybrid in all schools. The fourth project year also included summer and fall implementation activities for **Implementation Year 4** (e.g., providing STARI materials for Year 4).
- The last project period (October 2021 to June 2023) included coaching and other supports for STARI teachers during **Implementation Year 4**, as well as dissemination and reporting activities. Expenditures for the last project period were not available at the time of this report and are not included in the calculations presented in this appendix.

Measures and Calculations

Expenditures. The expenditures reported include all project expenditures (e.g., salaries for a SERP-hired coach and SERP staff, the cost of teacher trainings and stipends, and the cost of printing and delivering STARI materials, etc.) minus the costs associated with the independent evaluation.

Students served. The number of students served in each project year is based on the number of students for whom materials were ordered at the end of the project period. For example, the total number of students served in the third project year is based on the number of students for whom materials were ordered in the fall of Implementation Year 3.

Findings

Exhibit J.1 shows total expenditures and students served, overall and by project year. Expenditures were lowest in the first project year because the first several months were used for

planning, although as noted earlier, the first project year did include the summer training and fall activities for Implementation Year 1. Expenditures were highest in the second project year because this period includes school year supports for Year 1, as well as summer/fall implementation activities for Implementation Year 2. Expenditures decreased in the third and fourth project years because fewer schools were able to implement STARI due to the COVID-19 pandemic.

Across all project years, the cost per student was $1,333, but this number masks variation across projects years due to the pandemic. In the first and second project years, the cost per student was $840 and $857, respectively. In the third and fourth project years, the cost per student was higher, $3,649 and $2,663 respectively, because fewer schools and students were served due to the pandemic, yet resources still had to be expended to support schools and adapt STARI materials for virtual learning and teacher training.

In a non-pandemic context, the number of students served in Implementation Year 3 and Year 4 would have been higher, which would have resulted in a lower cost per student in later project years and overall.

Exhibit J.1. Expenditures and Cost per Student, by Project Year

Project Year	STARI implementation years	Project activities	Total expenditures	# students served	Cost per student
October 2017 to September 2018 (Planning and summer/fall activities for Implementation Year 1)	1	• Recruit districts and schools • Hire and train coaches • Summer teacher training • Provide materials for Implementation Year 1	$882,442	1050 (Fall of Implementation Year 1)	$840
October 2018 to September 2019 (Implementation Year 1 and summer/fall activities for Year 2)	1 and 2	• Teacher professional learning and coaching for Implementation Year 1 • Summer training for district coaches • Summer training for teachers • Provide materials for Implementation Year 2	$1,771,089	2066 (Fall of Implementation Year 2)	$857
October 2019 to September 2020 (Implementation Year 2 and summer/fall activities for Year 3)	2 and 3	• Teacher professional learning and coaching for Implementation Year 2 • Adapt STARI materials for online delivery • Create online PLC teacher training series • Provide materials for Implementation Year 3	$1,474,052	404 (Fall of Implementation Year 3)	$3,649

Project Year	STARI implementation years	Project activities	Total expenditures	# students served	Cost per student
October 2020 to September 2021 (Implementation Year 3 and summer/fall activities for Year 4)	3 and 4	• Teacher professional learning and coaching for Implementation Year 3 • Provide materials for Implementation Year 4	$1,131,875	423 (Fall of Implementation Year 4)	$2,663
Total (all project years)	**All**	**All**	**$5,259,459**	**3945**	**$1,333**

Notes: Total expenditures do not include evaluation costs. The number of students in each project year is based on the number of students for whom materials were ordered at the end of the project period. Expenditures for the last project year (October 2021 to June 2023) were not available at the time of this report.