



Evaluation of Accelerating Literacy through Self-Regulation, Collaboration, and Persistence: Reading Apprenticeship in North Carolina | Final Report

September 2026

Prepared for:

Heather Howlett
Director of Professional Learning
WestEd
hhowlet@wested.org

Prepared by:

Sara Rutherford-Quach
Rebecca Schmidt
Nonye Alozie
Madeline Coole
Santiago Navia Jaramillo
Kaily Yee



September 2026

Acknowledgments:

The findings reported here draw from the efforts of a larger team over nearly six years. The authors thank Nicole Arshan for her early leadership of the student as well as Candice Benge, Bladimir Lopez-Prado, Mary McCracken, Kerry Schellenberger, and Audra Wingard for their contributions to the evaluation. They also thank Sophie Burgess, Charles Harding, Roxanne Jones, Katie Roberts, and Katie Woodworth for their help with this report. The researchers would also like to thank WestEd's Reading Apprenticeship Project Team: Heather Howlett, Linda Friedrich, Diane Lee, Misty Sailors, Laurie Erby, William Lloyd and Alicia Ross. Your engagement and feedback in the evaluation contributed to a great partnership. Not least, the authors are grateful to all teachers and leaders who participated in the evaluation.

This report was prepared by SRI with funds provided by WestEd under an Education Innovation and Research (EIR) Mid-phase grant from the U.S. Department of Education (#S411B200016). The views expressed in this report do not necessarily reflect the views or policies of the U.S. Department of Education.

Suggested citation:

Rutherford-Quach, S., Schmidt, R. A., Alozie, N., Coole, M., Navia Jaramillo, S., & Yee, K. (2026). *Evaluation of accelerating literacy through self-regulation, collaboration, and persistence: Reading Apprenticeship in North Carolina. Final report.* SRI.

Contents

List of Exhibits	v
Abstract	1
Introduction	2
Overview of Reading Apprenticeship in North Carolina	4
Project Description.....	4
Logic Model	5
Mediators.....	7
Outcomes	7
Contextual Moderators.....	8
Evaluation Design and Methods.....	10
Study Districts, Schools, and Teachers	10
Research Questions	11
Implementation	11
Descriptive Outcomes.....	11
Impact Outcomes	12
Sustainability and Scaling	12
Cost Effectiveness	12
Data Sources.....	12
Independence of Evaluation	14
Implementation	15
Fidelity Measures	15
Fidelity Findings	19
Component 1: Participation in and Quality of Teacher Professional Learning.....	20
Component 2: Leadership Training for Sustainability.....	23
Component 3: Classroom Implementation of Core RA Practices	23
Facilitators and Barriers	24
Outcomes	29
Descriptive Outcomes	29
Methods.....	29
Findings	30
Impact Outcomes	32

Preregistration of the Study Design	32
Samples.....	32
Study Design.....	34
Impact Findings.....	38
Sustainability and Scaling.....	42
Strategy to Sustain and Scale	42
Sustainability and Scaling Findings	43
Cost-Effectiveness	46
Conclusion	47
References	49
Appendix A: Implementation	A-1
Appendix B. Descriptive Outcomes	B-1
Appendix C. Impact Outcomes	C-1

List of Exhibits

Exhibit	Page
Exhibit 1. Reading Apprenticeship in North Carolina Logic Model	9
Exhibit 2. Characteristics of Cohorts	11
Exhibit 3. Survey Response Rates by Cohort and School Year	13
Exhibit 4. Fidelity of Implementation Scoring System for Each Key Component in the Logic Model.....	17
Exhibit 5. Number of Districts, Schools, and Teachers in Implementation Study	19
Exhibit 6. Number of Districts Meeting Fidelity for Each Component	20
Exhibit 7. Districts Meeting Fidelity of Teacher Professional Learning.....	21
Exhibit 8. Inclusion of Key Activities in Teacher Professional Learning by Cohort and Implementation Year	22
Exhibit 9. Fidelity of Leadership Training.....	23
Exhibit 10. Fidelity of Classroom Implementation	24
Exhibit 11. Teacher-Reported Frequency with Which Students Practiced Self-Regulation, Collaboration, and Persistence in the Classroom	30
Exhibit 12. One-Year Impact Treatment Group Sample	33
Exhibit 13. Two-Year Impact Treatment Sample	34
Exhibit 14. Variables Used for Weighting and Matching and as Model Controls	35
Exhibit 15. Standardized Baseline Differences Between Treatment (T) and Comparison (C) Students	37
Exhibit 16. Impact of One Year of RA on Student Outcomes	39
Exhibit 17. Estimated Science Course Grades in Grade 9 and 10, for Treatment and Comparison Students	40
Exhibit 18. Impact of Two Years of RA on Student Outcomes	40
Exhibit 19. Predicted Science Course Grades from Two-Year Impact Models, for Treatment and Comparison Students.....	41
Exhibit 20. Cost of RA Implementation by Year	46
Exhibit 21. Student Impacts and Cost per Student per Year of RA	46

Abstract

Nationally, adolescent reading performance has continued to decline, even as younger students have begun to regain lost ground. Concern about adolescent reading performance is particularly acute in North Carolina, which has one of the nation's largest rural/nonrural gaps, with rural districts reporting lower grade 8 English language arts (ELA) scores and five-year graduation rates than nonrural districts.

To address these disparities, in 2021, WestEd received a six-year Education Innovation and Research (EIR) Mid-phase grant (which included a one-year extension) to implement Reading Apprenticeship in rural North Carolina school districts. The grant supported this expansion of Reading Apprenticeship, which provided sustained engagement in professional learning, coaching, and other supports to help grade 9 and 10 teachers integrate reading, writing and discussion into ELA, science, and social studies instruction. SRI conducted an independent evaluation of the project's implementation and effects on student perseverance and achievement.

This final report describes the implementation, descriptive, and impact studies of Accelerating Literacy Through Self-Regulation, Collaboration, and Persistence: Reading Apprenticeship in North Carolina, which was implemented in 19 North Carolina districts and expanded to 6 Pennsylvania districts. For the implementation study, SRI used programmatic, survey, interview, and observation data to assess the extent to which the key project components were implemented as intended and what factors facilitated or constrained implementation. The descriptive study used teacher reports of changes in student self-regulation, collaboration, and persistence in their classrooms over the two years of implementation. For the impact study, SRI relied on a quasi-experimental design, matching schools within Reading Apprenticeship districts (treatment group) to schools in other North Carolina districts (comparison group). This study compared perseverance and achievement – as measured by attendance, course grades, and end-of-course state standardized test scores in grades 9 and 10 – in the treatment and comparison schools after one and two years of Reading Apprenticeship implementation.

Overall, implementation fidelity was mixed, perhaps because of challenges associated with the post-pandemic climate. Teacher professional learning and leadership training reached higher levels of fidelity in some districts than others, while implementation of core RA classroom practices was consistently limited. Despite the mixed implementation findings, the impact study found positive effects of Reading Apprenticeship on the proximal outcomes of science course grades in both the one- and two-year analyses. This finding is consistent with previous research on Reading Apprenticeship that found positive effects on both teacher and student science outcomes (Fancsali et al., 2015; Greenleaf et al., 2011; Gulemetova et al., 2023). However, in this study, there were no impacts on more distal outcomes such as attendance and end-of-course standardized test scores.

Introduction

In recent years, literacy policy and practice have concentrated heavily on reading instruction in the early elementary grades, with considerably less attention devoted to adolescent readers in grades 4 through 12. Recent literacy achievement trends, however, underscore the need for a more sustained focus on adolescent readers. Although younger students have begun to recover some of the reading losses experienced before and during the COVID-19 pandemic, adolescent reading performance has shown little recovery and recently declined even further (Dewey, 2026). Results from the 2024 National Assessment of Educational Progress (NAEP) show that only 30 percent of students in grade 8 performed at or above proficiency in reading, and about a third scored below the NAEP Basic achievement level (National Center for Education Statistics [NCES], n.d.-b). Scores among lower- and middle-performing students remained below pre-pandemic levels, and lower-performing students experienced some of the steepest declines of the past decade (NCES, n.d.-c). These trends continued into high school: Grade 12 NAEP scores were low and had continued to decline (NCES, n.d.-a), while college-readiness indicators such as ACT scores reached their lowest levels since the 1990s (ACT, 2025).

Improving adolescent literacy requires a focus on reading and writing across the disciplines. As students progress through the secondary grades, they encounter increasingly complex texts and are expected to use reading to build knowledge, evaluate evidence, and solve problems in different subject areas. These demands extend beyond foundational literacy skills and vary by discipline: English language arts, social studies, and science differ in how knowledge is constructed, what counts as evidence, how arguments are developed, and how information is represented and communicated (Goldman et al., 2016; Shanahan & Shanahan, 2008). Disciplinary literacy instruction makes these practices explicit, helping students read, reason, and communicate in discipline-specific ways (Goldman et al., 2016; Pearson et al., 2020). Meeting these demands also requires rich vocabulary and background knowledge, access to complex grade-level texts, and opportunities for collaborative, text-based inquiry and authentic disciplinary work (Lesaux & Kieffer, 2010; Pearson et al., 2020; Snow & Biancarosa, 2003).

Reading Apprenticeship (RA) is an intervention designed to support secondary teachers across content areas to simultaneously develop students' discipline-specific literacy skills in order to improve student achievement (Greenleaf et al., 2023). The project challenges teachers to see themselves as teaching literacy as well as disciplinary content and prepares them to set up collaborative, inquiry-based classrooms in which thinking aloud about text is commonplace. Multiple studies have established the effectiveness of RA for both teachers and students. Most recently, researchers found that engaging in a hybrid model of RA, which included both in-person and online professional learning, led to large positive effects on teacher practice (Gulemetova et al., 2023). Moreover, earlier iterations of RA have been shown to improve students' reading comprehension, subject area knowledge, and disciplinary literacy (Fancsali et

al., 2015; Greenleaf et al., 2009; Greenleaf et al., 2011; Somers et al., 2010). Three of these studies also found positive impacts on students' perseverance as measured by attendance and course completion (Somers et al., 2010) and factors such as confidence and self-regulation (Fancsali et al., 2015; Greenleaf et al., 2009).

Recent research has pointed to a growing need to address self-regulation, collaboration, and persistence as components of reading and the meaning-making process (Duke & Cartright, 2021; Ho & Lau, 2018; Pearson et al., 2020). Successful reading requires students to manage their own engagement, use strategies to support comprehension of the text, and sustain the motivation needed to construct meaning. Motivation, engagement, and self-regulation during reading, particularly the use of comprehension strategies, have been shown to predict reading achievement (Cartwright et al., 2020; Steacy et al., 2016). While many of these factors, such as the use of comprehension strategies as a means of student self-regulation during reading, have long been part of the RA framework and were measured in some previous studies of RA, they were not as overtly addressed in the professional learning as other aspects of the model. In 2020, WestEd sought to enhance this aspect of RA and redesigned formative assessment tools and other teacher supports to center these aspects.

With the support of a federal Education Innovation and Research (EIR) Mid-phase grant, the RA Project Team at WestEd implemented this innovative version of RA – *Accelerating Literacy Through Self-Regulation, Collaboration, and Persistence* - with 25 primarily rural districts in five distinct cohorts over four school years:

- Cohort A: Five North Carolina districts during the 2021–22 and 2022–23 school years
- Cohort B: Two North Carolina districts during the 2022–23 and 2023–24 school years
- Cohort C: Six North Carolina districts during the 2023–24 and 2024–25 school years
- Cohort D: Six North Carolina districts during the 2024–25 school year
- Cohort E: Six Pennsylvania school districts in the 2024–25 school year, to examine the extent to which RA could be replicated in a different rural context

WestEd contracted with SRI to conduct an independent evaluation of this EIR-funded project. This report presents findings from SRI's evaluation, including evidence about the project's implementation and its effects on student outcomes. Following an overview of the RA model and the participating districts, the report describes the study methods and findings on implementation, outcomes, sustainability, and cost-effectiveness. Together, these analyses shed light on both the promise of RA and the conditions that shaped its implementation, offering lessons for educators and policymakers to consider as they seek to strengthen secondary teachers' disciplinary literacy instruction and improve adolescents' perseverance and literacy achievement.

Overview of Reading Apprenticeship in North Carolina

Schools in rural North Carolina face notable literacy challenges, particularly in rural, low-wealth districts. In 2019, students in rural districts in North Carolina had significantly lower English language arts (ELA) test scores in grade 8 when compared with students in nonrural districts. Likewise, students facing socioeconomic barriers in these rural districts performed well below the national average, and significantly below similar students in nonrural districts in North Carolina (Reardon et al., 2026). Five-year graduation rates, a measure of student perseverance, were also significantly lower in rural North Carolina areas than in nonrural areas (WestEd, 2019). Moreover, the rural/nonrural achievement difference has been more pronounced in North Carolina than almost any other state (Showalter et al., 2019).

The Reading Apprenticeship (RA) in North Carolina project was designed to simultaneously strengthen disciplinary literacy instruction and support teachers to develop students' collaborative skills, self-regulation, and persistence in rural, low-wealth high schools through sustained, evidence-based professional learning. The project facilitated professional learning sessions, small-group coaching, and supports for grade 9 and 10 teachers to integrate reading, writing, and discussion into content-area instruction across ELA, science, and social studies.

Project Description

Reading Apprenticeship is both a professional development model and an instructional framework. The instructional framework focuses on four interacting learning dimensions that are developed through extensive reading and metacognitive conversation: creating a collaborative classroom, building on student strengths, building reading skills, and developing content knowledge (Greenleaf et al., 2023). This framework grounds both the RA instructional practices at the heart of the model and this project's emphasis on self-regulation, collaboration, and persistence. Students in RA classrooms, for example, develop their capacity to self-regulate while reading by learning to use comprehension strategies productively through collaboration and metacognitive conversations.

Intensive and sustained professional learning has long been central to RA (Darling-Hammond et al., 2017; Greenleaf et al., 2009). During sequential PL sessions, expert facilitators model instructional practices, engage teachers in experiential learning, and support site-based teams to implement. RA facilitators also provide coaching throughout the school year to further support implementation and integration. During school-team or professional learning community (PLC) meetings led by teacher leaders, colleagues work together to analyze formative assessment and refine implementation.

Reading Apprenticeship in North Carolina sought to preserve these core elements of professional learning but offer a more efficient hybrid training series that fit the local context

and considered pandemic-era constraints. Traditionally, RA-trained teachers had participated in 10 days of professional learning in discipline-specific teams (e.g. science teachers), with five of these days being sequential summer training. Convening discipline-specific groups for consecutive PL days, however, did not prove scalable or cost effective, especially for rural districts still reeling from the COVID-19 pandemic. RA in North Carolina thus was designed to engage cross-disciplinary teams in shorter consecutive professional learning days (2–3) but include more frequent touchpoints throughout the year. Small-group coaching and professional learning communities were also conducted more frequently and online to increase access.

This specifically tailored RA professional learning series was implemented for two consecutive years with Cohorts A–C and one year for Cohorts D and E. Its final components included:

- Annual meetings of instructional and administrative leaders that support the integration of RA into districts’ core practices
- Blended professional learning, which included in-person sessions and online small-group coaching, for grade 9 and 10 ELA, social studies, and science teachers (54 hours for two years; 30 hours for one year)
- Leadership training for campus-based teacher leaders to support implementation and sustain the effective use of RA in participating districts
- Preparation of teacher-leaders to become certified RA facilitators to further expand RA in North Carolina and beyond

While all these PL elements were originally planned, some specific details changed during the first year of the grant. The number of required professional learning hours, for example, was lowered due to continued pandemic-related disruptions, weather events, and limitations that affected educator capacity at the beginning of the five-year grant.¹

Logic Model

Reading Apprenticeship facilitates changes in teacher practice and student learning through professional learning that builds on existing structural inputs such as the RA instructional framework and resources and includes support (key) components and direct components. Support components include the activities and preparation provided by the RA Project Team that support the participants (teachers) to implement the RA instructional model as intended (Weiss et al., 2014). Direct components, by contrast, involve direct interaction with the target population—in this case, grade 9 and 10 North Carolina students enrolled in ELA, science, and social classes taught by teachers engaged in the RA professional learning (Hill et al., 2023). These components and their embedded activities were designed to strengthen teachers’ knowledge, mindsets, and instructional practices (mediators) so they could better support

¹ All North Carolina districts were required to offer 30 hours of summer school during 2021 to mitigate pandemic learning loss. This requirement made it difficult for teachers to complete the RA summer sessions. To accommodate this requirement and the lack of substitute-teacher support across North Carolina, The RA Project Team revised the two-year professional learning series from 82 to 54 hours.

students' self-regulation, collaboration, persistence, and learning across content areas (outcomes).

The RA in North Carolina logic model (see Exhibit 1) operationalizes these relationships, showing how the RA instructional framework and professional development model is hypothesized to lead to improved student achievement. Specifically, the logic model illustrates the relationships between support components, direct components, mediators, short- and long-term outcomes, and contextual moderators.

Support Components

The logic model's two support components include (1) professional learning and small-group coaching for ELA, science, and social studies teachers in grades 9 and 10 and (2) leadership training for teacher leaders and administrators.

Professional Learning and Coaching for Teachers. The RA Project Team developed professional learning structures and materials for teachers to learn how to more effectively integrate disciplinary reading, writing, and discussion into content area instruction. Through face-to-face and online sessions, North Carolina teachers deepened their understanding of RA's instructional framework (four learning dimensions and the roles of metacognitive conversation and extensive reading) across disciplines. They examined their own and their students' reading practices, analyzed disciplinary texts, and learned routines and strategies for supporting students to make sense of challenging academic content.

Professional learning opportunities modeled the kinds of instructional experiences teachers were expected to create for students. Expert facilitators demonstrated RA routines and strategies and engaged teachers in active, collaborative learning. Teachers articulated their own reading comprehension challenges, worked together to make sense of difficult texts, analyzed and annotated videos of RA instruction, planned lessons with core RA routines, and collaborated across disciplines.

Small-group coaching supported teachers as they implemented RA in their classrooms. During coaching and school-team meetings, teachers received feedback from RA facilitators and teacher leaders, reflected on their implementation, refined lesson plans and instructional practices, and reviewed formative data about students' learning.

Leadership Training for Teacher Leaders and Administrators. Teacher leaders and school and district administrators, who can act as important levers for school and district implementation, also engaged in sustained, job-embedded professional learning (Darling-Hammond et al., 2017). Leadership training opportunities were tailored by role. Teacher leaders, who were practicing teachers nominated by administrators for their instructional knowledge and leadership capacity, participated in sessions focused on deepening their RA

knowledge, integrating RA practices² into existing curriculum, and coaching teachers with implementation. RA facilitators also trained teacher leaders to facilitate school and district team meetings. Administrators attended sessions focused on developing RA knowledge, supporting teacher leaders and teachers, and promoting sustainability. The trainings for both teacher leaders and administrators also provided opportunities for leaders to collaborate with peers from other schools and districts.

Direct Components

The RA in North Carolina model included one direct component: classroom implementation of target RA instructional practices.

Implementation of Core RA Instructional Practices. At the heart of the RA in North Carolina model was teachers using core RA practices. These practices included writing and implementing lessons that integrate RA, establishing community agreements, modeling and using collaboration routines, modeling and using think-alouds and other metacognitive routines, modeling and using comprehension strategies, and collecting and analyzing formative student data. These are the instructional practices that distinguish RA from other literacy approaches and interventions. Previous research found that the implementation of three of these practices—modeling and use of collaboration routines, metacognitive routines, and comprehension strategies—was associated with improvements in student achievement (Fancsali et al., 2015).

Mediators

Both the support and direct components were intended shift teachers' mindsets and instructional practices as a means to improve student outcomes. Specifically, the RA Project Team hypothesized that RA would lead to improved student outcomes via three mediators: deepening teachers' knowledge about disciplinary literacy; shifting their mindsets regarding students' capacities, learning as an inquiry process, and their role in teaching reading; and changing instructional practice in particular ways (shifting the intellectual load more to students, providing more text-based learning and peer-to-peer discussion opportunities, and intentionally embedding conversations about reading strategies into disciplinary lessons).

Outcomes

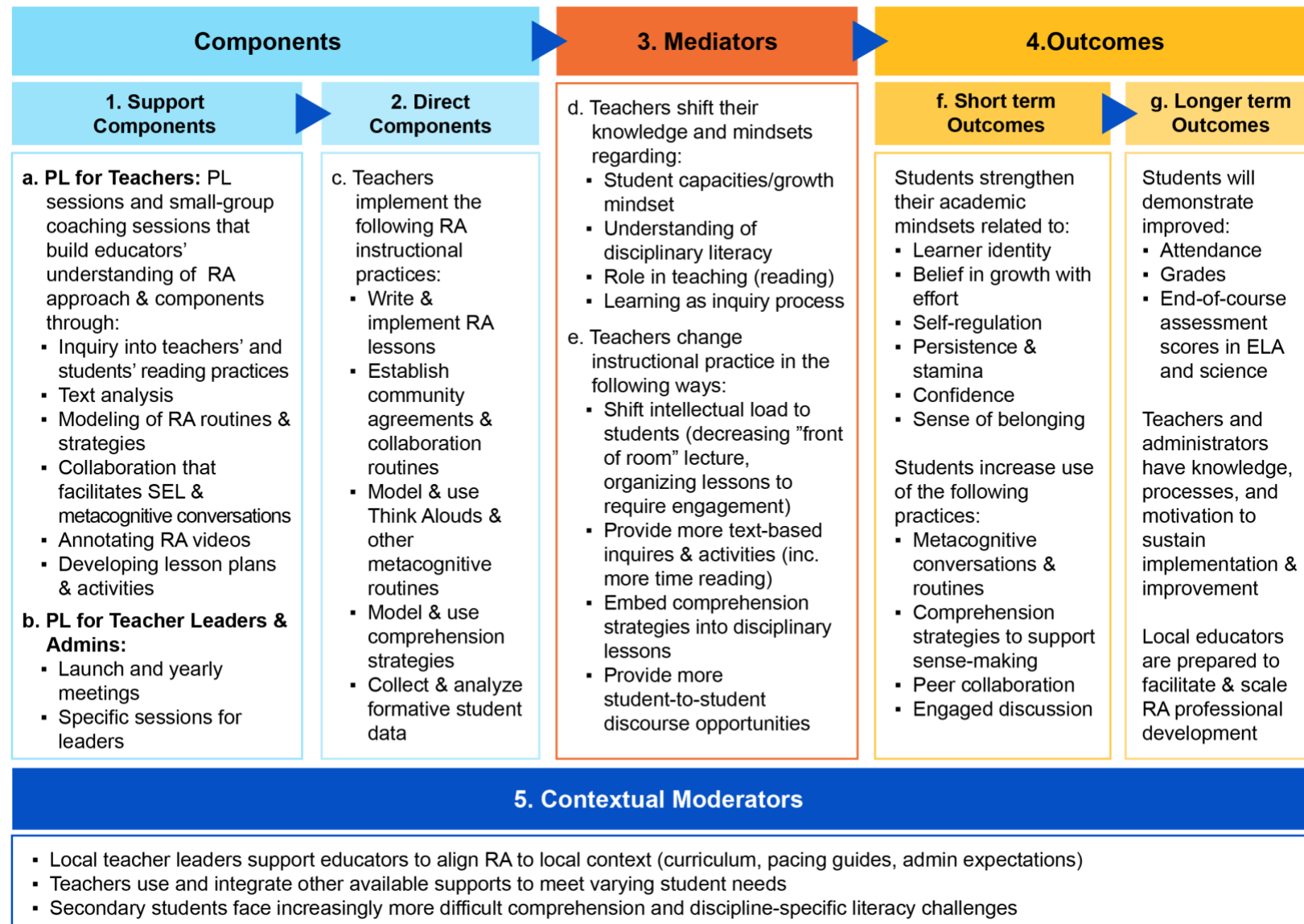
These teacher-level shifts were expected to influence two **short-term student outcomes**: strengthening students' academic mindsets and increasing students' use of core RA routines and strategies. Improvements in these short-term outcomes would ultimately lead to **longer-term student outcomes**: improved attendance, course grades, and end-of-course state standardized test scores in ELA and science. Improved student outcomes, along with the continued

² RA collaboration structures, metacognitive routines, and comprehension strategies are collectively referred to as "core RA practices" in this report.

engagement and commitment of teachers and leaders, would in turn enable sustained RA implementation and set the stage for scaling of the project, which comprised the **longer-term project outcomes**.

Contextual Moderators

The RA Project Team also worked with district partners to ensure that each participating school had adequate administrative and teacher-leader support for the project.

Exhibit 1. Reading Apprenticeship in North Carolina Logic Model

Evaluation Design and Methods

The evaluation of Reading Apprenticeship (RA) in North Carolina included examinations of implementation, outcomes, sustainability, scaling, and cost. For the implementation study, SRI used project survey, interview, and observation data to (1) examine the extent to which participating districts implemented RA as intended and (2) identify common facilitators and barriers to implementation. For the outcomes, SRI used a combination of descriptive methods and a quasi-experimental matching design to examine students' academic mindsets, use of RA strategies and routines, attendance, course grades, and end-of-course standardized test achievement.

Study Districts, Schools, and Teachers

The RA Project Team originally recruited nearly 30 rural, low-wealth North Carolina districts that had substantial percentages of grade 9 students entering high school reading below grade level to participate in the project in 2020. SRI's evaluation was initially designed to be a two-year randomized control trial to estimate the effects on teacher practice and student outcomes. This study was to be conducted with one large cohort of 25 treatment schools and 25 control schools. The treatment schools were to engage in RA professional learning during the 2021–22 and 2022–23 school years, and the control (or delayed treatment) schools were to participate in 2023–24 and 2024–25.

These original plans were disrupted by lingering effects of the COVID-19 pandemic, which contributed to teacher shortages, restrictive professional learning and school visitation policies, and an exhausted educator workforce in North Carolina. By the time the 2021–22 school year started, approximately 50 percent of the districts that had initially expressed interest in the project decided not to move forward because of the anticipated demands on teachers.

Given this situation, WestEd and SRI extended the timeline for treatment and revised plans for recruitment, professional learning, and the evaluation. Recruitment was expanded to include a small number of nonrural districts, and professional learning was further streamlined and—for the first cohort—changed to a completely online format. SRI also shifted the evaluation from a randomized to a quasi-experimental design, which eliminated the need to recruit schools willing to be randomized to a control condition. However, it also removed teacher practice and reduced the viability of student self-regulation and collaboration as outcomes, as they relied on student surveys and systematic classroom observations that were not feasible at the time.

SRI thus shifted the implementation study design to examine the new hybrid professional learning model, providing concentrated formative feedback on online professional learning during the first year of RA implementation and moving in-person site visits to subsequent study

years. With an eye toward scaling, the RA Project Team also added a final cohort in a new state context: Pennsylvania.

In total, the evaluation included five cohorts, 25 districts, and 47 schools. Exhibit 2 provides a summary of participating cohorts. Exhibit A-1 in Appendix A provides more detail on participating districts and schools.

Exhibit 2. Characteristics of Cohorts

Cohort	State	Total Districts	Percentage of Districts with Rural Locale	Total Participating Schools	Average HS Enrollment
Cohort A (2021–2023)	NC	5	80%	15	1,256
Cohort B (2022–2024)	NC	2	50%	2	1,210
Cohort C (2023–2025)	NC	6	83%	15	2,172
Cohort D (2024–2025)	NC	6	83%	10	1,619
Cohort E (2024–2025)	PA	6	83%	6	376

Within the RA districts, all schools serving grade 9 and 10 students and not adhering to an alternative school model were given the opportunity to participate in the project. District and school leaders identified the ELA, science, and social studies teachers and teacher leaders who were eligible to participate. However, not all eligible teachers participated because of capacity, district and school decisions, and teacher choice. The average teacher participation rate, across districts and cohorts, was 87 percent.

Research Questions

This report addresses research questions in five categories: implementation, descriptive outcomes, impact outcomes, sustainability and scaling, and cost-effectiveness. Mediators and moderators are not addressed here but will be examined in future reports.

Implementation

1. To what degree was Reading Apprenticeship (RA) implemented with fidelity?
2. What contextual factors facilitated or impeded implementation?

Descriptive Outcomes

3. To what extent did teachers perceive changes in students' academic mindsets and use of core RA practices?

Impact Outcomes

4. (Exploratory) What is the impact of one year in a school with RA-trained teachers on grade 9 and 10 students' achievement and attendance as compared with business as usual?
 - a. What is the impact of RA on students' course grades in literacy, social studies, and science?
 - b. What is the impact of RA on students' end-of-course state standardized test scores in ELA and science?
 - c. What is the impact of RA on students' daily attendance?
5. (Confirmatory) What is the impact of two years in a school with RA-trained instructors on grade 10 students' achievement and attendance as compared with business as usual?
 - a. What is the impact of RA on students' course grades in literacy, social studies, and science?
 - b. What is the impact of RA on students' end-of-course state standardized test scores in ELA and science?
 - c. What is the impact of RA on students' daily attendance?

Sustainability and Scaling

6. To what extent have RA-trained educators and schools developed the expertise, ownership, and plans needed to sustain RA?
7. How does WestEd replicate and scale RA while maintaining project fidelity?

Cost Effectiveness

8. What is the cost of implementing RA, relative to the project's impact on students?

Data Sources

To answer the research questions, SRI conducted primary data collection and collected extant data in participating districts.

Reading Apprenticeship Program Data. SRI collected program data from the WestEd RA Project Team for the 2021–22 through the 2024–25 school years. These data included teacher rosters; attendance data for the professional learning sessions, online small-group coaching, leadership team meetings, and the RA Administrators course; and teacher submission of artifacts (e.g., lesson plans) or reflection on implemented lessons.

Evaluations of Professional Learning. SRI collected evaluations of online professional learning during the first year of the grant (2021–22 school year) to support and inform WestEd’s shift to online delivery.

Teacher Surveys. In spring 2022 through spring 2025, SRI administered surveys to all teachers in their first or second year of RA professional learning (see Exhibit 3 for response rates). However, only teachers who said they had participated in at least one RA activity were asked to complete the full survey. The surveys asked teachers about their experience with the RA professional learning opportunities, their classroom implementation of RA practices, the challenges they faced in implementing RA, and perceived impact of participating in the RA professional learning.

Exhibit 3. Survey Response Rates by Cohort and School Year

Cohort		2021–22	2022–23	2023–24	2024–25
A	# teachers	97 of 112	85 of 118		
	Response rate	87%	72%		
B	# teachers		22 of 28	14 of 18	
	Response rate		79%	78%	
C	# teachers			122 of 153	103 of 155
	Response rate			80%	66%
D	# teachers				105 of 142
	Response rate				74%
Overall	# teachers	97 of 112	107 of 146	136 of 171	208 of 297
	Response rate	87%	73%	80%	70%

Interviews. During the first year of RA implementation, in the 2021–22 school year, SRI interviewed six teachers in Cohort A about their experience with different aspects of professional learning and implementation. In the 2023–24 school year, SRI interviewed eight district leaders, school administrators, and teacher leaders in fall 2023 and 11 teacher leaders and teachers in spring 2024, across three districts in Cohort C. In spring 2024, SRI also held 11 student focus groups in these districts; each focus group included 2–4 students. In the spring of the following school year (2024–25), SRI interviewed 18 school administrators, teacher leaders, and teachers across five districts.

During the interviews and focus groups, SRI used semi-structured protocols with questions aligned to the research questions and logic model, tailored to each respondent’s role. Interview topics included teacher reflection on their participation in RA; frequency and focus of the professional learning; perceived project impacts on teacher practice and student outcomes; and sustainability. Student focus group topics included students’ use of RA routines and comprehension strategies in the classroom and their academic mindsets.

Classroom and Professional Learning Observations. SRI observed both classrooms and RA professional learning in spring 2024 and 2025. Classroom observations were intended to

provide researchers with a better understanding of participating teachers' instructional practice and school context. Observers took notes on instructional practices using a semi-structured protocol aligned with key RA instructional framework dimensions. During the RA professional learning, researchers took notes on what was discussed during these activities and recorded examples of teacher ownership, collaboration, and modeling and discussion of instructional practices.

Extant Data. SRI collected statewide student, course, teacher, and school data from the North Carolina Education Research Data Center (NCERDC). We did not collect extant data from the Pennsylvania districts in Cohort E.

Independence of Evaluation

SRI conducted all evaluation activities independently, except for WestEd collecting professional learning attendance data. These activities included the collection of extant data, observations, surveys, and interviews; identification of the comparison sample; and all implementation, outcome, sustainability, and cost analyses.

Implementation

SRI assessed Reading Apprenticeship (RA) implementation fidelity across districts in Cohorts A through E and examined the conditions that supported or constrained implementation. Fidelity analyses drew on project data provided by WestEd and teacher survey responses, while interview data and open-ended survey responses offered insight into the implementation facilitators and barriers. This section begins by outlining SRI's approach to assessing implementation fidelity and then presents the implementation findings.

Fidelity Measures

SRI worked with the RA Project Team to specify indicators and thresholds for the support and direct components: (1) active, sustained participation in high-quality teacher professional learning; (2) participation in leadership training for sustainability; and (3) classroom implementation of core RA practices (Exhibit 4).³ For each indicator, SRI identified the appropriate unit of measurement and defined the thresholds for high fidelity at the unit level (e.g., teacher, administrator). SRI and the RA Project Team also established district-level fidelity thresholds for each indicator and component. Indicator language and acceptable thresholds for some indicators (e.g., teacher professional learning attendance, implementation of core practices) were adapted from prior iterations of RA that demonstrated positive impacts on student achievement (Fancsali et al., 2015). Exhibit 4 outlines each indicator and threshold for the three components. For the project-level component scores, SRI and the RA Project Team set the threshold for high fidelity at 75 percent of all study districts meeting the district-level threshold for high fidelity. The threshold for medium fidelity was 50 percent of study districts meeting district-level thresholds.

³ For Cohorts B and C, this component also included an additional artifact-based indicator, which was originally designed to track teachers' implementation of RA-influenced lesson plans. Artifacts were not collected in Cohort A because of the shift to primarily online learning, nor were they collected in Cohorts D and E because of the streamlined version of the professional learning.

Exhibit 4. Fidelity of Implementation Scoring System for Each Key Component in the Logic Model

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at District Level	Data Source
Component 1: Active, sustained participation in high-quality teacher professional learning (PL)				
1.1a PL Year 1	Teacher	Teachers attend 2 out of the 3 PL days by the end of Year 1	At least 75% of teachers meet threshold	RA attendance records
1.1b PL Year 2 (rollover)	Teacher	Teachers attend 4 out of the 5 PL days by the end of Year 2	At least 75% of teachers meet threshold	RA attendance records
1.2 Coaching	Teacher	Teachers attend 4 out of 6 of the two-hour online school-based coaching sessions (Years 1 & 2)	At least 75% of teachers meet threshold	RA attendance records
1.3 PL reading practices	Teacher	Teacher answers that RA training “almost always” or “often” involves inquiry into students’ and into teacher’s own reading practices (Years 1 & 2)	At least 75% of teachers meet threshold	Teacher survey
1.4 PL routines and strategies	Teacher	Teacher answers that WestEd training “almost always” or “often” involves demonstration of instructional routines and/or strategies (Years 1 & 2)	At least 75% of teachers meet threshold	Teacher survey
1.5 PL building self-regulation	Teacher	Teacher answers that WestEd training “almost always” or “often” focuses on building students’ self-regulation, personal engagement, collaboration, and persistence. (Years 1 & 2)	At least 75% of teachers meet threshold	Teacher survey
1.6 PL analysis of texts	Teacher	Teacher answers that WestEd training “almost always” or “often” involves analysis of texts across the disciplines (Years 1 & 2)	At least 75% of teachers meet threshold	Teacher survey
Component 2: Participation in leadership training for sustainability				
2.1a District leader (DL) annual meeting	District leader	District leaders	At least 1 district leader per site attends annual meeting each year	RA attendance records
2.1b Teacher leader (TL) annual meeting	Teacher leader (district)	N/A	At least 1 teacher leader per site attends annual project-wide leadership team meetings each year	RA attendance records

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at District Level	Data Source
2.2a District leader course	District leader	N/A	At least 1 district leader per site attends at least 3 sessions of the RA Administrators course by the end of Year 2	RA attendance records
2.2b School leader course	School	At least 1 school administrator per school completes at least 3 sessions of the RA Administrators course by the end of Year 2	75% of schools meet fidelity	RA attendance records
2.3 TL training	Teacher leader (district)	N/A	At least one teacher leader per district attends the full-day training and attends 2 out of 3 online meetings in Year 2.	RA attendance records
2.4 TL coaching	Teacher leader (district)	N/A	At least one teacher leader per site co-facilitates at least 1 coaching session or leads 1 PLC session in Year 2.	RA attendance records
Component 3: Classroom implementation of core RA practices				
3.1a Cohort B artifact	Teacher	Cohort B teachers submit at least 6 artifacts and/or reflections after 2 years of implementation	At least 75% of teachers meet fidelity	RA submission records
3.1b Cohort C artifact	Teacher	Cohort C teachers submit at least 4 artifacts and/or reflections each year	At least 75% of teachers meet fidelity	RA submission records
3.2 Collaboration opportunities	Teacher	Teachers respond that they provide students with opportunities to collaboratively discuss texts “daily or almost daily” or “2–3 times a week”	At least 50% of teachers meet fidelity (Y1) At least 75% of teachers meet fidelity (Y2)	Teacher survey
3.3 Metacognitive routines	Teacher	Teachers respond that they model, teach, or use metacognitive routines “daily or almost daily” or “2–3 times a week”	At least 50% of teachers meet fidelity (Y1) At least 75% of teachers meet fidelity (Y2)	Teacher survey
3.4 Comprehension strategies	Teacher	Teachers respond that they model and teach comprehension strategies “daily or almost daily” or “2–3 times a week”	At least 50% of teachers meet fidelity (Y1) At least 75% of teachers meet fidelity (Y2)	Teacher survey

Fidelity Findings

SRI measured implementation across all participating teachers, teacher leaders, administrators, and, when applicable, schools for Cohorts A, B, C, D, and E. Exhibit 5 provides details about the districts, schools, and teachers in each cohort. School-level results were rolled up to the district level, and district level results were rolled up to the project level. For indicators based on the teacher survey, SRI measured fidelity of implementation across the teachers who responded to the survey each year.

Exhibit 5. Number of Districts, Schools, and Teachers in Implementation Study

Cohort	School Year	Number of Participating Districts	Number of Participating Schools	Number of Participating Teachers	Number of Teachers Surveyed ^a	Number of Survey Respondents
Cohort A	2021–22	5	14	123	112	97 (87%)
	2022–23	5	15	152	118	85 (72%)
Cohort B	2022–23	2	2	28	28	22 (79%)
	2023–24	2	2	20	18	14 (78%)
Cohort C	2023–24	6	15	162	153	122 (80%)
	2024–25	6	15	152	155	103 (66%)
Cohort D	2024–25	6	10	95	93	80 (86%)
Cohort E	2024–25	6	6	49	49	25 (51%)
Full Sample	2021–22	5	14	123	112	97 (87%)
	2022–23	7	17	180	146	107 (73%)
	2023–24	8	17	182	171	136 (80%)
	2024–25	18	31	296	297	208 (70%)

^a Teachers who attrited from study schools between the start of the school year and survey administration were not surveyed but were included in non-survey fidelity-of-implementation calculations (e.g., attendance at professional learning).

To achieve high fidelity for teacher training (Component 1) and leadership training (Component 2), each *district* had to meet the indicator-level threshold for at least four of the six indicators (67 percent).⁴ To achieve high fidelity for classroom implementation of core practices (Component 3), each district must have met the indicator-level threshold for three out of four indicators.⁵

Exhibit 6 shows the implementation level for each component within each cohort for each year of the implementation study. Overall, implementation was mixed across the four years of the

⁴ Four of five indicators in Component 1 for Cohort A, Year 1, where there was no expectation for the in-person professional learning, and two of two indicators in Component 2 for Cohort A and B, Year 1, where leaders had until the end of Year 2 to complete the RA Administrators course and teacher leaders had until the end of Year 2 to attend leadership training and facilitate coaching sessions.

⁵ Two of three indicators in Cohorts A, D, and E, where teachers were not required to submit artifacts.

initiative: While the teacher professional learning (PL) and leadership training components were at high levels of fidelity in approximately half of the districts for Y1 or Y2, the classroom implementation component was consistently low. Exhibit A-2 in Appendix A details the implementation level for each component for each district across the cohorts. The sections that follow Exhibit 6 describe fidelity findings by component.

Exhibit 6. Number of Districts Meeting Fidelity for Each Component

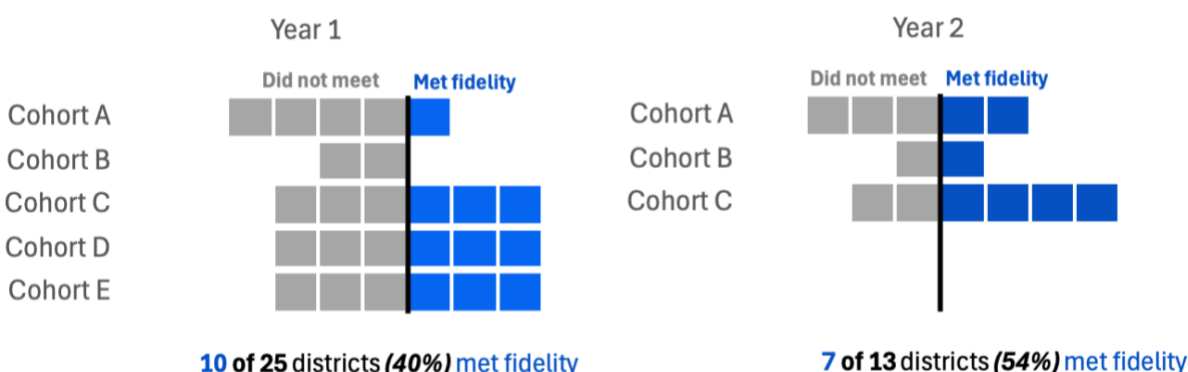
Year	Component	Cohort A	Cohort B	Cohort C	Cohort D	Cohort E	Full Sample
Year 1	Component 1: Participation in and quality of teacher professional learning	1 of 5 districts	0 of 2 districts	3 of 6 districts	3 of 6 districts	3 of 6 districts	10 of 25: Low
	Component 2: Leadership training for sustainability	4 of 5 districts	1 of 2 districts	6 of 6 districts	1 of 6 districts	3 of 6 districts	15 of 25: Mid
	Component 3: Classroom implementation of core RA practices	2 of 5 districts	0 of 2 districts	1 of 6 districts	0 of 6 districts	0 of 6 districts	3 of 25: Low
Year 2	Component 1: Participation in and quality of teacher professional learning	2 of 5 districts	1 of 2 districts	4 of 6 districts	N/A	N/A	7 of 13: Mid
	Component 2: Leadership training for sustainability	3 of 5 districts	1 of 2 districts	2 of 6 districts	N/A	N/A	6 of 13: Low
	Component 3: Classroom implementation of core RA practices	0 of 5 districts	0 of 2 districts	0 of 6 districts	N/A	N/A	0 of 13: Low

Component 1: Participation in and Quality of Teacher Professional Learning

Although fidelity increased from Year 1 to Year 2 for Cohorts A through C, RA teacher professional learning was not implemented with high fidelity in any year.

When aggregated across indicators and districts, RA did not meet the 75 percent project threshold in either Year 1 or Year 2 for teacher PL. However, there was some improvement over time: The percentage of districts meeting the threshold increased from the first to the second years for every two-year cohort. Cohort C came the closest to reaching high fidelity, with four of six of districts (67 percent) achieving high fidelity in Year 2. Overall, 40 percent of districts in Year 1 (10 of 25) and 54 percent of districts in Year 2 (7 of 13) reached the threshold for high fidelity of implementation in teacher training (Exhibit 7).

Exhibit 7. Districts Meeting Fidelity of Teacher Professional Learning



Attendance at professional learning sessions (1.1) and coaching (1.2) varied by cohort and was highest in Cohort C.

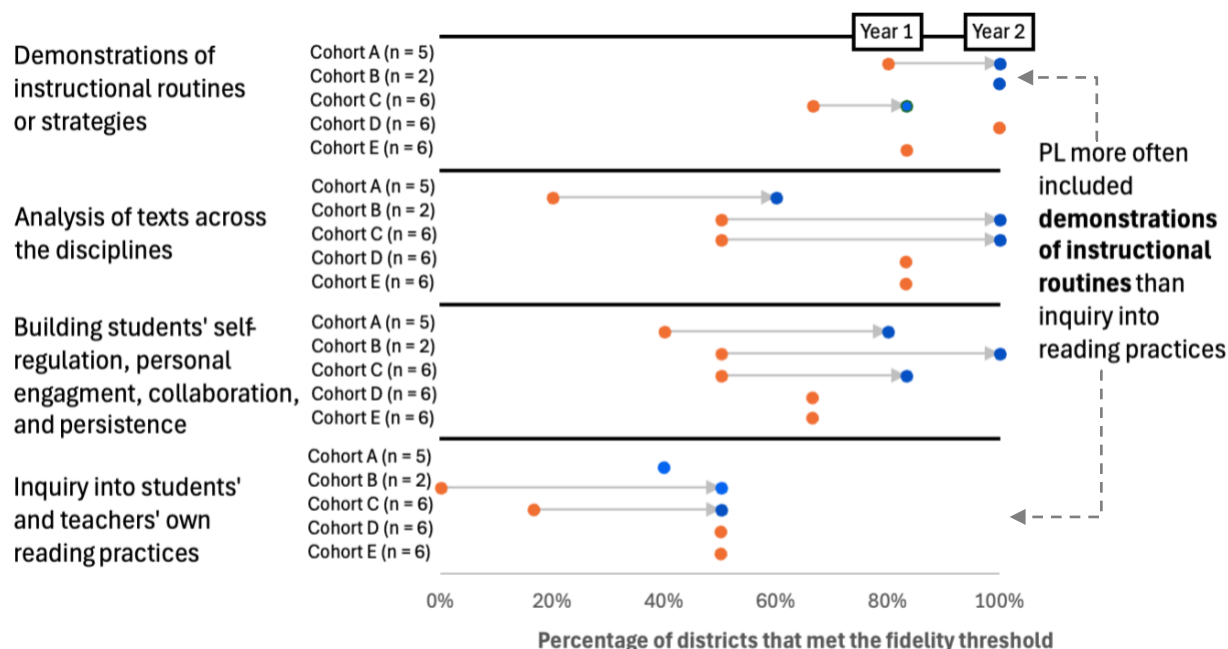
As a result of the COVID-19 pandemic, Cohort A's first year of implementation had no strict expectations for in-person PL attendance; instead, teachers were required to attend four of five PL sessions (online or in person) by the end of Year 2. Despite the flexibility, none of the Cohort A districts met the threshold for attending PL (1.1) or coaching sessions (1.2) in either year. Cohort B had mixed results, with one of the two districts meeting fidelity on the attendance indicators in Year 2. Cohort C had the highest fidelity: Five of six districts met the threshold for in-person PL attendance in Year 1. However, Cohort C PL attendance fell in the second year, and coaching attendance also slightly decreased.

Cohorts D and E had only one year of online coaching sessions, and all five of their in-person PL sessions were condensed into that year as well. Only a quarter of these districts (3 of 12) met the threshold for attending in-person PL or online coaching sessions in this condensed form, a rate lower than the rates for Cohort B or C in Year 1.

Teachers reported that professional learning included demonstrations of instructional routines or strategies (1.4) more often than inquiry into students' and teachers' reading practices (1.3).

The annual teacher survey included questions about the content and quality of the PL that teachers participated in. To meet the threshold for high fidelity, 75 percent of the teachers who attended PL in each district had to report that the PL they received "often" or "almost always" included the key activities referenced in Exhibit 8. Teachers were most likely to report that the PL contained *demonstrations of instructional routines or strategies*, and in most districts the percentage increased between Year 1 and Year 2 of implementation. Fidelity was lowest with respect to *inquiry into students' and teachers' reading practices* in PL, with between 17 and 50 percent of districts meeting the threshold.

Exhibit 8. Inclusion of Key Activities in Teacher Professional Learning by Cohort and Implementation Year



Across all cohorts, teachers' perceptions regarding the frequency and quality of the key professional learning activities (1.3–1.6) improved from the first to the second implementation year.

Exhibit 8 also captures a general increase in fidelity across cohorts for the indicators focused on PL content. The biggest change that teachers reported between the two years of PL offerings was in *building students' self-regulation, personal engagement, collaboration, and persistence*. In Year 1, 56 percent of districts met the threshold for teachers reporting that the PL included this practice “often” or “almost always,” and 85 percent of all districts met the threshold in Year 2.

Teachers in Cohort D and E districts, who engaged in fewer hours of PL across only one year, reported participating in these PL activities more frequently than teachers in Cohorts A through C in their first year of the RA project.

The survey also asked teachers to rate the quality, relevance, and impact of the PL opportunities. Although these items did not feed into indicators of implementation fidelity, they provide valuable information regarding teachers' assessment of the PL. In all cohorts and across most items, teachers' ratings of the PL were substantially higher in their second year of implementation than in their first year. This finding may indicate an overall improvement in the quality of the PL over time. Furthermore, Cohort D and E teachers rated the PL as high or higher in their one year of implementation than the other cohorts did in their second year.

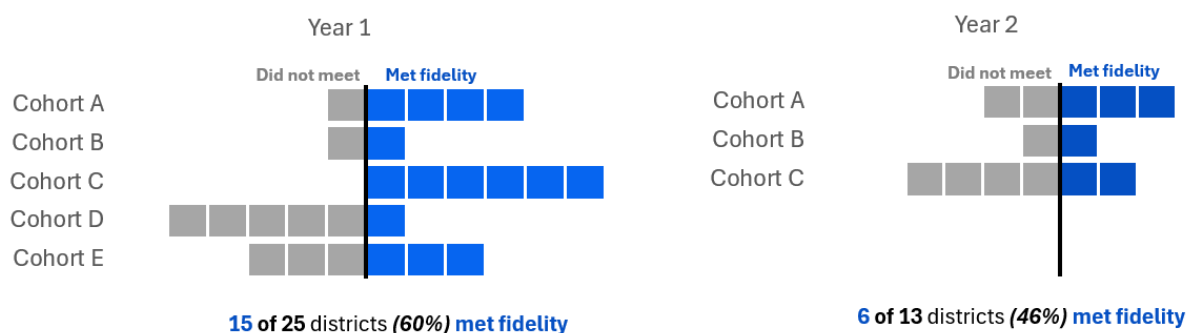
Component 2: Leadership Training for Sustainability

Project-level implementation of leadership training was in the mid-range in Year 1 and was low in Year 2 of implementation.

In Year 1, implementation of leadership training was mixed across cohorts: Cohorts A and C met the threshold for high fidelity across the indicators, but only one of two Cohort B districts, one of six Cohort D districts, and three of six Cohort E districts met the threshold in Year 1. Thus, overall project-level implementation was in the mid-range, with 60 percent of districts having high fidelity on this component in Year 1 (Exhibit 9).

In Year 2, the number of fidelity indicators required increased from two to six, making it more challenging to reach high fidelity. Although a similar percentage of districts met the two common indicators in Year 1 and Year 2, the percentage of districts with high fidelity in leadership training in Year 2 was 46 percent, lower than in Year 1. The most common reason districts missed achieving fidelity in leadership training was that district and school leaders did not complete the RA Administrator course by then end of Year 2.⁶ (District administrators completed the course in 28 percent of districts, and school administrators completed in 32 percent of districts.) Notably, only 20 percent of Cohort D districts (1 of 6) implemented leadership training with fidelity in their only year of implementation, while half of Cohort E districts (3 of 6) reached high fidelity in this component.

Exhibit 9. Fidelity of Leadership Training



Component 3: Classroom Implementation of Core RA Practices

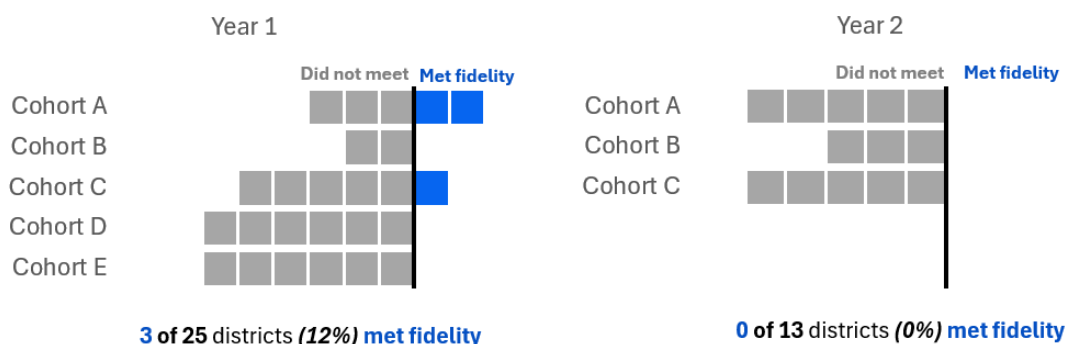
Very few districts met fidelity for implementation of core RA practices.

Overall, only 12 percent of districts (3 of 25) met the threshold for high fidelity on the direct component in Year 1 of implementation, and none met the threshold in Year 2 (Exhibit 10). This finding indicates that classroom implementation of core RA practices was *lower* than the implementation of teacher PL or leader training—the support components. The decrease in

⁶ Leaders in Cohorts D and E were required to complete the administrator course by the end of their first and final year.

fidelity from Year 1 to Year 2 can also be explained by the higher expectations of implementation over time: To meet fidelity, a higher percentage of teachers in each district had to report modeling and teaching comprehension strategies, collaboration, and metacognitive routines in Year 2 (75 percent) than in Year 1 (50 percent).

Exhibit 10. Fidelity of Classroom Implementation



On average, teachers did not implement core RA practices as often as expected, and there was little change between implementation years.

On the survey, teachers reported, on average, that they supported collaboration, modeled and used metacognitive routines, and modeled and used comprehension strategies between one and three times a week. This frequency was below the fidelity threshold for both Year 1 and Year 2. Across cohorts, between 36 and 54 percent of teachers reported supporting collaboration in their classrooms 2–3 times a week or more (the fidelity threshold). About the same percentages of teachers reported regularly modeling and using comprehension strategies. Teachers were least likely to model and use metacognitive routines at the expected frequency (31–43 percent). Cohort D and E teachers did not implement core RA practices at rates substantially lower than Cohort A, B, and C teachers.

Although districts did not meet the classroom implementation thresholds, most teachers reported implementing each of the core RA practices at least once a week. Across cohorts and survey years, 73–95 percent of teachers reported supporting collaboration at least weekly, 71–95 percent reported modeling and using metacognitive routines, and 73–94 percent reported modeling and using comprehension strategies.⁷ Thus, the fidelity findings indicate that teachers were implementing the practices, but generally less frequently than expected.

Facilitators and Barriers

Interview, observation, and survey data provided insight into the conditions that supported or impeded RA implementation. Teachers across cohorts were represented in the survey, whereas

⁷ Cohort A, which was most impacted by pandemic-related constraints and only engaged in virtual professional learning during their first year, consistently had the lowest rates of implementation.

only select Cohort A and C teachers were interviewed in spring 2022, 2024, and 2025. SRI interviewed some of the same Cohort C teachers in 2024 and 2025 to learn about their experiences across the two years of implementation. Interviews addressed both facilitators and barriers, while the open- and closed-ended survey questions focused specifically on challenges. Because the site visit samples may overrepresent educators with stronger uptake of RA, the qualitative findings should be interpreted as illustrative rather than representative of all participating districts. The findings below describe how four areas—the design and execution of professional learning, school-based leadership and collaboration, integration into existing instruction, and perceptions of student engagement—facilitated or impeded implementation.

Experiential, in-person professional learning helped teachers connect RA to their instruction, while sustained engagement and collaboration were more difficult in online small-group coaching.

Participating teachers valued opportunities to experience RA as learners while facilitators modeled the instructional practices and RA routines they were expected to use with students. They also appreciated discussing implementation with colleagues at their school sites and gradually revising ideas as their understanding of RA developed. Some teachers reported that concepts introduced early in the professional learning became more meaningful after they had experience applying RA in their classrooms.

I wish I could go back to the first days of training and reprocess it because I would see it differently now. (Teacher, interview, 2024)

Most participating teachers preferred these in-person experiences to online small-group coaching. Participants enumerated what they believed to be barriers to meaningful participation in the online sessions: technology problems, distractions, scheduling conflicts, frequent transitions involving breakout rooms and resources, and fewer opportunities to interact with colleagues or obtain feedback from RA facilitators. Some also found the online sessions confusing or difficult to apply to their classroom practice.

Teachers noted a preference for more in-person contact, clearer advance information about session content, fewer transitions, and more opportunities to receive feedback on their implementation. Taken together, teachers' accounts suggest that the professional learning was most useful when it allowed them to see RA core practices, particularly metacognitive routines, modeled, discuss them with colleagues, and receive feedback as they applied them in their classrooms.

Teacher leaders and collaborative school structures supported implementation, although their influence depended on clear expectations and administrative support.

Across the three years of interviews, educators consistently identified support from school leaders, teacher leaders, and colleagues as an important facilitator. Professional learning

communities (PLCs), joint planning sessions, and informal conversations gave teachers opportunities to share experiences, work through challenges, and integrate RA core practices in ways that fit their contexts. School leaders supported implementation when they understood the demands placed on teachers, protected time for collaboration and planning, and communicated clear expectations.

Teacher leaders played a particularly important role in supporting deeper implementation of RA within their schools. They frequently demonstrated the most integrated classroom implementation and helped colleagues connect individual RA routines to broader shifts in instructional practice. They modeled RA routines and how to integrate them into existing curriculum, facilitated collaborative learning, provided individual assistance, and developed resources that teachers could use in their classrooms. Their influence was stronger when administrators actively supported RA and gave them time and authority to work with colleagues.

Our principal has been incredibly supportive. ... He has done extremely well making sure that RA is happening during the school day, and to help us make the teachers see that it's not extra. (Teacher leader, interview, 2025)

In other settings, unclear expectations, insufficient time, or leadership turnover constrained the role of teacher leaders. A few teacher leaders also hesitated to view themselves as leaders or experts and reporting needing additional guidance about how to support their peers.

Teachers integrated RA more deeply when they had time, flexibility, and support to connect it to their existing curricula; implementation was more limited when they experienced RA as an additional demand or a collection of supplemental strategies.

Teachers described a substantial initial workload associated with finding appropriate texts, adapting lessons, reflecting on their practice, and helping students become accustomed to unfamiliar routines. Competing professional learning initiatives, testing demands, and limited planning or instructional time compounded these challenges. Some social studies and science teachers also had difficulty seeing how RA applied to their disciplines, which contributed to less consistent buy-in outside of ELA.

For the first few PDs in person and online, I was still a bit confused. Wasn't quite sure what to do in a science classroom. ... I was trying to work out how to apply it to science. (Teacher, interview, 2024)

Over time, however, teachers' experience often changed as they became more familiar with the RA approach. In 2024, interview participants described beginning RA with limited choice and sometimes substantial skepticism but reported that their buy-in generally increased throughout the school year. Among teachers interviewed across both years of implementation, some said that RA clicked during the second year, when they stopped viewing it as additional content and more fully integrated it into their existing curricula. These teachers described broader changes

in their instructional stance, including placing more responsibility for meaning-making on students.

As I started getting a little better with these strategies that I've been using, I've seen myself talking less and the students are doing more ... They were more responsible when it came to doing work. (Teacher, interview, 2024)

By contrast, teachers who continued to view RA primarily as a collection of supplemental strategies tended to implement isolated routines without substantially changing classroom discourse or students' intellectual roles.

Teachers perceived student engagement as a consistent challenge, although they also described ways that RA supported engagement and belonging.

A perceived lack of student engagement was the most frequently reported challenge in open-ended survey responses, appearing in 34 percent of responses in 2021–22, 42 percent in 2022–23, 47 percent in 2023–24, and 35 percent in 2024–25. Interview participants described students who were reluctant to read aloud, write, collaborate, share their thinking, or take on greater intellectual responsibility.

They are so ready to let me think for them, and they just sit there and receive the knowledge. I'm saying to them, "No. I'm going to be quiet. I'm going to step back, and I want you to think and you to talk it through." (Teacher, interview, 2025)

Teachers linked some of this disengagement to attention-diverting technology and students' discomfort with unfamiliar or open-ended routines.

Teachers generally described these engagement challenges as broader concerns related to reading and school rather than reactions to RA itself. In fact, some teachers said that RA's collaborative routines and emphasis on making thinking visible helped them build stronger relationships, establishing more welcoming classroom environments, and increase students' willingness to participate.

It's changed the classroom culture, especially with the class norms and things of that nature. It makes for a safer environment for, I think, my kids to take that risk. (Teacher leader, interview, 2025)

Thus, low student engagement sometimes made RA routines more difficult to implement, while those same routines could support greater participation as teachers developed the relationships, scaffolds, and classroom norms.

Taken together, the findings suggest that implementation depended on more than participating teachers' exposure to RA routines. Deeper integration was supported when professional learning modeled the intended instruction, educators had opportunities to learn with colleagues, teacher

leaders received organizational support, and teachers understood RA as a way to teach their existing content. Implementation remained more limited when online professional learning was difficult to engage with, local leadership structures or support was unclear, or teachers perceived RA as an additional demand that competed with existing priorities.

Outcomes

The overall mid- to low-level implementation of RA at the thresholds set by the study suggested that the impact on students would likely be low as well. Nonetheless, we examined whether the model as implemented had effects on students. In future analyses, we will examine whether impacts vary based on the thresholds set for implementation (e.g., implementing key practices once a week, rather than 2–3 times a week).

In this section, we discuss first descriptive methods and findings related to short-term changes in student behavior, as reported by study teachers. Then, we present impact findings related to longer-term changes in student proximal (course grades) and distal (test scores and attendance) outcomes.

Descriptive Outcomes

Given the switch from a randomized controlled trial to a quasi-experimental design (see the Introduction), SRI did not have data on comparison students' behaviors, nor did SRI have direct observations of either teacher practice or student short-term outcomes. Therefore, Research Question 3 (RQ 3) was addressed using descriptive statistics and qualitative analysis of teacher self-report survey items.

Methods

For RQ 3, SRI examined teacher-reported changes in students' classroom behaviors using an annual teacher survey. Three banks of items asked teachers about their students' classroom behaviors; these item banks were mapped to three constructs represented in the short-term outcomes of the RA logic model (Exhibit 1): Self-regulation and persistence (academic mindsets) and collaboration (use of core RA practices). For each construct, SRI combined teachers' responses to the items into a scale using a simple mean approach.⁸ Additionally, open-ended items asked teachers about what changes they saw in their students, and responses addressed other aspects of student short-term outcomes, such as metacognition and engagement.

The sample for RQ 3 included all five study cohorts but was limited to teachers who responded to the annual survey. To address RQ 3, SRI examined the scale mean for each of the three scales (self-regulation, collaboration, and persistence) in the first year of implementation and compared it with the second year of implementation. Researchers used *t*-tests to establish whether these differences were statistically significant.

To analyze open-ended questions, SRI engaged in an iterative coding process, identifying a priori and deductive codes for individual items, coding all participant responses, and noting thematic patterns within and across items.

⁸ The full item text of these survey items is shown in Exhibit B-1. Item-level responses are shown in Exhibit B-2 and scale alpha reliability coefficients are provided in Exhibit B-3 Appendix B.

Findings

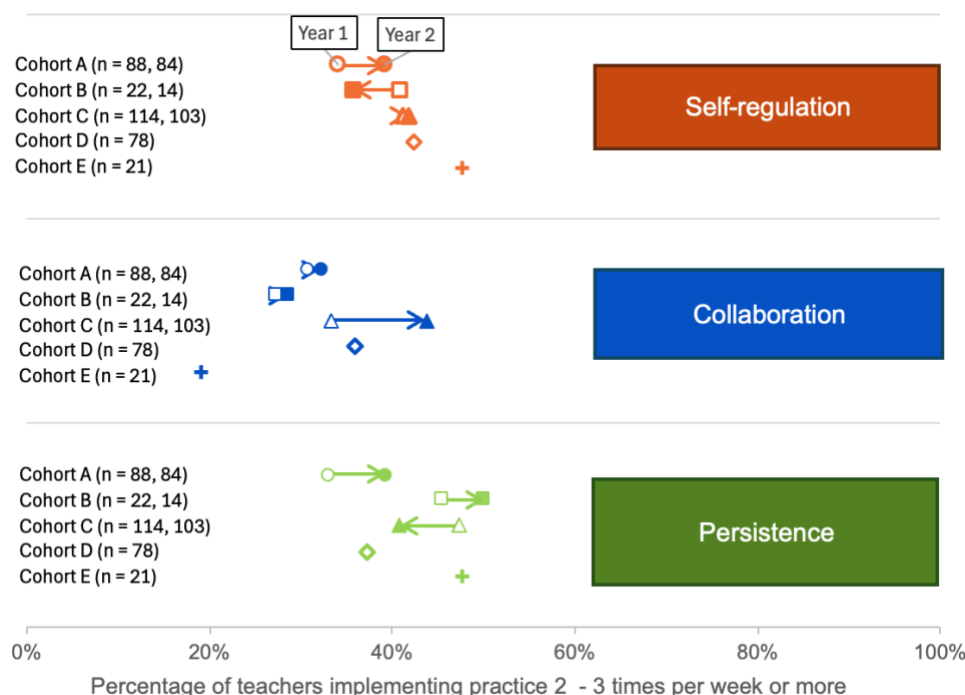
The survey included closed-ended items about students' activities and behaviors in the classroom as well as open-ended items about changes teachers perceived in their students as a result of RA.

Teachers reported that students practiced self-regulation and persistence more often than collaboration.

Teachers most commonly reported that their students practiced self-regulation and persistence, with 33–57 percent of teachers reporting self-regulation behaviors 2–3 times a week and 33–50 percent of teachers reporting the same for persistence behaviors (Exhibit 11). Teachers less often reported that their students engaged in collaboration activities, with 19–44 percent of teachers reporting collaborative activities 2–3 times a week or more. Although teachers' reports of most of these behaviors did not change significantly between their first and second year of implementation,⁹ the overall direction was positive for most items and in most cohorts.

Cohort D, which implemented RA for only one year, reported these student activities at similar rates to those of teachers in other cohorts. Cohort E's reporting was more varied, with higher reports of self-regulation and persistence behaviors and lower reporting of collaboration.

Exhibit 11. Teacher-Reported Frequency with Which Students Practiced Self-Regulation, Collaboration, and Persistence in the Classroom



Source. Teacher survey items WEQ18, WEQ14, and WEQ16 (see Exhibit B-1 for item text).

⁹ See Exhibit B-4 in Appendix B for *t*-tests for individual items and scales.

Teachers reported improvements in student comprehension and metacognition, collaboration, and engagement with reading.

The survey provided respondents with an opportunity to describe any changes they had seen in students that they attributed to RA. Across all four years, roughly two thirds of teachers (ranging from 67 to 79 percent) who responded to this open-ended question reported some form of improvement in students. The most salient areas of growth that teachers described were student comprehension and metacognition, collaboration, and engagement with reading.

The most common theme across all four years was improvements in student comprehension or activities related to metacognition and comprehension. Nearly a third of respondents across all years described students gaining a deeper understanding of texts or practicing metacognitive activities that may facilitate comprehension (e.g., talk-to-the-text or annotation) with more frequency. This theme continued in Year 4, when 29 percent of respondents saw improvement in this area.

I've noticed meaningful changes in my students that I can directly attribute to Reading Apprenticeship. One student ... finds herself talking to the text and making annotations even when reading books for fun. She said she can't read the same way anymore because she is constantly engaging in metacognitive conversations with the text. This is a clear sign that the RA strategies are becoming internalized and transferable beyond the classroom. It's encouraging to witness students developing lasting habits of mind regarding reading. (Social studies teacher, open-ended survey response, 2025)

Other teachers also noted that RA-related practices improved student collaboration. The frequency of this theme increased over time, with 9 percent of respondents reporting it in 2021–22, 13 percent in 2022–23 and 2023–24, and 21 percent in 2024–25. Teachers described students who were more willing to speak to peers, more willing to ask questions in large-group discussions, and more likely to hold each other accountable for their learning.

I do see [that] the students are quicker to volunteer to read aloud now. They also are better at trying to collaborate with each other and hold each other accountable for what they are expected to learn. (ELA teacher, open-ended survey response, 2025)

Finally, teachers also thought their students displayed higher engagement in reading because of RA. As with student collaboration, this theme grew in prevalence across the years, with 7 percent of teachers mentioning it in 2021–22, 17 percent in 2022–23, 20 percent in 2023–24, and 19 percent in 2024–25.

I've seen a noticeable shift in student engagement and confidence. Students are more willing to take academic risks, especially when sharing their thinking aloud or participating in peer conversations. (ELA teacher, open-ended survey response, 2025)

Other themes also emerged over the years. Several respondents in 2021–22 and a few in 2024–25 observed that students had increased confidence when reading. Also, a few respondents in 2023–24 and 2024–25 saw improvements in student fluency or stamina. Relatedly, a few respondents in 2024–25 observed students becoming increasingly comfortable with struggling while reading.

Conversely, across all years, a subset of respondents (ranging from 21 to 33 percent) did not see changes in their students that they could attribute to RA. When giving a reason for this determination, a few respondents in this subset reported a lack of student buy-in, while others described how they simply had insufficient time to implement RA (see Facilitators and Barriers section for additional information on challenges).

Impact Outcomes

To address RQs 4 and 5, the impact study featured a quasi-experimental optimal matching approach in which schools that implemented RA were matched with schools with similar students that did not implement RA.

Preregistration of the Study Design

The confirmatory research question (RQ 4) and impact study design were preregistered with the Registry of Efficacy and Effectiveness Studies (REES) with Registry ID 19740.1.

Samples

The study samples varied by research question. The confirmatory research question (RQ4) focused on the impact of RA after two full years of implementation and was limited to Cohorts A, B, and C. The exploratory research question (RQ 5) examined the impact after only one year and included Cohort D (North Carolina schools) as well.¹⁰

One-Year Impact Sample

The one-year impact sample included four cohorts of schools that participated in the RA project for at least one year (Cohorts A, B, C, and D; see Exhibit 12) and their matched comparison schools.

Treatment was offered to all teachers of grade 9 and 10 mainstream ELA, social studies, and science courses (i.e., not special education or English learner pull-out courses) in study schools. Therefore, within these schools, the student sample included grade 9 and 10 students in these courses who had valid outcome data. (See Exhibit C-1 in Appendix C for the number of students excluded by these criteria.)

The student sample was *not* limited to students in courses with teachers who took up the intervention (i.e., those who attended RA professional learning and coaching as described in the

¹⁰ Pennsylvania schools (Cohort E) are excluded from both impact samples.

Implementation Study section). The North Carolina Education Research Data Center (NCERDC) provided de-identified data, so SRI was unable to separate teachers who were offered the intervention (grade 9 and 10 teachers in ELA, social studies, and/or science) from those who fully implemented the project (e.g., attended PL). Thus, this analysis functions as an intent-to-treat approach. However, considering the average size of the participating districts and schools, the percentage of eligible 9th and 10th grade ELA, science, and social studies teachers who participated in RA, and the teacher participation rate of 87 percent, it is likely that most students had at least one if not several teachers who received at least some RA training.

For the one-year impact study, the treatment group was also limited to students who had the potential to be in an RA-trained teacher's classroom for *at most* one year (not the full two years of the intervention). This group included grade 9 and 10 students in the study schools' first year of implementation and grade 9 students in the schools' second year of implementation. The grade 9 students in the second year of implementation were new to the treatment high schools, so they could not have been exposed to RA-trained teachers in the previous year, when they were in middle schools. Exhibit 12 shows the groups of treatment students, by cohort and grade level, who were included in the analyses. The control group was defined as students who were in matched comparison schools and in focal courses in the same years as treatment students.

Exhibit 12. One-Year Impact Treatment Group Sample

Cohort	2020–21	2021–22	2022–23	2023–24	2024–25
Cohort A	Grade 8 (Baseline)	Grade 9 (outcome)			
	Grade 9 (Baseline)	Grade 10 (outcome)			
		Grade 8 (baseline)	Grade 9 (outcome)		
Cohort B		Grade 8 (baseline)	Grade 9 (outcome)		
		Grade 9 (Baseline)	Grade 10 (outcome)		
			Grade 8 (baseline)	Grade 9 (outcome)	
Cohort C			Grade 8 (baseline)	Grade 9 (outcome)	
			Grade 9 (Baseline)	Grade 10 (outcome)	
				Grade 8 (baseline)	Grade 9 (outcome)
Cohort D				Grade 8 (baseline)	Grade 9 (outcome)
				Grade 9 (Baseline)	Grade 10 (outcome)

Two-Year Impact Sample

The two-year impact sample included three cohorts of schools who participated in the RA project for two consecutive years (Cohorts A, B, and C; see Exhibit X) and their matched comparison schools.

As in the one-year outcomes analysis, the student sample included only students in mainstream ELA, social studies, and/or science courses (i.e., not special education or English learner pull-out courses). However, the two-year impact analysis included only students who had been in treatment schools and focal courses in both grade 9 (the first year of implementation for their school) and grade 10 (the second year of implementation for their school). Thus, these students had the potential to be exposed to RA-trained teachers for the full two years of the intervention. Exhibit 13 shows the groups of treatment students, by cohort and grade level, who were included in the two-year impact analyses.

Exhibit 13. Two-Year Impact Treatment Sample

Cohort	2020-21	2021-22	2022-23	2023-24	2024-25
Cohort A	Grade 8 (Baseline)		Grade 10 (outcome)		
Cohort B		Grade 8 (baseline)		Grade 10 (outcome)	
Cohort C			Grade 8 (baseline)		Grade 10 (outcome)

As in the one-year impact analysis, the student sample could not be limited to students with teachers who took up the intervention. The comparison group was defined as students who were in matched comparison schools in both years. Thus, the comparison students had no exposure to the intervention.

As in the one-year impact sample, the sample for each analysis included only students who had valid outcome data. (See Exhibit C-3 in Appendix C for the number of students excluded by each criterion.)

Study Design

The study design for RQs 4 and 5 used a variation on the optimal propensity score matching approach (Zubizarreta & Keele, 2017), in which students are matched first and then the schools with the best matches are selected to serve as the comparison group. This approach results in balanced covariates at both the school and student levels.

For each outcome sample, SRI first used logistic regression to fit propensity scores at the student level, starting with the variables shown in Exhibit 14.

Exhibit 14. Variables Used for Weighting and Matching and as Model Controls

Variable		Definition
Student variables	Baseline achievement	For grade 9 outcomes, the grade 8 GPA in ELA, social studies, and science For grade 10 outcomes, the grade 9 GPA in ELA, social studies, and science
	Race/ethnicity	Binary variables created from race/ethnicity: • African American or Black non-Hispanic • Asian non-Hispanic • Hispanic, any race • Other race non-Hispanic • White non-Hispanic serves as the comparison group
	Gender	Male vs. female ^a
	Special education status	Identified as a student with disabilities
	English learner status	Identified as an English learner
	Socioeconomic status	Flagged as “economically disadvantaged”
Teacher variables	Certified in grade 9 or 10	Teacher license area code includes a certification for students in grade 9 and/or grade 10
	Years of experience	Years of experience in North Carolina schools
	Highest degree	Teacher’s highest documented degree is Masters or Doctorate vs. Bachelors
School variables	Percent white	Calculated percentage of students identified as “white” race/ethnicity in the data provided
	Percent male	Calculated percentage of students identified as “male” gender in the data provided
	Percent receiving special education services	Calculated percentage of students identified as “students with disabilities” in the data provided
	Percent English learners	Calculated percentage of students identified as English learners in the data provided
	Percent low-income	Calculated percentage of students identified as “economically disadvantaged” in the data provided

^a The datasets from NCERDC included only these two genders.

SRI then ranked all nontreatment schools in North Carolina by both the average propensity score and the total propensity score across all of their students. Finally, we selected the comparison schools with the highest rank (highest total and average propensity score) at a rate of three comparison schools for every one treatment school. The matching rates are shown in Exhibit C-2 and C-4 in Appendix C.

The propensity scores used to identify matched comparison schools were also transformed into student-level weights using the following equation:

$$\frac{pr}{1 - pr}$$

These weights were applied in the final models, and the variables used to create them were entered as covariates.

Measures

RQs 4 and 5 examined impacts on student attendance and two measures of student achievement (end-of-course state standardized tests and course grades). Students' attendance percentage was constructed by dividing the total days present during the school year by the total days enrolled and then multiplying by 100. In North Carolina, students take state standardized end-of-course assessments in English II and Biology, most commonly at the end of grade 10. This study used students' scale scores on those assessments as the test score outcomes.

Course grades were transformed from their original format to use a five-point scale ranging from 0 (F or fail) to 4 (A). Exhibit C-5 in Appendix C shows a crosswalk between the original values found in the variable in the NCERDC data and the value used for the constructed course grade variable.

Missing Data

Students were missing data in outcomes, baseline measures, and covariates, with the proportion missing varying from 0 to 49 percent, depending on the measure and the student group. Each outcome sample was defined as students who were not missing the outcome measure. To minimize bias associated with outcome imputation, SRI did not impute for missing outcomes. For missing baseline and demographic variables, SRI used mean imputation and entered dummy variables for each imputed variable into the impact models.

Impact Models

The functional form for the impact models varied by outcome. Ordinal logistic regression models were used for course grades, linear regression was used for end-of-course test scores, and censored linear regression (Tobit) models were used for daily attendance, with a lower limit of zero and an upper limit of 100. All impact models started with two levels (students nested within schools).¹¹

Broadly, the models took the following form:

$$y_{jk} = \gamma_{00} + \gamma_{10}Treatment_k + \gamma_{20}School_k + \gamma_{30}Teacher_k + \gamma_{01}Student_{jk} + e_k + u_{jk}$$

In this model, y_{jk} represents the outcome of interest (course grades, test scores, or attendance), and it is regressed on a binary treatment indicator at the school level ($Treatment_k$) and vectors of school-, teacher- and student-level covariates ($School_k$, $Teacher_k$, and $Student_{jk}$, respectively)

¹¹ SRI also explored three-level models, using district as a nesting variable. However, those models would not converge for any outcomes. Additionally, the two-level models for science test scores would not converge and thus used a single-level approach.

detailed in Exhibit 14.¹² School and student error terms are represented as e_k and u_{jk} . All models also applied the student-level propensity scores created in the first step of the matching approach as weights, to account for any remaining preexisting differences between students in the matched schools.

Baseline Equivalence

To meet What Works Clearinghouse (WWC) standards with reservations, each analysis must demonstrate equivalence on one or more baseline measures. For daily attendance, the baseline measure was the same metric at baseline as at outcome. However, the same approach was not possible for course grades or end-of-course test scores.

High school students in North Carolina are usually assessed in ELA and science only in grade 10 (the English II and Biology end-of-course assessments). Therefore, there was no parallel baseline measure for students in the one-year test-score impacts, as those students usually did not have ELA or science end-of-course test scores in grade 9. For the two-year impacts, SRI explored using grade 8 end-of-grade test scores as baseline, but extensive missing data made this approach unfeasible. Likewise, student course-grade outcomes also suffered from high levels of missing data in course grades in the prior year, due to poor course-grade reporting. Therefore, for test scores and course grades, SRI used student GPA as the baseline measure and also assessed equivalence on an additional characteristic: student race/ethnicity (white, non-Hispanic).

As shown in Exhibit 15, all analysis samples met the baseline equivalence requirement, with baseline differences less than 0.25 standard deviations. Additional information on baseline equivalence is shown in Exhibits C-9 through C-18 in Appendix C.

Exhibit 15. Standardized Baseline Differences Between Treatment (T) and Comparison (C) Students

Grade Level	Outcome	Standardized Difference Between T & C at Baseline		Number of Students
		Baseline Measure	White, Non-Hispanic	
Grade 9 one-year outcomes	ELA course grades	0.025	0.014	61,426
	Science course grades	0.006	0.040	58,409
	Social studies course grades	0.095	0.001	72,976
	Attendance	0.012	0.054	159,363

¹² Because of differences in the samples, the final set of covariates varied by outcome. Final covariates for each analysis are shown in Exhibit C-6 through Exhibit C-8 in Appendix C.

Grade Level	Outcome	Standardized Difference Between T & C at Baseline		Number of Students
		Baseline Measure	White, Non-Hispanic	
Grade 10 one-year outcomes	ELA course grades	0.001	0.004	115,977
	Science course grades	0.003	0.042	101,049
	Social studies course grades	0.059	0.225	153,789
	ELA test scores	0.181	0.077	131,977
	Science test scores	0.143	0.108	88,581
	Attendance	0.007	0.064	234,584
Grade 10 two-year outcomes	ELA course grades	0.155	0.072	75,586
	Science course grades	0.169	0.076	57,963
	Social studies course grades	0.205	0.074	85,519
	ELA test scores	0.225	0.112	103,130
	Science test scores	0.051	0.143	58,352
	Attendance	0.028	0.148	124,100

Note. Differences are shown in standard deviation units (Cohen's *d* for the continuous achievement variable and Cox's index for the binary low-income indicator).

Impact Findings

This section describes findings for both the confirmatory and exploratory student impact analyses. As discussed in the Implementation section, these impacts are based on a mid- to low-level of implementation of the RA model.

One-Year Impacts

One year in a school participating in RA in North Carolina had a positive, statistically significant impact on student course grades in science in both grade 9 and grade 10 (Exhibit 16). These impacts were moderate, with effect sizes close to a third of a standard deviation.

However, there were no significant impacts on ELA or social studies course grades, nor were there impacts on the more distal outcomes of attendance or test scores, and the direction of the coefficients was not consistent.

Exhibit 16. Impact of One Year of RA on Student Outcomes

Outcome	Estimate ^a	SE	p	Effect Size ^b	N Schools	N Students
Grade 9 ELA course grades	1.5	0.41	.150	0.24	138	61,426
Grade 10 ELA course grades	1.1	0.18	.507	0.07	158	115,977
Grade 9 science course grades	1.7*	0.39	.012	0.34	135	58,409
Grade 10 science course grades	1.6**	0.27	.004	0.29	156	101,049
Grade 9 social studies course grades	0.9	0.19	.814	-0.03	139	82,952
Grade 10 social studies course grades	0.9	0.11	.670	-0.03	168	153,789
ELA test scores	0.45	0.653	.494	0.05	168	131,977
Science test scores	-0.30	0.841	.723	-0.01	164	88,581
Grade 9 attendance	-0.36	0.630	.570	-0.03	168	159,363
Grade 10 attendance	0.25	0.822	.761	0.02	168	234,584

^a Course grades estimates are expressed as odds ratios. Test score estimates are in the original scale of the test. Attendance estimates are in percentage points.

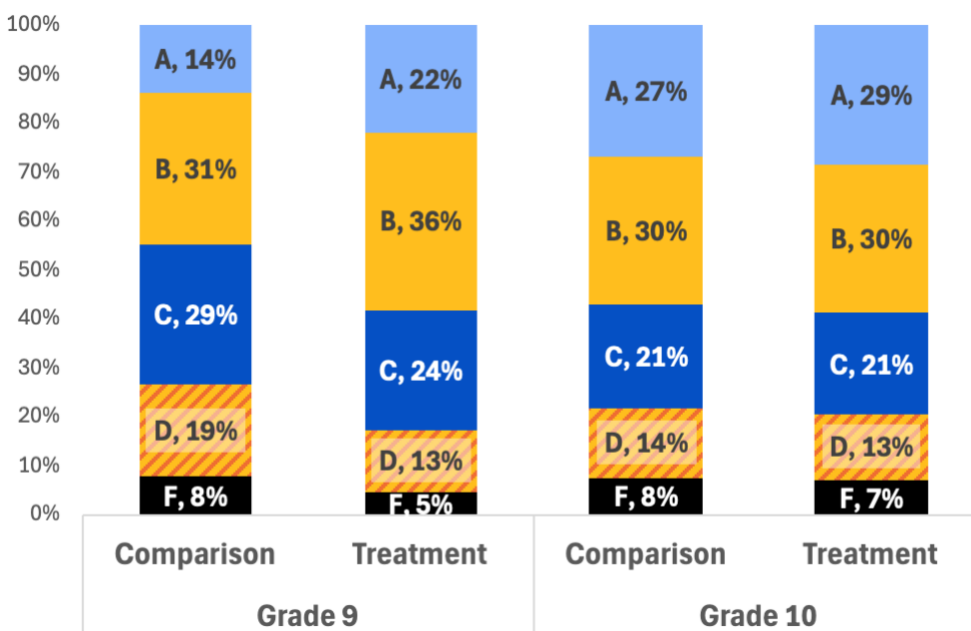
^b Effect sizes for test scores and attendance are Cohen's d, (estimate is divided by the pooled standard deviation of the outcome). Course grade effect sizes are Cox's index (the natural log of the odds ratio divided by 1.65).

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

Another way to understand the impact of RA on students' course grades is to examine the marginal effects for each grade category. Exhibit 17 shows these estimates in grade 9 and 10 science for treatment and comparison students. Twenty-two percent of grade 9 students in the treatment group and 14 percent of grade 9 students in the comparison group had an estimated course grade of A. Likewise, in grade 10, 29 percent of treatment students and 27 percent of comparison students had an estimated course grade of an A.¹³ Additionally, slightly lower percentages of treatment students had estimated grades of D or F, compared with comparison students. This finding demonstrates that, although the effect sizes are moderate and the differences are statistically significant, the effect of this low-level implementation of RA on course grades is relatively small in practice. In other words, a single year of RA, implemented at a lower level than intended, has a small impact on the most proximal outcome for which this study had impact data.

¹³ The relationship between the odds ratios reported in the Exhibit 16 and the percentages in Exhibit 17 is roughly the following: odds ratio = $(pt/(1 - pt))/(pc/(1 - pc))$, where pt is the proportion of As in the treatment group and pc is the proportion of As in the comparison group. In grade 9, the odds for treatment students of getting an A is $0.22/(1 - 0.22) = 0.28$. The odds for comparison students getting an A is $0.14/(1 - 0.14) = 0.16$. The ratio of these odds is $0.28/0.16 = 1.7$.

Exhibit 17. Estimated Science Course Grades in Grade 9 and 10, for Treatment and Comparison Students



Two-Year Impacts

Two years of RA in North Carolina also had a moderate, statistically significant impact on students' course grades in science, and no impact on other course grades or on test scores or attendance. As shown in Exhibit 18, the impact on science course grades is similar to the impact of one year of RA, further indicating that low to mid-range implementation has a small impact on the outcome most within teachers' control (their students' course grades) but does not have a significant impact on more distal outcomes.

Exhibit 18. Impact of Two Years of RA on Student Outcomes

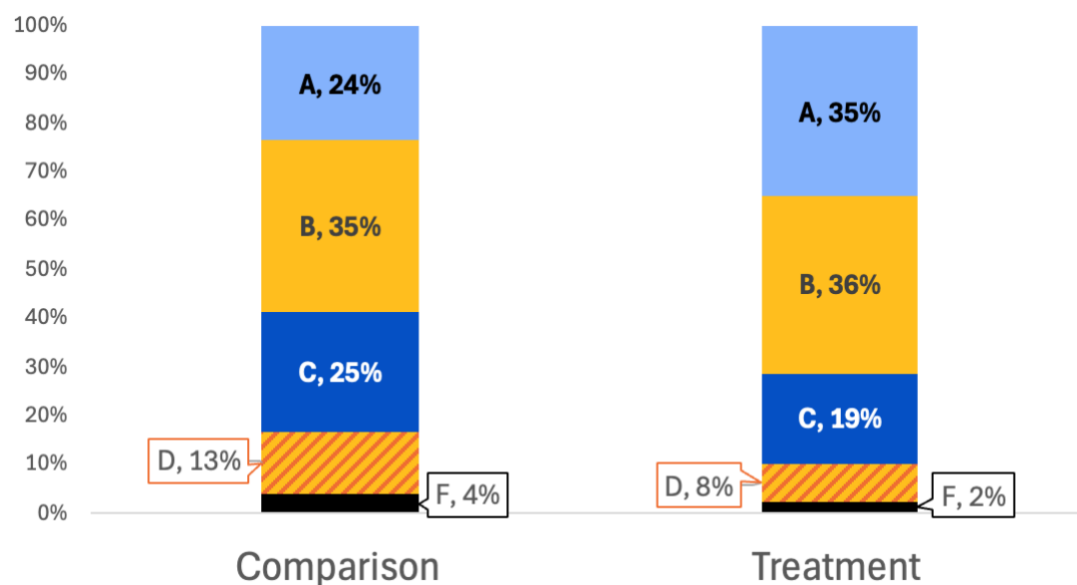
Outcome	Estimate	SE	p	Effect Size	N Schools	N Students
ELA course grades	1.4	0.25	.087	0.19	120	75,586
Science course grades	1.8	0.32	.002	0.35	100	57,963
Social studies course grades	0.8	0.16	.374	-0.11	112	85,519
ELA test scores	-0.48	0.64	.453	-0.052	128	103,130
Science test scores	-0.62	1.27	.628	-0.014	124	58,372
Attendance	1.42	1.01	.157	0.862	128	124,100

** $p < .01$.

As in the one-year impacts, the two-year impact of RA on student science course grades was largely in the estimated percentage of students getting an A. More than a third of treatment students (35 percent) were estimated to get an A in science, compared with 24 percent of comparison students (Exhibit 19). Treatment students also had a lower estimated probability of

getting a C (19 percent compared with 25 percent of comparison students) or a D (8 percent compared with 13 percent of comparison students).

Exhibit 19. Predicted Science Course Grades from Two-Year Impact Models, for Treatment and Comparison Students



Sustainability and Scaling

Alongside assessing implementation and measuring outcomes, the evaluation examined whether participating districts were positioned to continue RA after grant-funded support ended and whether the model could be adapted for broader implementation. This section first describes the RA Project Team's strategy to sustain and scale RA, followed by findings on local plans and capacity, the condensed one-year professional learning model, and implementation in Pennsylvania.

Strategy to Sustain and Scale

The RA Project Team sought to sustain RA in North Carolina by preparing local educators to lead the work to continue implementation after the end of the grant in their rural, low-wealth districts. The original design called for one teacher leader in each participating district to support classroom implementation and facilitate professional learning with colleagues. Teacher leaders received additional training and coaching from the project team to support them in their role. Administrators were also expected to develop some RA expertise and make plans for continuing the project after the grant ended. The original strategy for sustainability also included annual leadership meetings for administrators and teacher leaders and partnerships with North Carolina's Regional Education Service Alliances (RESAs) to support RA beyond the grant period.

The RA Project Team also created a pathway through which select teacher leaders could become certified RA facilitators. Participants completed additional training, observed RA professional learning with attention to facilitation, and then began co-facilitating with expert RA facilitators. The design called for at least 10 teacher leaders to complete this additional training and join the national RA facilitator network. This approach was intended to give districts access to local facilitators who could provide support that fit their local contexts.

The RA Project Team revised the teacher leader strategy as implementation progressed. In Cohorts A and B, teacher leader development began during participating districts' second year. However, in response to district and evaluation feedback, the project team moved the start of teacher leader training and engagement to the first year for Cohort C. The project team also encouraged each district to form a leadership team made up of a district leader, school administrator, and teacher leader. This team was intended to be a backbone of support across the district.

The RA scaling strategy also changed in response to recruitment and implementation challenges. When recruitment in North Carolina fell short of the original targets as a result of pandemic-related challenges, The RA Project Team revised the recruitment timeline, adding an additional North Carolina cohort, expanded participation to (non-study) teachers in grades 11 and 12 in some districts, and recruited a scaling cohort made up of rural, high-need districts in

Pennsylvania. The project team also condensed the professional learning for the final North Carolina and Pennsylvania cohorts into a single year consisting of five in-person days and online coaching. The expansion provided an opportunity to implement a streamlined professional learning model in both the original context and a second rural state context and to examine whether the project could maintain fidelity while scaling.

Sustainability and Scaling Findings

Interviewed educators described growing local expertise and plans to continue RA, but evidence that participating districts were broadly prepared to sustain the work was limited.

Nearly all participants interviewed in 2024–25, the final year of interviews, expressed a desire to continue RA. They expected the RA framework’s emphasis on collaboration, an asset-focused classroom culture, and building on students’ experiences with reading to remain part of their instructional practice.

I see people using RA [when the grant ends]. I don’t know that they’re necessarily going to get their book out and keep using that. But hopefully the strategies that they have learned, and the social aspect and the engagement, will continue to be a part of their planning and thinking. (Teacher, interview, 2025)

The strongest evidence of local expertise came from teacher leaders, who demonstrated integrated implementation and supported colleagues by modeling practices, facilitating collaborative learning, and developing resources tailored to their schools. These accounts suggest substantial ownership and capacity in some settings, but they do not indicate how prevalent that readiness was across all participating educators and schools.

Interviewed educators also identified concrete structures that could support continued implementation. Respondents in every district included in the interviews viewed PLCs and other protected opportunities for collaboration as important for helping teachers learn from one another, work through challenges, and remain accountable for integrating RA. Proposed plans to continue RA included beginning-of-year training, brief refresher sessions, mentoring and individual support from teacher leaders, and onboarding materials for new teachers. Some respondents also hoped to extend RA into earlier grades or additional subject areas to establish greater instructional coherence.

The broader evidence, however, supports a more cautious picture of sustainability. The teacher survey did not directly measure educators’ expertise, ownership, or plans to continue RA, and the fidelity findings showed uneven participation in leadership training and limited implementation of core RA practices across districts. Interviewed educators also expressed uncertainty about whether districts could provide teacher leaders with sufficient time and authority, fund training, or maintain commitment amid staff and leadership turnover.

I know that once the grant ends, it will be harder for us as a district to participate. ... I don't know how much we will be able to do that, just with the financial constraints that come with being such a small district. (School leader, interview, 2025)

Thus, SRI's evaluation identified promising local capacity and planning but could not establish that most participating schools and districts were prepared to sustain RA after the grant ended.

The one-year professional learning model appeared to preserve key features of RA and levels of implementation comparable to the two-year model.

The RA Project Team condensed the professional learning for Cohorts D and E into one year, with five in-person days and online coaching. Teachers in these cohorts reported that the professional learning included key RA activities—such as modeling instructional routines, supporting student self-regulation and collaboration, and analyzing disciplinary texts—at rates comparable to or higher than those reported by teachers during the first year of the two-year model. They also rated the quality and relevance of the professional learning as highly as or higher than teachers in earlier cohorts did during their second year. Participation was more limited, however; only six of the 12 districts across the two cohorts met the fidelity threshold for teacher professional learning.

Teachers in Cohorts D and E also reported participating in key professional learning activities (1.1–1.2) at rates higher than the Year 1 averages but lower than the Year 2 averages for Cohorts A through C. However, Cohort D and E teachers perceived the professional learning as including essential elements (1.3–1.6) at the same or higher rates during their one year of implementation than the other cohorts did in their second year. Cohort D and E teachers also implemented core RA practices at similar rates to those of Cohort A, B, and C teachers. Together, these results suggest that the condensed RA professional learning, despite lower teacher attendance, did not compromise content, quality, or implementation in the classroom, which further strengthens the evidence in favor of making RA more sustainable.

Implementation in rural Pennsylvania districts (Cohort E) appeared broadly similar to implementation in rural North Carolina districts, but evidence of fidelity was limited.

The RA Project Team scaled the one-year RA model in six rural, high-need Pennsylvania districts. Pennsylvania teachers who responded to the survey generally rated the content and quality of the professional learning positively and reported using core RA practices at rates similar to those of teachers in the North Carolina cohorts. At the district level, three of the six Pennsylvania districts met the fidelity threshold for teacher professional learning, the same proportion as the North Carolina districts participating in the one-year model.

The available data, however, provide only a partial picture of implementation in Pennsylvania districts. Only 25 of the 49 participating teachers completed the survey, and SRI did not conduct

interviews or classroom observations in these districts. Therefore, SRI could not determine whether survey respondents' experiences represented implementation across their districts or assess the depth with which teachers integrated RA into their instruction. Overall, reported implementation patterns in Pennsylvania did not appear substantially different from those in North Carolina, but the available data were insufficient to assess how consistently or deeply RA was implemented across all six districts.

Cost-Effectiveness

Amidst low- to mid-level implementation, limited evidence of sustainability, and small impacts on student outcomes, SRI examined the costs of Reading Apprenticeship (RA) as implemented, relative to the number of students who benefited from the project.

For costs, SRI took the grant costs (including nonfederal funds, or “matching” dollars) in each year students were served, subtracted the costs of the evaluation, and then summed this amount over the four years of RA implementation. This approach provides a rough estimate that incorporates costs that may not be relevant to districts implementing the project in the future (i.e., grant-specific costs such as recruitment) and omits some costs that were borne by the districts (e.g., administrator time). Additionally, it includes the costs of implementing RA in Pennsylvania, which are not included in the impact estimates. This approach yielded a total cost of \$5.9 million (Exhibit 20).

Exhibit 20. Cost of RA Implementation by Year

Year	Federal + Non-Federal Funds	Evaluation Cost	Project Cost
2021 (half year of implementation – Cohort A)	\$1,178,255.16	\$347,146.24	\$831,108.92
2022 (Cohorts A and B)	\$914,245.77	\$203,900.56	\$710,345.21
2023 (Cohorts B and C)	\$1,414,564.62	\$345,278.41	\$1,069,286.21
2024 (Cohorts C, D, and E)	\$2,172,713.67	\$356,043.41	\$1,816,670.26
2025 (half year of implementation – Cohorts D and E)	\$1,817,742.05	\$327,883.42	\$1,489,858.63
Total across all four years	\$7,497,521.27	\$1,580,252.04	\$5,917,269.23

SRI calculated the number of students who benefited from the RA project as the number of grade 9 and 10 students in the one-year attendance analyses. These analyses had the least missing data and thus represents the largest group of students who may have benefited. This estimate was 57,446 students. Thus, the average annual cost per student was \$109.93 (Exhibit 21).

Exhibit 21. Student Impacts and Cost per Student per Year of RA

Impact	Effect Size	Cost per Student per Year
Grade 9 science course grades after 1 year	0.56	\$109.93
Grade 10 science course grades after 1 year	0.48	\$109.93
Grade 10 science course grades after 2 years	0.57	\$109.93

Conclusion

The RA Project Team implemented Reading Apprenticeship (RA) in 19 mostly rural, high-need North Carolina districts and later extended a condensed version of the model to six Pennsylvania districts. The project operated during a challenging period marked by pandemic recovery, competing instructional demands, staff and leadership turnover, and, in several districts, severe weather disruptions.

Within this context, implementation was mixed. Teacher professional learning and leadership training reached higher levels of fidelity in some districts, but implementation of core RA practices was consistently limited. Surveyed and interviewed educators nevertheless described meaningful changes in their understanding and instruction, particularly when professional learning modeled core RA practices, teachers had opportunities to collaborate, and local leaders supported teachers in connecting RA to their existing curricula. Teacher leaders often demonstrated some of the deepest implementation and helped extend the work to colleagues, although their influence depended on sufficient time, authority, and administrative support.

Despite partial implementation, the impact study found a consistent positive impact for science course grades. One year in a North Carolina school participating in RA had a statistically significant positive effect on science course grades in grades 9 and 10, and two years of participation had a significant positive effect on science course grades in grade 10. These effects were small to moderate in practical terms but appeared across both the one-year and two-year analyses. The study did not find statistically significant effects on ELA or social studies course grades, ELA or science test scores, or attendance. Thus, the evidence of impact was promising but subject-specific and concentrated in course grades—the outcome most proximal to teachers' classroom instruction.

The positive effects on science course grades add to a broader pattern in prior evaluations of RA in science classrooms, which have reported positive findings for science instruction and student achievement (Fancsali et al., 2015; Greenleaf et al., 2011; Gulemetova et al., 2023). Although these studies examined different versions of RA and different outcomes, the recurring pattern raises the possibility that science instruction may be particularly responsive to the RA approach. RA asks students to examine disciplinary texts, self-regulate and monitor using comprehension strategies, make their thinking visible, discuss interpretations and confusion, and collaboratively construct meaning—practices that may align especially well with inquiry-based science instruction. Science course grades may also be more responsive than standardized assessments or attendance to changes in teachers' daily instructional practice.

However, this study did not directly test these explanations. Moreover, the impact analyses measured the effect of attending an RA school rather than the effect of specifically receiving instruction from a teacher known to have implemented RA with depth. Given the limited implementation of core RA practices, the results should be interpreted as the effects of offering

RA under the conditions observed in this study, not as an estimate of what fully implemented RA might accomplish.

The sustainability and scaling findings reinforce the importance of implementation conditions. Interviewed educators described substantial commitment, emerging local experience, and concrete plans to continue RA. Moreover, the condensed one-year model appeared to retain core professional learning content and teachers reported implementing RA at similar rates to those of the two-year model. Findings from Pennsylvania also suggest that the model could be extended to rural settings in a different state context, although the limited data preclude firm conclusions about fidelity. At the same time, the evaluation could not establish that most participating districts had the structures and resources needed to sustain RA.

These findings suggest that future implementation efforts should pair streamlined professional learning with protected collaboration time, clearly defined teacher-leader roles, and sustained administrative involvement. Future research should examine whether these conditions lead to deeper implementation and broader student impacts, investigate why effects were concentrated in science course grades, and assess how impacts vary by implementation depth, instructional context, and student characteristics. Such work is needed to understand RA's potential when implemented more consistently in rural, high-need districts.

References

- ACT. (2025). Profile report – National: Graduating class 2025.
<https://www.act.org/content/dam/act/unsecured/documents/2025-act-profile-report-us.pdf>
- Cartwright, K.B., Bock, A.M., Clause, J.H., Coppage August, E.A., Saunders, H.G., & Schmidt, K.J., (2020). Near- and far-transfer effects of an executive function intervention for 2nd to 5th grade struggling readers. *Cognitive Development*, 56, Article 100932.
<https://doi.org/10.1016/j.cogdev.2020.100932>
- Darling-Hammond, L., Hyler, M. E. (2017). Effective teacher professional development. Research Brief. Learning Policy Institute. <https://eric.ed.gov/?id=ED606743>
- Dewey, D. C., Fahle, E., Kane, T. J., Reardon, S., & Staiger, D. O. (2026). *From learning recession to learning recovery: Understanding the sources of U.S. K–12 improvement*. Harvard University, Center for Education Policy Research.
<https://cepr.harvard.edu/resource/learning-recession-learning-recovery-understanding-sources-us-k-12-improvement>
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, 56(S1), S25–S44. <https://doi.org/10.1002/rrq.411>
- Fancsali, C., Abe, Y., Pyatigorsky, M., Ortiz, L., Chan, V., Saltares, E., Toby, M., Schellinger, A., & Jaciw, A (2015). *The impact of the Reading Apprenticeship Improving Secondary Education (RAISE) Project on academic literacy in high school: Report of a randomized experiment in Pennsylvania and California schools*. Empirical Education.
<https://www.empiricaleducation.com/pdfs/RAISEfr.pdf>
- Goldman, S.R., Britt, M.A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C.D., Shanahan, C., & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246. <https://psycnet.apa.org/doi/10.1080/00461520.2016.1168741>
- Greenleaf, C.L., Hanson, T., Herman, J., Litman, C., Madden, S., Rosen, R., Boscardin, C., Schneider, S., & Silver, D. (2009). *Integrating literacy and science instruction in high school biology: Impact on teacher practice, student engagement, and student achievement*. Final report to National Science Foundation. <https://eric.ed.gov/?id=ED625915>
- Greenleaf, C.L., Litman, C., Hanson, T.L., Rosen, R., Boscardin, C.K., Herman, J., & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of Reading Apprenticeship professional development. *American Education Research Journal*, 48(3), 647–717. <https://doi.org/10.3102/0002831210384839>
- Greenleaf, C.L., Schoenbach, R., Friedrich, L. Murphy, L., & Hogan, N. (2023). *Reading for understanding: How Reading Apprenticeship improves disciplinary learning in secondary and college classrooms* (3rd ed.). Jossey-Bass.
- Gulemetova, M., Elkins, R., Sanchez, R., Barach, I., Chan, V., & Joshi, M. (2024). *An evaluation of Reading Apprenticeship in science classrooms during COVID 19*. AIR.
<https://www.air.org/sites/default/files/2024-05/SETDI-Evaluation-Reading-Apprenticeship-Science-Classrooms-COVID-November-2023.pdf>

- Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). *Conducting implementation research in impact studies of education interventions: A guide for researchers* (NCEE 2023-005). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/use-work/resource-library/resource/tooltoolkit/conducting-implementation-research-impact-studies-education-interventions-guide-researchers>
- Ho, E.S.C., & Lau, K. (2018). Reading engagement and reading literacy performance: Effective policy and practices at home and in school. *Journal of Research in Reading*, 41(4), 657–670. <https://doi.org/10.1111/1467-9817.12246>
- Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, 47(3), 596–632. <https://doi.org/10.3102/0002831209355469>
- National Center for Education Statistics. (n.d.-a). *NAEP report card: Grade 12 reading*. U.S. Department of Education, Institute of Education Sciences. <https://www.nationsreportcard.gov/reports/reading/2024/g12/>
- National Center for Education Statistics. (n.d.-b). *NAEP report card: Reading. Grade 8*. https://www.nationsreportcard.gov/reports/reading/2024/g4_8/?grade=8
- National Center for Education Statistics. (n.d.-c). National trends and student skills: Grade 8. In *NAEP report card: Reading*. https://www.nationsreportcard.gov/reports/reading/2024/g4_8/national-trends/?grade=8
- Pearson, P. D., Sullivan Palincsar, A., Biancarosa, G. & Berman, A. I. (2020). *Reaping the rewards of the Reading for Understanding Initiative*. National Academy of Education. <https://naeducation.org/reaping-the-rewards-of-reading-for-understanding-initiative/>
- Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & Saliba, J. (2026). Stanford Education Data Archive (Version 6.0). <https://purl.stanford.edu/xh833nn4025>
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40–59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Showalter, D., Hartman, S., Johnson, J., & Klein, B. (2019). Why Rural Matters. Why Rural Matters 2018-2019. <https://doi.org/https://files.eric.ed.gov/fulltext/ED604580.pdf>
- Snow, C. E., & Biancarosa, G. (2003). *Adolescent literacy and the achievement gap: What do we know and where do we go from here?* Carnegie Corporation of New York. <https://carnegie.org/publication/adolescent-literacy-and-the-achievement-gap-what-do-we-know-and-where-do-we-go-from-here/>
- Somers, M. A., Corrin, W., Sepanik, S., Salinger, T., Levin, J., & Zmach, C. (with Wong, E.). (2010). *The Enhanced Reading Opportunities Study final report: The impact of supplemental literacy courses for struggling ninth-grade readers* (NCEE 2010-4021). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/use-work/resource-library/report/evaluation-report/enhanced-reading-opportunities-final-report-impact-supplemental-literacy-courses-struggling-ninth>

- Steacy, L.M., Elleman, A.M., Lovett, M.W., & Compton, D.L. (2016). Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: A new approach for testing intervention elements. *Scientific Studies of Reading*, 20(4), 283–295. <https://doi.org/10.1080/10888438.2016.1178267>
- Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778–808. <https://www.jstor.org/stable/24033389>
- WestEd, Learning Policy Institute & Friday Institute for Educational Innovation at North Carolina State University. (2019). *Sound basic education for all: An action plan for North Carolina*. WestEd. <https://wested2024.s3.us-west-1.amazonaws.com/wp-content/uploads/2024/07/11170223/Sound-Basic-Education-for-All-An-Action-Plan-for-North-Carolina.pdf>
- Zubizarreta, J. R., & Keele, L. (2017). Optimal multilevel matching in clustered observational studies: A case study of the effectiveness of private schools under a large-scale voucher system. *Journal of the American Statistical Association*, 112(518), 547–560. <https://doi.org/10.1080/01621459.2016.1240683>



SRI Education, a division of SRI, is helping federal and state agencies, school districts, major foundations, nonprofit organizations, and international and commercial clients tackle some of the most complex issues in education to help students succeed. Our mission is **to reduce barriers and optimize outcomes for all children, youth, and families**. We do this by conducting high-quality research, supporting use of data and evidence, helping to strengthen state and local systems, and developing tools that improve teaching and accelerate and deepen learning. Our work covers a range of topics, including early learning and development, student behavior and well-being, teaching quality, digital learning, STEM and computer science, literacy and language arts, and college and career pathways.

© 2026 SRI International. All rights reserved. SRI International® is a registered trademark, and SRI Education™ is a trademark of SRI International. All other trademarks are the property of their respective owners.

Silicon Valley

(SRI headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
1.650.859.2000

Washington, DC

1100 Wilson Boulevard
Suite 2700
Arlington, VA 22209
1.703.524.2053

education@sri.com

www.sri.com/education-learning/