The Impact of the Reading Apprenticeship Improving Secondary Education (RAISE) Project on Academic Literacy in High School

A REPORT OF A RANDOMIZED EXPERIMENT IN PENNSYLVANIA AND CALIFORNIA SCHOOLS

December 2015

Cheri Fancsali, IMPAQ International

Yasuyo Abe, IMPAQ International

Mikhail Pyatigorsky, IMPAQ International

Lorena Ortiz, IMPAQ International

Vincent Chan, IMPAQ International

Eliana Saltares, IMPAQ International

Megan Toby, Empirical Education Inc.

Adam Schellinger, Empirical Education Inc.

Andrew Jaciw, Empirical Education Inc.

ACKNOWLEDGEMENTS

The randomized control trial reported here was supported by an Investing in Innovation (i3) Validation grant #U39B100255 awarded to WestEd. The study was conducted by IMPAQ International under a subcontract to Empirical Education Inc., which assisted in research design, data collection, and additional exploratory analysis of implementation and had overall responsibility to WestEd for the evaluation. We are grateful for the support and feedback from our Evaluation Advisory Panel members James Kemple, Johannes Bos, Cynthia Coburn, Elizabeth Moje, John Sabatini; our i3 Technical Assistance providers Jennifer Hamilton and Mike Puma; and staff at WestEd. The authors also express deep gratitude to the district staff, teachers, and administrators in the Pennsylvania and California schools for their assistance and cooperation in conducting this research.

ABOUT EMPIRICAL EDUCATION INC.

Empirical Education Inc. is a Silicon Valley-based research company that provides tools and services to help K-12 school systems make evidence-based decisions about the effectiveness of their programs, policies, and personnel. The company brings its expertise in research, data analysis, engineering, and project management to customers that include the U.S. Department of Education, educational publishers, foundations, leading research organizations, and state and local education agencies.

Executive Summary

Nationally, two-thirds of high school students are unable to read and comprehend complex academic materials, think critically about texts, and synthesize information from multiple sources, or communicate what they have learned (NAEP, 2013). Without a substantial change in their academic literacy, U.S. high school students face continued academic problems in high school and college because they are unable to handle the quantity and complexity of assigned reading (ACT, 2012). Further, literacy instruction that fosters the skills and dispositions required for reading comprehension of complex materials is seldom found in U.S. high schools (Duschl, Schweingruber, & Shouse, 2007; Reisman, 2011). Recent research suggests that disciplinary literacy and reasoning skills are rarely a focus of secondary instruction (ACT Inc., 2009, 2013a, 2013b). Teachers report that little time is devoted to supporting reading comprehension (Ness, 2008, 2009; Vaughn et al., 2013). Instead, literacy instruction and activities tend to center on using texts for basic reading comprehension and summary of information (Kiuhara, Graham, & Hawken, 2009), rather than as a core resource for constructing new knowledge (Banilower et al., 2013; Smith & Ochoa-Angrino, 2012).

The Reading Apprenticeship instructional framework was developed by WestEd's Strategic Literacy Initiative (SLI) two decades ago to help teachers provide the literacy support students need to be successful readers in the content areas. It has since reached over 100,000 teachers in schools across the country, at the middle school, high school, and college levels. The Reading Apprenticeship framework focuses on four interacting dimensions of classroom learning culture: Social, Personal, Cognitive, and Knowledge-Building. These four dimensions are woven into subject-area teaching through metacognitive conversation—conversations about the thinking processes students and teachers engage in as they read. The context in which this all takes place is extensive reading—increased in-class opportunities for students to practice reading complex academic texts in more skillful ways. Teachers also work with students on explicit comprehension strategy instruction, vocabulary and academic language development techniques, text-based discussion, and writing. Reading Apprenticeship is designed to help teachers create classroom cultures in which students feel safe to share reading processes, problems, and solutions.

In 2010, WestEd received a "Validation" grant from the Department of Education's Investing in Innovation Fund (i3) competition to scale-up and conduct a randomized controlled trial of the intervention through a project called Reading Apprenticeship Improving Secondary Success (RAISE). RAISE took place in California, Michigan, Utah, Pennsylvania, and Indiana and worked with nearly 2,000 teachers who served approximately 630,000 students during the grant period. This report presents findings from the randomized controlled trial conducted in two of those states: California and Pennsylvania.

Overview of the Intervention

For the RAISE project, WestEd's Strategic Literacy Initiative (SLI) provided high school teachers of English language arts, science, and history in the study with 65 hours of inquiry-based Reading Apprenticeship professional development over the course of 12 months. The professional development was designed to transform teachers' understanding of their role in adolescent literacy development and to

build enduring capacity for literacy instruction in the academic disciplines. These changes in teacher attitudes and instructional approaches are hypothesized to result in changes in student attitudes, motivation, and behavior while at the same time building skills and knowledge for subject-specific literacy tasks, strengthening students' view of themselves as readers and learners, and yielding gains in student achievement.

SLI developed a number of new elements for the RAISE project to support the dissemination and implementation of the Reading Apprenticeship intervention at a broad scale. Specifically, they 1) recruited and trained a cadre of professional development facilitators, 2) appointed state site coordinators to provide support and resources to schools, 3) recruited teacher leaders at each school who held monthly school-based meetings to provide support to teachers throughout implementation, and 4) provided support and resources to school administrators including an on-line course on the framework.

Research Design

The i3 impact evaluation of RAISE, conducted by IMPAQ International and Empirical Education Inc., employed a cluster randomized controlled trial (RCT) in which 42 schools were randomly assigned to a treatment group (22 schools) or a control group (20 schools). English Language Arts (ELA), science, and history teachers recruited from treatment schools received 65 hours of professional development and ongoing support, while control schools conducted business as usual. The schools were recruited and randomized in two waves. In wave 1, 32 schools (17 treatment, 15 control) in California and Pennsylvania were recruited and randomized in 2011, with implementation beginning in fall 2011. We collected three years of data from these wave 1 schools. In wave 2, an additional 10 schools (5 treatment, 5 control) in California were recruited in 2012 to increase the number of schools in our sample serving English learners; implementation began in fall 2012. We collected two years of data from these wave 2 schools.

This was an intent-to-treat design, with impact estimates generated by comparing average outcomes in schools randomly assigned to treatment status with average outcomes in schools assigned to control group status, regardless of the level of participation in or implementation of RAISE instructional approaches after random assignment.

This report presents key implementation and impact findings from the i3 impact evaluation of the RAISE project. Most of the findings in this report are from the sample of students and data collected during teachers' second year in the study, after treatment teachers had received the full "dose" of professional development delivered over 12 months and could therefore be expected to fully implement Reading Apprenticeship. We used the data from the first and third years to conduct supplemental analyses.

Data sources for this report include principal, teacher, and student surveys; professional development observations and attendance records; school district student records; and an assessment of students' literacy skills.

Key Findings about RAISE Implementation

Implementation fidelity and contextual factors that may have facilitated or hindered implementation of RAISE were measured through professional development observations and attendance records, teacher surveys, and principal surveys. These data indicated that RAISE professional development and in-school support was delivered as intended.

- Over 85% of the observed sessions exhibited the key professional development design characteristics including: a focus on practices and collaboration that facilitate metacognitive inquiry and conversations, content focused on disciplinary literacy, and active learning for teachers.
- More than three quarters of teachers met the fidelity threshold set by SLI for attending the RAISE professional development; however, the teachers who met this threshold tended to be clustered in the same schools. Ten out of the 22 (45%) RAISE schools did not meet the school-level professional development attendance fidelity threshold.
- Over 90% of the RAISE schools had a RAISE-trained teacher leader who facilitated monthly team meetings and provided on-site support. A total of 67% of RAISE teachers attending at least 4 of the 10 on-site team meetings per year met the fidelity threshold set by SLI.
- While the program-level fidelity thresholds were met for attendance at the RAISE on-site monthly meetings, attendance varied greatly at the school level, suggesting that building coherence and communities of practice may have been more challenging at certain schools.

Feedback on the training was positive, with teachers who attended reporting that it prepared them to implement the Reading Apprenticeship approach.

- Over 90% of teachers who responded to survey questions about the RAISE professional development felt that it "moderately", "more than moderately", or "completely" prepared them to use the set of literacy practices modeled during the training.

Treatment teachers reported more support for literacy instruction than their control peers and generally held positive views of Reading Apprenticeship and its efficacy. Their survey responses indicated buy-in and commitment to implementing the framework.

- RAISE teachers reported receiving support for literacy instruction at a greater frequency than control teachers, and they rated this support as "very" or "more than moderately" helpful at higher levels than control teachers.
- Over 50% of teachers across subject areas reported believing that Reading Apprenticeship would be "highly" or "more than moderately" effective at improving students' reading comprehension.
- 61% of teachers reported being fully committed to Reading Apprenticeship at the end of year 2.

However, implementation was not without challenges, with most teachers (over 60%) reporting experiencing competing priorities that hampered implementation, such as standardized test preparation or addressing content standards. Contextual factors may also have challenged implementation in some schools. For example, five schools (three treatment, two control) were reorganized into a single school under one principal. Though we do not have any evidence that the reorganization caused "contamination" between treatment and control schools, the disruption likely affected student and teacher data response rates and may have hindered treatment teachers' ability to implement Reading Apprenticeship.

Key Findings on Teacher Mediating Outcomes

Monthly teacher surveys measured the extent to which RAISE had an impact on teacher mediating outcomes including shifts in instructional practice and confidence in literacy instruction. Measured during the second year of implementation, RAISE had statistically significant impacts on teachers' use of core

Reading Apprenticeship practices and on their confidence in delivering literacy instruction with effect sizes ranging from 0.41 to 0.62. The following were areas of impact.

- Employing practices that foster student independence
- Providing opportunities for students to practice metacognitive conversations
- Providing opportunities for students to practice comprehension strategies
- Providing opportunities for student collaboration
- Teacher confidence in literacy instruction

The analyses of teacher survey data suggest RAISE had an impact on reported attitudes and instructional practices in key areas emphasized by the Reading Apprenticeship framework. These areas of impact indicate a substantive shift in teachers' practices away from the tendency to focus on basic reading comprehension and summary of information to focus on close reading and deep engagement with texts to build knowledge—the type of complex disciplinary literacy instruction envisioned by the Common Core State Standards. RAISE teachers were more likely than control teachers to encourage student-directed learning by using practices that foster student independence, providing opportunities for students to practice various reading strategies, and offering opportunities for peer-to-peer learning and collaboration. There were positive, but not statistically significant, differences in two other areas of practice: 1) providing extensive reading opportunities that reflect a variety of genres and text types and 2) promoting and employing instruction that promotes engagement, student-centered learning, and inquiry-based learning.

Among science teachers, we found an additional area of impact on instructional practices emphasized by Reading Apprenticeship: teachers modeling comprehension strategies. Further, in each of the areas where we found positive impacts of RAISE, the effect size for the impacts was larger for science teachers than for ELA and history teachers.

We hypothesize that the additional area of impact and larger effect sizes for science teachers are related to the fact that ELA and history teachers were likely employing some of these practices prior to the intervention, to a greater extent than their science educator peers. Thus, for science teachers, the uptake of Reading Apprenticeship required a larger transformation in their instructional and pedagogical approach, and yielded a larger effect size. Supporting this conjecture, we found that science teachers in the control group did, in fact, report less frequent use of practices indicative of the Reading Apprenticeship approach than ELA and history teachers.

The size of the effects on teacher practice increased between year 1 and 2, especially for teachers' confidence in providing literacy instruction, suggesting that the additional professional development received by teachers in the summer following their first year of implementation, along with the on-site support during year 2, increased teachers' comfort level and ability to implement Reading Apprenticeship.

Key Findings on Student Mediating Outcomes

Changes in teacher practices as a result of RAISE are hypothesized to change students' classroom experiences, attitudes, and behaviors. These mediating student outcomes were measured through a year-end student survey.

- RAISE produced positive and statistically significant impacts on the full sample of students in the following two student mediating outcome domains that are hallmarks of the Reading Apprenticeship framework.
    - Increased integration of reading instruction into content-area teaching
    - Increased metacognitive inquiry
- The size of the impacts on student mediating outcomes increased over time.

The effect sizes of the impacts were 0.21 and 0.18 respectively. Impacts in other areas were positive but not statistically significant including outcomes related to collaboration in a community of readers and writers; use of comprehension strategies; reader identity; and participation in metacognitive conversations.

There was also a statistically significant impact on participation and contribution to class discussions, class time spent reading among science students, and variety of reading material among history students. The effects on ELA students were smaller and not statistically significant.

Key Findings on Student Literacy Achievement Outcomes

Student literacy achievement was measured through an online, scenario-based assessment developed by Educational Testing Service (ETS) for this study. The assessment was designed to measure the strategic reading processes that are primary targets of Reading Apprenticeship and closely aligned with the Common Core State Standards. The assessment was designed to be a more rigorous measure of complex reading comprehension than typical state ELA tests.

- By the end of the second year of implementation, RAISE had a positive and statistically significant impact on student literacy in science classes. The effect size of the impact was 0.32.

This effect size translates into an improvement index of 12.6 percentage points: that is, we would expect control students to move from the 50th percentile to the 62.6th percentile if they were exposed to RAISE. Results for the other two subjects were not statistically significant but with a meaningful effect for ELA classrooms (effect size = 0.22) and a non-meaningful result for history classrooms.

The impact in science is particularly impressive given that implementing the Reading Apprenticeship framework may require a more dramatic change in science teachers' core practices and routines than is needed by ELA and history teachers.

- For the full sample and for key subgroups, including English language learners, low-income students, low prior performers, non-white students, and students in Pennsylvania schools, we found positive but not statistically significant impacts, with effect sizes ranging from 0.11 to 0.25. These results may reflect the study's limited ability to detect a modest size effect.

CONCLUSIONS

Findings from this study demonstrate the success of the RAISE project in providing teachers with training and support at scale to help them change their instructional practices in order to foster metacognitive inquiry and support comprehension, particularly in science. These findings are consistent with positive findings from other studies of Reading Apprenticeship. The primarily positive, yet not statistically significant results for the full sample and subgroups of students, including English language learners, indicate that the study's sample size may not have been large enough to detect a modest size impact.

The results from this study point to several areas in need of further investigation. Specifically, the differences in impact by subject area and state need to be better understood. Further, SLI and the larger field would benefit from additional research on those factors that support bringing the model to scale and generating meaningful classroom-level changes in instruction, particularly for ELA and history teachers. Overall, the study's findings demonstrate the potential of RAISE to address the paucity of content-specific reading instruction in U.S. secondary schools—especially in science, where the need may be greatest.

Table of Contents

Reference this report: Fancsali, C., Abe, Y., Pyatigorsky, M., Ortiz, L., Hunt, A., Chan, V., Saltares, E., Toby, M., Schellinger, A., & Jaciw, A. P. (2015). *The Impact of the Reading Apprenticeship Improving Secondary Education (RAISE) Project on Academic Literacy in High School: A Report of a Randomized Experiment in Pennsylvania and California Schools*. (Empirical Education Rep.No. Empirical_ RAISE-7019-FR1-O.2). Palo Alto, CA: Empirical Education Inc. Retrievable at http://empiricaleducation.com/pdfs/RAISEfr.pdf

Introduction

CONTEXT OF THE STUDY

The Strategic Literacy Initiative (SLI) at WestEd developed Reading Apprenticeship in 1995 to help teachers provide the literacy support students need to be successful readers in the content areas. Reading Apprenticeship has since reached over 100,000 teachers in schools across the country, at the middle school, high school, and college levels. In 2010, WestEd received a validation grant from the U.S. Department of Education's Investing in Innovation Fund (i3) competition to scale-up and study the Reading Apprenticeship Improving Secondary Education (RAISE) project. Through RAISE, WestEd served 1,964 teachers and approximately 630,000 students from 274 schools across five states: California, Indiana, Michigan, Pennsylvania, and Utah. The focus of the RAISE project was supporting teachers to change their classroom practice by integrating active literacy learning into their disciplines (e.g., history, science, and English). Over the grant period, four cohorts of cross-disciplinary school teams across the five RAISE states participated. A number of new elements were developed in RAISE to support the dissemination and implementation of the Reading Apprenticeship intervention at this broad scale.

As part of this project, IMPAQ International and Empirical Education Inc. conducted a rigorous, large-scale randomized controlled trial (RCT) to evaluate the efficacy of RAISE. The RCT took place in two of the five states: California and Pennsylvania. It was designed to test the impact of the Reading Apprenticeship instructional and professional development model under conditions necessary for dissemination at scale. This report presents findings from the impact and implementation study conducted through this RCT. An associated formative evaluation, the "scale-up study" was conducted across four of the five states involved in the project including Pennsylvania, though in schools not participating in the RCT. Schools not included in the RCT are referred to as "scale-up schools."

THE READING APPRENTICESHIP FRAMEWORK AND PROFESSIONAL DEVELOPMENT MODEL

RAISE was an ambitious project developed to address the need for high school instruction that focuses on reading in the content areas. Nationally, two-thirds of high school students are unable to read and comprehend complex academic materials, think critically about texts, synthesize information from multiple sources, or communicate what they have learned (NAEP, 2013). Without a substantial change in their academic literacy, U.S. high school students face continued academic problems in high school and college because they are unable to handle the quantity and complexity of assigned reading (ACT, 2012). Further, literacy instruction that fosters the skills and dispositions required for reading comprehension of complex materials is seldom found in U.S. high schools (Duschl et al., 2007; Reisman, 2011). Teachers report that little time is devoted to supporting reading comprehension (Ness, 2008, 2009; Vaughn et al., 2013), particularly in content-areas. Instead, literacy instruction and activities tend to center on using texts for basic reading comprehension and summary of information (Kiuhara, Graham, & Hawken, 2009), rather than to acquire and construct new knowledge (Banilower et al., 2013; Smith & Ochoa-Angrino, 2012).

Reading Apprenticeship is an instructional framework that helps teachers support discipline-specific literacy and learning in their varied content areas by attending to four interacting dimensions of

classroom learning culture: social, personal, cognitive, and knowledge-building. The **social dimension** involves building community. The classroom becomes a safe environment where students see other students and their teacher as resources for learning. The **personal dimension** includes drawing on students' understandings and experiences, as well as developing students' identities as competent readers, building their awareness of their purposes and goals for reading, and connecting current academic tasks to future career or educational goals. The **cognitive dimension** involves developing students' mental processes, including their text-based problem-solving strategies. The **knowledge-building dimension** includes building students' knowledge—not only of the content of the text, but also of language and word construction, genre and text structure, and discipline-specific discourse practices. The framework targets learning dispositions, as well as literacy skills and knowledge. At the center of Reading Apprenticeship is an ongoing metacognitive conversation—conversations about the thinking processes students and teachers engage in as they read. This conversation is carried on both internally—through metacognitive reading and reasoning routines—and externally—as teachers and students talk about their personal relationships to reading, the social environment and resources of the classroom, their affective responses and cognitive activity, and the knowledge required to make sense of complex texts. This metacognitive conversation takes place through extensive reading, including increased in-class opportunities for students to practice reading complex academic texts in more skillful ways as they collaborate to make meaning of these texts for learning purposes. As implied by the model's name, the core pedagogical stance of Reading Apprenticeship does not involve teachers imparting knowledge to students. Rather, teachers facilitate learning through instruction, modeling, and opportunities to practice in a collaborative social context.

Reading Apprenticeship's inquiry-based professional development is designed to transform teachers' understanding of their role in adolescent literacy development and to build enduring capacity for literacy instruction in the academic disciplines. The inquiry-based professional development model engages teachers in learning about the complexity of literacy and learning with disciplinary texts through the following.

- Experiential learning that mirrors the instructional environment and practices of the framework
- Learning how the framework supports students' literacy and learning
- Practicing specific pedagogies
- Carrying out formative assessment focused on student reading, thinking and learning

Through the i3-funded RAISE project, SLI staff and consultants provided Reading Apprenticeship professional development to up to nine teachers in each of the schools in the RCT, including up to three teachers from each subject area: science, history, and English language arts (ELA). Each teacher was offered 10 days (65 hours) of subject-specific professional development over 12 months: RAISE 5-Day Foundation Institute in the first summer, 2-Day Calibration Institute in the following winter, and a final 3-Day Springboard Institute in the next summer. SLI also developed a cadre of "teacher leaders," with at least one at each school site, who were expected to convene team meetings at their schools to support teachers' implementation of Reading Apprenticeship.

RAISE's key components, its hypothesized teacher and student mediating outcomes, long-term student outcomes, and contextual factors that may facilitate or hinder implementation are shown in the logic model in Table 1. In this model, RAISE professional development and support should lead to changes in teacher instructional approaches and practices, which result in changes in students' attitudes and behaviors, ultimately yielding gains in student achievement.[1]

[1] The achievement outcomes listed in the logic model are those that are hypothesized to increase as a result of RAISE. The i3 validation study did not collect data on all of these outcomes. The literacy assessment designed specifically for this study is the primary achievement outcome in this report.

TABLE 1. LOGIC MODEL OF READING APPRENTICESHIP

1. Inputs	2. Mediating outcomes and outputs		3. Outcomes
A. Teachers and teacher leaders receive 65 hours of Reading Apprenticeship professional development characterized by: Content focused on disciplinary literacy Collective participation Active learning Coherence Inquiry-based professional development on practices and collaboration that facilitates metacognitive inquiry and conversations *B. Teachers participate in follow-up support*: At least monthly on-site teacher meetings facilitated by teacher leaders *C. Teacher leaders are recruited:* One teacher leader in each school *D. Online resources and training* available to RAISE school administrators *E. State coordinators provide school-based support*	*A. Teacher Mediating Outcomes* A1. Teacher leaders support teacher development and implementation of Reading Apprenticeship: At least monthly on-site meetings or opportunities for teacher community and collaboration A2. Teachers increase use of Reading Apprenticeship strategies: ▪ Providing reading opportunities that reflect breadth in genres and text types, frequency, volume, and accountability for reading ▪ Supporting student effort to comprehend disciplinary text ▪ Fostering metacognitive inquiry into reading and thinking processes ▪ Providing explicit instruction and modeling of reading comprehension routines, tools, strategies, and processes ▪ Fostering and supporting student collaboration ▪ Employing instruction that promotes engagement, student-centered learning and inquiry-based learning	*B. Student Mediating Outcomes* B1. Increased collaboration in a community of readers and writers B2. Increased use of comprehension strategies B3. Increased metacognitive inquiry B4. Improved reader identity B5. Improved student identity B6. Increased reading of a variety of texts B7. Increased engagement	*A. Increased achievement, especially among high-need students* A1. Increased disciplinary literacy in science, ELA, and U.S. history, as measured by student literacy assessments A2. Increased content knowledge in science and U.S. history A3. Improved course performance: improved grades; increased number credits earned in core courses A4. Increased promotion and retention: increased likelihood of on-time promotion; increased probability of retention in school

TABLE 1. LOGIC MODEL OF READING APPRENTICESHIP

(decreased dropouts)

4. Factors that facilitate or hinder implementation	
A. Sense of commitment and purpose related to the initiative	F. Support for implementation at site: administrative, social, and material
B. School cohesion and community	G. Burden on teachers (Reading Apprenticeship can be more work for teachers, being responsive to learners in the moment, responding to new goals, new routines, and a higher level of cognitive complexity)
C. Understanding and knowledge of disciplinary ways of thinking	H. Misalignment of district policies with the initiative; curriculum constraints
D. Teacher sense of self-efficacy, confidence	
E. Reducing risk for teachers, especially in evaluation of new practices	

Previous RCTs have tested the efficacy of the Reading Apprenticeship framework and the professional development model in studies with more closely monitored implementation. These studies have demonstrated strong positive effects on teacher practice resulting from the Reading Apprenticeship professional development—most notably, teachers' increased use of reading comprehension strategy instruction, metacognitive inquiry routines, and collaborative learning structures in their classrooms. The studies also showed positive effects on students' literacy and content-area achievement, as well as students' comprehension strategies, identity, motivation, and engagement; English language learners particularly benefited from Reading Apprenticeship instruction (Greenleaf et al., 2011a, 2011b; Kemple et al., 2008; Somers et al., 2010).

Several new components developed to provide Reading Apprenticeship professional development and support for implementation at the necessary scale for the RAISE i3 grant differentiate this RCT study from prior ones. The new components include the following.

- **Training and apprenticing 85 professional development facilitators**—primarily teachers participating in the early cohorts of the intervention themselves—to deliver the revised, discipline-based Reading Apprenticeship professional development series to meet the scale demand for the i3 grant. As part of this, SLI developed extensive and detailed materials, protocols, and assessments to support facilitator development.

- **Recruiting and supporting teacher leaders** for each school team with the expectation that they convene and facilitate monthly on-site team meetings, using protocols provided by the program developers. Teacher leaders were often volunteer teachers participating in the study, but also included curriculum coordinators or school administrators with primary responsibility for supporting teachers, but who were not implementing Reading Apprenticeship in a classroom. Teacher leaders were offered the same 65 hours of RAISE professional development provided to teachers in the initiative and attended an additional teacher leader webinar in the first year, with three face-to-face meetings per year in subsequent years.

- **Establishing and expecting monthly meeting participation** from all participating teachers to discuss Reading Apprenticeship implementation

- **Appointing state-level RAISE coordinators** to provide locally knowledgeable support to RAISE school teams. State coordinators carried out a number of functions necessary to implement grant activities at a distance from the central SLI office. These functions included the following.
    - Communicating and coordinating state grant activities
    - Convening and facilitating cross-state Teacher Leader meetings
    - Working directly with school administrators to enhance their support for RAISE
    - Promoting RAISE work in regional and state-level venues to build sustainability
    - Conducting inquiry into and facilitating conversations about the model with the central SLI office staff and other state coordinators

A state coordinator served the Pennsylvania schools participating in the RCT, as well as the scale-up schools. No state coordinator was appointed in California, however, as that role was played by SLI staff in the California offices at WestEd.

- **Providing RAISE school administrator programs and materials.** SLI provided opportunities during the teacher professional development sessions for RAISE school administrators to share their ideas, needs, and perspectives on their school teams' implementation, as well as the opportunity to participate in an online administrator's course.

Methods

RESEARCH DESIGN

The i3 impact evaluation of RAISE employed a cluster RCT in which 42 schools were randomly assigned to a treatment group (22 schools) or a control group (20 schools). Teachers recruited from treatment schools were offered 65 hours of professional development and ongoing support, while control schools conducted business as usual. The schools were recruited and randomized in two waves: In wave 1, 32 schools (17 treatment, 15 control) were recruited and randomized in 2011, with implementation beginning in fall 2011. We collected three years of data from these wave 1 schools. In wave 2, an additional 10 schools (5 treatment, 5 control) in California were recruited in 2012 to increase the number of schools in our sample serving English learners; implementation began in fall 2012. We collected two years of data from these wave 2 schools.

This report presents key implementation and impact findings from the i3 evaluation. The evaluation used multiple sources of qualitative and quantitative data: principal, teacher, and student surveys; professional development observations and attendance records; school district student records; and an assessment of students' literacy skills. The evaluation addresses the questions outlined below.

Student Impact Questions Regarding Achievement Outcomes

- What are the effects of RAISE on student literacy achievement?
- What are the effects of RAISE on the literacy achievement of English Language Learners?
- What are the effects of RAISE on student literacy by subject area (ELA, history, science)?
- What are the effects of RAISE on student literacy by key student subgroups, including:
    - Students with low prior achievement and/or weak prior performance?
    - Economically disadvantaged students?
    - Minority students?
- What are the effects of RAISE on student literacy by state (California, Pennsylvania)?

Teacher Meditating Impact Question

- What are the effects of RAISE on teacher practices and teacher attitudes?

Student Mediating Impact Questions

- What are the effects of RAISE on student engagement, reading attitudes, and behaviors?
- What are the effects of RAISE on student engagement, reading attitudes, and behaviors of student subgroups, including:
    - English language learners?
    - Students with low prior achievement and/or weak prior performance?

- Economically disadvantaged students?
- Minority students?

Implementation Questions

- To what extent is RAISE implemented in a way that is consistent with the program model and underlying theory of action?
- What are the contextual factors that support or hinder RAISE implementation?

Our confirmatory impact analyses[2] were conducted on the full sample of students and the English language learner subgroup in the second year of implementation. They answered the question: What are the effects of RAISE on student literacy achievement? In addition to this, we conducted a series of exploratory analyses. Specifically, noting that the implementation and contextual factors were different in California and Pennsylvania, we conducted exploratory subgroup analyses by state. We also conducted analyses by subject area, because the developers hypothesized that the impact may be larger in science where Reading Apprenticeship represents a larger pedagogical shift from more traditional or typical science classroom instruction.

The program effects on students and teachers are estimated as the differences in outcomes between treatment and control groups. The impact was estimated as "intent-to-treat" effects of the intervention, which meant that all randomly assigned schools and their study-eligible cohorts of students and teachers were included in the analysis samples, regardless of the level of actual exposure to the intervention (even those who did not end up actually receiving the intervention were included, as long as they were initially "intended" to receive it). We estimated a two-level model, with individual students or teachers nested within schools. The use of this statistical model allowed us to account for clustering effects, as well as to control for baseline covariates, so as to improve the precision of the impact estimates. Additional details on the estimation model and methods are provided in Appendix A. Impact Estimation Model

The implementation analysis provides context for assessing and understanding the measured impacts of RAISE on student and teacher outcomes. Evaluators and program developers created a system of numerical thresholds to evaluate the fidelity of implementation of the RAISE core components over the course of the research study. The core program components—as identified in the logic model (see Table 1)—include: delivery of the RAISE professional development content, attendance at the RAISE professional development, recruitment of a RAISE teacher leader, and attendance at the RAISE team monthly meetings. Each of these core components were assigned a teacher-, school-, and/or program-level threshold that defines adequate implementation as intended by the program developers. Data from teacher surveys and professional development observations and attendance rosters were used to calculate fidelity scores at each level. Data from teacher and principal surveys were used to supplement the fidelity of implementation findings and describe the implementation in more detail.

[2] 'Confirmatory' impact analyses address the primary research questions of the study. I3 evaluators were required to identify the confirmatory research questions prior to data collection and analysis. Other research questions and analyses, while substantively important, are considered 'exploratory' or secondary to the confirmatory questions.

STUDY SAMPLE

The schools in the RCT study of RAISE are a subset of schools that participated in the larger i3 validation project. This section of the report explains the recruitment and random assignment process and then describes the characteristics of the RCT study schools, teachers, and students.

Recruitment and Random Assignment

With the assistance of SLI, the evaluation team recruited 42 schools and randomly assigned them to treatment or control groups. To be included in the study, schools needed to (1) serve grades 9–12 in a single building and (2) have teachers who were willing to participate in the study who taught at least one regular education course in ninth grade ELA, tenth grade biology, or eleventh grade U.S. history.[3] Schools were not eligible if they had any of the following characteristics.

- Specific admissions criteria related to academic achievement, such as eighth grade attendance or test score requirements
- Student population predominately receiving special education services
- Non-traditional population of students, such as single-gender schools or schools with students who had previously dropped out of or were expelled from high school

In addition, study schools could not already be implementing Reading Apprenticeship. The study team intentionally recruited schools that served high proportions of students who: were eligible for free and reduced-price lunch, were English language learners, and/or had low prior achievement.

Though this was a school-level random assignment study, Reading Apprenticeship was not implemented on a school-wide basis. Prior to random assignment, the study team recruited volunteer teachers from each subject area of interest in each school. To be eligible for the study, teachers needed to teach at least one class in one of the following subject areas: ninth grade ELA, tenth grade biology, or eleventh grade U.S. history. We excluded special education teachers in self-contained classrooms, teachers co-teaching a common group of students in the same classroom, and teaching facilitators who were not designated to lead class instruction. Schools were not required to have participation from all teachers in the target subject areas. We attempted to recruit at least one, and up to three, teachers per subject area per school. In total, we recruited 252 teachers into the study prior to random assignment. Once the volunteers agreed to participate, we blocked schools based on key organizational units (e.g., district), baseline characteristics of the schools (percent of students eligible for free lunch, race/ethnicity, percent English language learners, prior average academic performance), and predicted academic performance. We randomly assigned schools to treatment and control status within these blocks. The random assignment of schools resulted in 22 schools and 130 teachers in the treatment group and 20 schools and 122 teachers in the control group.

[3] Once recruited, the grade-levels and subject areas taught changed for many teachers. These teachers and their students remained in the analytic sample, even though they may have been teaching another science class, e.g., physics, instead of biology, or another history class, e.g., world history, instead of US history.

The student sample consisted of students enrolled in target subject area classes instructed by the participating teachers at the time of outcome data collection.[4] Students were not identified prior to random assignment; they 'joined' the study by enrolling in a study teacher's class after random assignment. However, students and their parents were unlikely to know the treatment/control status of the school or teachers in the study when enrolling in the school or a particular teacher's class, and therefore, any bias to our sample stemming from their joining status is considered minimal. In total, three cohorts of students were included in the study, representing teachers' first, second, and third years of Reading Apprenticeship implementation (see Table 2). The student sample included 14,383 students in the first year of implementation, 14,747 in the second year, and 9,194 in the third year.[5]

Most of the findings in this report are from the sample of students and data collected during teachers' second year in the study, after treatment teachers had received the full "dose" of professional development delivered over 12 months and could therefore be expected to fully implement Reading Apprenticeship. We used the data from the first and third years to conduct supplemental analyses. The first-year data were used to report interim findings based on short-term results, and the third-year data were used to explore potential longer-term effects.

TABLE 2. IMPLEMENTATION YEAR

Wave	Schools	2011-2012	2012-2013	2013-2014
1	22 schools in CA and PA: 17 RAISE, 15 control	year 1	year 2	year 3
2	10 schools in CA: 5 RAISE and 5 control	n/a	year 1	year 2

Baseline Study Sample Characteristics

To evaluate whether the random assignment resulted in statistically equivalent groups at baseline (i.e., prior to implementing Reading Apprenticeship), we compared the school-level and individual-level characteristics of the treatment and control groups. Table 3 shows the characteristics of the schools in this RCT study. There were no statistically significant differences between the treatment and control schools at baseline.

[4] If a teacher had both regular and special education or advanced classes, we included only their regular classes in our sample. If a teacher changed subject areas and no longer taught in one of the three target areas, we included students in those non-target area classes in our sample, on the hypothesis that, once trained, teachers could implement the Reading Apprenticeship approach in any subject.

[5] As noted earlier, year 3 data were collected from wave 1 schools only.

TABLE 3. SELECTED BASELINE CHARACTERISTICS OF STUDY SCHOOLS

Characteristic	Total (*N* = 42)	Treatment (*n* = 22)	Control (*n* = 20)	*p* value
Prior achievement, FRPL, and ELL				
Students proficient in 11th grade ELA /reading at baseline	55.5%	56.1%	54.9%	.854
Students eligible for free and reduced-price lunch	39.5%	41.6%	37.3%	.624
English language learners	10.4%	11.3%	9.4%	.590
Student race and ethnicity				
Nonwhite	49.3%	49.6%	49.0%	.960
Hispanic	33.4%	33.6%	33.1%	.966

Source. IMPAQ staff calculations based on the National Center for Educational Statistics Common Core of Data and district-provided student records. The records from the year prior to RAISE implementation are used as baseline data (2010-11 for wave 1 schools and 2011-12 for wave 2 schools).

Table 4 shows the characteristics of teachers who participated in the study. Study teachers were predominantly female and white. On average, they had 10 years of teaching experience, 8 of which were in the target subject area for the study. There were no statistically significant differences between the treatment and control teachers.

TABLE 4. SELECTED BASELINE CHARACTERISTICS OF STUDY TEACHERS

	Total (*N* = 252)	Treatment (*n* = 130)	Control (*n* = 122)	*p* value
Gender and race				
Female	57.8%	56.3%	59.3%	.674
Nonwhite	26.6%	26.8%	26.3%	.958
Subject				
ELA	33.9%	34.2%	33.6%	.930
U.S. history	33.5%	32.5%	34.5%	.748
Science	32.6%	33.3%	31.9%	.815
Experience				
Average years teaching	9.94	10.35	9.54	.649
Average years teaching the target subject	7.99	8.28	7.70	.650
Percent certified as reading specialist	2.7%	3.6%	1.8%	.403
Highest education level				
Bachelor's degree	37.2%	33.0%	41.2%	.269
Master's degree	55.8%	60.7%	50.9%	.167
Professional diploma or specialty	2.7%	3.6%	1.8%	.675
Doctorate	0.9%	0.0%	1.8%	.307
Other degree or credential	3.5%	2.7%	4.4%	.489

Source. IMPAQ staff calculations based on teacher responses to year 1 study surveys

Note. The total teacher sample size is 252. Actual number of respondents for each subgroup varies because of missing data: 225–236 for the total sample, 113–116 for the treatment group, and 111–120 for the control group.

Table 5 shows the characteristics of students in the baseline sample, that is, all study-eligible students for whom baseline data were collected.[6] A total of 11% of students in the baseline sample were classified as receiving special education services; 14% were classified as English language learners. Over half (56%) were eligible for free and reduced-price lunch. The sample included a relatively high proportion of low-performing students compared with students in their respective states: 38.3% of students in our sample scored in the bottom third on the eighth-grade state reading or ELA test. A majority (60.8%) were nonwhite. There were no statistically significant differences between the treatment and control students in the baseline sample on these characteristics. (See Appendix D. Analytic Sample Baseline Equivalence for baseline equivalence of the analytic samples.)

[6] The baseline sample differs from the analytic samples. The analytic samples include all students with outcome data who were in our impact analyses and for whom impacts were estimated.

TABLE 5. SELECTED CHARACTERISTICS OF STUDENT BASELINE SAMPLE

Student characteristic	Total (*N* = 14,747)	Treatment (*n* = 7,783)	Control (*n* = 6,964)	*p* value
Female	48.7%	49.5%	47.8%	.146
Special education	11.3%	8.6%	14.1%	.204
English language learner	13.8%	13.6%	14.0%	.978
Eligible for free or reduced-price lunch	56.2%	56.7%	55.7%	.846
Scored in bottom third on 8th grade state ELA or reading test	38.3%	39.9%	36.6%	.266
Nonwhite	60.8%	63.4%	57.9%	.636

Source. IMPAQ staff calculations on demographic data collected from study school districts.

Note. The total student sample size is 14,747, but demographic records were not reported for all students in the sample. Actual sample sizes vary by characteristic, depending on the completeness of the administrative data for each student in our sample.

Data Sources and Measures

The findings presented in this report used multiple sources of qualitative and quantitative data, collected over three years for wave 1 schools and two years for wave 2 schools. Table 6 shows the sources and timing for the data collected.[7]

TABLE 6. DATA SOURCES AND DATA COLLECTION TIMEFRAME

Data source	Collection timeframe
Teacher surveys	Monthly during three school years, September 2011– May 2012; September 2012- May 2013; September 2013 – May 2014 (wave 1 only): 27 survey collections in total
Student record data	Yearly, 2011–2014
Student literacy assessments	Yearly at the end of spring semester, 2012–2014
Student surveys	Yearly at the end of spring semester, 2012–2014
Principal surveys	Yearly
Professional development observations	Summer 2011, winter 2012, summer 2012, winter 2012, summer 2013
Professional development attendance rosters	Summer 2011, winter 2012, summer 2012, winter 2013, summer 2013

STUDENT RECORDS DATA

To track student achievement outcomes, the evaluation team worked with each school and district in the study to collect individual-level data for all students in treatment and control classrooms. In addition to data on background variables, such as race/ethnicity, gender, special education status, and English language learner status, we also collected state standardized test scores, where available.

STUDENT LITERACY ASSESSMENTS

Online assessments of literacy achievement were administered to students at the end of each year of the study. The assessments are general literacy assessments with a disciplinary focus. They were developed by Educational Testing Services (ETS) as part of the Reading for Understanding (RfU) grant funded by the Institute for Education Sciences. The scenario-based assessments were designed to measure how well students read and reason about text sources in a discipline where they have been exposed to content and strategies for understanding text (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014). The forms

[7] The evaluation data collection also included classroom observations. Findings from those data are presented in other reports.

were designed without specific knowledge of the Reading Apprenticeship intervention, and with no input from the researchers and developers beyond a general description of the study. Further, the forms were never shared with the researchers, developers, or implementers of Reading Apprenticeship. ETS has used the same forms in multiple other data collection efforts at the secondary level.

There were three forms in the assessment designed to assess literacy in the context of each of three subjects: biology, history, and literature (ELA). All used a similar structure. The biology form included texts on ecosystems and invasive species, the history form included texts on U.S. immigration and Ellis Island, and the literature form included a piece by Langston Hughes and text on the Harlem Renaissance. It assessed a variety of purposeful literacy activities in which students are expected to read multiple texts for understanding. The scenarios organized the assessment around a theme and goal for reading; for example, students were asked to imagine they were studying for an exam or preparing for a presentation. They were then asked to participate in a sequence of tasks that would lead to a final goal, such as identifying important ideas and meaning, evaluating sources, or integrating information across multiple sources. The assessments were not designed to assess or be dependent on specific content knowledge in any of the three subject areas, but rather to assess student literacy skills in the context of each subject.

The scenario-based assessment was pilot tested in fall 2011 and spring 2012 to collect evidence of its psychometric properties. The assessment displayed adequate reliability for each of the subject-area forms (r = 0.84 for biology, 0.85 for history, and 0.88 for ELA). For students with available data, we examined the relationship between the state standardized ELA tests (the Keystone Literature Tests for PA and California Standards Test for CA) and the literacy assessment used in this study. The correlation coefficient between state test z-scores and the literacy assessment scores was 0.69 overall: 0.70 for California (n = 2,612) and 0.66 for Pennsylvania (n = 1,060). This modestly high correlation suggests that the literacy assessment captured some of the same underlying constructs related to reading comprehension as those state tests. It also suggests that the ETS assessment captured other aspects of reading comprehension—such as disciplinary reading strategies emphasized by the Common Core and Reading Apprenticeship—not measured through the state tests. Finally, psychometric testing also showed sufficient range and variability in scores, with no evidence of ceiling or floor effects (O'Reilly et al., 2014).

Additional information on the ETS-developed literacy assessment, including basic statistics of the scores by state and by student characteristics, is provided in Appendix E. Student Literacy Assessment.

STUDENT SURVEYS

The logic model for RAISE (Table 1) posits that changes in teacher attitudes and instructional approaches result in changes in student attitudes, dispositions, stamina, and persistence with respect to reading; while at the same time, building reading comprehension skills and knowledge for subject-specific literacy tasks and strengthening students' view of themselves as readers and learners. Data on student attitudes and behaviors related to reading and literacy were collected through a student survey administered to all treatment and control students in the spring of each school year. We used 130 items from several previously validated surveys: the Tripod Project Survey developed for the Gates Foundation-funded Measuring Effective Teaching project; the Metacognitive Awareness of Reading Strategies Inventory; the Reading Apprenticeship Opportunity to Learn Surveys, developed by WestEd; and the National

Assessment of Educational Progress student survey. Table 7 describes the domains and constructs developed from the student survey. (See Appendix B. Student Survey Constructs for a list of survey items comprising each construct.)

TABLE 7. STUDENT SURVEY DOMAINS AND CONSTRUCTS

Domain/ construct number	Construct description	Possible construct range		Reliability
		Min	Max	Alpha
Domain 1: Increased collaboration in a community of readers and writers (2 constructs)				
1.1	**Participation in and contribution to class discussions** Frequency of contribution to and participation in class discussions concerning vocabulary, content, and interpretation of texts	1	4	0.82
1.2	**Conferring** Extent of discussion and exchange of information in the classroom regarding how activities are assigned and completed	1	5	0.75
Domain 2: Increased use of comprehension strategies (4 constructs)				
2.1	**Use of global reading strategies** Student use of generalized, intentional reading strategies allowing a global analysis of a text before and during reading activities	1	5	0.86
2.2	**Use of problem-solving strategies** Student use of localized, focused problem-solving strategies to better understand a text while reading	1	5	0.81
2.3	**Use of support reading strategies** Student use of practical strategies aimed at sustaining responses to a text such as notetaking, summarizing, and discussing materials during and after reading	1	5	0.83
2.4	**Integration of content and literacy activity** Frequency of activities that encourage the integration of content knowledge and reading strategies by summarizing, interpreting, and identifying themes in a text	1	4	0.79
Domain 3: Increased metacognitive conversations (1 construct)				
3.1	**Metacognitive conversations** Students discuss and inquire into their own and others' reading processes and report teacher's instruction of strategies that assist with reading comprehension	1	4	0.81
Domain 4: Improved reader identity (1 construct)				
4.1	**Reader identity** Increased student awareness of reading processes, habits, strengths, weaknesses, attitudes, and preferences	1	4	0.75

TABLE 7. STUDENT SURVEY DOMAINS AND CONSTRUCTS

Domain/ construct number	Construct description	Possible construct range Min	Possible construct range Max	Reliability Alpha
Domain 5: Improved student identity (1 construct)				
5.1	**Student identity** Students' emphasis on spending class time learning and understanding concepts; confidence in their ability to complete class work regardless of difficulty; and identifying themselves as serious, capable learners	1	5	0.88
Domain 6: Increased reading of a variety of texts (3 individual items)				
6.1	**Class time spent reading** Amount of time in classroom spent reading	1	4	NA
6.2	**Variety of reading material** Frequency of incorporation of graphs, charts, tables, and illustrations in reading activities	1	4	NA
6.3	**Pages of reading per day** Student report of number of pages read daily in class and for homework	1	5	NA
Domain 7: Increased academic engagement (3 constructs)				
7.1	**Effort to learn** Effort to learn through regularly attending class and thorough completion of classwork	1	5	0.71
7.2	**Happiness and belonging** Student happiness in the classroom and satisfaction with academic achievements	1	5	0.73
7.3	**Engaging instruction** Student report of engaging instructional practices that make learning enjoyable and interesting	1	5	0.83
Source. IMPAQ staff calculations based on student responses to year 1, 2, and 3 study surveys				

TEACHER SURVEYS

Collecting valid and reliable data on teacher practices using a survey can be challenging. Survey responses are often subject to bias due to recall difficulty, a tendency to provide socially desirable responses, and variation in teachers' daily practice that is not captured by measures from a single point in time. To address these concerns, we collected monthly log-like surveys that gathered detailed information about the nature, frequency, and mode of instruction during a specific week of each month. The surveys also asked questions related to attitudes and beliefs toward Reading Apprenticeship (for the treatment group only) and about teachers' confidence in their ability to teach literacy. A total of 27 surveys were

administered to wave 1 teachers, and 18 surveys were administered to wave 2 teachers (nine per year of the study). Across the surveys, a core group of questions asked teachers about their teaching practices. Each month, the survey asked about teaching practices during a single class (the "target class") one week during the month. For example, the survey asked, "During the week of October 31 through November 4, how many of your target class periods included the following practices?" The target class remained consistent throughout the year.[8] For the treatment group teachers, the surveys also asked questions about participation in monthly RAISE team meetings, attitudes and beliefs toward Reading Apprenticeship, supports received from literacy instruction, and confidence in teaching literacy.

The RAISE logic model (see Table 1) hypothesizes that the intensive 65 hours of professional development and the ongoing support provided by teacher leaders through site-based monthly meetings will have an impact on teachers' instructional practices and routines. Informed by this model, we used the teacher survey data to create 12 constructs intended to capture the effects of Reading Apprenticeship on the following dimensions of teacher behavior and attitudes (see Table 8).

- Providing extensive reading opportunities that reflect a variety of genres and text types (measured by construct 1)
- Supporting student effort to comprehend disciplinary text (measured by construct 2)
- Fostering metacognitive inquiry into reading and thinking processes (measured by constructs 4-6)
- Providing explicit instruction and modeling of reading comprehension routines, tools, strategies, and processes (measured by constructs 7-9)
- Fostering and supporting student collaboration (measured by construct 10)
- Employing instruction that promotes engagement, student-centered learning, and inquiry-based learning (measured by construct 11)
- Confidence in delivering literacy instruction (measured by construct 12)

Construct 3, measuring the extent to which teachers employed traditional instructional strategies such as lecture and using quizzes to assess comprehension, represents a contrast to the Reading Apprenticeship approach. Therefore, we did not expect RAISE to have an impact on these strategies.

For each construct, we calculated the mean score over the first year of implementation (surveys 1–9), the mean score over the second year of implementation (surveys 10–18), and the mean score over the first and second years of implementation (surveys 1–18). For wave 1 schools, we also created the mean score over the third year of implementation (surveys 19–27). For each of the 12 constructs, Table 8 provides the description, the minimum and maximum range for the construct score, and the construct's internal reliability statistic based on surveys from the second year of implementation (Cronbach's Alpha). (See Appendix C. Teacher Survey Constructs for a list of survey items comprising each construct.)

[8] In a few cases, teachers taught semester-long courses rather than year-long courses. In these cases, the target class changed when the semester changed.

TABLE 8. TEACHER SURVEY CONSTRUCTS

Construct number	Construct description	Possible construct range		Reliability
		Min	Max	Alpha
	Variety of Text Types			
1	Total number of text types that a teacher asked students to work with over a week, in or outside of class (e.g., newspapers, textbooks, historical documents)	0	7	0.54
	Fostering Student Independence			
2	Total number of minutes over a week that a teacher uses practices to foster independence, such as providing guided practice of reading comprehension strategies and having students teach other students	0	12	0.66
	Traditional Instructional Strategies			
3	Total number of minutes over a week that a teacher employs traditional strategies, such as direct instruction and giving quizzes to assess comprehension.	0	12	0.74
	Teachers Instructing Metacognitive Inquiry			
4	Total number of metacognitive inquiry strategies in which teachers provided instruction over a week (e.g., asking questions about the text, writing to clarify understanding, discussing meaning of texts)	0	4	0.65
	Teachers Modeling Metacognitive Inquiry			
5	Total number of metacognitive inquiry strategies that teachers modeled during their class over a week (e.g., asking questions about the text, writing to clarify understanding, discussing meaning of texts)	0	4	0.54
	Students Practicing Metacognitive Inquiry			
6	Total number of metacognitive inquiry strategies that students practiced during class over a week (e.g., asking questions about the text, writing to clarify understanding, discussing meaning of texts)	0	4	0.66
	Teachers Instructing Comprehension Strategies			
7	Total number of comprehension strategies (e.g., setting a reading purpose, annotating text, choosing a reading approach that fits the purpose) in which teachers provided instruction over a week	0	8	0.74
	Teachers Modeling Comprehension Strategies			
8	Total number of comprehension strategies (e.g., setting a reading purpose, annotating text, choosing a reading approach that fits the purpose) that teachers modeled during the class over a week	0	8	0.70

TABLE 8. TEACHER SURVEY CONSTRUCTS

Construct number	Construct description	Possible construct range		Reliability
		Min	Max	Alpha
	Students Practicing Comprehension Strategies			
9	Total number of comprehension strategies (e.g., setting a reading purpose, annotating text, choosing a reading approach that fits the purpose) that students practiced during class over a week	0	8	0.75
	Student Collaboration			
10	Total number of minutes over a week that teachers had students work on reading and writing activities in pairs, in small groups, and as a whole class	0	15	0.67
	Student Engagement			
11	Total of teachers' ratings on the proportion of students in their class completing homework, paying attention in class, and participating in class activities	0	15	0.71
	Teacher Self-Confidence in Literacy Instruction			
12	Total of teachers' ratings on their confidence in their ability to provide literacy instruction, such as providing opportunities for reading a variety of texts of different genres and teaching students to analyze their own thinking about texts	0	55	0.91

Source. IMPAQ staff calculations on the teacher response to study surveys during the second year of implementation.

PRINCIPAL SURVEYS

A principal survey was administered at the end of each year of implementation. The purpose of the principal survey was to better understand the principals' perspectives on the school climate—including collaboration and support—and their opinions of new instructional programs, including Reading Apprenticeship.

PROFESSIONAL DEVELOPMENT OBSERVATIONS

For wave 1 professional development, all five days of the initial Reading Apprenticeship training were observed in Pennsylvania, and one day was observed in California. According to the program developers, there is a developmental progression throughout the five days, during which teachers develop their understanding of Reading Apprenticeship and progressively gain the knowledge, beliefs and instructional practices targeted by the professional development. The purpose of observing all five days was to document this progression. An additional full day in the second site was observed in order to document any major site-to-site differences. The two-day training held in the winter of the first year of implementation was observed in both Pennsylvania and California. The final three-day training, held in the summer following the first year of implementation, was observed in California only. For wave 2 professional development, all 10 days were observed in California.

At each training, selected sessions were observed by the evaluators to record the content delivered and the instructional methods employed by the facilitators. The agenda for a single day of professional development is generally divided into two or three modules in the morning and two or three modules in the afternoon. Evaluators tried to balance observation of sessions across the separate ELA, biology, and U.S. history trainings.

PROFESSIONAL DEVELOPMENT ATTENDANCE ROSTERS

Teacher attendance at the RAISE professional development was obtained from two primary sources: teacher-reported attendance on the monthly surveys and WestEd's attendance rosters obtained through daily sign-in sheets.

RESPONSE RATES

All original 42 schools that were randomly assigned remained in the study over the three-year study period.. However, three schools did not have any eligible ELA teachers, three did not have any eligible history teachers, and two did not have any eligible science teachers at random assignment. Further, only 31 of the 42 schools served English language learners and were included in subgroup analyses of this population. Thus, as shown in Table 9, the school sample sizes for subject-level analyses vary. Table 9 also shows the response rates for year 2 teacher surveys, student survey, and student literacy assessment. The lower than desired response rates for the student literacy assessment reflects the difficulty some schools and teachers had in internet connectivity challenges and scheduling access to computers to take the online assessment at the end of the school year, in addition to normal student mobility. See Appendix F. Sample Attrition for more detailed information on sample attrition.

TABLE 9. SCHOOL, STUDENT, AND TEACHER SAMPLE SIZES AND RESPONSE RATES

	Total		Treatment		Control	
School sample	*N*		*n*		*n*	
Total number of schools	42		22		20	
Schools in California	21		11		10	
Schools in Pennsylvania	21		11		10	
Schools with ELA teachers	39		20		19	
Schools with history teachers	39		20		19	
Schools with science teachers	40		22		18	
Schools in ELL sample	31		16		15	
Teacher sample	*N*	%	*n*	%	*n*	%
Total number of teachers	252	100%	130	100%	122	100%
Responded to all nine year 2 surveys	178	70.6%	90	69.2%	88	72.1%
Responded to at least one year 2 survey	208	82.5%	105	80.8%	103	84.4%

TABLE 9. SCHOOL, STUDENT, AND TEACHER SAMPLE SIZES AND RESPONSE RATES

	Total		Treatment		Control	
Student sample	*N*	%	*n*	%	*n*	%
Total number of students	14,747	100%	7,783	100%	6,964	100%
Responded to year 2 survey	12,563	85.2%	6,612	85.0%	5,951	85.5%
Completed year 2 literacy assessment	10,173	69.0%	5,531	71.1%	4,642	66.7%

Source. IMPAQ staff calculations on data collected for the evaluation

Implementation of RAISE

This section addresses the following research questions.

1. To what extent is RAISE implemented in a way that is consistent with the program model and underlying theory of action?
2. What are the contextual factors that support or hinder RAISE implementation?

The findings related to these research questions provide context for assessing and understanding the measured impacts of RAISE on student and teacher outcomes. This section uses descriptive statistics from RAISE professional development attendance records and observations, and from teacher and principal survey data,[9] to provide context for assessing and understanding the measured impacts of RAISE on student and teacher outcomes We have also included comparisons between the RAISE and control groups, by state, and by subject area, to give further context to RAISE implementation.

IMPLEMENTATION OF CORE PROGRAM COMPONENTS

As described in the methods section, fidelity of implementation was measured for each of the core program components against teacher-, school-, and/or program-level thresholds. The core RAISE components include: delivery of the RAISE professional development content, attendance at the RAISE professional development, recruitment of a RAISE teacher leader, and attendance at the RAISE team monthly meetings. While the results of these analyses have been reported as a concise record of program implementation (see Appendix A. Impact Estimation Model), evaluators collected additional information from participants that enables a closer inspection of RAISE activities over the study period (see Appendix K. Context for Program Implementation for detailed tables of survey responses).

RAISE Professional Development

Key findings related to the RAISE professional development include the following.

- The RAISE professional development was delivered as intended: Over 85% of the observed professional development sessions exhibited the five key design characteristics.
- More than three quarters of teachers met the fidelity threshold for attending the RAISE professional development; however, the teachers who met this threshold tended to be clustered in the same schools. Ten out of the 22 (45%) RAISE schools did not meet the school-level professional development attendance fidelity threshold.
- Over 90% of teachers (n = 96) who responded to survey questions about the RAISE professional development felt that it "moderately", "more than moderately", or "completely" prepared them to use the set of literacy practices modeled during the training.

[9] The sample size for fidelity of implementation results is based on the sample of teachers in the treatment group only for each year of implementation: n = 113 for Year 1, n = 105 for Year 2, and n = 69 for Year 3. For survey results, the sample size varies depending on the number of respondents to the particular question and survey. Sample sizes for surveys in Year 1 and 2 ranged from n = 211 to n = 241 (representing the number of teachers in both conditions), constituting between 85% and 96% of total study participants.

Based on observations of the RAISE Institute sessions, the professional development was delivered in a manner consistent with the theory of action. Over 85% of the sessions observed (n = 45) exhibited content that was inquiry-based, and focused on disciplinary literacy, collective participation, active learning, and coherence.

Eighty-seven out of 113 (77%) RAISE teachers attended all five of the first five days of the RAISE professional development and at least four of the last five days, meeting the teacher-level attendance target as intended by the developers. School-level fidelity of implementation thresholds set by SLI were not met for participation in RAISE professional development. Only 12 out of 22 (55%) schools had 75% (or more) of teachers meeting the threshold of attending 9 out of 10 days. Of the teachers who did not meet the threshold, most missed one or more of the three sessions in their entirety (5-Day Foundation, 2-Day Calibration, or 3-Day Springboard).

Teachers who met fidelity tended to be clustered in schools. Within the 12 schools where school-level fidelity was met, 58 out of 62 (94%) teachers met fidelity, but within the 10 schools where school-level fidelity was not met, only 29 out of 51 (57%) teachers met fidelity. For 7 of the 10 schools that did not meet fidelity, the school was within one teacher of meeting the fidelity threshold; however, within those seven schools, 11 out of 12 teachers (92%) missed all of at least one of the three sessions: 5-Day Foundation, 2-Day Calibration, or 3-Day Springboard. There were no significant differences by state or subject area for attendance at the professional development.

Following the full 10 days of the RAISE professional development, teachers were asked to rate their level of preparation on a set of key literacy strategies modeled during the professional development.

A. Supporting students in working on reading or writing activities collaboratively by setting norms, creating safety, providing prompts that promote collaboration, and providing guidance and feedback on student participation
B. Modeling or demonstrating reading comprehension strategies such as setting a reading purpose, previewing text, chunking, or visualizing
C. Supporting students in their attempts to understand disciplinary texts such as challenging literature, textbooks, primary documents, or scientific articles
D. Providing explicit instruction on reading comprehension strategies such as setting a reading purpose, previewing texts, chunking, and visualizing
E. Teaching students to analyze their own thinking about reading texts
F. Providing students with opportunities for reading a variety of texts of different types and genres
G. Employing routines or assignments that are open-ended—such as group discussion or free choice in reading materials—enabling all students to feel comfortable participating and have some measure of success
H. Structuring lessons that hold students accountable for reading, for example, so that students have to do the assigned reading in order to succeed
I. Facilitating students' active engagement in learning through the use of inquiry-based instructional methods
J. Asking students to pose questions and problems about course readings

K. Giving students roles that make them responsible for making sense of texts, for example, by asking students to lead discussions or make arguments based on their interpretations of texts

As shown in Figure 1, teachers felt most prepared to 1) support students working collaboratively on reading and writing activities and 2) model or demonstrate reading comprehension strategies. Very few teachers reported feeling less than moderately prepared to implement any of the strategies.

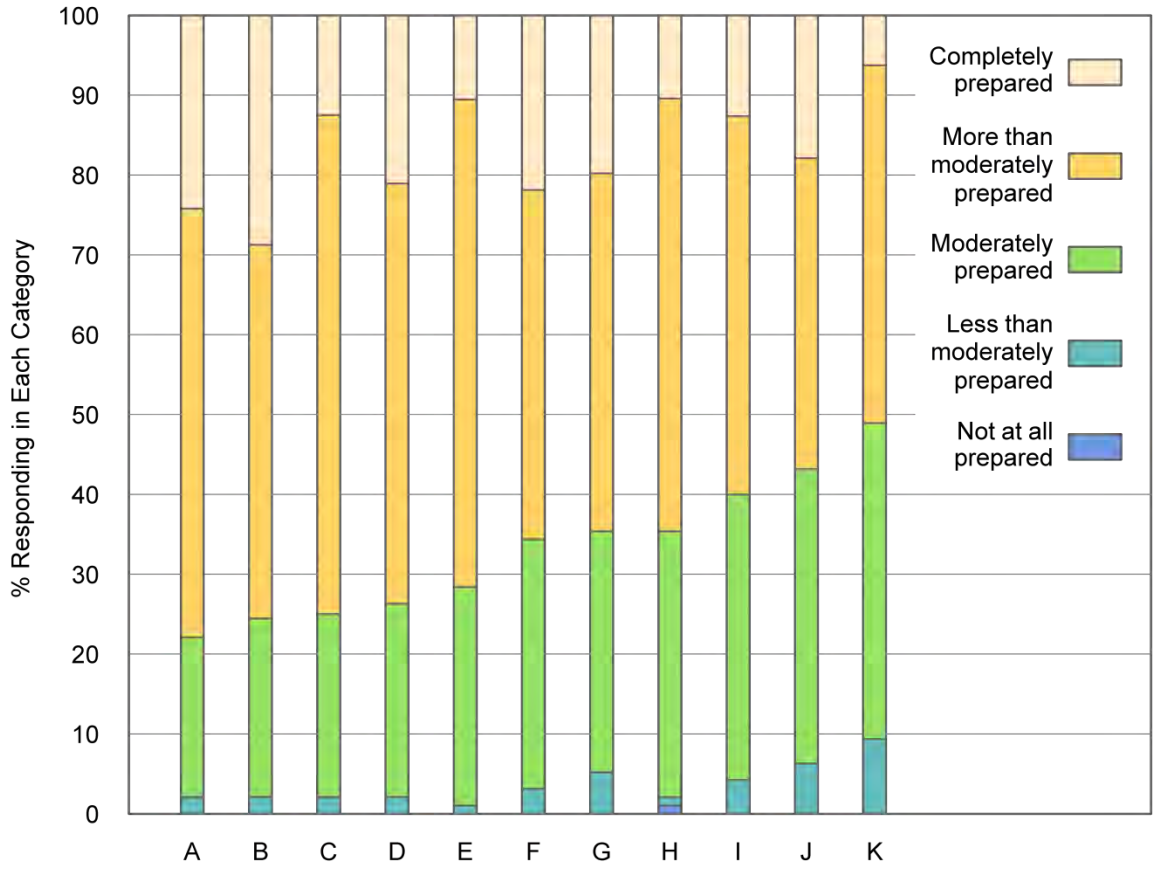


FIGURE 1. TEACHER REPORTED LEVEL OF PREPARATION AFTER RAISE PD

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

n = 94-96 for each strategy

RAISE Teacher Leaders and Monthly Team Meetings

Key findings related to RAISE teacher leaders and monthly team meetings include the following.

- Over 90% of the RAISE schools had a RAISE-trained teacher leader during each year of implementation.

- While the program-level fidelity thresholds were met for attendance at the RAISE monthly meetings, attendance varied greatly at the school level, suggesting that building coherence and communities of practice may have been more challenging at certain schools.

To facilitate ongoing learning, SLI recruited teacher leaders at each school, who received further support and training from the RAISE state site coordinators and coordinated and led the monthly team meetings in their school. All but one of the 22 RAISE schools had a teacher leader in years 1 and 2, and all but two schools had a teacher leader in year 3.

The RAISE monthly team meetings were intended to be a key mechanism for support and collaboration among RAISE teachers. Teachers were expected to attend at least four (out of 10 possible meetings, between August and May) RAISE team meetings per year to meet the school-level fidelity threshold. To meet program-level fidelity, 80% of schools had to have at least half of their teachers attend four or more meetings per year. In the first year of implementation, teachers averaged 5.4 meetings attended over the year and 18 of the 22 (82%) schools met the school-level fidelity threshold. During year 2, 70 out of 105 (67%) RAISE teachers met this fidelity threshold of attending at least four team meetings, and 19 of the 22 (86%) RAISE schools had at least half of their teachers meet this threshold. Also during the second year of implementation, the percentage of RAISE teachers who reported attending a monthly meeting within a particular month ranged from a low of 40% to a high of 60%, and on average, teachers attended 4.5 meetings. There are two possible hypotheses for the decline in average attendance at the meetings: there was a decrease in participation in RAISE in the later years, or teachers were finding alternative ways to collaborate and support their Reading Apprenticeship implementation.

As expected, meeting attendance varied by school, suggesting that building cohesion and a community of practice may be more challenging in certain schools. Of the 31 teachers who did not meet fidelity but had enough survey data in order for us to make a determination, 17 were concentrated in three schools that essentially stopped holding meetings, with teachers in these schools averaging less than one meeting attended over the year. The teachers who met the fidelity threshold attended an average of 6.5 team meetings during the second year of implementation, while those teachers who did not meet fidelity attended an average of 1.2 team meetings during the year. While there were no differences in monthly meeting attendance in either year by subject area, teachers in California in year 2 averaged more meetings attended than teachers in Pennsylvania, 5.4 versus 3.8 ($p < .01$).

The most common reason selected for not attending meetings was that they were not offered. The survey design did not call for teacher leaders to report specifically on whether or not a meeting was offered in a given month, and teacher reports within schools did conflict, indicating the difficulty of coordinating meetings across RAISE school teams. While over 80% of principals who responded to surveys ($n = 14$) in RAISE schools reported that time and space were allocated for monthly team meetings, responses from teachers to open-ended questions indicate that the difficulty of coordinating teacher schedules and school resources made meeting regularly a challenge. Of those teachers who attended, at least 80% reported that the monthly meetings were at least moderately helpful in each month (see Figure 2).

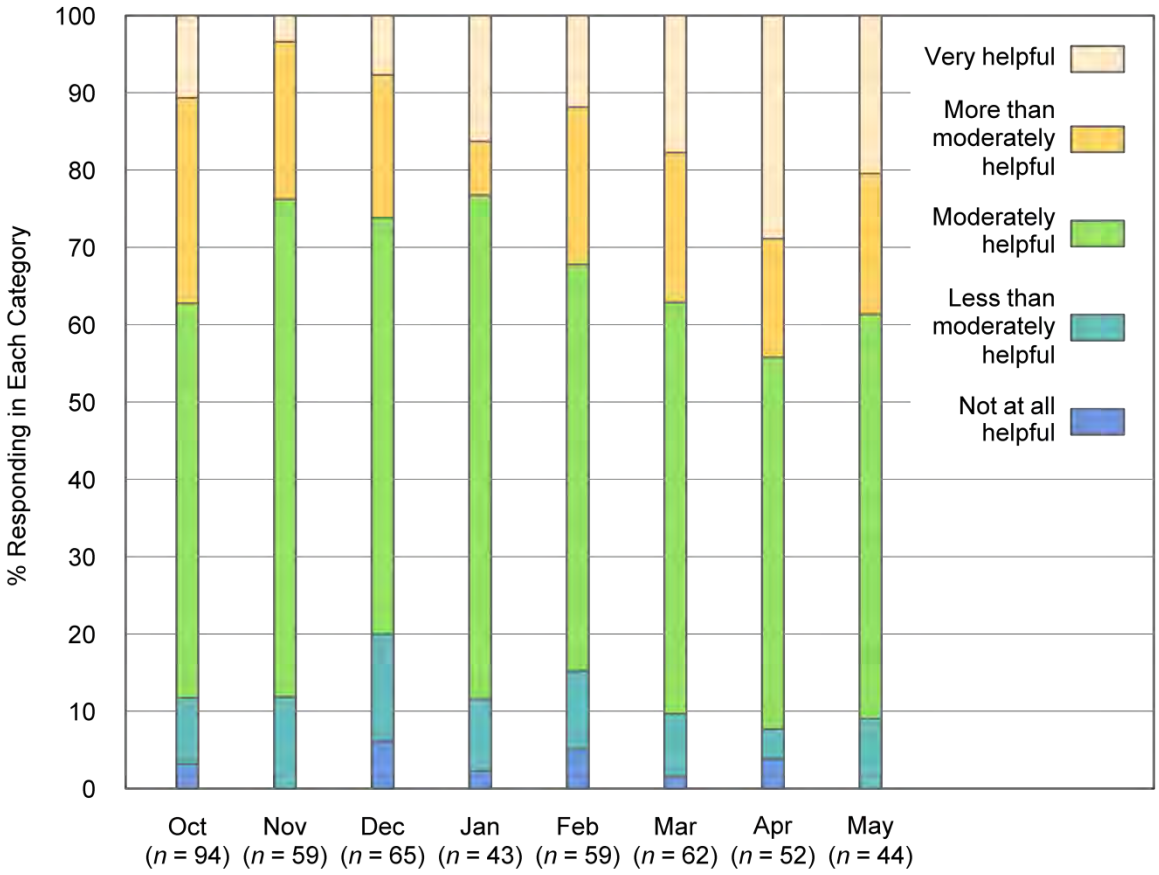


FIGURE 2. TEACHER REPORTED HELPFULNESS OF RAISE MONTHLY TEAM MEETINGS

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

CONTEXTUAL FACTORS OF IMPLEMENTATION

As shown in the logic model (Table 1), the program developers hypothesized that teachers would be supported in their implementation of RAISE by contextual factors outside of the core components, such as their sense of commitment and purpose related to the initiative; school cohesion and community; and administrative, social, and material support for implementation at site. This section covers the types of support for literacy instruction that teachers reported receiving outside of RAISE professional development sessions and monthly meetings and how helpful they perceived this support to be. We have also included information on reported challenges and barriers to implementation, and overall impressions of RAISE (see Appendix K. Context for Program Implementation for full results). Key findings related to the support and barriers to RAISE implementation include the following.

- RAISE teachers reported receiving support for literacy instruction (outside of official RAISE professional development and meetings) at a greater frequency than control teachers, and they

rated this support as "very" or "more than moderately" helpful at higher levels than control teachers. This was significant both overall and specifically for science teachers.

- The primary challenge to implementing Reading Apprenticeship was competing priorities, such as standardized test preparation or addressing content standards.
- Over 50% of teachers across subject areas felt that Reading Apprenticeship would be "highly" or "more than moderately" effective at improving students' reading comprehension.
- More ELA teachers than non-ELA (science and history) teachers reported that Reading Apprenticeship was "very well aligned" with their classroom goals and standards.
- 61% of teachers reported being fully committed to Reading Apprenticeship at the end of year 2.

Support for Literacy Instruction

Four times during the second year of implementation, teachers in both RAISE and control schools were asked to indicate which (if any) types of support for implementing literacy instruction they received during the prior month. RAISE teachers were explicitly instructed to exclude activities during monthly RAISE team meetings as a source of support. Teachers could select any of the following options: informal collaboration with other teachers, coaching and mentoring, model lessons, observation and feedback, resources, classroom management help, political support (for example, someone "backed them up" in a conflict over implementation of literacy instruction), a change in school or district policy that was relevant to literacy instruction, or "other". We looked at how frequently RAISE and control teachers reported any of type of support across the year. Across all subjects, RAISE teachers reported receiving more frequent support for literacy instruction compared to control teachers, 44% to 32% ($p < .01$). We further explored this finding by subject area (Figure 3). It is not surprising that fewer control science teachers reported receiving support for literacy instruction than their ELA or history counterparts. The difference in the average reported receipt of support for literacy instruction between the RAISE and control science teachers is statistically significant ($p < .05$). As hypothesized in the logic model, it is encouraging that RAISE teachers reported more support for literacy instruction—either through informal collaboration with their peers or seeking out and receiving more materials/resources.

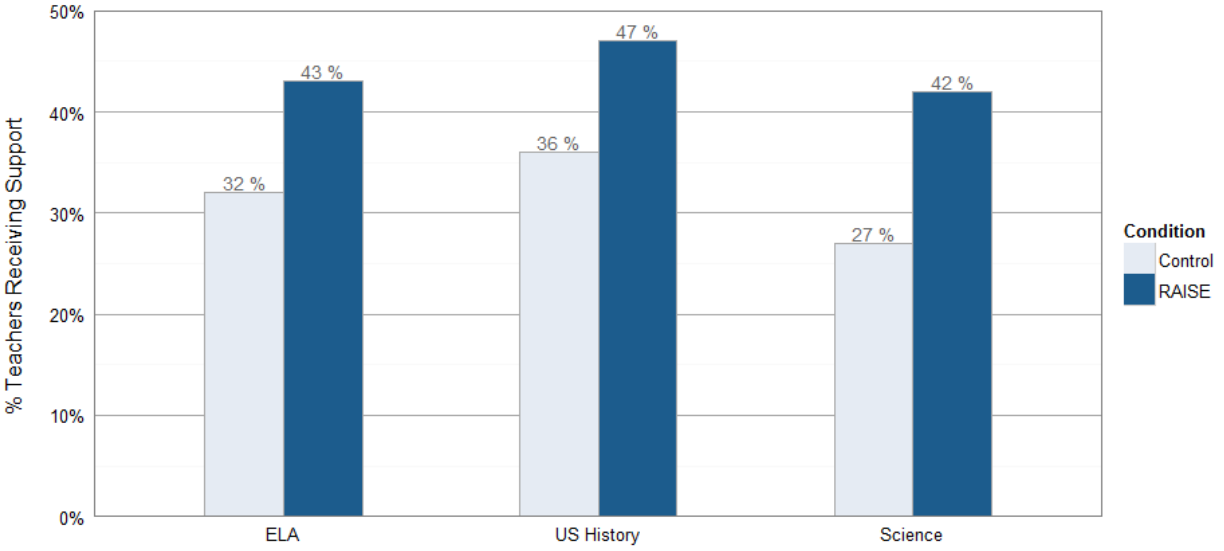


FIGURE 3. TEACHERS REPORT RECEIVING SUPPORT FOR LITERACY INSTRUCTION

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

n =110 for Control, *n* = 108 for RAISE

Helpfulness of Support Received by Teachers

Teachers who reported receiving the above support were asked, in general, how helpful the support was for improving literacy instruction in their classroom. Teachers rated the support on a 5-point Likert scale. On average, RAISE teachers were more likely to rate the support they received for literacy instruction (outside of the monthly RAISE meetings) as very helpful or more than moderately helpful compared to control teachers ($p < .01$). This finding was also significant for science teachers, as well as ELA teachers (both $p < .05$).

Challenges to Reading Apprenticeship Implementation

Every three months, teachers were asked what challenges they faced in implementing Reading Apprenticeship. Competing priorities was the most commonly selected response during year 2, with just over 60% of teachers selecting it, on average. This was echoed by principals, with 65% of respondents (n = 14) selecting it in the year 2 principal survey. Many of the open-ended responses suggested that the pressures of standardized tests created difficulty for teachers in implementing Reading Apprenticeship. The next most commonly selected responses were student behavior and student ability, selected by 34% and 31% of teachers on average, respectively, with open-ended teacher responses indicating that lack of student motivation inhibited the implementation of Reading Apprenticeship.

At the same time points, RAISE teachers were asked whether or not there were any school district policy constraints that made implementing Reading Apprenticeship difficult. The responses remained fairly consistent across the school year, with only 10% of teachers indicating that they believed district policy interfered with implementation of Reading Apprenticeship. The teachers who reported facing district policy constraints were then given an opportunity to explain their answer, with most of these responses highlighting logistical challenges: teachers mentioned obstacles such as photocopying limits or having trouble finding a meeting time that worked for all teachers.

Alignment with Classroom Goals and Content Standards

In May of year 2, RAISE teachers were asked to think back over their experience and determine how well Reading Apprenticeship aligned with the content standards and goals of their classroom (see Figure 4 and Figure 5). While overall, 86% (n = 106) of teachers reported that Reading Apprenticeship was very well aligned or somewhat well aligned with both their classroom goals and content standards, researchers expected that perceived alignment would be higher for ELA teachers and this was confirmed. ELA teachers reported higher alignment with both content standards and classroom goals than either science or history teachers. Nearly 80% (n = 27) of ELA teachers said Reading Apprenticeship is very well aligned with their content standards, while 59% (n = 20) of history teachers and 44% (n = 17) of science teachers stated the same opinion. In addition, 82% (n = 27) of ELA teachers found Reading Apprenticeship to be very well aligned with classroom goals, while 61% (n = 20) of history teachers and 51% (n = 20) of science teachers agreed. When we examined the difference of ELA teachers vs. non-ELA teachers, we found a significant difference for the alignment with both content standards and classroom goals ($p < .01$).

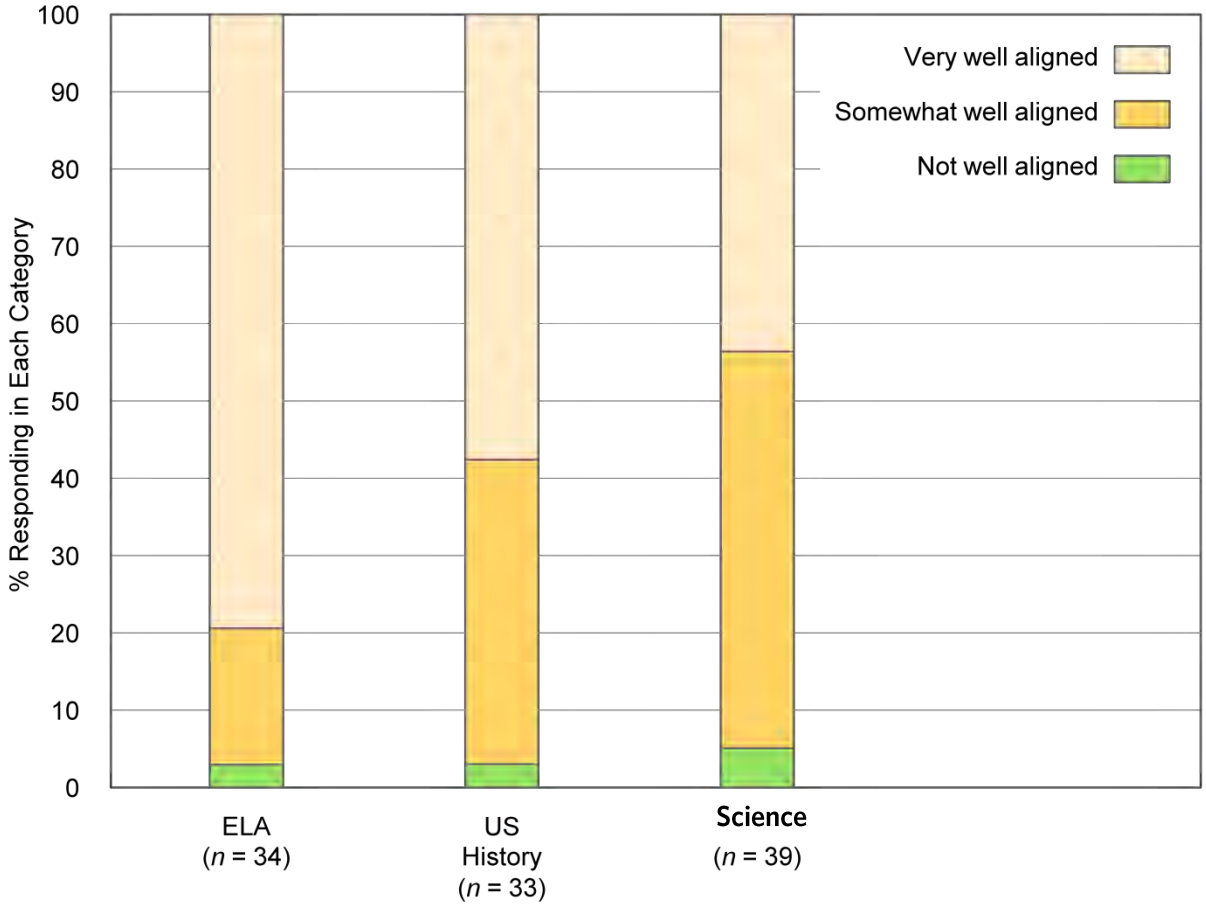


FIGURE 4. TEACHER REPORTED ALIGNMENT OF READING APPRENTICESHIP WITH CLASSROOM CONTENT STANDARDS

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

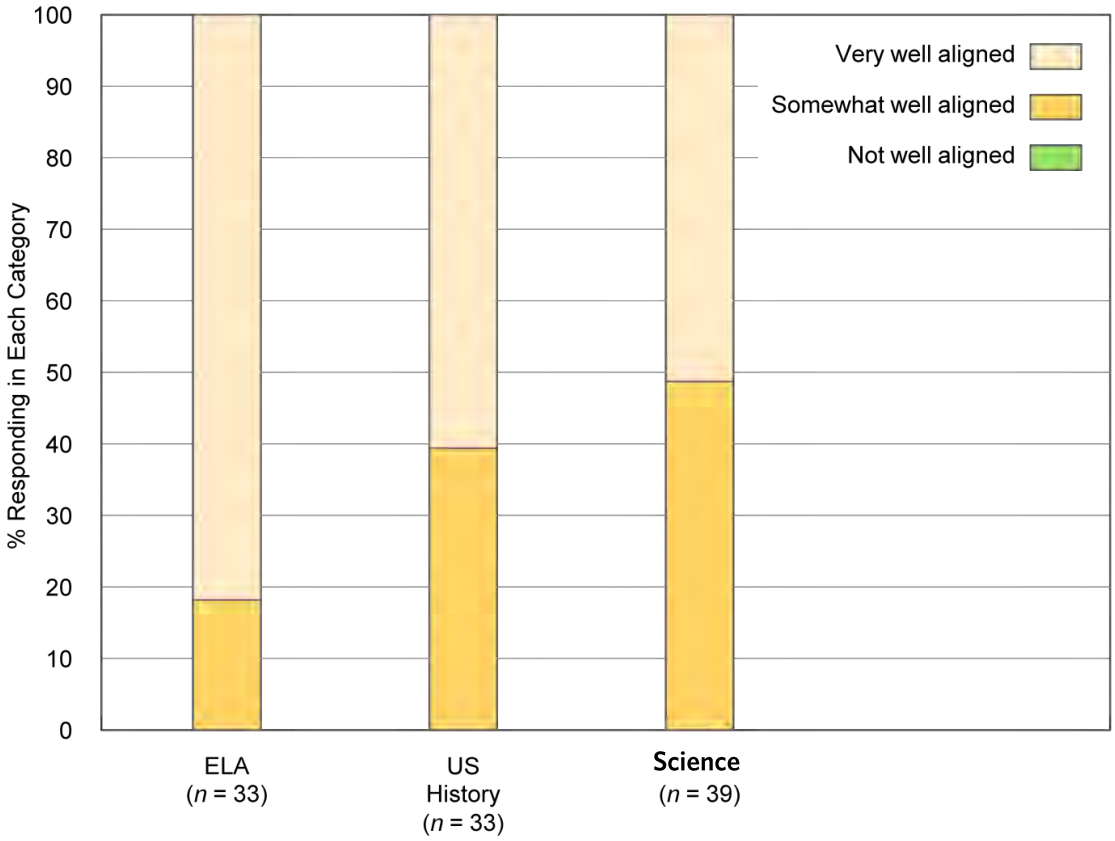


FIGURE 5. TEACHER REPORTED ALIGNMENT OF READING APPRENTICESHIP WITH CLASSROOM GOALS

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

Other Literacy-related Professional Development and State Context

In May of the second year of implementation, teachers in both RAISE and control schools were asked if they participated in other literacy-related professional development during the year. Thirty-three percent of respondents (n = 196)—split nearly equally between RAISE and control teachers—reported that they had received other literacy-related professional development, with open-ended responses almost unanimously pointing to school, district, or Intermediate Unit (IU)-led Common Core workshops in Pennsylvania. Only six teachers in California reported other literacy-related professional development, with most responses mentioning school-level curriculum workshops. In one California district where seven of the study schools were located (four treatment and three control), the study team discovered another literacy-based professional development initiative, employing similar strategies and approaches to Reading Apprenticeship, was operating during the second and third years of implementation. No study teachers reported participating in this initiative, though we do not know if this initiative weakened the contrast between RAISE and control schools. In another California district, five of the study schools (three treatment and two control) were co-located on a single campus. In the second year of implementation,

those five schools were combined into a single school under the leadership of a single principal. We have no evidence of contamination based on these schools merging from teacher survey responses; however, the disruption endemic of such large-scale school reorganization likely had an effect on school operations including teachers' ability to implement Reading Apprenticeship.

Finally, while California did not require standardized testing in 2013-2014, the landscape in Pennsylvania featured both the accountability of the Keystone exams and the groundswell towards Common Core. Teachers in Pennsylvania frequently praised Reading Apprenticeship for integrating well with the new standards, with one remarking that "the transition into Common Core with [Reading Apprenticeship] strategies is very seamless." In fact, the alignment of Reading Apprenticeship to the Common Core was one of the key selling points made by SLI when recruiting districts and schools for the project.

Commitment to Reading Apprenticeship and Overall Impressions

By the end of their second year of implementation, 63% (n = 68) of RAISE teachers reported being fully committed to Reading Apprenticeship, another 36% (n = 39) reported being willing to give it a try, and only 1% of teachers reported that it was not a priority. Additionally, teachers were asked how well they understood the Reading Apprenticeship framework.[10] As shown in Figure 6, we found that 61% (n = 65) of RAISE teachers reported that they "get" the Reading Apprenticeship model and use it often as they plan and reflect on their teaching, with another 31% (n = 33) reporting that it is starting to make more sense as they work to integrate it into their daily practice.

[10] The full-text of the three response options: 1) "I get it and am referring to it often as I plan and reflect on my teaching"; 2) "It is starting to make more sense to me as I work with the approach to integrate it into my daily practice"; 3) "I understand some aspects of it, but I do not understand how it would translate into daily practice."

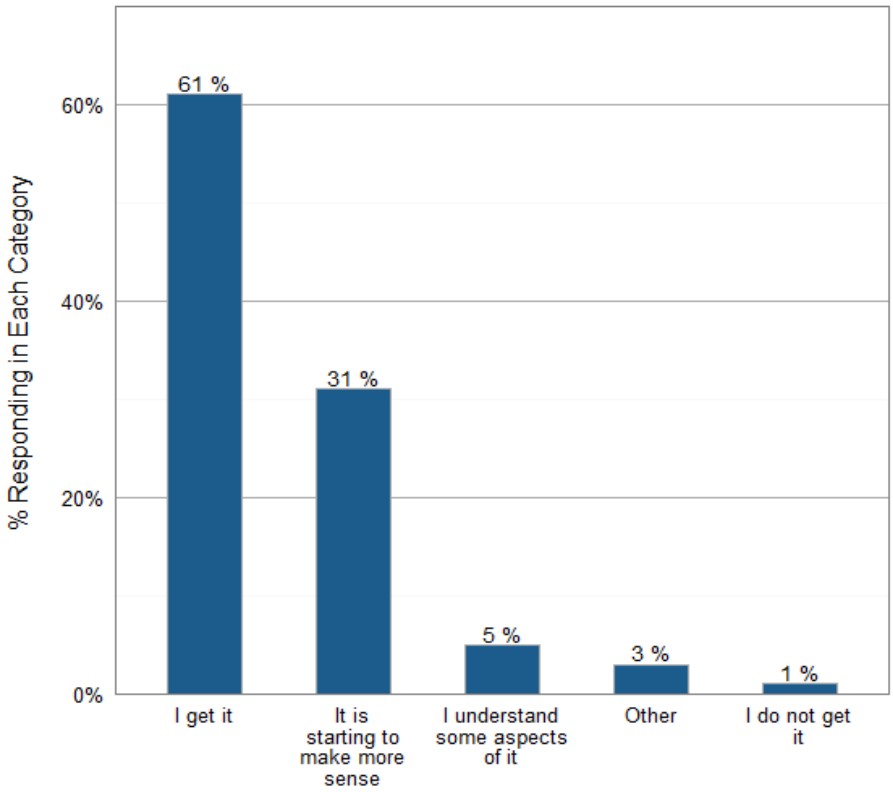


FIGURE 6. TEACHER REPORTED LEVEL OF UNDERSTANDING OF READING APPRENTICESHIP BY END OF YEAR 2

Source. Empirical Education staff calculations based on teacher responses to year 2 study surveys

(*n* = 108)

Impact of RAISE on Teacher Mediating Outcomes

In this section, we present findings on the impact of RAISE on teacher mediating outcomes in year 2. Our analyses address the research question: What are the effects of RAISE on teacher practices and teacher attitudes?

The RAISE theory of action (Table 1) posits that, as a result of the 10 days of professional development and the ongoing support of teacher leaders and on-site meetings, RAISE teachers will increase their use of practices promoted in the Reading Apprenticeship framework, such as providing extensive reading opportunities and fostering metacognitive inquiry. The model also hypothesizes that RAISE will improve teachers' confidence in delivering literacy instruction.

Information on these practices and beliefs comes from a teacher survey. We conducted all impact analyses on the "intent-to-treat" analytic sample, that is, all teachers and students in the study sample, regardless of their exposure to the RAISE intervention. We estimated impacts using the two-level hierarchical model in Appendix A. Impact Estimation Model. As noted earlier, we focus on the second year of implementation, when treatment teachers had had the opportunity to receive the full 10 days or 65 hours of professional development.[11] However, we also summarize notable findings from year 1 and year 3 and describe subject-area differences; full results from those analyses are in Appendix G. Additional Impact Analyses for Teacher Mediating Outcomes.

Key findings on teacher mediating outcomes include the following.

- RAISE had statistically significant impacts on teachers' use of core Reading Apprenticeship practices and on their confidence in delivering literacy instruction.
- The size of impact estimates increased from the first to the second year of implementation, suggesting that the ongoing support strengthened teachers' implementation of Reading Apprenticeship.
- The impact of RAISE was greater on science teachers' use of instructional practices and confidence than teachers in the other subject areas.

Impact on Teacher Practices and Confidence

Treatment group teachers were more likely than control group teachers to report that they implemented classroom practices that were promoted by Reading Apprenticeship. As shown in Table 10, RAISE had statistically significant impacts on four out of six domains of Reading Apprenticeship practice, listed below with the survey constructs that indicated impact.

- Support student effort to comprehend disciplinary text: Construct 2, Fostering student independence
- Foster metacognitive inquiry into reading and thinking processes: Construct 6, Students practicing metacognitive conversations

[11] We also focus on Year 2 because it includes the full sample of 42 schools, while year 3 includes only wave 1 schools.

- Provide explicit instruction and modeling of reading comprehension routines, tools, strategies, and processes: Construct 9, Students practicing comprehension strategies
- Foster and support student collaboration: Construct 10, Student collaboration

Statistically significant impact also emerged in teachers' confidence in literacy instruction, Construct 12.

The effect sizes for these impacts were moderate to large, ranging from 0.41 to 0.62. Findings from the teacher survey suggest that RAISE teachers made significant changes in their instructional approach and stance, embodying the Reading Apprenticeship framework. On average, RAISE teachers were significantly more likely to use practices that foster student independence and offer opportunities for peer-to-peer learning and collaboration. RAISE teachers provided instruction and modeling on metacognitive and comprehension skills at about the same level as control teachers. However, they were more likely to provide opportunities for students to *practice* those skills in class.

Teachers did not report a statistically significant impact in two domains of Reading Apprenticeship practice: providing extensive reading opportunities that reflect a variety of genres and text types (construct 1) and student engagement (a measure of the logic model element: employing instruction that promotes engagement, student-centered learning, and inquiry-based learning, construct 11).

Several items on the teacher survey asked about traditional practices that are not emphasized by Reading Apprenticeship, such as lecture, videos, and quizzes (construct 3). As expected, when looking across the full sample of teachers, these practices were not affected by RAISE.

Impact on Teacher Mediating Outcomes over Time

Impact estimates for teacher mediating outcomes increased from the first to the second year of implementation by varying degrees. (See Appendix G. Additional Impact Analyses for Teacher Mediating Outcomes.) A particularly large increase in effect size emerged between years 1 and 2 on teachers' ratings of self-confidence. This finding suggests that, over time and with the full course of professional development, teachers became more comfortable with the Reading Apprenticeship approach and with integration of Reading Apprenticeship practices and routines into their instruction. The greatest difference between what treatment teachers and control teachers reported was in their confidence structuring lessons so that students have to do the assigned reading in order to be successful, supporting students in understanding disciplinary text, modeling comprehension strategies, and supporting collaborative work.

TABLE 10. IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES, YEAR 2

Outcome measure (Domain/Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (impact)	Standard error	*p* value	Effect size	*N*
Provide extensive reading opportunities that reflect a variety of genres and text types							
1. Variety of text types	2.76	2.70	0.05	0.213	.798	0.04	206

TABLE 10. IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES, YEAR 2

Outcome measure (Domain/Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (impact)	Standard error	*p* value	Effect size	*N*
Support student effort to comprehend disciplinary text							
2. Fostering student independence	4.36	3.34	1.02**	0.309	< .001	0.51	206
Traditional instructional strategies (not expected to be affected by RAISE)							
3. Traditional instructional strategies	4.35	4.16	0.18	0.315	.562	0.09	206
Foster metacognitive inquiry into reading and thinking processes							
4. Teachers instructing metacognitive inquiry	0.76	0.84	- 0.07	0.114	.528	- 0.09	206
5. Teachers modeling metacognitive inquiry	0.96	0.88	0.08	0.096	.422	0.11	206
6. Students practicing metacognitive inquiry	1.91	1.49	0.42**	0.131	.001	0.46	206
Provide explicit instruction and modeling of reading comprehension routines, tools, strategies, and processes							
7. Teachers instructing comprehension strategies	1.49	1.57	- 0.08	0.230	.719	- 0.06	206
8. Teachers modeling comprehension strategies	1.84	1.52	0.32	0.192	.096	0.23	206
9. Students practicing comprehension strategies	3.42	2.41	1.00**	0.238	< .001	0.62	206
Foster and support student collaboration							
10. Student collaboration	4.38	3.31	1.07**	0.405	.008	0.47	206
Employ instruction that promotes engagement, student-centered learning and inquiry-based learning							
11. Student engagement	12.09	12.01	0.08	0.270	.760	0.05	206
Teacher attitudes							
12. Teachers' self-confidence in literacy instruction	39.31	36.67	2.63*	1.066	.014	0.41	206

Source. IMPAQ staff calculations based on teacher responses to year 2 study surveys

* Significant at 5%

** Significant at 1%

Variation in Impact on Teacher Practices by Subject Area

The impact on teacher practices varied by subject area. RAISE had a statistically significant impact on science teachers in four of the five areas of impact found for the full sample. The fact that the effect sizes for science teachers (0.56 to 0.97) were larger than for the full sample (0.41 to 0.62) suggests that the impacts on the full sample were largely driven by the effects on science teachers. The four areas in which RAISE achieved statistically significant impacts on science teachers are in supporting students to comprehend disciplinary text, fostering metacognitive inquiry, teaching and modeling comprehension strategies, and confidence in literacy instruction.[12] In the fifth area in which RAISE had an impact on the full sample, student collaboration, the impact on science teachers was also positive, though not statistically significant.

The effects on ELA and history teachers were less striking. RAISE had statistically significant impacts for ELA teachers in three areas: metacognitive inquiry, comprehension strategies, and student collaboration. There were no statistically significant impacts for history teachers. (See Appendix G. Additional Impact Analyses for Teacher Mediating Outcomes for details.)

We speculate that RAISE had a larger impact on science teachers than ELA and history teachers because implementing Reading Apprenticeship requires a greater pedagogical shift for science teachers. ELA and history teachers may already have been using some Reading Apprenticeship practices before participating in RAISE professional development. For example, the ELA and history teachers might have employed collaborative learning structures and taught inquiry and comprehension skills, such as annotating text, before RAISE. Science teachers may have been less likely to use these instructional practices. In that case, implementing the Reading Apprenticeship framework would require a greater change in instructional practices for science teachers than for the others.

To explore this hypothesis, we examined the reported experiences of control teachers to see if science teachers were generally less likely than ELA or history teachers to use practices such as integrating content and literacy activities. If non-RAISE science teachers were less likely than non-RAISE ELA and history teachers to use Reading Apprenticeship practices, then implementing the Reading Apprenticeship framework might be expected to have a larger impact on their instruction and a relatively smaller effect on the practices of ELA and history teachers.

The results, shown in Table 11, bear out this hypothesis. Compared with ELA and history teachers, science teachers in control schools tended to report lower implementation of the practices and classroom activities that were influenced by RAISE. Specifically, they reported lower use of classroom practices that foster student independence and less frequent opportunities for students to practice metacognitive inquiry and comprehension strategies. They also reported lower levels of confidence in providing literacy instruction. The one area in which control science teachers' scores were similar to those of ELA and history teachers was student collaboration. This is also the one area in which RAISE had a positive, but not statistically

[12] We also found that RAISE had a statistically significant impact on science teachers' reported use of traditional instructional practices such as lectures and video. These practices were included in the survey as 'distractor' items—i.e., items for which we did not hypothesize an impact caused by RAISE.

significant impact on science teachers' practices. Table 11 further shows that science teachers reported lower use of other practices and activities on which the effects of RAISE were not statistically significant.

TABLE 11. PRACTICES AND CLASSROOM ACTIVITIES REPORTED BY CONTROL GROUP TEACHERS, FOR THE SECOND YEAR OF IMPLEMENTATION

	Science (n = 32)		ELA (n = 33)		History (n = 37)		All (N = 102)	
	Average	SD	Average	SD	Average	SD	Average	SD
Teacher practices/classroom activities found to be positively impacted by RAISE								
2. Fostering student independence	1.97	1.946	4.16	1.705	3.49	1.621	3.23	1.956
6. Students practicing metacognitive inquiry	0.82	0.842	1.85	0.902	1.61	0.750	1.44	0.929
9. Students practicing comprehension strategies	1.47	1.538	2.87	1.408	2.58	1.250	2.33	1.504
10. Student collaboration	3.29	2.833	3.07	2.044	3.40	2.116	3.26	2.324
12. Teachers' self-confidence in literacy instruction	33.84	5.689	39.17	7.587	36.09	6.187	36.38	6.813
Other teacher practices/classroom activities								
1. Variety of text types	3.19	1.059	1.38	0.967	3.27	0.982	2.63	1.321
3. Traditional instructional strategies	2.88	2.160	4.50	1.734	4.61	1.885	4.03	2.065
4. Teachers instructing metacognitive inquiry	0.39	0.487	1.08	0.955	0.98	0.752	0.83	0.807
5. Teachers modeling metacognitive inquiry	0.45	0.510	1.14	0.719	0.94	0.644	0.85	0.686
7. Teachers instructing comprehension strategies	0.97	1.101	1.82	1.563	1.76	1.144	1.53	1.327
8. Teachers modeling comprehension strategies	0.97	1.128	1.80	1.349	1.54	1.106	1.45	1.233
11. Student engagement	11.96	1.742	11.96	1.643	11.98	1.915	11.97	1.759

Source. IMPAQ staff calculations based on teacher responses to year 2 study surveys

Impact of RAISE on Student Outcomes

In this section, we examine the impact of RAISE on student outcomes. Our logic model (Table 1) hypothesizes that changes in teacher practices as a result of RAISE will change students' classroom experiences, attitudes, and behaviors. Ultimately, these changes will lead to improvements in student literacy outcomes. Student mediating outcomes were measured through a student survey administered at the end of each study year. Student literacy outcomes were measured through an online literacy assessment, also administered at the end of each study year. As with the teacher mediating outcome analyses, we estimated the impact of RAISE on student mediating outcomes and literacy achievement for year 2 based on the two-level hierarchical model described in Appendix A. Impact Estimation Model.[13]

STUDENT MEDIATING OUTCOMES

This section addresses the following research question: What are the effects of Reading Apprenticeship on student engagement and on reading attitudes and behaviors?

Key findings include the following.

- RAISE produced positive and statistically significant impacts in two student mediating outcome domains that are hallmarks of the Reading Apprenticeship framework, listed below with the survey constructs that indicated impact.
    - Increased use of comprehension strategies; construct 2.4: integration of content and literacy activity
    - Increased metacognitive inquiry; construct 3.1: metacognitive conversations
- The size of the impacts on student mediating outcomes increased over time.

Impact on Student Attitudes and Behaviors

As shown in Table 12, RAISE had an impact on integration of content and literacy activity (Construct 2.4) and metacognitive conversations (Construct 3.1) in year 2. The first reflects an increase in using comprehension strategies and measured the extent to which students reported that their teachers' instructional practices fostered integration of content and literacy activities, for example, by summarizing and interpreting the meaning of passages, identifying the main themes, and working in small groups with other students to practice reading comprehension strategies. In contrast with typical high school instruction, which seldom focuses on reading comprehension in content-area courses, integrating reading instruction *into* content-area teaching is a key principle of the instruction promoted by the Reading Apprenticeship framework. Reading Apprenticeship works from the core stance that to be proficient content-area readers, students need to be taught discipline-based reading strategies in order to comprehend content-specific structures, vocabulary, methodologies, perspectives, interpretations, and biases.

[13] Missing outcome data were handled using listwise deletion, while the dummy variable adjustment method was applied to address missing covariate data.

The second area of impact was students' engagement in metacognitive conversations. Measures of metacognitive conversations include the extent to which students reported learning from each other's different ways of reading and thinking, listening and responding to each other's ideas, and exploring different ways of understanding reading. Metacognitive conversations are the foundation of Reading Apprenticeship, undergirding the four dimensions of the framework: social, personal, cognitive, and knowledge-building factors. Class discussions that focus not only on the content of texts, but also on how to read science and history materials—and why people read these materials in the ways they do—support students to clarify content, practice comprehension strategies, make connections to other related texts and topics, and make visible the often invisible process of reading (Greenleaf et al., 2011a).

These findings from the student survey support findings on teachers' practices: Students in treatment classrooms experienced a different approach to instruction that centered on providing disciplinary reading instruction and fostering metacognitive inquiry. The analyses also showed that RAISE had positive but not statistically significant impacts on other measured constructs in year 2, including participation in class discussions and class time spent reading. The impact on measures of attitude and academic disposition such as reader identity, belongingness and effort to learn were positive but very small.

Impact on Student Mediating Outcomes over Time

Impact analysis for the first year of implementation showed no statistically significant effects for any of the student constructs (see Appendix H. Additional Impact Analyses for Student Mediating Outcomes). The effect sizes for the two constructs found to be statistically significant in the second year were larger than the estimate impacts for other constructs in year 1 but were not statistically significant. This finding may be explained by the increase in effect sizes for teacher practices in year 2. Teachers' implementation of Reading Apprenticeship practices seems to have increased and improved over time, leading to better student outcomes in year 2. However, as reported below, the pattern was not consistent for all three subject areas.

TABLE 12. IMPACT ESTIMATES FOR STUDENT OUTCOMES, YEAR 2

Domain/Construct	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*N*
Increased collaboration in a community of readers and writers							
1.1: Participation in and contribution to class discussions	2.93	2.84	0.09	0.054	.102	0.12	11,398
1.2: Conferring	3.37	3.30	0.07	0.056	.194	0.09	11,400
Increased use of comprehension strategies							
2.1: Use of global reading strategies	2.94	2.92	0.03	0.039	.507	0.03	11,553
2.2: Use of problem-solving strategies	3.36	3.37	0.00	0.034	.916	0.00	11,534
2.3: Use of support reading strategies	2.59	2.53	0.05	0.051	.292	0.06	11,543
2.4: Integration of content and literacy activity	2.78	2.64	0.14**	0.052	.009	0.18	11,456
Increased metacognitive conversations							
3.1: Metacognitive conversations	2.94	2.81	0.13**	0.045	.004	0.21	11,463
Improved reader identity							
4.1: Reader identity	2.36	2.33	0.03	0.045	.533	0.04	11,419
Improved reader identity							
5.1: Student identity	3.30	3.30	0.00	0.039	.999	0.00	11,426
Increased reading of a variety of texts							
6.1: Class time spent reading	2.73	2.62	0.10	0.079	.191	0.11	11,421
6.2: Variety of reading material	2.69	2.64	0.05	0.043	.238	0.05	11,396
6.3: Pages of reading per day	2.06	2.03	0.03	0.125	.782	0.03	10,831
Increased academic engagement							
7.1: Effort to learn	3.49	3.49	0.01	0.037	.865	0.01	11,390
7.2: Happiness and belonging	3.40	3.38	0.01	0.060	.835	0.01	11,386
7.3: Engaging instruction	3.50	3.48	0.02	0.083	.834	0.02	11,401

Source. IMPAQ staff calculations based on student responses to year 2 study surveys

** Significant at 1%

Variation in Impact on Student Mediating Outcomes by Subject Area

As was the case of teacher outcomes, impact on students' experiences, attitudes, and behaviors varied by subject area. Students in both science and history classes in the treatment schools reported statistically significantly higher levels of integration of content and literacy activity and metacognitive conversations than did students in the control schools, with science students reporting the largest impacts. The difference among ELA students was smaller and not statistically significant. (See Appendix H. Additional Impact Analyses for Student Mediating Outcomes for details.) Science students also showed statistically significant impacts in additional constructs: participation in and contribution to class discussions and class time spent reading. History students reported statistically significant improvement in the extent to which they engaged in a variety of reading material in class.

STUDENT LITERACY ASSESSMENT OUTCOMES

This section addresses the following research questions.

- What are the effects of RAISE on student literacy achievement?
- What are the effects of RAISE on the student literacy achievement on English Language Learners?
- What are the effects of RAISE on student literacy by subject area (English language arts, history, science)?
- What are the effects of RAISE on student literacy by key student sub-groups, including:
    - students with low prior achievement and/or weak prior performance?
    - economically disadvantaged students?
    - minority students?
- What are the effects of RAISE on student literacy by state (California, Pennsylvania)?

As described earlier, we used a summative reading comprehension assessment to measure disciplinary literacy in ELA, U.S. history, and biology among high school students in our sample.

Table 13 reports the year 2 result for students in the "intent to treat" analytic sample.

Key findings on the impact on student literacy include the following.

- RAISE had a statistically significant positive impact on student literacy in science classes (effect size = 0.32). The effect of RAISE on the literacy scores of the full sample of students in treatment classes was positive but not statistically significant.
- Similarly, the effect of RAISE on key subgroups—including English language learners, low prior performers, students eligible for free and reduced-price lunch, and nonwhite students—was also positive, but not statistically significant.
- The effect on treatment group students was, as expected, greater in year 2 than in year 1.

Impact on Student Literacy for the Full Sample and by Subject Area

RAISE had positive but not statistically significant impacts on literacy achievement for the full sample of students enrolled in the study schools in the second year of implementation, as shown in Table 13. The improvement index in the second-to-last column of Table 13 shows the expected change in percentile rank on the literacy assessment as a result of exposure to Reading Apprenticeship. The improvement index of 5.6% for the full sample means that control group students who scored at the 50th percentile would score in the 55.6th percentile if they were exposed to Reading Apprenticeship. The finding is robust across alternative model specifications and estimation methods, meaning the results were consistent when tested using different statistical models and methods. The fact that the results are consistently positive, but not statistically significant, may indicate that the study sample was not large enough to detect a modest-sized impact.

Because RAISE focuses on discipline-specific rather than generic reading skills, we also investigated impact on literacy assessments by subject area. The results, also presented in Table 13, are similar to those for teacher and student mediating outcomes. Students in treatment science classes had a statistically significantly higher literacy achievement scores, with an estimated effect size of 0.32. This finding is also robust across alternative model specifications and estimation methods (see Appendix I. Additional Impact Analyses for Student Literacy). The improvement index of 12.6% indicates that control group students in the 50th percentile would have improved their percentile ranking to 62.6 as a result of exposure to RAISE . The impact on literacy achievement among students in ELA classes was positive but smaller, with an effect size of 0.22, and no statistical significance. For students in history classes, the estimated impact, at an effect size of 0.08, was negative but small and not statistically significant.

TABLE 13. IMPACT ESTIMATE FOR LITERACY ASSESSMENT SCORES, FULL SAMPLE BY SUBJECT AREA, YEAR 2

	Treatment group mean	Control group mean	Difference (impact)	Standard error	*p* value	Effect size	Improvement index	*N*
Literacy achievement assessment score	- 0.02	- 0.16	0.13	0.100	.184	0.14	5.6%	10,173
Students in science classes	0.01	- 0.29	0.31*	0.120	.010	0.32	12.6%	4,360
Students in ELA classes	- 0.04	- 0.24	0.20	0.135	.148	0.22	8.7%	2,936
Students in history classes	- 0.06	0.02	- 0.08	0.151	.618	- 0.08	- 3.2%	3,449

Source. ETS-designed literacy assessment

Note. As the literacy assessment measure, we used standardized scores calculated by ETS, which can take negative values. For the second year analytic sample (*n* = 10,173), for example, the unadjusted mean of the score ranged from a score of -2.45 to 2.73, with the mean of -.06 and the standard deviation of 0.937. The negative values on the score are valid values. See Appendix E. Student Literacy Assessment for additional information on the literacy assessment

measure.

Impact on Student Literacy for Key Subgroups

The instructional framework of Reading Apprenticeship is expected to improve the performance, not only of high school students generally, but also of disadvantaged and high-need students. Accordingly, we examined the effect of RAISE on literacy scores of nonwhite students, students who were eligible for free or reduced-price lunch, English language learners, and students with low prior performance on state ELA tests. Previous studies (Greenleaf et al., 2011a, 2011b; Kemple et al., 2008) have shown that Reading Apprenticeship is particularly helpful for English language learners and students with prior low ELA performance. Table 14 shows for the results for these subgroups.

The study did not find statistically significant impacts among disadvantaged students in the second year of implementation. For English language learners, the estimated impact was positive but modest, with an effect size of 0.15, and was not statistically significant. The English language learner subsample size was particularly small, with 1,156 students in 30 schools; modest impacts are not expected to be detectable at such a small sample size. For nonwhite students, the estimated positive effect size was also modest, at 0.11, and was not statistically significant. Among the students who were eligible for free or reduced-price lunch, the positive effect size was larger, 0.23, and approached statistical significance. The estimated effect for each of these subgroups was about the same size as the effect for their complimentary group (e.g., non-ELLs, white, not-eligible for free or reduced-price lunch), and we did not find evidence for differential impacts by these subgroups.

We also examined literacy score results by state. Though the result is not statistically significant, Pennsylvania schools showed a medium-sized positive effect of 0.25 and an improvement index of nearly 10%. The effect size for California schools was near zero. As noted in the implementation findings earlier in this report, the contexts for these two states differed. California schools served more students who were low income, English language learners, and poor prior performers, and therefore more high-needs than Pennsylvania students. Though teachers' participation in the professional development and onsite teacher meetings did not vary significantly by state, other factors such as competing initiatives and school reorganization seemed to have affected California more than Pennsylvania schools. These dynamics may have contributed to different outcomes for students in these states.

TABLE 14. IMPACT ESTIMATES FOR LITERACY ASSESSMENT SCORES BY STUDENT SUBGROUPS, YEAR 2

Group	Treatment group mean	Control group mean	Difference (impact)	Standard error	*p* value	Effect size	Improvement index	*N*
Full sample	- 0.02	- 0.16	0.13	0.100	.184	0.14	5.6%	10,173
English language learners	- 0.75	- 0.84	0.09	0.117	.428	0.15	6.0%	1,156

TABLE 14. IMPACT ESTIMATES FOR LITERACY ASSESSMENT SCORES BY STUDENT SUBGROUPS, YEAR 2

Group	Treatment group mean	Control group mean	Difference (impact)	Standard error	*p* value	Effect size	Improvement index	*N*
Students with low prior performance	- 0.59	- 0.71	0.12	0.095	.207	0.18	7.1%	2,745
Eligibility for free / reduced-price lunch	- 0.20	- 0.39	0.20	0.116	.089	0.23	9.1%	4,776
Nonwhite students	- 0.19	- 0.29	0.10	0.102	.333	0.11	4.4%	5,365
California students	- 0.15	- 0.19	0.04	0.127	.765	0.04	1.6%	6,440
Pennsylvania students	0.16	- 0.08	0.24	0.175	.171	0.25	9.9%	3,733

Notes. * significant at 5 percent ** significant at 1 percent

As the literacy assessment measure, we used standardized scores calculated by ETS, which can take negative values. For the second year analytic sample (*n* =10,173), for example, the unadjusted mean of the score ranged from a score of -2.45 to 2.73, with the mean of -.06 and the standard deviation of 0.937. The negative values on the score are valid values. See Appendix C for additional information on the literacy assessment measure.

Source. IMPAQ staff calculations on ETS assessment data

Impact on Student Literacy over Time

In addition to testing the effect of RAISE at the end of the second year of implementation, we examined the literacy assessment outcomes for students in the first and third years of implementation.[14] When the student literacy assessment was administered at the end of the first school year, treatment teachers had been exposed to only 7 of the 10 days of professional development. As expected, RAISE had less impact in year 1 than in year 2. We found no statistically significant improvement on the student literacy assessment, either for the full sample or for any of the subgroups discussed above.

[14] The data do not lend themselves to conclusions about dosage or trends in program effects over the three years. The composition of the student samples changed from one year to the next; for example, less than 25% of the year 2 student sample was also in the year 1 sample. Furthermore, the year 3 sample is limited to the first wave of schools and their students, and therefore, is much smaller than the first two years' samples. Thus, we cannot draw conclusions about any growth or decline in program effects over the three years.

Summary and Conclusions

PROJECT OVERVIEW

The project evaluated in this randomized controlled trial was an instance of the Reading Apprenticeship model designed for scale. The RAISE project served nearly 2,000 teachers and 630,000 high school students in five states (California, Indiana, Michigan, Pennsylvania, and Utah). Teachers in 42 schools in two of those states participated in the random assignment study.

To support an implementation that would not require ongoing direct involvement of the staff of SLI, the developers used the i3 funding to develop and field an innovation that aimed to build local capacity at the school, district, and state levels. Fundamental to this approach was fostering a community of teachers within the schools, each supported by a teacher leader who could convene monthly team meetings. To encourage this community in high schools, which are often highly departmentalized, the project chose to support three major content areas: ELA, history and science. Each content area had been the subject of previous studies, demonstrating effect sizes ranging from 0.14 to 0.26 (Greenleaf et al., 2011a, 2011b; Kemple et al., 2008; Somers et al., 2010). For this project, an important purpose for including multiple content areas was to enhance the teacher team at each school and to develop a school-wide commitment to the approach. Including multiple subjects at different grades would also increase the chance that students may receive multiple exposures to the approach. Developers designed content-specific professional development, encompassing a total of 10 days of training over the course of a year.

THE IMPACT OF RAISE ON STUDENT ACHIEVEMENT

Student achievement outcomes were measured in the spring of the second year of the intervention. The subject-specific literacy achievement test developed by ETS provided a common scale across the three content areas, allowing both subject-specific and full-sample results. The largest difference in literacy achievement between treatment and control groups emerged among students in science classes, with a standardized effect size of 0.32. This effect size translates into an improvement index of 12.6%; that is, control students in the 50th percentile would be expected to score in the 62.6th percentile if they were taught by RAISE teachers. Results for ELA classes were positive, but not statistically significant, with an effect size of 0.22. The effect on history classes was negligible and was not statistically significant. The positive effect of RAISE on the full sample of students (all subjects combined) also was not significant, at an effect size of 0.14.

The positive, yet not statistically significant, results for the full sample and for ELA suggest that the sample may not have been large enough to detect meaningful impacts. The study was designed to detect an effect size of 0.16 for the full sample and 0.24 for subgroups. The small subsample sizes may help to explain the lack of statistical significance of findings on key subgroups including English language learners, low-income students, and students with low prior achievement. The effects of RAISE on all of these subgroups were positive, with effect sizes ranging from 0.11 to 0.23, but they are not statistically significant. The results for English language learners, with a non-significant but positive effect of 0.15, may be particularly affected by a small subsample size, as this was based on only 30 of the 42 schools in the sample.

The results by state show that RAISE had a positive but not statistically significant impact on students in Pennsylvania, with an effect size of 0.25. Even though this finding is not statistically significant, given its size, this effect is considered "substantively important" by the What Works Clearinghouse (WWC, 2014). In contrast, the effect size on students in the California sample was effectively zero. Contextual factors that affected implementation may explain this difference.

THE IMPACT OF RAISE ON TEACHER AND STUDENT MEDIATING OUTCOMES

RAISE teachers were more likely than control teachers to implement practices that foster student independence and collaboration and develop students' metacognitive and comprehension skills and strategies: hallmarks of the Reading Apprenticeship approach. RAISE teachers were also more likely to give students time to practice metacognitive and comprehension skills in class. With the exception of the area of student collaboration, these effects were particularly strong for science teachers.

RAISE's positive impacts on mediating student outcomes, measured by a student survey, were also stronger in science classes than in the full sample. Across subject areas, RAISE had a statistically significant positive effect in two key areas on the student survey: metacognitive conversations and practices integrating subject-area content and literacy activities. In science classes, RAISE had statistically significant positive impacts on two additional outcomes: students' participation and class time spent reading.

Examination of data from teacher and student surveys offers possible explanations of the difference in effects between science and the two other subjects. In keeping with the RAISE logic model (Table 1), the impacts on science teacher and student mediating outcomes are congruent with impacts in science students' literacy achievement. One explanation for the greater impact on science may be that the Reading Apprenticeship framework requires a larger pedagogical shift for science teachers than for ELA and history teachers. Examination of the control group shows that ELA and history teachers were more likely than science teachers to implement some of the literacy practices characteristic of the Reading Apprenticeship framework. This finding suggests that implementing the Reading Apprenticeship approach had a greater effect on science teachers and students because it required a larger change in core practices for science teachers than for ELA and history teachers. One could also postulate that science teachers, because they are less apt to employ the type of core instructional practices that are congruent with the Reading Apprenticeship model, may be more resistant to or have more difficulty implementing the framework than history and ELA teachers. The results of this study suggest that this is not the case. Rather, science teachers appeared to embrace the model. Changing the science teachers' core practices to include a measurable focus on disciplinary literacy is a noteworthy accomplishment. It demonstrates the potential of RAISE to address the dearth of content-specific reading instruction in U.S. secondary schools (Ness, 2008, 2009; Vaughn et al., 2013)—even in science, where the need may be greatest.

In other areas of the subject-specific results, the logic model's chain of effects, from teacher mediating outcomes to student mediating outcomes to student achievement, is less strong. As expected, the effects of RAISE on student literacy scores by subject are congruent with the effect on teacher practices in those subjects: The largest impact is in science, followed by ELA. History shows no significant impact in either teacher practice or student achievement. However, the results for *student* mediating outcomes are

different. They show positive and significant impacts on student behaviors and attitudes in both science and history, but the impact in ELA is smaller and not statistically significant. There could be a variety of reasons for this apparent lack of congruence, including measurement error. For example, because the literacy assessment is content-specific, different subject tests may have different properties; although ETS's psychometric testing did not indicate any significant difference in the subject-area forms (O'Reilly et al., 2014). Another explanation is that the items on the student survey may have been understood differently by ELA students than by history and science students.

CONTEXTUAL FACTORS RELATED TO RAISE IMPLEMENTATION

The extent to which RAISE was implemented faithfully in treatment schools should, according to the logic model (Table 1), support the positive and significant impacts that the professional development had on teacher and student outcomes. This study's implementation analyses showed that RAISE teachers reported a number of factors that supported implementation of Reading Apprenticeship. Nearly all of the teachers (over 90%) who responded to survey questions about the RAISE professional development felt that it "moderately", "more than moderately", or "completely" prepared them to use the literacy practices modeled during the training. RAISE teachers reported receiving support for literacy instruction at a greater frequency than control teachers. A large majority of RAISE teachers (86%) also reported that Reading Apprenticeship was aligned with their classroom goals and content standards. By the end of the second year of implementation, over 90% of the RAISE teachers reported that they were continuing to integrate the framework into instruction. Over three-fifths reported being fully committed to making Reading Apprenticeship work.

However, implementation of RAISE was not without challenges. The most commonly cited challenge was competing initiatives. Another issue was participation in professional development. Though a majority of RAISE teachers (77%) participated in 9 of the 10 days of professional development, fidelity of implementation in some schools did not reach the level hypothesized by the program developers to be "successful" (defined as having three-fourths of participating teachers attend 9 days of professional development). Just over half (55%) of the treatment schools met this threshold. A third issue was schools' ability to build a community of RAISE teachers through ongoing participation in monthly RAISE team meetings, another critical support intended to foster adoption and sustainability of Reading Apprenticeship at scale. The fidelity threshold for monthly meeting attendance was met. However, meeting attendance declined in the second year of implementation; the percentage of RAISE teachers who reported attending a meeting in a particular month ranged from a low of 40% to a high of 60%. Though this decline may suggest that teachers were less engaged in RAISE, it may alternatively suggest that teachers found other ways to collaborate and support implementation of Reading Apprenticeship.

Implementation challenges may have contributed to the large difference between impacts on student achievement in California (effect size = 0.04) and Pennsylvania (effect size = 0.25). Two contextual factors may have impeded implementation in California: the reorganization of five treatment and control schools and concurrent implementation of a similar literacy-based professional development initiative in seven treatment and control schools. Another possible factor is accountability standards. Though both states experienced intense focus on implementation of the Common Core standards, California schools received

waivers from state standardized testing requirements, while Pennsylvania schools implemented new science end-of-course tests required for graduation. Several Pennsylvania teachers commented on surveys that Reading Apprenticeship supported their transition to new standards and assessment systems. The new accountability pressures in Pennsylvania, particularly for science, could have positively influenced teachers' willingness to embrace the Reading Apprenticeship framework.

LIMITATIONS TO THE STUDY

As with any research, this study had some limitations, including measurement challenges, insufficient sample size, and attrition.

One measurement challenge is reliance on teacher and student self-reports of behaviors and beliefs. Self-reported information can be subject to recall errors; for example, teachers' memory of the amount of class time they spent on specific activities last week may be inaccurate. Recall error should, however, affect treatment and control teachers and students equally, so it would not necessarily bias the impact estimate. A larger issue in the reliability of self-reports may be the tendency to provide socially desirable responses. Some teachers who received RAISE professional development may have over-reported their use of Reading Apprenticeship strategies simply because they knew they were expected to use these strategies. The research team attempted to address these issues by constructing survey questions that did not rely either on jargon that would be more familiar to RAISE teachers or on descriptions of activities that would clearly signal a Reading Apprenticeship approach. We also surveyed teachers monthly rather than yearly to increase reliability and to reduce that chance that reports would reflect idiosyncratic variations in school schedules such as field trips or testing days. As is common in randomized control trials, neither teacher nor student survey measures were administered at the start of the evaluation (prior to "treatment"); therefore, we are not able to rule out pre-existing differences between the groups in the way they respond to surveys. A final issue is with the timing of the student survey, which was fielded only once per year, and thus students' responses may not accurately capture the impact of all practices employed by their teachers throughout the year.

Another possible measurement limitation is that the impact on student literacy was assessed using a specialized, subject-specific assessment developed for this project. The assessment developers took care to avoid over-aligning the instrument to the Reading Apprenticeship model; for example, they did not use Reading Apprenticeship terminology or concepts that are not ordinarily used in classrooms, and they took the reading passages from a widely used test (the National Assessment of Educational Progress).

The study's sample size presented additional limitations. The evaluation did not have a large enough sample to detect effects that would typically fall into the "modest" category. This limitation particularly affected our ability to analyze the effects of the intervention on subgroups of students, such as English language learners. Even the ability to detect statistically significant effects by subject area may have been affected by study school samples. At random assignment, three schools did not have any eligible ELA teachers, three did not have any eligible history teachers, and two did not have any eligible science teachers. This variation may have limited our ability to detect statistically significant effects, explaining, for example, the effect size of 0.22 for literacy achievement among ELA students—an effect that, although is not statistically significant, is likely to be educationally meaningful.

A third limitation of the study is significant student attrition: 31% of students did not complete the literacy assessment in year 2 (though the sample of treatment and control students who did have data was statistically equivalent in terms of their demographic characteristics). The effect of the missing data on the results is unclear. Missing data was less of an issue for the teacher and student surveys, where the missing data rates fell within the range of what is commonly considered acceptable: Year 2 survey data were not available for 15% of students and 18% of teachers.

FUTURE DIRECTIONS

The RAISE project represents an ambitious effort by SLI to build self-sustaining supports for implementing Reading Apprenticeship and bringing it to scale. The schools in this study represent 15% of all schools affected by RAISE. SLI developed a scale-up model to reach hundreds of schools across states and contexts in order to support academic literacy instruction. Findings from this study demonstrate the success of the project in providing teachers with training and support to help them change their instructional practices in order to foster metacognitive inquiry and support comprehension in the content areas.

However, the results also point to several areas where further investigation could help SLI achieve an even greater impact. Specifically, investigation into the reasons that RAISE had a greater impact in science could inform refinements to the professional development and support that may increase the positive impact in ELA and history classes. Similarly, the difference in impact between Pennsylvania and California calls for investigation into the implementation and contextual differences that may have influenced these results, so that SLI can discover how to address such obstacles in future implementations. Overall, the study's findings demonstrate the potential of RAISE to address the paucity of content-specific reading instruction in U.S. secondary schools—especially in science, where the need may be greatest.

References

ACT, Inc. (2013a). *ACT National Curriculum Survey 2012: English Language Arts*. Iowa City, IA.

ACT, Inc. (2013b). *ACT National Curriculum Survey 2012: Science*. Iowa City, IA: ACT, Inc.

ACT, Inc. (2012). *Catching Up to College and Career Readiness*. Iowa City, IA: ACT, Inc.

ACT, Inc. (2009). *ACT National Curriculum Survey 2009*. Iowa City, IA: ACT, Inc.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education.* Chapel Hill, N.C.: Horizon Research.

Common Core Standards Initiative. (2012). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects.

Duschl, R., Schweingruber, H., & Shouse, A. (2007). *Taking science to school: Learning and teaching science in grades K-8* (National R). Washington D.C.: The National Academies Press.

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., & Jones, B. (2011a). Integrating literacy and science in science: Teaching and learning impacts of Reading Apprenticeship professional development. *American Educational Research Journal*, *48*(3), 647–717.

Greenleaf, C.L., Hanson, T., Herman, J., Litman, C., Rosen, R., Schneider, S., & Silver, D. (2011b). *A Study of the Efficacy of Reading Apprenticeship Professional Development for High School History and Science Teaching and Learning*. Final report to Institute for Education Sciences, National Center for Education Research, Teacher Quality/Reading and Writing, Grant # R305M050031

Kemple, J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., & Drummond, K. (2008). *The Enhanced Reading Opportunities Study: Early Impact and Implementation Findings* (NCEE 2008-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136.

Ness, M. K. (2008). Supporting secondary readers: When teachers provide the "what," not the "how." *American Secondary Education*, *37*(1), 80–95.

Ness, M. K. (2009). Reading comprehension strategies in secondary content area classrooms: Teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons*, *49*(2), 143–166.

O'Reilly T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing Reading Comprehension Assessments for Reading Interventions: How a Theoretically Motivated Assessment Can Serve as an Outcome Measure. *Educational Psychology Review, 26*(3), 403-424.

Reisman, A. (2011). The "Document-Based Lesson": Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers. *Journal of Curriculum Studies*, *44*(2), 233–264.

Smith, C., & Ochoa-Angrino, S. (2012, November). *High school students' emotional responses to science reading and academic reading engagement: Relationships to science achievement*. Presented at the LRA 62nd Annual Conference, Sheraton San Diego Hotel and Marina, San Diego, CA.

Somers, M. A., Corrin, W., Sepanik, S., Salinger, T., Levin, J., & Zmach, C. (2010). *The Enhanced Reading Opportunities Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers*. NCEE 2010-4021. National Center for Education Evaluation and Regional Assistance.

U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). (2013). *Reading Assessment, Data Explorer Tool*. Retrievable from http://nces.ed.gov/nationsreportcard/naepdata

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, *48*(1), 77–93.

What Works Clearinghouse (WWC). (2014). *WWC Procedures and Standards Handbook (version 3.0).* Princeton, NJ: U.S. Department of Education, Institute of Education Sciences.

Appendix A. Impact Estimation Model

The program impacts were estimated as the differences in outcomes between program and control groups. The analyses produced intent-to-treat estimates of the intervention, based on samples that included all randomly-assigned schools and their study-eligible cohorts of teachers and students, regardless of the level of actual participation in the intervention.

We estimated the effects of Reading Apprenticeship using a two-level linear model, with individuals (students or teachers) nested within schools.

Level 1: Student or Teacher

(Eq. 1) $$Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + \varepsilon_{ij}$$

where Y_{ij} denotes the outcome for student/teacher i (i = 1, 2, … N) in school j (j = 1, 2, … K), X_{qij} represents individual-level covariates and ε_{ij} is the error term specific to student/teacher i in school j.

Level 2: School

(Eq. 2) $$\beta_{0j} = \gamma_0 + \gamma_1(Treatment_j) + \sum_{s=1}^{S} \eta_s W_{sj} + u_j$$

(Eq. 3) $$\beta_{qj} = \delta_{q0}$$

where *Treatment* is a binary variable indicating whether school j is randomly assigned to receive Reading Apprenticeship (*Treatment* = 1) or not (Treatment = 0) and its coefficient, γ_1, represents the estimated effects of Reading Apprenticeship on the student (or teacher) outcome. W_{sj} represents school-level covariates and u_j is the error term specific to the j-th school, representing the random school effects.

In practice, we estimate the reduced-form of the two-level random-intercept model, which can be expressed as follows:

(Eq. 4) $$Y_{ij} = \gamma_0 + \gamma_1(Treatment_j) + \sum_{q=1}^{Q} \delta_q X_{qij} + \sum_{s=1}^{S} \eta_s W_{sj} + u_j + \varepsilon_{ij}$$

where variables are as described above. Equation 4 is derived by combining equations 1 through 3. Coefficients δ_q and η_s are estimators for marginal effects of individual and school level covariates, respectively. We assume that individual-level effects do not vary across blocks (schools).

To evaluate the impact of Reading Apprenticeship training, we tested the null hypothesis that there was no difference between the groups (H_0: $\gamma_1 = 0$). If the null hypothesis was rejected by a two-tailed test at the 5% significance level, we concluded that the outcome was different between students (teachers) in schools exposed to Reading Apprenticeship versus students (teachers) in schools that were not exposed to Reading Apprenticeship.

The level of inference for this study is the school, which allows us to interpret the impact estimate as reflecting a combination of the intervention's effects on the outcomes of individuals and its effect on the composition of the students (or teachers) in the treatment and comparison schools. However, as shown below, the study provided evidence that the two groups of students (and teachers) in the analytic sample were equivalent at baseline. This suggests that the estimated effect largely reflects the intervention's effect on outcomes for these students (or teachers). The study thus assesses whether there was a difference in the

average outcomes between students enrolled (or teachers employed) in the schools randomly assigned to Reading Apprenticeship and those that were randomly assigned to the control group. This interpretation of results is consistent with What Works Clearinghouse (WWC) standards.

COVARIATES

For the analysis of impact of Reading Apprenticeship on student literacy, we selected covariates from a set of potential explanatory variables suggested in our original i3 evaluation design document, using the following guidelines.

- Include blocking variables (school-level)
- Include variables for which baseline equivalence is not established
- Limit variables to those that best fit the model
- Keep a set of covariates that would empirically explain the variation in the outcome well

We examined potential covariates systematically, according to the guidelines outlined above. Based on these analyses, we selected the following, in addition to assignment block variables. [15]

- School-level baseline state test results (percent proficient on 11th grade state ELA/Reading test)
- Student pretest scores (state-standardized z-scores of the 8th grade state ELA/Reading test)
- Student gender
- Student race/ethnicity (Asian, Black, Hispanic, other non-White, White)
- Student ELL status
- Student special education status
- Student free/reduced price lunch status

We used the same covariates in the student survey model. For impact analysis on teacher survey responses, we applied similar guidelines to the data available for teachers. Below are the resulting covariates.

- Blocking variables (school-level)
- School-level baseline state test results (percent proficient on 11th grade state ELA/Reading test)
- Teacher gender
- Teacher race/ethnicity (non-White, White)
- Teacher education (BA, more than BA)
- Teacher's total years of teaching
- Teacher's years of teaching the subject taught during the study
- Teacher's reading specialist status

[15] We conducted sensitivity analyses based on alternative covariates specifications. These include an unadjusted model with only blocking variables, as well as models with additional school-level demographic and achievement controls.

Appendix B. Student Survey Constructs

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
Domain 1: Increased collaboration in a community of readers and writers		**0.859**	**0.870**
1.1 - Frequency of contribution to and participation in class and small group discussions	In this class, how often do you discuss new or difficult vocabulary?	0.783	0.815
	In this class, how often do you explain what we have read?		
	In this class, how often do you discuss different interpretations of what we have read?		
	In this class, how often do you have a class discussion about something that the whole class has read?		
1.2 - Effective and respectful collaboration with peers, drawing on each other's knowledge	How much does this class include working together to figure out the meaning of the readings?	0.723	0.752
	How much does this class include listening and responding to each other's ideas?		
	Thinking about this class, how often have you and your classmates worked with partners or groups on reading assignments in class?		
	Thinking about this class, how often have you and your classmates shared difficulties and ways you solved reading comprehension problems?		
1.3 - Extent of discussion and exchange of information	Students get to decide how activities are done in this class.	0.747	0.753
	Students speak up and share their ideas about class work.		
	My teacher wants us to share our thoughts.		
	My teacher respects my ideas and suggestions.		
	My teacher gives us time to explain our ideas.		

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
Domain 2: Increased use of comprehension strategies		**0.931**	**0.931**
2.1 - Use of global reading strategies	I have a purpose in mind when I read.	0.840	0.861
	I think about what I know to help me understand what I read.		
	I preview the text to see what it's about before reading it.		
	I skim the text first by noting characteristics like length and organization.		
	I use tables, figures, and pictures in the text to increase my understanding.		
	I use context clues to help me better understand what I'm reading.		
	I use typographical aids like boldface and italics to identify key information.		
	I check my understanding when I come across conflicting information.		
	I try to guess what the material is about when I read.		
	I decided what to read closely and what to ignore.		
	I critically analyze and evaluate the information presented in the text.		
	I think about whether the content of the text fits my reading purpose.		
	I check to see whether my guesses about the text are right or wrong.		

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
2.2 - Use of problem-solving strategies	I read slowly but carefully to be sure I understand what I'm reading. When the text becomes difficult I pay closer attention while reading. I stop from time to time to think about what I'm reading. I try to picture or visualize information to help me understand what I read. When the text becomes difficult I reread to increase my understanding. I try to guess the meaning of unknown words or phrases. When text becomes difficult, I read aloud to help me understand what I read. I adjust my reading speed according to what I'm reading. I try to get back on track when I lose concentration.	0.815	0.818
2.3 - Support reading strategies	I take notes while reading to help me understand what I read. I summarize what I read to reflect on important information about the text. I underline or circle information in the text to help me remember it. I paraphrase (restate ideas in my own words) to better understand what I read. I go back and forth in the text to find relationships among ideas in it. I ask myself questions I like to have answered in the text. I use reference material, such as a dictionary, to help me understand what I read. I discuss what I read with others to check my understanding.	0.809	0.826

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
2.4 - Integration of content and literacy activity	Thinking about this class, how often has your teacher worked with students or small groups while they practiced reading comprehension strategies?	0.784	0.789
	In this class, when reading a story, article, or other passage, how often does your teacher ask you to summarize the passage?		
	In this class, when reading a story, article, or other passage, how often does your teacher ask you to interpret the meaning of the passage?		
	In this class, when reading a story, article, or other passage, how often does your teacher ask you to identify the main themes of the passage?		
Domain 3: Increased metacognitive inquiry		**0.617**	**0.639**
3.1 - Students discuss and inquire into their own and others' reading processes	How much does this class include learning from each other's different ways of reading and thinking?	0.617	0.639
	Thinking about this class, how often has your teacher taught different ways to help students understand reading better (for example: thinking aloud, analyzing sentences or chunks of text, questioning, summarizing)?		
	Thinking about this class, how often has your teacher talked about what is going on in the teacher's mind while the teacher reads materials?		
Domain 4: Improved reader identity		**0.751**	**0.749**
4.1 - Increased student awareness of reading processes, habits, strengths, weaknesses, attitudes and preferences	How much has your experience in this class helped you in understanding yourself better as a reader and learner?	0.751	0.749
	How much has your experience in this class helped you in making you curious to read about other things in this subject area (e.g., science, history, or English)?		
	How much has your experience in this class helped you in seeing yourself as a reader?		

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
Domain 5: Student identity		**0.887**	**0.883**
5.1 - Student identity	One of my goals in this class has been to learn as much as I can.	0.887	0.883
	In this class, it is important to me to thoroughly understand my class work.		
	I care about pleasing my teacher in this class.		
	How much has your experience in this class helped you in thinking about your future educational goals?		
	How much has your experience in this class helped you in being a more serious student?		
	How much has your experience in this class helped you in thinking of yourself as a capable student?		
	How much has your experience in this class helped in feeling like you can succeed in more challenging classes?		
	How much has your experience in this class helped you in being willing to tackle challenging reading materials?		
	Even if the work in this class is hard, I can learn it.		
	I'm certain I can master the skills taught in this class.		
	I can do almost all the work in this class if I don't give up.		
	I have been able to figure out the most difficult work in this class.		

TABLE B1. SURVEY OUTCOME DOMAIN, CONSTRUCTS, SURVEY QUESTIONS, AND RELIABILITY COEFFICIENTS

Construct	Item	Cronbach's α Year 1 (2011-12)	Cronbach's α Year 2 (2012-13)
Domain 6: Increased reading of a variety of texts			
6.1 - Class time spent reading	Thinking about this class, how often have you and your classmates spent class time reading?	--	--
6.2 - Frequency of incorporation of graphs, charts, tables, and illustrations in reading	Thinking about this class, how often has your teacher taught ways to read charts, graphs, tables and illustrations?	--	--
6.3 - Pages of reading per day	About how many pages a day do you have to read in school and for homework for this class?	--	--
Domain 7: Academic engagement		**0.886**	**0.890**
7.1 - Increased effort to learn	I have pushed myself hard to completely understand my lessons in this class. When doing schoolwork for this class, I try to learn as much as I can and I don't worry about how long it takes. I have done my best quality work in this class all year long. For this class, I try hard to be on time and not to be absent.	0.719	0.713
7.2 - Increased efficacy, happiness, and satisfaction	This class is a happy place for me to be. This class feels like a happy family. I feel respected in this class. Being in this class makes me feel angry. I am satisfied with what I have achieved in this class.	0.744	0.741
7.3 - Student report of engaging instruction	This class does not keep my attention—I get bored. My teacher makes learning enjoyable. My teacher makes lessons interesting. I like the ways we learn in this class.	0.835	0.838

Appendix C. Teacher Survey Constructs

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Variety of Text Types	Please select the types of texts that students in your target class worked with during the week (0 = No; 1 = Yes): • Newspaper/ magazine articles (including articles on-line) Textbook • Graphs/ charts/ images/ diagrams • Historical documents • Literature • Illustrations • Reference text • Lab procedures	0.61	0.48	0.54
Fostering Student Independence	Over the entire week, how many minutes did you spend using each of the following approaches to help your students understand text (# minutes): • Guided practice of reading comprehension strategies • Students teach other students • During the week, which of the following strategies did students learn to help them understand text (0 = No; 1 = Yes): • Discussing confusing parts of text- Teacher instructs • Discussing confusing parts of text- Teacher models • Discussing confusing parts of text- Student practice	0.67	0.65	0.66

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Traditional Reading Strategies	Over the entire week, how many minutes did you spend using each of the following approaches to help your students understand text (# minutes): • Direct instruction (e.g. presentation, summary, background info on topic, mini-lecture) • Video • Quizzes • Asked oral questions about details of the text to check student understanding	0.70	0.68	0.74
Teachers Instructing Metacognitive Inquiry	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0= No; 1 = Yes): • Working in groups to discuss meaning of texts • Asking questions about the text • Writing to clarify understanding • Previewing long or challenging texts to identify strategies for dealing with them (Selected Teacher instructs)	0.61	0.59	0.65
Teachers Modeling Metacognitive Inquiry	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0 = No; 1 = Yes): • Working in groups to discuss meaning of texts • Asking questions about the text • Writing to clarify understanding • Previewing long or challenging texts to identify strategies for dealing with them	0.57	0.55	0.54

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Students Practicing of Metacognitive Inquiry	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0 = No; 1 = Yes): • Working in groups to discuss meaning of texts • Asking questions about the text • Writing to clarify understanding • Previewing long or challenging texts to identify strategies for dealing with them	0.69	0.63	0.66
Teachers Instructing Comprehension Strategies	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0 = No; 1 = Yes): • Setting a reading purpose • Choosing a reading approach that fits the reading purpose • Visualizing what the author is describing or representing content in drawings • Making sense of graphs and other visuals • Predicting • Annotating text (e.g. making notes in the margins of text) • Re-reading • Taking on different roles to make sense of the text (e.g. presenter, note taker)	0.71	0.68	0.74

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Teachers Modeling Comprehension Strategies	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0 = No; 1 = Yes): • Setting a reading purpose • Choosing a reading approach that fits the reading purpose • Visualizing what the author is describing or representing content in drawings • Making sense of graphs and other visuals • Predicting • Annotating text (e.g. making notes in the margins of text) • Re-reading • Taking on different roles to make sense of the text (e.g. presenter, note taker)	0.7	0.66	0.7
Students Practicing Comprehension Strategies	During the week, which of the following strategies did students learn to help them understand text? Please indicated whether you provided instruction, modeled (i.e., presented an example of a behavior that students can emulate or learn from), and /or asked students to practice while you monitored progress (0 = No; 1 = Yes): • Setting a reading purpose • Choosing a reading approach that fits the reading purpose • Visualizing what the author is describing or representing content in drawings • Making sense of graphs and other visuals • Predicting • Annotating text (e.g. making notes in the margins of text) • Re-reading • Taking on different roles to make sense of text (e.g. presenter, note taker)	0.75	0.71	0.75

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Student Collaboration (Survey 2, 4, 6, 8 only)	During the week, how many minutes did your target class students spend working in class on reading activities and writing activities in the following situations (# minutes): • Reading in pairs • Reading in small groups • Writing in pairs • Writing in small groups • Writing as a class	0.63	0.64	0.67
Student Engagement	What portion of students in the target class did the following occur (1 = none; 2 = some; 3 = about half; 4 = most; 5 = Nearly all): • Completed their homework • Paid attention in class • Actively participated in class activities	0.64	0.68	0.71

TABLE C1. TEACHER SURVEY CONSTRUCT NAME, SURVEY QUESTIONS, AND RELIABILITY SCORE

Construct Name	Survey Question	Cronbach's α Yr 1 Report Appendix, 12 Mos Wave 1	Cronbach's α 12 Months, Both Waves	Cronbach's α 24 Months, Both Waves
Teacher Self-Confidence in Literacy Instruction (Survey 3, 8 only)	• Please rate your level of confidence in your ability to do the following (classroom instruction, 1 = very low; 2 = low; 3 = moderate; 4 = high; 5 = very high): • Provide opportunities for reading a variety of texts of different types/genres • Teach students to analyze their own thinking about texts • Structure lessons so that students have to do the assigned reading in order to be successful • Support students in their attempts to understand disciplinary text (e.g. challenging literature, textbooks, primary documents, scientific articles) • Provide explicit instruction around reading comprehension strategies (e.g. setting a reading purpose, previewing text, chunking, visualizing) • Model/demonstrate reading comprehension strategies (e.g. setting a reading purpose, previewing text, chunking, visualizing) • Support students in working on reading and writing activities in groups (small groups or whole class), (i.e. setting norms, creating safety, providing prompts that promote collaboration, and providing guidance/feedback) • Give students roles that make them responsible for making sense of texts (e.g. asking students to lead discussions or make arguments based on their interpretations of texts) • Facilitate students' active engagement in learning through the use of inquiry-based instructional methods (i.e., where students learn by questioning and problem-solving) • Ask students to pose questions and problems about course readings • Employ routines or assignments that are open-ended (e.g. group discussion; free choice in reading materials) so that all students feel comfortable participating and can have some measure of success	0.88	0.91	0.91

Appendix D. Analytic Sample Baseline Equivalence

In order to examine the baseline equivalence of our analytic samples, we regressed each student and teacher covariate included in the year 2 impact analyses on (a) the treatment status (indicator variable that takes the value of 1 for schools that were randomly assigned to receive Reading Apprenticeship and 0 for schools that were not) and (b) on blocking variables. We applied the same two-level random-intercept methodology[16] as was used for impact analysis. Additionally, we evaluated the equivalence of school-level prior achievement (percent of students in the school who were proficient on the 11th grade state ELA/Reading test in the baseline year) using a two-sample t-test with unequal variances.[17] All equivalence checks were performed at the level of the underlying impact analysis, i.e., at the student level for all student outcomes and at the teacher level for teacher outcomes; they were performed separately for each analytic sample.

Table D1 presents equivalence results for teacher survey analyses, based on the full sample of teachers. The analytic sample was identical for all constructs. For the teacher-level variables, adjusted difference column shows the regression coefficients for the treatment variables; for the school-level variable, difference is computed as a simple difference between unadjusted means. The *p* value column indicates the level of significance of the difference in either case. Table D1 shows that equivalence between the treatment and control groups was achieved for all covariates.[18]

TABLE D1. BASELINE EQUIVALENCE FOR TEACHER SURVEY ANALYSES, FULL ANALYTIC SAMPLE

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
School-Level Variable								
Prior Achievement in 11th Grade ELA/Reading (Percent Proficient Schoolwide) in Base Year	104	0.60	0.192	102	0.59	0.179	0.02	.515
Teacher-Level Variables								
Nonwhite	104	0.26	0.441	102	0.23	0.420	0.03	.719
Female	104	0.59	0.495	102	0.59	0.495	-0.01	.942
Reading Specialist	104	0.04	0.193	102	0.01	0.099	0.03	.182
Completed education beyond Bachelor's Degree	104	0.65	0.478	102	0.58	0.496	0.07	.380

[16] We used the Stata command *mixed*, estimated using residual maximum likelihood.
[17] School and treatment group intercepts cannot be simultaneously estimated when the outcome variable is measured at the school level.
[18] Teacher-level results shown are for imputed variables. Equivalence analyses on variables without imputation produced similar results.

TABLE D1. BASELINE EQUIVALENCE FOR TEACHER SURVEY ANALYSES, FULL ANALYTIC SAMPLE

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
Number of years teaching specific target subject prior to start of 2nd year	104	8.45	5.777	102	7.55	6.285	0.81	.428
Number of years teaching prior to start of 2nd year	104	10.49	6.274	102	9.39	6.476	0.81	.488
Note. Inference based on clustering								

Table D2 presents equivalence results for student survey analyses, based on the full sample of students who responded to questions comprising Construct 3.1 (Metacognitive Inquiry). While the analytic samples were not identical for all constructs, response rates were similar and the choice of sample did not affect the conclusions of the analysis. Table D2 shows that equivalence between treatment and control groups was achieved for all student-level covariates.[19] However, the difference in school-level prior achievement, while small, is statistically significant.[20] The difference is 1 percentage point, which translates into a standardized difference (Hedges' g) of 0.076 standard deviations—well below the What Works Clearinghouse criteria of 0.25 standard deviations. We included the school-level prior achievement variable in the impact analyses as a covariate control for observed differences in average baseline student performance.

TABLE D2. BASELINE EQUIVALENCE FOR STUDENT SURVEY ANALYSES, CONSTRUCT 3.1 METACOGNITIVE INQUIRY

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
School-Level Variable								
Prior Achievement in 11th Grade ELA/Reading (Percent Proficient Schoolwide) in Base Year	6,045	0.58	0.185	5,418	0.56	0.177	0.01**	<.001

[19] Student-level results shown are for imputed variables. Equivalence analyses on variables without imputation produced similar results.

[20] The significance of the difference is determined in part by the size of the sample and the number of observations in each school. Thus, while the difference is not significant in the teacher sample, it is statistically significant in the larger student sample.

TABLE D2. BASELINE EQUIVALENCE FOR STUDENT SURVEY ANALYSES, CONSTRUCT 3.1 METACOGNITIVE INQUIRY

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
Student-Level Variables								
ELL/LEP (current as of 12 mos.)	6,045	0.11	0.311	5,418	0.12	0.321	-0.01	.741
Special Education at 12 mos.	6,045	0.07	0.260	5,418	0.13	0.337	-0.07	.133
Receives Free or Reduced Price Lunch	6,045	0.47	0.499	5,418	0.49	0.500	-0.02	.817
Race								
Asian	6,045	0.06	0.239	5,418	0.09	0.291	-0.01	.691
Hispanic	6,045	0.40	0.490	5,418	0.37	0.483	0.03	.811
Black	6,045	0.07	0.247	5,418	0.05	0.214	0.02	.656
Other nonwhite race [a]	6,045	0.01	0.103	5,418	0.03	0.159	-0.04	.359
Female	6,045	0.44	0.497	5,418	0.45	0.497	-0.01	.878
Prior achievement (8th grade state test) Z-score	6,045	-0.11	0.787	5,418	-0.04	0.855	-0.09	.393

Note. * = significant at 5 percent, ** = significant at 1 percent. Inference based on clustering.

[a] This category captures any other specified race other than White, Asian, Hispanic/Latino, or Black/African American.

Tables D3 through D5 presents equivalence results for student literacy analyses, based on the full sample of students as well as the English language learner and science classroom subsamples. All analyses followed the methodology described above. The tables show that student-level covariates were equivalent in all samples,[21] while school-level achievement was statistically significantly different in the full sample (treatment - control = 1%) and in the science subsample (treatment - control = -1%). As was discussed above, we included the prior achievement as a covariate in the impact model to control for the observed baseline differences.

[21] Equivalence analyses on variables without imputation produced similar results.

TABLE D3. BASELINE EQUIVALENCE FOR STUDENT LITERACY ANALYSES, FULL ANALYTIC SAMPLE

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
School-Level Variable								
Prior Achievement in 11th Grade ELA/Reading (Percent Proficient Schoolwide) in Base Year	5,531	0.59	0.176	4,642	0.57	0.174	0.01**	<.001
Student-Level Variables								
ELL/LEP (current as of 12 mos.)	5,531	0.11	0.314	4,642	0.12	0.321	-0.01	.812
Special Education at 12 mos.	5,531	0.07	0.251	4,642	0.13	0.341	-0.07	.145
Receives Free or Reduced Price Lunch	5,531	0.46	0.499	4,642	0.48	0.500	-0.03	.785
Race								
Asian	5,531	0.06	0.246	4,642	0.10	0.294	-0.01	.794
Hispanic/Latino	5,531	0.40	0.491	4,642	0.35	0.477	0.03	.767
Black/African American	5,531	0.05	0.218	4,642	0.05	0.213	0.02	.696
Other Nonwhite Race [a]	5,531	0.01	0.101	4,642	0.03	0.170	-0.04	.354
Female	5,531	0.45	0.497	4,642	0.45	0.497	-0.01	.901
Prior Achievement (8th Grade State Test) Z-Score	5,531	-0.09	0.789	4,642	-0.03	0.858	-0.09	.439

Note. * = significant at 5 percent, ** = significant at 1 percent. Inference based on clustering.

[a] This category captures any other specified race other than White, Asian, Hispanic/Latino, or Black/African American.

TABLE D4. BASELINE EQUIVALENCE FOR STUDENT LITERACY ANALYSES, ENGLISH LANGUAGE LEARNER SUBSAMPLE

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
School-Level Variable								
Prior Achievement in 11th Grade ELA/Reading (Percent Proficient Schoolwide) in Base Year	615	0.46	0.124	541	0.45	0.134	0.01	.070
Student-Level Variables								
Special Education at 12 mos.	615	0.17	0.375	541	0.18	0.382	-0.02	.762
Receives Free or Reduced Price Lunch	615	0.79	0.411	541	0.77	0.423	0.04	.732
Race								
Asian	615	0.09	0.281	541	0.14	0.346	0.02	.752
Hispanic/Latino	615	0.74	0.437	541	0.68	0.468	0.08	.518
Black/African American	615	0.00	0.070	541	0.01	0.086	-0.002	.887
Other Nonwhite Race [a]	615	0.00	0.070	541	0.03	0.159	-0.04	.396
Female	615	0.40	0.491	541	0.39	0.487	0.07	.228
Prior Achievement (8th Grade State Test) Z-Score	615	-0.84	0.686	541	-0.82	0.624	-0.03	.792

Note. * = significant at 5 percent, ** = significant at 1 percent. Inference based on clustering.

[a] This category captures any other specified race other than White, Asian, Hispanic/Latino, or Black/African American.

TABLE D5. BASELINE EQUIVALENCE FOR STUDENT LITERACY ANALYSES, SCIENCE CLASS SUBSAMPLE

Characteristic	Treatment *n*	Unadjusted treatment group mean	Unadjusted treatment SD	Control *n*	Unadjusted control group mean	Unadjusted control SD	Adjusted difference	*p* value
School-Level Variable								
Prior Achievement in 11th Grade ELA/Reading (Percent Proficient Schoolwide) in Base Year	2,284	0.57	0.173	2,076	0.58	0.158	-0.01*	.006
Student-Level Variables								
ELL/LEP (current as of 12 mos.)	2,284	0.13	0.342	2,076	0.13	0.339	0.01	.872
Special Education at 12 mos.	2,284	0.06	0.229	2,076	0.12	0.327	-0.09	.087
Receives Free or Reduced Price Lunch	2,284	0.44	0.496	2,076	0.49	0.500	0.02	.852
Race								
Asian	2,284	0.07	0.259	2,076	0.09	0.291	-0.01	.747
Hispanic/Latino	2,284	0.36	0.480	2,076	0.34	0.475	0.11	.339
Black/African American	2,284	0.05	0.212	2,076	0.05	0.222	0.001	.977
Other Nonwhite Race [a]	2,284	0.01	0.119	2,076	0.03	0.166	-0.04	.315
Female	2,284	0.43	0.495	2,076	0.45	0.497	0.00	.998
Prior Achievement (8th Grade State Test) Z-Score	2,284	-0.13	0.811	2,076	-0.03	0.829	-0.10	.414

Note. * = significant at 5 percent, ** = significant at 1 percent. Inference based on clustering.

[a] This category captures any other specified race other than White, Asian, Hispanic/Latino, or Black/African American.

Appendix E. Student Literacy Assessment[22]

CONSTRUCTION OF STUDENT LITERACY OUTCOME MEASURE

In order to measure the effects of the intervention on academic literacy, we collaborated with Educational Testing Service (ETS) to develop an instrument designed to measure the strategic reading processes that are primary targets of Reading Apprenticeship, without over-aligning the test to the model. Based on the Global Integrated Scenario-Based Assessment, ETS designed the assessment to measure how well students read and reason about text sources in a discipline where they have been exposed to content and strategies for understanding text (O'Reilly et al., 2014). This assessment was administered on-line at the end of each study implementation year to students in study teachers' classrooms. The assessment has three forms, corresponding to the subject areas of focus: ELA, Biology, and US History. Students enrolled in more than one Reading Apprenticeship study teacher's class may have taken more than one form. Each form included closed- and open-ended questions.

The ETS-designed assessment yielded different sets of scale scores—a unidimensional, simple structure, and bifactor scale scores. The unidimensional scale score uses items across all three forms that represent a measure of general (not discipline specific) literacy skills. This score assumes the three forms are measuring a unidimensional factor of literacy. The simple structure scores are based on items from a single form (one scale score for each disciplinary form a student takes) and represents a measure of discipline specific literacy (e.g., the science form measures reading in science). Two bi-factor scale scores are generated for each student per form taken, a general literacy scale score and a discipline-specific (ELA, history, or science) scale score.[23] Figure E1 provides a visual representation of the analytical models used to estimate these scores. The scale scores were then computed using the estimated item parameters.

[22] The information provided in this appendix includes information from internal memos provided by Educational Testing Services.

[23] ETS used the two-parameter logistic (2PL) model and generalized partial credit model (GPCM) on the multiple choice and constructed response items, respectively, for the unidimensional scales. The multidimensional 2PL and GPCM were also used for the bifactor model. The item parameters were estimated concurrently using a multiple-group model; the item parameters were constrained to be equal across groups. Using the estimated item parameters, expected *a posteriori* (EAP) abilities were estimated for each student under each model. For the bifactor model and dimension-specific unidimensional models, students did not receive scores for tests that they did not take.

FIGURE E1. VISUAL REPRESENTATION OF ANALYTIC MODELS OF LITERACY SCORES

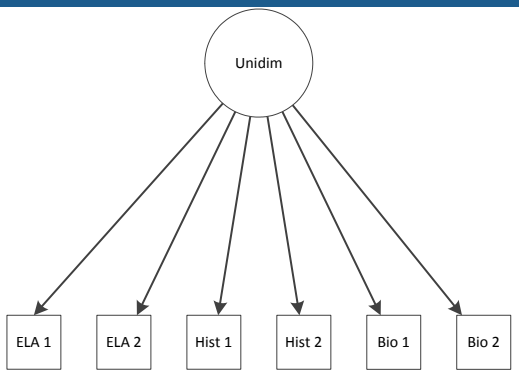
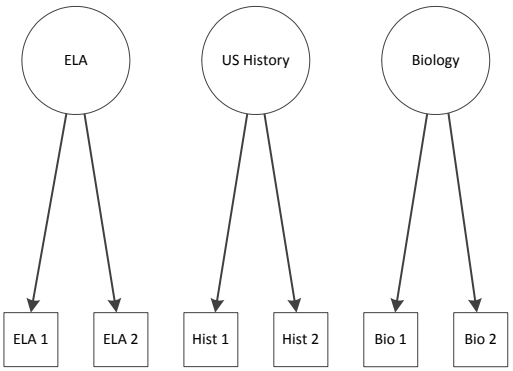
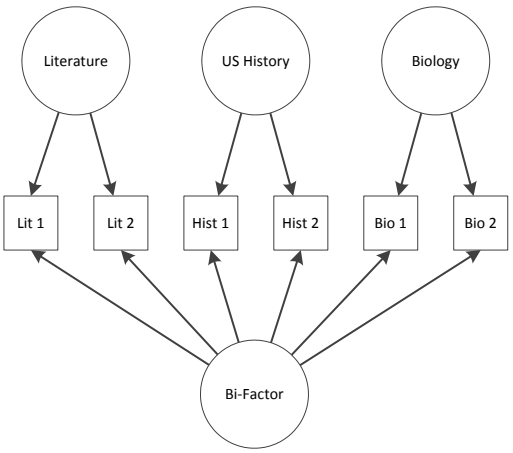
Score description	Score visual representation
Unidimensional score • Based on all items on the test • A common construct/theme underlying three subjects (ELA/Biology/US History) • One interpretation of this score is that it captures academic literacy skills supporting performance across all three subjects.	
Subject Scores (ELA, Biology, History) • Each subject score (ELA, US History, or Biology) is measured by corresponding subject specific items • One interpretation of these scores is that they capture the more discipline specific literacy skills.	
Bifactor general factor (or bifactor general) score • A "residual" score after removing the effects of the subject score (ELA, Biology and History) • The bifactor general score captures a common construct underlying general skill reflected on all three test forms.	

Table E1 provides the correlation among the five outcome scores, and Table E2 provides the marginal reliability of each score.[24] The unidimensional and bifactor general scores are highly correlated with each

[24] Conceptually, IRT marginal reliabilities are similar to Cronbach's alpha in classical theory.

other, as well as with other variables. Correlations among ELA, biology, and history scores are moderate. The reliability of bifactor general scores is lower than other scores, but still acceptably high.

We used the unidimensional scores as our primary outcome measure for impact analyses. The unidimensional score, which uses items from the three forms in a single score, assumes a common construct for general literacy skills that support performance in all three subjects, is highly reliable, and is highly correlated with bifactor general scores.

In using the unidimensional score, we make an implicit assumption that the three subject tests are equivalent, or, even if not, the sample is balanced in terms of subject area tests, so that the subject-test specific variation will not bias the impact estimates. To test the robustness of the findings based on the unidimensional scores, we also conducted a sensitivity analysis using the bifactor model scale as the dependent variable. We found that the results based on bifactor general scores were consistent with those based on unidimensional scores.

TABLE E1. CORRELATION AMONG ETS OUTCOME SCORES

	Unidimensional	Bifactor-general	Science	History	Literature
Unidimensional	1.00	0.94	0.98	0.98	0.98
Bifactor-General		1.00	0.97	0.92	0.85
Science			1.00	0.75	0.73
History				1.00	0.68
Literature					1.00

Source. IMPAQ staff calculations

TABLE E2. MARGINAL RELIABILITIES

Unidimensional	Bifactor-general	Science	History	Literature
0.877	0.786	0.861	0.878	0.872

Source. Educational Testing Service

DESCRIPTIVE STATISTICS OF THE LITERACY OUTCOME MEASURE

Table E3 provides basic unadjusted univariate statistics of the student outcome measure by assignment group, state, and students' demographic characteristics. As expected, the scores were lower among ELL students, nonwhite students, students who were eligible for free or reduced price lunch, and students with lower prior achievement scores (based on state ELA/Reading tests). The unidimensional scores were lower, on average, in California than in Pennsylvania, likely reflecting higher ELL concentrations in California. The tests were not grade-specific, and students in higher grades did not score higher on the assessment.

TABLE E3. A SUMMARY OF THE UNIDIMENSIONAL SCORES FOR THE YEAR 2 SAMPLE

Variable	Obs	Mean	Std. dev.	Min	Max
Total analytic student sample at Second Year of Implementation	10,173	-0.06	0.937	-2.45	2.73
Assignment					
Treatment	5,531	-0.002	0.911	-2.45	2.73
Control	4,642	-0.14	0.962	-2.42	2.69
Grade					
9th grade	4,764	-0.01	0.946	-2.45	2.73
10th grade	2,787	-0.20	0.908	-2.42	2.67
11th grade	2,310	-0.01	0.931	-2.45	2.59
12th grade	312	-0.09	0.982	-2.02	2.57
State					
California	6,440	-0.13	0.898	-2.45	2.59
Pennsylvania	3,733	0.05	0.990	-2.37	2.73
ELL status (current status)					
ELL/LEP	1,156	-0.78	0.618	-2.45	1.20
Race					
White	3,891	0.14	0.992	-2.37	2.73
Economic status					
FRPL	4,776	-0.29	0.859	-2.45	2.59
8th grade state reading/ELA test					
Higher prior achievement (z-score > -0.5)	7,428	0.14	0.936	-2.45	2.73
Lower prior achievement (z-score <= -0.5)	2,745	-0.62	0.672	-2.45	1.73
Total analytic student sample at First Year of Implementation	9,376	-0.04	0.903	-2.56	2.88
Total analytic student sample at Third Year of Implementation	6,856	0.08	0.928	-2.49	2.77
Source. IMPAQ staff calculations					

The ETS-developed literacy assessment is moderately correlated with state tests, based on a sample of students for whom we had both the ETS and state test results from the second year of the implementation. The state tests refer to the ELA scores from the California State Test (CST) and literature scores from Pennsylvania's Keystone Test. While the ETS assessment is specifically developed to capture the effects of reading strategies emphasized by Reading Apprenticeship, the state tests measure broader skills and knowledge in language and literature. Therefore, the tests are not expected to be directly comparable. However, they are correlated to the extent that these tests are all expected to reflect basic literacy levels of students. Because we were not able to collect individual-level state test scores from many study schools,

the sample with both tests was relatively small. Table E4 summarizes correlation coefficients between the state tests (z-scores) and the unidimensional scores for the second year of implementation.

TABLE E4. CORRELATION BETWEEN ETS UNIDIMENSIONAL OUTCOME SCORES AND STATE TESTS

	Correlation with state tests (Combined CA and PA)	Correlation with California state tests (CST ELA)	Correlation with PA Keystone tests (Keystone literature)
Unidimensional Score	0.6939	0.7046	0.6603
Number of observations	3,672	2,612	1,060

Source. IMPAQ staff calculations

Appendix F. Sample Attrition

FIGURE F1. TEACHER SAMPLES AT THE END OF THE SECOND YEAR OF IMPLEMENTATION

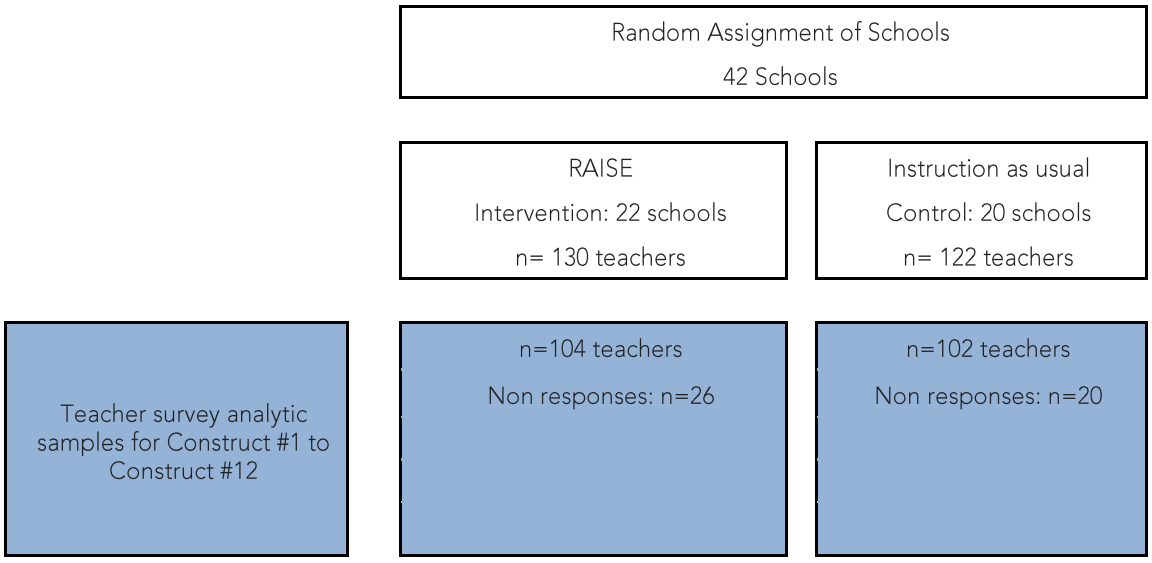


FIGURE F2. STUDENT SAMPLES AT THE END OF THE SECOND YEAR OF IMPLEMENTATION, FULL SAMPLE

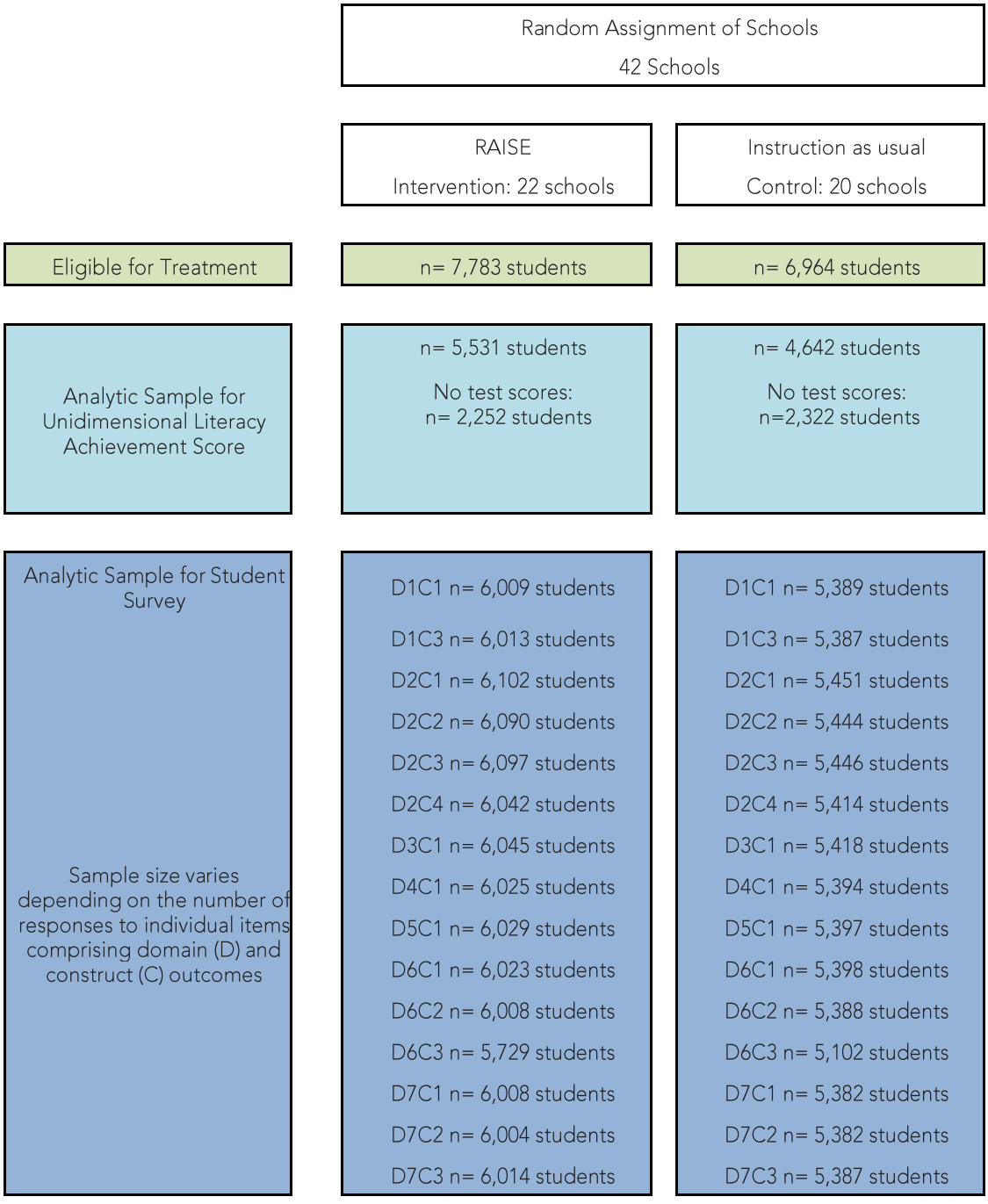
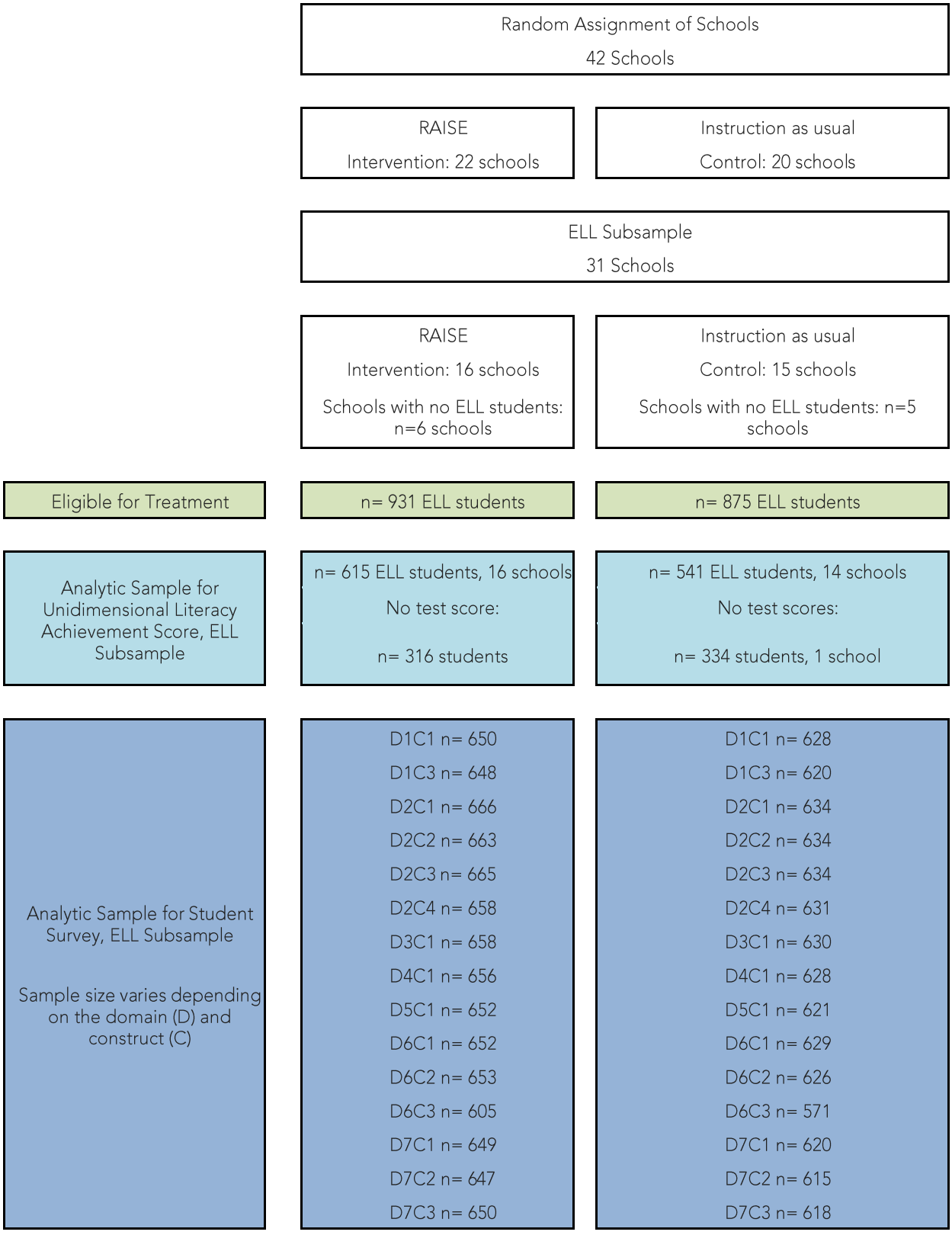


FIGURE F3. STUDENT SAMPLES AT THE END OF THE SECOND YEAR OF IMPLEMENTATION, ELL SUBSAMPLE



Appendix G. Additional Impact Analyses for Teacher Mediating Outcomes

See Appendix A for description of model and covariates.

IMPACT ANALYSIS RESULTS FOR YEAR 1 AND YEAR 3

TABLE G1. FULL SAMPLE IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES, YEARS 1 AND 3

Outcome measure (Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
1. Variety of Text Types							
1st year average	2.70	2.81	-0.11	0.180	.545	-0.09	217
3rd year average	2.52	2.57	-0.05	0.225	.812	-0.04	133
2. Fostering Student Independence							
1st year average	4.74	3.84	0.90**	0.281	.001	0.37	217
3rd year average	3.60	3.24	0.36	0.351	.309	0.18	133
3. Traditional Instructional Strategies							
1st year average	4.71	4.62	0.09	0.282	.744	0.04	217
3rd year average	4.08	4.07	0.01	0.412	.972	0.01	133
4. Teachers Instructing Metacognitive Inquiry							
1st year average	0.90	1.06	-0.16	0.117	.169	-0.21	217
3rd year average	0.62	0.71	-0.09	0.141	.525	-0.12	133
5. Teachers Modeling Metacognitive Inquiry							
1st year average	1.10	1.03	0.07	0.109	.529	0.09	217
3rd year average	0.85	0.82	0.03	0.131	.821	0.04	133
6. Students Practicing Metacognitive Inquiry							
1st year average	1.98	1.58	0.40**	0.125	.001	0.45	217
3rd year average	1.75	1.51	0.23	0.177	.192	0.24	133
7. Teachers Instructing Comprehension Strategies							
1st year average	1.80	1.80	0.00	0.190	.999	0.00	217
3rd year average	1.33	1.31	0.02	0.233	.921	0.02	133
8. Teachers Modeling Comprehension Strategies							
1st year average	1.97	1.78	0.19	0.179	.290	0.14	217
3rd year average	1.52	1.41	0.11	0.218	.620	0.08	133
9. Students Practicing Comprehension Strategies							
1st year average	3.18	2.36	0.82**	0.209	<.001	0.55	217
3rd year average	3.05	2.45	0.60*	0.278	.031	0.37	133
10. Student Collaboration							
1st year average	5.06	3.76	1.31**	0.389	<.001	0.43	215

TABLE G1. FULL SAMPLE IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES, YEARS 1 AND 3

Outcome measure (Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
3rd year average	4.06	3.20	0.86*	0.388	.027	0.37	131
11. Student Engagement							
1st year average	11.75	11.70	0.05	0.261	.857	0.02	217
3rd year average	11.67	11.98	-0.31	0.340	.360	-0.16	133
12. Teachers' Self-confidence in Literacy Instruction							
1st year average	36.28	35.72	0.57	1.193	.636	0.08	213
3rd year average	40.99	36.50	4.49**	1.611	.005	0.61	129

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on RAISE teacher survey data

YEAR 2 IMPACT ANALYSIS RESULTS BY SUBJECT

TABLE G2. IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES BY SUBJECT, YEAR 2

Outcome measure (Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
1. Variety of Text Types							
ELA	1.41	1.44	-0.03	0.275	.924	-0.03	67
Science	3.48	3.05	0.43	0.311	.165	0.45	70
History	3.36	3.45	-0.09	0.295	.751	-0.08	69
2. Fostering Student Independence							
ELA	4.97	4.07	0.90	0.493	.067	0.48	67
Science	4.01	2.08	1.93**	0.489	<.001	0.97	70
History	4.13	3.82	0.30	0.529	.566	0.16	69
3. Traditional Instructional Strategies							
ELA	4.14	4.48	-0.34	0.451	.445	-0.21	67
Science	4.55	3.13	1.42**	0.503	.005	0.67	70
History	4.42	4.93	-0.51	0.621	.408	-0.28	69
4. Teachers Instructing Metacognitive Inquiry							
ELA	0.96	1.01	-0.05	0.238	.841	-0.05	67
Science	0.52	0.39	0.13	0.145	.376	0.23	70
History	0.87	1.04	-0.16	0.210	.436	-0.20	69
5. Teachers Modeling Metacognitive Inquiry							
ELA	1.31	1.13	0.18	0.191	.357	0.24	67
Science	0.74	0.49	0.26	0.144	.077	0.45	70

TABLE G2. IMPACT ESTIMATES FOR TEACHER SURVEY OUTCOMES BY SUBJECT, YEAR 2

Outcome measure (Teacher survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
History	0.84	1.02	-0.19	0.196	.343	-0.27	69
6. Students Practicing Metacognitive Inquiry							
ELA	2.30	1.80	0.50*	0.215	.020	0.61	67
Science	1.62	0.88	0.74**	0.224	.001	0.82	70
History	1.82	1.75	0.07	0.223	.739	0.08	69
7. Teachers Instructing Comprehension Strategies							
ELA	1.63	1.74	-0.11	0.406	.788	-0.07	67
Science	1.20	0.97	0.23	0.283	.410	0.21	70
History	1.79	1.87	-0.08	0.400	.842	-0.05	69
8. Teachers Modeling Comprehension Strategies							
ELA	2.32	1.78	0.54	0.454	.233	0.36	67
Science	1.61	0.97	0.64*	0.303	.035	0.56	70
History	1.57	1.79	-0.23	0.369	.541	-0.16	69
9. Students Practicing Comprehension Strategies							
ELA	3.94	2.83	1.11*	0.437	.011	0.73	67
Science	3.08	1.55	1.52**	0.391	<.001	0.98	70
History	3.31	2.82	0.50	0.470	.289	0.31	69
10. Student Collaboration							
ELA	4.49	3.03	1.46*	0.593	.014	0.65	67
Science	4.57	3.25	1.32	0.813	.103	0.51	70
History	4.03	3.56	0.47	0.532	.376	0.23	69
11. Student Engagement							
ELA	12.24	11.89	0.35	0.486	.468	0.20	67
Science	12.30	11.73	0.57	0.508	.263	0.30	70
History	12.01	12.09	-0.08	0.489	.867	-0.05	69
12. Teachers' Self-confidence in Literacy Instruction							
ELA	41.78	39.16	2.62	1.893	.166	0.40	67
Science	39.44	33.14	6.30**	1.884	<.001	0.94	70
History	37.68	36.39	1.29	1.817	.478	0.24	69

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on RAISE teacher survey data

Appendix H. Additional Impact Analyses for Student Mediating Outcomes

See Appendix A for description of model and covariates.

IMPACT ANALYSIS RESULTS FOR YEAR 1

TABLE H1. FULL SAMPLE IMPACT ESTIMATES FOR STUDENT SURVEY OUTCOMES, YEAR 1

Outcome measure (Student survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
1.1 - Participation/contribution to class discussions	2.91	2.90	0.01	0.050	.906	0.01	11,415
1.3 - Conferring	3.32	3.32	0.00	0.068	.991	0.00	11,406
2.1 - Use of global reading strategies	2.93	2.92	0.01	0.039	.734	0.02	11,616
2.2 - Use of problem-solving strategies	3.37	3.35	0.02	0.032	.433	0.03	11,605
2.3 - Support reading strategies	2.56	2.51	0.05	0.052	.356	0.05	11,617
2.4 - Integration of content and literacy activity	2.78	2.69	0.09	0.052	.069	0.12	11,458
3.1 – Metacognitive inquiry	2.90	2.85	0.06	0.049	.239	0.09	11,481
4.1 - Reader identity	2.35	2.33	0.02	0.040	.675	0.02	11,446
5.1 - Student identity	3.27	3.31	-0.04	0.045	.340	-0.06	11,480
6.1 - Class time spent reading	2.71	2.65	0.06	0.063	.357	0.06	11,428
6.2 - Variety of reading material	2.66	2.69	-0.03	0.051	.611	-0.03	11,410
6.3 - Pages of reading per day	1.98	2.05	-0.07	0.103	.481	-0.06	10,880
7.1 - Increased effort to learn	3.46	3.48	-0.02	0.040	.635	-0.02	11,384
7.2 - Happiness/belonging	3.38	3.43	-0.05	0.071	.484	-0.06	11,382
7.3 - Engaging instruction	3.46	3.57	-0.11	0.104	.295	-0.11	11,400

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on RAISE student survey data

YEAR 2 IMPACT ANALYSIS RESULTS BY SUBJECT

TABLE H2. IMPACT ESTIMATES FOR STUDENT SURVEY OUTCOMES BY SUBJECT, YEAR 2

Outcome measure (Student survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
1.1 - Participation/contribution to class discussions							
ELA	2.99	3.00	-0.02	0.118	.881	-0.02	3,246
Science	2.85	2.67	0.18*	0.079	.026	0.23	4,889
History	3.00	2.93	0.07	0.083	.418	0.09	3,888
1.3 - Conferring							
ELA	3.38	3.36	0.03	0.090	.782	0.03	3,243
Science	3.27	3.23	0.04	0.085	.618	0.05	4,892
History	3.53	3.35	0.17	0.112	.124	0.23	3,889
2.1 - Use of global reading strategies							
ELA	2.79	2.87	-0.08	0.066	.226	-0.10	3,288
Science	2.96	2.91	0.05	0.046	.290	0.06	4,968
History	3.00	3.01	0.00	0.047	.956	0.00	3,926
2.2 - Use of problem-solving strategies							
ELA	3.29	3.38	-0.09	0.054	.089	-0.11	3,283
Science	3.36	3.33	0.03	0.040	.492	0.04	4,956
History	3.39	3.42	-0.03	0.054	.620	-0.04	3,922
2.3 - Support reading strategies							
ELA	2.39	2.50	-0.11	0.080	.182	-0.13	3,287
Science	2.62	2.54	0.07	0.047	.114	0.09	4,960
History	2.66	2.63	0.03	0.055	.609	0.03	3,925
2.4 - Integration of content and literacy activity							
ELA	2.89	2.83	0.06	0.096	.549	0.08	3,262
Science	2.65	2.47	0.18*	0.074	.017	0.22	4,919
History	2.89	2.73	0.16*	0.076	.040	0.21	3,899
3.1 – Metacognitive inquiry							
ELA	2.97	2.88	0.09	0.078	.271	0.14	3,262
Science	2.88	2.74	0.14*	0.065	.033	0.22	4,926
History	3.02	2.85	0.18**	0.062	.005	0.30	3,900
4.1 - Reader identity							
ELA	2.41	2.43	-0.02	0.072	.811	-0.02	3,248
Science	2.26	2.25	0.01	0.056	.883	0.01	4,901
History	2.43	2.39	0.05	0.070	.494	0.06	3,894

TABLE H2. IMPACT ESTIMATES FOR STUDENT SURVEY OUTCOMES BY SUBJECT, YEAR 2

Outcome measure (Student survey construct)	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
5.1 - Student identity							
ELA	3.28	3.33	-0.05	0.060	.405	-0.07	3,253
Science	3.25	3.23	0.03	0.060	.629	0.04	4,893
History	3.37	3.34	0.03	0.073	.673	0.05	3,897
6.1 - Class time spent reading							
ELA	3.01	2.90	0.11	0.118	.336	0.13	3,252
Science	2.58	2.37	0.21**	0.080	.008	0.25	4,901
History	2.75	2.67	0.08	0.105	.452	0.09	3,893
6.2 - Variety of reading material							
ELA	2.40	2.42	-0.01	0.067	.845	-0.01	3,242
Science	2.78	2.81	-0.03	0.071	.695	-0.03	4,891
History	2.78	2.59	0.18*	0.094	.049	0.20	3,888
6.3 - Pages of reading per day							
ELA	2.84	2.50	0.34	0.222	.127	0.24	3,088
Science	1.85	1.83	0.02	0.094	.859	0.01	4,619
History	1.85	1.83	0.02	0.119	.876	0.02	3,726
7.1 - Increased effort to learn							
ELA	3.45	3.50	-0.04	0.055	.440	-0.05	3,245
Science	3.49	3.45	0.03	0.060	.566	0.04	4,881
History	3.54	3.53	0.01	0.073	.843	0.02	3,888
7.2 - Happiness/belonging							
ELA	3.36	3.37	-0.01	0.086	.920	-0.01	3,240
Science	3.30	3.30	-0.01	0.096	.952	-0.01	4,881
History	3.59	3.50	0.09	0.096	.340	0.11	3,890
7.3 - Engaging instruction							
ELA	3.46	3.44	0.02	0.116	.865	0.02	3,245
Science	3.35	3.40	-0.05	0.130	.674	-0.06	4,889
History	3.75	3.60	0.15	0.164	.352	0.16	3,893

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on RAISE student survey data

Appendix I. Additional Impact Analyses for Student Literacy

See Appendix A for description of model and covariates.

IMPACT ANALYSIS RESULTS FOR YEAR 1 AND YEAR 3

TABLE I1. IMPACT ESTIMATES FOR UNIDIMENSIONAL LITERACY ASSESSMENT SCORES BY STUDENT SUBGROUP, YEARS 1 AND 3

Sample	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
Full Sample							
Year 1	-0.01	-0.11	0.11	0.081	.189	0.12	9,376
Year 3	0.05	0.03	0.02	0.107	.824	0.03	6,856
ELL Subsample							
Year 1	-0.78	-0.82	0.04	0.110	.713	0.07	1,014
Year 3	-0.92	-0.74	-0.18	0.197	.355	-0.29	218
Subject - Science							
Year 1	-0.10	-0.18	0.08	0.090	.376	0.09	4,133
Year 3	0.12	-0.04	0.16	0.185	.397	0.17	2,288
Subject - ELA							
Year 1	0.03	-0.17	0.19	0.128	.132	0.22	2,387
Year 3	0.03	0.03	0.00	0.159	.990	0.00	2,586
Subject - History							
Year 1	0.07	-0.03	0.10	0.146	.502	0.11	3,305
Year 3	-0.03	0.03	-0.06	0.160	.699	-0.07	2,629

Source. IMPAQ staff calculations on ETS assessment data

ALTERNATIVE STUDENT LITERACY SCORING (YEAR 2)

TABLE I2. FULL SAMPLE IMPACT ESTIMATES FOR LITERACY ASSESSMENT SCORES BY SCORE TYPE, YEAR 2

Outcome measure	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
Unidimensional Score	-0.02	-0.16	0.13	0.100	.184	0.14	10,173
Bi-Factor Score	-0.03	-0.15	0.11	0.085	.179	0.13	10,168
History Assessment	-0.06	-0.05	0.00	0.151	.981	0.00	3,891
Biology Assessment	0.07	-0.25	0.32**	0.109	.003	0.34	4,367
Literacy Assessment	-0.03	-0.23	0.20	0.139	.143	0.22	3,576

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on ETS assessment

ROBUSTNESS CHECKS (YEAR 2)

The benchmark student literacy impact model was estimated using mixed-effect estimation (using residual maximum likelihood, REML) and included blocking variables, student characteristics, and school characteristics. Below we present results of selected robustness checks (alternative covariates and estimation methods) for the year 2 unidimensional literacy impact estimates. Additional robustness checks were performed on year 2 scores, as well as on year 1 and year 3 scores (not shown). Results are generally consistent across alternative specifications.

Model 1 includes only blocking variables and student pretests as covariates. Model 2 includes all covariates included in the benchmark model, as well as all other available school-level characteristics: school-level 11th-grade proficiency in the year prior to the baseline, average 8th-grade ELA/Reading scores, and average student demographics. Model 3 is equivalent to the benchmark model but without imputation of missing data (and with listwise deletion of observations).

Models 4 and 5 use two common alternative estimation methods for the benchmark model: Model 4 is a mixed-effects model estimated using maximum likelihood (MLE); Model 5 is estimated using ordinary least squares (OLS) with robust standard errors clustered at the school level. Maximum likelihood method is the default optimization technique for mixed-level models in some statistical packages, However, REML, used as our benchmark estimation method, adjusts the degrees of freedom downward when estimating the variance (standard error) components, while MLE does not. As a result, REML produces more conservative estimates of standard errors of coefficients and is often preferred to MLE. OLS with cluster-robust standard errors is also commonly used in a variety of fields when handling clustered data (such as teachers or students nested in schools). Unlike mixed-effect models, cluster-robust OLS makes no explicit assumptions about the distribution of between-cluster variation in outcomes. In large samples, mixed-effect methods can produce gains in efficiency if the distribution of the errors is correctly specified.[25] Both methods, however, are susceptible to small sample bias, i.e., when there are few (generally, less than 50) clusters.[26] In practice, which method is more conservative may not be known *a priori*, although multiple studies have found mixed-effect methods to produce more conservative estimates of standard errors than cluster-robust OLS.[27]

[25] Cameron, A. C., & Miller, D. L. (2010). Robust inference with clustered data. *Handbook of empirical economics and finance*, 1-28.
[26] Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.
[27] Yasuyo, A., & Gee, K.A. (2014). Sensitivity analyses for clustered data: An illustration from a largescale clustered randomized controlled trial in education. *Evaluation and Program Planning*, 47(2014), 26-34.

TABLE I3. ALTERNATIVE IMPACT ESTIMATES FOR UNIDIMENSIONAL LITERACY ASSESSMENT SCORES BY STUDENT SUBGROUP, YEAR 2

Sample	Adjusted treatment group mean	Adjusted control group mean	Difference (Impact)	Standard error	*p* value	Effect size	*n*
1. Blocking covariates and student pretests only							
Full Sample	-0.03	-0.13	0.11	0.096	.262	0.11	10,173
ELL Subsample	-0.77	-0.79	0.02	0.100	.849	0.03	1,156
Science	0.02	-0.26	0.28*	0.123	.023	0.29	4,360
ELA	-0.11	-0.21	0.10	0.142	.496	0.11	2,936
History	-0.07	0.02	-0.09	0.143	.515	-0.10	3,449
2. Benchmark model plus additional school-level covariates							
Full Sample	-0.01	-0.15	0.13	0.129	.302	0.14	10,173
ELL Subsample	-0.76	-0.79	0.04	0.183	.846	0.06	1,156
Science	0.03	-0.27	0.29*	0.135	.031	0.31	4,360
ELA	-0.01	-0.25	0.24	0.172	.165	0.27	2,936
History	0.03	0.01	0.02	0.179	.897	0.02	3,449
3. Benchmark model without imputation of missing data							
Full Sample	0.05	-0.15	0.20	0.124	.109	0.21	6,345
ELL Subsample	-0.76	-0.81	0.05	0.132	.718	0.08	756
Science	0.07	-0.24	0.31*	0.141	.027	0.32	2,608
ELA	0.02	-0.13	0.15	0.161	.365	0.17	2,000
History	0.17	-0.02	0.19	0.162	.251	0.20	2,001
4. Mixed-effects model estimated using maximum likelihood							
Full Sample	-0.02	-0.16	0.13	0.082	.101	0.14	10,173
ELL Subsample	-0.73	-0.84	0.10	0.081	.201	0.17	1,156
Science	0.01	-0.29	0.31**	0.097	.001	0.32	4,360
ELA	-0.05	-0.24	0.19	0.109	.083	0.21	2,936
History	-0.05	0.02	-0.08	0.116	.517	-0.08	3,449
5. OLS with cluster-robust standard errors							
Full Sample	0.00	-0.15	0.15	0.076	.054	0.16	10,173
ELL Subsample	-0.70	-0.87	0.18*	0.078	.032	0.29	1,156
Science	0.02	-0.27	0.29**	0.094	.003	0.31	4,360
ELA	-0.03	-0.21	0.18	0.108	.099	0.20	2,936
History	0.02	0.02	-0.01	0.103	.953	-0.01	3,449

Note. * = significant at 5 percent, ** = significant at 1 percent

Source. IMPAQ staff calculations on ETS assessment data

Appendix J. Fidelity of Implementation Summary

As part of the National Evaluation of Investing in Innovation (NEi3) grant requirements, evaluators must report FOI for each key component of inputs in the logic model. FOI is measured to indicate whether the program was implemented as intended by the developer. The evaluator specifies thresholds for meeting fidelity and then collects data, assesses, and reports whether fidelity was met at the program-level for *each key component* of inputs in the logic model, at least once each year for two or more years.

The evaluators worked with WestEd's Strategic Literacy Initiative (SLI) to identify the key components and thresholds for FOI. We identified two teacher-level, one school-level, and one program-level component—a total of four key components—for inclusion in the FOI measure. Fidelity for the two teacher-level components (teacher participation in professional development and teacher participation in monthly meetings) is determined first at the teacher level, then aggregated to the school level, and finally rolled up to the program-level. The school-level component is the recruitment of teacher leaders, which is then aggregated to the program-level. The program-level component is the presence of five characteristics of RAISE professional development.

TABLE J1. FIDELITY OF IMPLEMENTATION RESULTS

			Fidelity results		
Component	Operational definition	Implementation fidelity thresholds	Year 1	Year 2	Year 3
A.1: Teacher participation in RA professional development	10 days of training across 3 separate sessions (5 days, then 2 days, then 3 days)	Teacher-level: All 5 days of initial training and at least 4 of the following 5 School-level: at least 75% of teachers meet fidelity Program-level: at least 80% of schools meet fidelity	0	N/A	N/A
B.1: Teacher participation in Monthly Team Meetings	Teacher attended monthly meeting	Teacher level: 4 or more meetings/year School-level: at least 50% of teachers meet fidelity Program-level: at least 80% of schools meet fidelity	1	1	0
C.1: Teacher leader (TL) recruitment	SLI will recruit a TL for each school	School-level: teacher leader recruited Program-level: at least 80% of schools have a teacher leader	1	1	1
D.1: PD content was delivered as planned	PD exhibits content focused on disciplinary literacy, collective participation, active learning, coherence, inquiry-based	Program-level: all 5 of the characteristics were present in at least 75% of the sessions	1	N/A	N/A

Note. PD occurs only in Year 1, so components A.1 and D.1 are measured only in Year 1

RAISE PROFESSIONAL DEVELOPMENT ATTENDANCE

We found that while 87/113 (77%) of teachers attended all five of the first five days and at least four of the last five days (Figure H1), only 12/22 (55%) of schools had at least 75% of teachers meeting this threshold (Figure H2); therefore, program-level fidelity was not met for this component. We examined the distributions of teachers in schools more closely and found the following:

- Within the 12 schools where fidelity was met, 58/62 (94%) teachers met fidelity, with seven schools with 100% of teachers meeting fidelity.
- Within the 10 schools where fidelity was not met, only 29/51 (57%) teachers met fidelity.
- For 7 of the 10 schools that did not meet fidelity, the school was within one teacher of meeting fidelity, but within those seven schools, 11/12 (92%) teachers missed all of at least one of the three training sessions (5-day, 2-day or 3-day institute).
- Among the entire sample, teachers who met fidelity averaged 9.96 days of training, while teachers who did not meet fidelity averaged 5.4 days of training.

While it appears that several schools were on the edge of meeting fidelity, only a significant change of the teacher or school level thresholds would result in a change in fidelity at the program-level (i.e., either changing the teacher level threshold from 9 days to 6 days or changing the school level threshold of teachers meeting fidelity in a school from 75% to 66% would be sufficient to meet fidelity at the program-level).

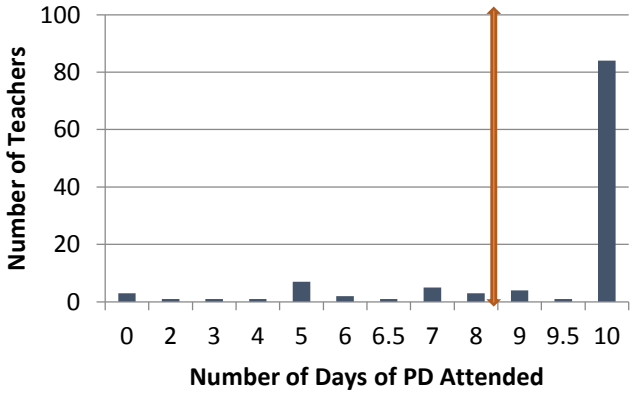


FIGURE J1. DAYS OF PD

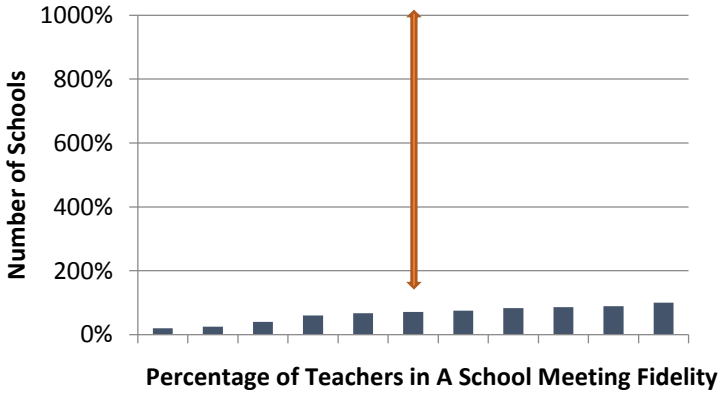


FIGURE J2. PERCENTAGE OF TEACHERS

RAISE MONTHLY TEAM MEETING ATTENDANCE

Data used to determine fidelity for component B (monthly meeting attendance) was collected through monthly online teacher surveys. In year 1, 86/113 (76%) teachers met fidelity by attending at least 4 meetings, with 18/22 (82%) schools meeting fidelity (at least 50% of teachers attending at least 4 meetings); program-level fidelity was met for this component. Eight teachers had insufficient survey data to determine fidelity for this component. None of these teachers met fidelity for component A (PD attendance) and five of these teachers did not respond to any of the surveys. The teachers who met fidelity attended an average of 6.6 team meetings during the year, while those teachers who did not meet fidelity attended an average of 1.7 team meetings during the year. Of the 26 teachers who did not meet fidelity or had insufficient data to make a determination in year 1, only four met fidelity in year 2, with eight having insufficient data to make a determination. The results of the group of teachers with insufficient data to make a determination would not have had an effect either way on fidelity at the program-level.

In year 2, only 70/105 (66.7%) teachers met fidelity, with 19/22 (86%) schools meeting fidelity; therefore, program-level fidelity was met for this component. During year 2, there were fewer teachers in the sample and fewer teachers met fidelity, but those teachers were clustered in schools that did not meet fidelity. Of the 31 teachers that did not meet fidelity but had enough survey data in order for us to make a determination, 17 were concentrated in three schools that essentially stopped holding meetings (these schools averaged less than one meeting per teacher). Similar to year 1, the teachers who met fidelity attended an average of 6.5 team meetings during year 2, while those teachers that did not meet fidelity attended an average of 1.2 team meetings during the year.

In year 3, nearly all schools reported little to no attendance at monthly meetings. Only 23/69 (33.3%) teachers met fidelity, and 3/15 (20%) schools met fidelity for this component; therefore program-level fidelity was not met for this component. The average number of meetings reported across the sample was 2.9 meetings per teacher. The few teachers who met fidelity attended an average of 7.3 team meetings during year 3, while those teachers that did not meet fidelity attended an average of 0.7 team meetings during the year.

IDENTIFICATION OF TEACHER LEADERS

In years 1 and 2, 21/22 (95%) of schools identified a teacher leader, and in year 3, 13/15 (87%)[28] schools identified a teacher leader; therefore, program-level fidelity was met for this component each year.

PROFESSIONAL DEVELOPMENT CONTENT

Researchers developed a protocol and observed a sample of the RAISE Institute to determine if the professional development was implemented as intended. Each of the five characteristics was present in over 85% of sessions, thereby meeting program-level fidelity for this component.

Over the entire sample of session observed, 69% had all five characteristics present, 26% of sessions had four of five characteristics present, and 6% of sessions had only three characteristics present.

FURTHER ANALYSIS: FINDINGS FROM VARIATION AMONG EXPERIMENTAL BLOCKS

To get additional insight into the role of fidelity as related to impact, we considered whether impacts on the ETS assessment varied across randomized blocks. Each of the 11 blocks can be thought of as a "mini-experiment," therefore if we observed heterogeneity in impact across the blocks, we could examine the relationship between impact and achieved fidelity. With only 11 blocks, the analysis of variation in impact was highly underpowered, still we considered it as an exploratory step. We observed no differences across randomized blocks in impact on the ETS assessment, and therefore did not proceed to explore differences in implementation among blocks.

DETAILED THRESHOLDS

Table J2 presents the sample sizes for each year of reporting, detailing the two cohorts of teachers staggered across consecutive years. Cohort 1 began in the summer of 2011 and was followed for all three years of the study, while teachers in Cohort 2 began in the summer of 2012 and were followed for the last two years of the study.

TABLE J2. YEARS OF IMPLEMENTATION FOR EACH COHORT WITH SAMPLE SIZES

Year of Implementation	Cohort 1	Cohort 2	Schools	Teachers
1	2011-2012	2012-2013	22	113
2	2012-2013	2013-2014	22	105
3	2013-2014	N/A	15	69

Monthly meetings and teacher leader recruitment are reported for all three years of implementation, while PD attendance and PD content are reported only for the first year of implementation (PD attendance across the full 10 days is aggregated together). At the program-level, each component will be reported to NEi3 as either 1 or 0: met fidelity or not. In summary, program fidelity was met for PD content but not PD attendance. Fidelity was met for teacher leader recruitment in all three years of the study and in the first two years for teacher attendance at monthly meetings, but not met in the third year.

[28] Year 3 only includes the first cohort of teachers/schools. Between Year 2 and Year 3 of the study, three RAISE schools combined into one school.

TEACHER PARTICIPATION IN PROFESSIONAL DEVELOPMENT

Participating teachers and teacher leaders are offered 65 hours of RAISE professional development (PD) through the 5-day foundational training, the 2-day turnaround training, and the 3-day summer springboard training. Attendance at the initial 5-day training was determined by the program developers to be a more important factor in adequate implementation than the follow-up 2-day or 3-day trainings. Therefore, the 5-day training carries relatively more weight in determining fidelity at the teacher level. Training attendance was gathered from attendance rosters and teacher surveys.

To meet fidelity, teachers must attend all five days of the initial summer training and at least four of the five days offered in the winter and the following summer. For a school to meet fidelity, 75% of their RAISE teachers must have adequate training attendance. To meet program-level fidelity, 80% of the schools must meet fidelity for this component.

TEACHER PARTICIPATION IN MONTHLY MEETINGS

Participating RAISE teachers are supposed to attend *at least* monthly onsite meetings of RAISE teachers facilitated by the teacher leader. These meetings provide support to teachers in their professional development, assist them with problem solving, and provide them with tools to facilitate implementation. Meetings may include sharing and reviewing student work, discussing problematic lessons and problem solving, sharing successful lessons, and exploring Reading Apprenticeship tools and protocols. Monthly teacher surveys collected self-reported data on participation in the monthly onsite team meetings.

The number of onsite team meetings teachers attended throughout the school year is calculated by summing responses to questions on the monthly teacher survey about participation (0=did not attend meeting, 1=attended meeting). Program developers determined that participation in at least four meetings per year meets teacher-level fidelity. For a school to meet fidelity, 50% of its teachers have to meet fidelity. For the program to meet FOI, 80% of the schools have to meet fidelity.

RECRUITMENT OF TEACHER LEADERS

SLI recruits and trains one teacher leader per school to support Reading Apprenticeship implementation and hold the monthly meetings. Teacher leaders are often volunteer teachers participating in the study, but also consisted of curriculum coordinators or school administrators with primary responsibility for supporting teachers, but who is not implementing Reading Apprenticeship in a classroom. Teacher leaders received the same 65 hours of RAISE professional development provided to teachers in the initiative and attended an additional teacher leader webinar in the first year, with three face-to-face meetings per year in subsequent years.

A teacher leader is coded as either present or absent for each school site. In order to meet program-level fidelity, at least 80% of schools must have identified a teacher leader.

CHARACTERISTICS OF RAISE PROFESSIONAL DEVELOPMENT

RAISE PD sessions should be delivered in the manner characterized in the logic model, specifically, in terms of the following.

Content focused on disciplinary literacy, as demonstrated by:

- Use of disciplinary texts

- Discussion of unique challenges of disciplinary texts, such as discipline-specific vocabulary; structural features, such as features of scientific journal articles; specialized ways in which specific terminology is used in the discipline; or discipline-specific forms of discourse, such as those supporting scientific inquiry
- Discussion of how students acquire knowledge about texts, language, and disciplinary discourse and practices
- Discussion of how teachers can anticipate problems that need addressing before students can understand particular disciplinary text

Collective participation, as demonstrated by teachers:

- Working as part of groups
- Taking on different roles to support each other's learning processes
- Engaging in joint problem solving activities

Active learning, as demonstrated by:

- Session organization and flow mimicking a Reading Apprenticeship classroom (think pair share, working in pairs, turn-taking, explicitly sharing metacognitive processes with others)
- Teachers practicing Reading Apprenticeship classroom routines (e.g., practice and model a think-aloud, think-pair-share)
- PD presented through inquiry-based instruction where participants are asked to raise questions and investigate instructional modes and methods
- Participants collaboratively investigating instruction and methods; for example, through watching, analyzing, discussing, and asking questions about videos of classrooms or printouts of student work

Coherence, as demonstrated by:

- Facilitators asking participants for examples of strategies they have used or to brainstorm how to incorporate Reading Apprenticeship strategies into their practices
- Use of participants' prior knowledge as jumping-off points for learning new Reading Apprenticeship concepts (e.g., "write down everything you know about acids and bases")
- Facilitators incorporating participants' own materials (texts/curricula) into PD sessions

Metacognitive inquiry, as demonstrated by:

- Facilitators modeling a metacognitive process (e.g., thinking aloud, talking to the text)
- Participants exploring their own thinking while reading a text (e.g., noticing/identifying/ writing about their thinking process, reading strategies used, confusions)
- Participants sharing their metacognitive processes in groups (e.g., think-alouds, sharing confusions and strategies used to overcome challenges)
- Participants watching videos of students sharing their metacognitive processes and discussing insights into students' thinking and learning processes

During each observed session, the evaluator coded whether each of the five PD characteristics was "present" or "absent." Upon completing data collection, researchers calculated the percent that each of

the five characteristics appeared in the number of overall sessions, producing a single percentage for each characteristic. Program developers determined that each of the five characteristics must be present in 75% or more of observed sessions to meet the criteria for FOI. If any one of the five characteristics appears in fewer than 75% of sessions overall, the program does not meet fidelity.

Appendix K. Context for Program Implementation

In this appendix, we include additional results from year 2 teacher surveys on RAISE implementation components, supports and barriers to implementation, and overall impressions of RAISE.

TABLE K1. RAISE TEACHERS REPORTING ATTENDANCE AT MONTHLY MEETINGS

Month	Responded to survey	Attended monthly meeting
Sep	115	68
Oct	113	54
Nov	113	59
Dec	110	65
Jan	110	43
Feb	108	59
Mar	108	62
Apr	108	52
May	108	44

TABLE K2. RAISE TEACHERS REPORTING ACTIVITIES AT MONTHLY MEETINGS

	Reviewing student work	Discussing problematic lessons	Problem solving	Sharing successful lessons	Exploring Reading Apprenticeship tools/protocols	Other
Oct (n = 94)	36.2%	33.0%	44.7%	70.2%	58.5%	12.8%
Nov (n = 59)	42.4%	42.4%	61.0%	74.6%	44.1%	6.8%
Dec (n = 65)	41.5%	30.8%	55.4%	72.3%	44.6%	12.3%
Jan (n = 42)	35.7%	33.3%	35.7%	52.4%	33.3%	19.0%
Feb (n = 59)	44.1%	30.5%	47.5%	66.1%	47.5%	10.2%
Mar (n = 62)	48.4%	32.3%	51.6%	64.5%	50.0%	12.9%
Apr (n = 52)	48.1%	34.6%	40.4%	69.2%	40.4%	21.2%
May (n = 44)	36.4%	29.5%	40.9%	54.5%	34.1%	20.5%

TABLE K3. RAISE TEACHERS REPORTING REASONS FOR NOT ATTENDING MONTHLY MEETINGS

	Prefer not to respond	Not Offered	Other obligations	Not interested	Other
Oct (*n* = 19)	0.0%	89.5%	0.0%	0.0%	21.1%
Nov (*n* = 54)	3.7%	63.0%	14.8%	0.0%	20.4%
Dec (*n* = 45)	6.7%	60.0%	17.8%	0.0%	17.8%
Jan (*n* = 67)	1.5%	58.2%	17.9%	0.0%	31.3%
Feb (*n* = 49)	2.0%	59.2%	20.4%	2.0%	22.4%
Mar (*n* = 46)	4.3%	65.2%	13.0%	0.0%	23.9%
Apr (*n* = 56)	0.0%	67.9%	25.0%	0.0%	16.1%
May (*n* = 63)	4.8%	60.3%	22.2%	0.0%	15.9%

TABLE K4. RAISE TEACHERS REPORTING HELPFULNESS OF MONTHLY MEETINGS

	Not at all helpful	Less than moderately helpful	Moderately helpful	More than moderately helpful	Very helpful
Oct (*n* = 94)	3.2%	8.5%	51.1%	26.6%	10.6%
Nov (*n* = 59)	0.0%	11.9%	64.4%	20.3%	3.4%
Dec (*n* = 65)	6.2%	13.8%	53.8%	18.5%	7.7%
Jan (*n* = 42)	2.3%	9.3%	65.1%	7.0%	16.3%
Feb (*n* = 59)	5.1%	10.2%	52.5%	20.3%	11.9%
Mar (*n* = 62)	1.6%	8.1%	53.2%	19.4%	17.7%
Apr (*n* = 52)	3.8%	3.8%	48.1%	15.4%	28.8%
May (*n* = 44)	0.0%	9.1%	52.3%	18.2%	20.5%

TABLE K5. RAISE TEACHERS REPORTING LEVEL OF PREPARATION AFTER PD

	Not at all helpful	Less than moderately helpful	Moderately helpful	More than moderately helpful	Very helpful
A	0.0%	3.1%	31.3%	43.8%	21.9%
B	0.0%	1.1%	27.4%	61.1%	10.5%
C	1.0%	1.0%	33.3%	54.2%	10.4%
D	0.0%	2.1%	22.9%	62.5%	12.5%
E	0.0%	2.1%	24.2%	52.6%	21.1%
F	0.0%	2.1%	22.3%	46.8%	28.7%
G	0.0%	2.1%	20.0%	53.7%	24.2%
H	0.0%	9.4%	39.6%	44.8%	6.3%
I	0.0%	4.2%	35.8%	47.4%	12.6%
J	0.0%	6.3%	36.8%	38.9%	17.9%
K	0.0%	5.2%	30.2%	44.8%	19.8%

Note. *n* = 94-96 for each option.

See corresponding list of strategies in "Implementation of Core Program Components"

TABLE K6. TEACHERS REPORTING SUPPORT FOR LITERACY INSTRUCTION

		None	Coaching/ mentoring	Model lessons	Observation/ feedback	Resources	Classroom management help	Political support	Policy	Other
Oct	Control (*n* = 103)	61.2%	13.6%	7.8%	9.7%	16.5%	1.0%	0.0%	6.8%	5.8%
	RAISE (*n* = 110)	45.5%	14.5%	11.8%	20.9%	29.1%	1.8%	3.6%	2.7%	2.7%
Dec	Control (*n* = 109)	67.9%	9.2%	10.1%	15.6%	13.8%	3.7%	0.9%	0.9%	5.5%
	RAISE (*n* = 110)	57.3%	17.3%	10.0%	14.5%	21.8%	2.7%	4.5%	2.7%	4.5%
Feb	Control (*n* = 110)	57.3%	15.5%	10.9%	13.6%	16.4%	3.6%	0.9%	2.7%	6.4%
	RAISE (*n* = 110)	50.9%	25.5%	18.2%	20.9%	20.0%	4.5%	3.6%	3.6%	3.6%
Apr	Control (*n* = 108)	73.1%	10.2%	10.2%	16.7%	5.6%	0.9%	0.9%	2.8%	1.9%
	RAISE (*n* = 108)	48.1%	25.9%	11.1%	27.8%	18.5%	2.8%	0.9%	0.9%	4.6%

TABLE K7. TEACHERS REPORTING FROM WHOM THEY RECEIVED SUPPORT FOR LITERACY INSTRUCTION

		Principal	District personnel	Teacher leader	Department head	Other teachers	Parents	Program representative	Other
Oct	Control (n = 40)	27.5%	25.0%	30.0%	10.0%	37.5%	0.0%	12.5%	5.0%
	RAISE (n = 60)	26.7%	13.3%	28.3%	13.3%	61.7%	3.3%	28.3%	6.7%
Dec	Control (n = 35)	45.7%	20.0%	22.9%	17.1%	45.7%	2.9%	2.9%	17.1%
	RAISE (n = 47)	31.9%	10.6%	42.6%	12.8%	63.8%	2.1%	19.1%	2.1%
Feb	Control (n = 47)	31.9%	31.9%	25.5%	14.9%	40.4%	0.0%	2.1%	12.8%
	RAISE (n = 54)	20.4%	9.3%	37.0%	9.3%	46.3%	0.0%	27.8%	11.1%
Apr	Control (n = 29)	37.9%	20.7%	27.6%	24.1%	51.7%	0.0%	0.0%	13.8%
	RAISE (n = 56)	21.4%	7.1%	32.1%	5.4%	41.1%	0.0%	35.7%	12.5%

TABLE K8. TEACHERS REPORTING HELPFULNESS OF SUPPORT FOR LITERACY INSTRUCTION

		Not at all helpful	Less than moderately helpful	Moderately helpful	More than moderately helpful	Very helpful
Oct	Control (n = 40)	2.4%	19.0%	45.2%	26.2%	7.1%
	RAISE (n = 60)	0.0%	4.8%	47.6%	31.7%	15.9%
Dec	Control (n = 35)	8.6%	8.6%	48.6%	8.6%	25.7%
	RAISE (n = 47)	0.0%	6.4%	48.9%	29.8%	14.9%
Feb	Control (n = 47)	6.5%	17.4%	45.7%	19.6%	10.9%
	RAISE (n = 54)	0.0%	2.0%	49.0%	35.3%	13.7%
Apr	Control (n = 29)	0.0%	20.7%	51.7%	17.2%	10.3%
	RAISE (n = 56)	1.9%	5.6%	38.9%	18.5%	35.2%

TABLE K9. RAISE TEACHERS REPORTING LEVEL OF CONFIDENCE IN LITERACY INSTRUCTION STRATEGIES, APRIL OF YEAR 2

	Not at all helpful	Less than moderately helpful	Moderately helpful	More than moderately helpful	Very helpful
A	3.8%	6.6%	36.8%	34.0%	18.9%
B	0.0%	3.7%	39.3%	38.3%	18.7%
C	0.0%	1.9%	23.4%	47.7%	27.1%
D	0.0%	1.9%	28.0%	46.7%	23.4%
E	0.0%	3.7%	30.8%	42.1%	23.4%
F	0.0%	4.7%	23.4%	48.6%	23.4%
G	0.9%	3.7%	29.9%	43.0%	22.4%
H	0.0%	16.8%	42.1%	32.7%	8.4%
I	0.9%	7.5%	42.1%	35.5%	14.0%
J	0.9%	4.7%	36.4%	43.0%	15.0%
K	2.8%	7.5%	34.9%	41.5%	13.2%

Note. *n* = 106-107 for each option

See corresponding list of strategies in "Implementation of Core Program Components"

TABLE K10. RAISE TEACHERS REPORTING LEVEL OF UNDERSTANDING OF READING APPRENTICESHIP

	I do not get it	I understand some aspects of it, but I do not understand how it would translate into daily practice	It is starting to make more sense to me as I work with the approach to integrate it into my daily practice	I get it and am referring to it often as I plan and reflect on my teaching	Other (please explain):
CA (*n* = 48)	2.1%	4.2%	29.2%	60.4%	4.2%
PA (*n* = 59)	0.0%	5.1%	32.2%	61.0%	1.7%
Total (*n* = 107)	0.9%	4.7%	30.8%	60.7%	2.8%

TABLE K11. RAISE TEACHERS REPORTING CHALLENGES IN IMPLEMENTING READING APPRENTICESHIP

	Lack of materials	Lack of parent support	Lack of administrative support	RA is too much work	Competing priorities	Student behavior	Student ability	Lack of understanding of how to implement Reading Apprenticeship	Not enough training on Reading Apprenticeship	None	Other
Oct (*n* = 111)	17.1%	6.3%	6.3%	34.2%	62.2%	27.0%	30.6%	3.6%	3.6%	12.6%	8.1%
Jan (*n* = 107)	21.5%	5.6%	5.6%	32.7%	60.7%	33.6%	28.0%	7.5%	5.6%	14.0%	8.4%
Apr (*n* = 107)	24.3%	12.1%	14.0%	33.6%	63.6%	41.1%	32.7%	8.4%	6.5%	9.3%	6.5%

TABLE K12. RAISE TEACHERS REPORTING LEVEL OF COMMITMENT TO READING APPRENTICESHIP

	Not a priority	Willing to give it a try	Fully committed
CA (*n* = 49)	0.0%	30.6%	69.4%
PA (*n* = 59)	1.7%	40.7%	57.6%
Total (*n* = 108)	0.9%	36.1%	63.0%

TABLE K13. RAISE TEACHERS REPORTING ALIGNMENT OF READING APPRENTICESHIP WITH CLASS STANDARDS

	Not well aligned	Somewhat well aligned	Very well aligned
ELA (*n* = 34)	2.9%	17.6%	79.4%
Science (*n* = 39)	5.1%	51.3%	43.6%
U.S. History (*n* = 34)	2.9%	38.2%	58.8%
Total (*n* = 107)	3.7%	36.4%	59.8%

TABLE K14. RAISE TEACHERS REPORTING ALIGNMENT OF READING APPRENTICESHIP WITH CLASS GOALS

	Not well aligned	Somewhat well aligned	Very well aligned
ELA (*n* = 33)	15.4%	15.4%	69.2%
Science (*n* = 39)	7.1%	45.2%	47.6%
U.S. History (*n* = 34)	19.0%	31.0%	50.0%
Total (*n* = 106)	13.8%	30.9%	55.3%

TABLE K15. RAISE TEACHERS REPORTING LEVEL READING APPRENTICESHIP WILL IMPROVE STUDENT ACHIEVEMENT

	Not at all effective	Less than moderately effective	Moderate effective	More than moderately effective	Highly effective
ELA (*n* = 34)	0.0%	5.9%	26.5%	26.5%	41.2%
Science (*n* = 39)	0.0%	5.1%	46.2%	25.6%	23.1%
U.S. History (*n* = 35)	0.0%	2.9%	45.7%	40.0%	11.4%
Total (*n* = 108)	0.0%	4.6%	39.8%	30.6%	25.0%