Teacher Teams and School Processes in Scaling-up a Content Literacy Innovation in High Schools

FINAL REPORT: THE EVALUATION OF THE SCALE-UP OF READING APPRENTICESHIP THROUGH THE READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION (RAISE) PROJECT

December 2015

Denis Newman
Jenna Zacamy
Valeriy Lazarev
Li Lin
Andrew P. Jaciw
Empirical Education Inc.

Whitney Hegseth
University of Michigan

Acknowledgements

The evaluation reported here was supported by an Investing in Innovation (i3) Validation grant #U39B100255 awarded to WestEd. We are grateful for the support and feedback from Cynthia Greenleaf and Ruth Schoenbach, co-directors of WestEd's Strategic Literacy Initiative and the developers of Reading Apprenticeship, and from our study advisor, Cynthia Coburn. The authors also express deep gratitude to the teachers and administrators in the Indiana, Michigan, Pennsylvania and Utah schools for their assistance and cooperation in conducting this research.

About Empirical Education Inc.

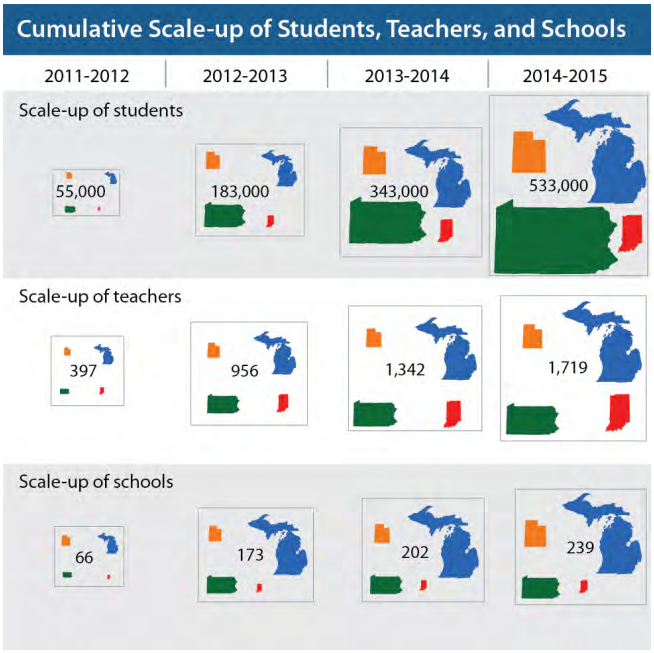
Empirical Education Inc. is a Silicon Valley-based research company that provides tools and services to help K-12 school systems make evidence-based decisions. The company brings its expertise in research, data analysis, engineering, and project management to customers that include the U.S. Department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

Executive Summary

We report on the scaling up of a high school content literacy intervention, Reading Apprenticeship, over a period of four years as part of the independent evaluation of an Investing in Innovation grant from the U.S. Department of Education to WestEd's Strategic Literacy Institute (SLI). The goal of our scale-up study was to understand the school processes that support successful implementation and promote scaling of an innovation. We focused on teachers, principals and other school-level program leadership who were also the focus for SLI's innovation designed to support scale. The logic model that our work was guided by built on research literature that provided insights into scale-up as increasing local ownership and depth of commitment, as well as growth in numbers. We provided formative feedback to SLI while contributing to the empirical methods and evidence for studying scale-up processes.

Study Design

The scale-up study was conducted in parallel to a RCT in which 22 treatment and 20 control schools from Pennsylvania and California participated for three years. In the scale-up portion of the evaluation, a total of 239 schools in four cohorts adopted the program in Pennsylvania, Michigan, Indiana, and Utah. Our study focused on the scale-up teachers and schools but also compared these to teachers and schools in the RCT treatment group. Surveys of teachers and principals, as well as records of participation in project activities provided a rich longitudinal dataset. In addition, four schools served as case studies where school leaders and program support staff were interviewed over two years to offer a concrete picture of decision processes and resources. We used descriptive statistics from surveys of the larger sample to track changes over time in features of program implementation, as well as resulting changes in attitudes and commitment among teachers. Regression methods were used to identify predictors of increased or decreased participation. We used these predictors in comparing the scale-up and RCT as contexts for the implementation and impact of Reading Apprenticeship.



Findings

Over the five years of the project, the scale-up of RAISE reached approximately 533,000 students, and 1,719 teachers in 239 schools in four states. During the first year of RAISE implementation, teachers and administrators in the Scale-up schools reported high levels of buy-in and commitment to the initiative. Teachers were successfully integrating Reading Apprenticeship practices into their instruction and cited collaboration and support from other teachers as the most effective means at building their capacity to implement what they learned during the RAISE professional development. Teachers and administrators identified competing initiatives as a primary challenge to implementing and sustaining RAISE long term and, as we found through our case study work, schools faced unique and localized challenges as they navigated their way through the scale-up. In the second and third year of implementation we found that trends in the uptake of RAISE activities and enthusiasm for the initiative decreased or leveled off by the third year. We also found a greater spread in

responses across schools by the third year, calling for further investigation into school-level processes and the characteristics of schools that may be associated with the gain or loss of RAISE participants over time.

Through this investigation, we found that teacher participation in team meetings during the first year and school-wide commitment (rather than individual commitment and actual classroom usage by the teachers) predicted the increase in number of teachers participating in a school. Such increases were less likely where the school did not see regular participation in team meetings in the first year and there was less commitment among teachers to school-wide success. Increases were unrelated to demographics and resources available to the school and unrelated to many characteristics and actions of administrators. We hypothesized that a process, consistent with SLI's innovation for scaling Reading Apprenticeship, which involved a cross-disciplinary teacher team within the school, led to increased and sustained program participation. As a final step, we compared the characteristics of Scale-up and RCT treatment schools. Scale-up schools had more of the characteristics associated with growth suggesting greater long term sustainability than in RCT schools. We raise the issue of whether the implementation under the constraints of the RCT is a good model for measuring impact in the context of scaling up an innovation.

TABLE OF CONTENTS

Reference this report: Newman, D., Zacamy, J., Lazarev, V., Lin, L. , Jaciw, A.P., & Hegseth, W. (2015). *Teacher Teams and School Processes in Scaling-up a Content Literacy Innovation in High Schools. Final Report: The Evaluation of the Scale-up of Reading Apprenticeship through the Reading Apprenticeship Improving Secondary Education (RAISE) Project.* (Empirical Education Rep. No. Empirical_i3SU-7019-FR1-O.2) Palo Alto, CA: Empirical Education Inc. Retrievable from http://empiricaleducation.com/pdfs/SUfr.pdf

Chapter 1: Study Overview

This report presents findings from a multi-year study of Reading Apprenticeship, an approach to improving academic literacy, as it was scaled-up to 274 high schools in five states. Our work, in collaboration with the program developers, WestEd's Strategic Literacy Initiative (SLI), was a formative experiment in which analyses of our surveys, interviews, and participant observations provided feedback to SLI as they rolled out Reading Apprenticeship over successive cohorts. Our goals in this project were, first, to provide data to SLI to support implementation of an innovative model designed to support school-based teacher teams that would facilitate local ownership of Reading Apprenticeship. Second, building on Coburn's (2003) insights, our goal was to measure SLI's success. Going beyond just measuring the numbers of students, teachers and schools reached, we wanted a metric that would be indicative of the program taking hold in schools, and ultimately, the scalability of the program.

EVALUATION OF STRATEGIC LITERACY INITIATIVE'S INVESTING IN INNOVATION FUND GRANT

SLI began developing Reading Apprenticeship in 1995 and has since reached over 100,000 teachers in schools across the country, at the middle school, high school and college levels. In 2010, SLI received a "Validation" grant from the Department of Education's Investing in Innovation Fund (i3) competition to scale-up and study the *Reading Apprenticeship Improving Secondary Education* (RAISE) project.[1] For this five-year project, SLI focused on three secondary school content areas: English Language Arts, History, and Biology. From the grant, SLI not only funded an independent randomized control trial (Fancsali et al., 2015) but also a parallel effort, reported here, to study the schools outside of the RCT that were to receive the same professional development and other supports in implementing Reading Apprenticeship. SLI's goals in the scale-up part of the grant were to build local education agency capacity to disseminate, support, and sustain academic literacy improvement in high school subject areas. While i3 and similar funding can "prime the pump," the project must build the capacity to disseminate, support, and sustain the innovation. Ultimately, adoption by school systems and evidence of success will keep it going. In Chapter 2, we provide an overview of how the RAISE initiative was scaled-up and present a conventional approach to "counting" scale in terms of the number of participants (teachers, schools, and districts) reached over time.

To support inquiry into the scale-up process, we developed an unconventional spiraling logic model described below (and included in Appendix A in more detail), which was inspired by the effort of putting the SLI approach to scaling up nationally together with Coburn's (2003) insights in the processes of buy-in and commitment that make an innovation self-sustaining. The logic model pointed to activities that potentially mediated between the RAISE program, changes in educator attitudes, and local adoption of Reading Apprenticeship as the model for literacy instruction.

As detailed in chapter 3, our formative evaluation of RAISE scale-up collected and analyzed data on the number of trainings, the reach of the program, the program elements that were taken up or not by

[1] Throughout the report, "RAISE" refers to the i3 project and to the project activities and Reading Apprenticeship is the instructional framework that is used to inform instruction in the classroom.

participants, and the processes that may have affected the uptake and sustainability of the initiative. We tracked participation in the scale-up through sign-in sheets at each of the training events and linked the data by school, district, and state each year. We surveyed the teachers and principals, participated in many of the project meetings and training events, and conducted case studies in four schools to get a richer, more contextualized understanding of the scale-up process. As cohorts of schools and teachers were added to the project each year, these data were collected from a total 239 schools in four states (Utah, Michigan, Indiana, and Pennsylvania). By 2014-15, approximately 1,720 teachers received training in the scale-up study side of the overall i3 project. We also collected some of the same data from schools participating in the RCT allowing the comparison discussed in Chapter 8.

Chapter 4 examines survey responses to gauge adoption of Reading Apprenticeship, buy in by teachers and school administrators along with measures of participation in program activities. These analyses provided useful insights for program developers and leaders and suggestions about malleable factors that could be the focus of future improvements. Chapter 5 presents a longitudinal analysis of the uptake of RAISE components and commitment of teachers to the program over three years. In Chapter 6, we take an up-close look at the strengths and challenges, and decisions that four case study schools faced as they scaled up the initiative over a two-year period.

In Chapter 7 we go back to the numbers and connect the quantitative evidence of scaling-up (a measure of the growth and loss of RAISE participation within states, districts, and schools), and contextual factors and indictors of scale-up processes within a school that predict this growth or loss. Chapter 8 explores the unique opportunity that we have to compare data on RAISE implementation and commitment between the treatment schools in the RCT and the scale-up schools. Chapter 9 summarizes our findings and draws out the implications for studying interventions at scale and the scale-up processes.

Before going into these findings, we first need to describe the program SLI was scaling up through the RAISE grant.

READING APPRENTICESHIP FRAMEWORK AND THE RAISE INTERVENTION AND LOGIC MODEL

What Reading Apprenticeship Is and Prior Research

Reading Apprenticeship is an instructional framework that helps teachers support discipline-specific literacy and learning in their varied content areas (Schoenbach, Greenleaf, & Murphy, 2012). It is not a curriculum, list of strategies, or set of materials. Instead, it gets teachers to attend to four interacting dimensions of classroom learning culture.

- **Social:** this involves building community. The classroom becomes a safe environment where students see other students and their teacher as resources for learning.
- **Personal**: this 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.
- **Cognitive**: this involves developing students' mental processes, including their text-based problem-solving strategies.

- **Knowledge-Building:** this 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.

At the center of Reading Apprenticeship is what the developers describe as an ongoing metacognitive conversation carried on both internally through metacognitive reading and reasoning routines and externally, as teacher 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 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. The framework targets learning dispositions as well as literacy skills and knowledge.

The Reading Apprenticeship intervention is inquiry-based professional development intended to transform teachers' understanding of their role in adolescent literacy development. This PD engages teachers in the following.

- learning about the complexity of literacy and learning with disciplinary texts
- learning how the framework supports students' literacy and learning
- practicing specific pedagogies
- carrying out formative assessment focused on student reading, thinking and learning

Previous RCTs have tested the efficacy of the Reading Apprenticeship framework and the professional development model in closely monitored efficacy studies that demonstrated strong positive effects on teacher practice—most notably, teachers' increased use of reading comprehension strategy instruction, metacognitive inquiry routines, and collaborative learning structures in their classrooms. They also show positive effects on students' literacy and content-area achievement (in science, ELA, and history), motivation, and engagement, and that English learners particularly benefited from Reading Apprenticeship instruction (Greenleaf et al., 2011a,b; Kemple et al., 2008; Somers et al., 2010). Two of these studies used an intensive professional development model consisting of 10 days of discipline-specific inquiry-based professional development. This model was embedded in RAISE, but is different from the other mode in which Reading Apprenticeship is often introduced as cross-subject professional development.

Innovations for the Investing in Innovation Fund Implementation at Scale

The i3 validation project that this scale-up study followed, RAISE, proposed to provide Reading Apprenticeship professional development and supports for implementation for approximately 2800 teachers in 306 schools, across four states (a fifth state, California, was added later as part of the RCT).

In addition to 10 days of inquiry-based discipline-specific professional development, as in two of the prior RCT studies, the RAISE project included several innovations to scale to this level.

1. **Onsite RAISE team meetings** involving all teachers who attended the RAISE professional development were to meet monthly on-site to support each others' implementation of Reading Apprenticeship routines

2. **Teacher leaders** recruited for each school from the participating team of teachers facilitated team meetings; in 2nd through 4th years of the project teacher leaders were offered additional professional development to help them in this role
3. **State-level RAISE Coordinators** appointed to provide locally knowledgeable support to RAISE school teams. State coordinators carried out a number of functions necessary to implement project activities at a distance from the west coast SLI office. These activities included:
    - communicating and coordinating project activities at the state and local level
    - convening and facilitating cross-site Teacher Leader meetings
    - working with SLI to coordinate and plan the RAISE Institute in their state
    - working with school administrators to enhance their support for the Reading Apprenticeship framework and RAISE project
    - promoting RAISE work in regional and state-level venues to build sustainability
    - conducting inquiry into and facilitating conversations about the model with site administrators, central SLI office staff, and other state coordinators
4. **85 Professional Development Facilitators** trained and apprenticed to deliver the revised, discipline-based 10-day Reading Apprenticeship professional development series. As part of this, SLI developed materials, protocols, online and face-to-face training, and assessments to support facilitator development. In additional, SLI "certified" facilitators to use the professional development materials for site based trainings (i.e. outside of the RAISE 10-day Institute).
5. **RAISE administrator programs and materials.** Although not initially in the proposal, in response to field requests, SLI developed an online administrator course and provided opportunities during the teacher professional development sessions for RAISE.
6. **Website for teacher exchange.** This web-based portal, called *Thinking Aloud*, from the initial RAISE proposal was supported for a semester in Year 2 and then discontinued due to low levels of use.

Scale-up Research vs. Scaling-up

In the RAISE project, the parallel studies involving, on the one hand, an RCT, and on the other hand, a study of the scaling up of the intervention without the goal of measuring impact, highlights a distinction in the research literature between "scale-up research" (McDonald, Keesler, Kauffman, & Schneider, 2006) and research on the processes of scale-up (e.g., Adelman and Taylor, 1997; Coburn, 2003). Scale-up research is the discipline of conducting large-scale effectiveness trials where internal validity based on random assignment is challenged by the potential for attrition, cross-over and the like, while external validity is challenged by contamination, subgroup differences, failures of adequate implementation, etc. Large scale RCTs involve a power analysis to predict the number of units (schools or teachers or classrooms) that will be needed to test the hypotheses and then typically a major recruiting effort goes into lining up enough units willing to participate in the research, which could involve being assigned to not doing the program for several years, although generous stipends may assist with participant motivation. (The recent interest in opportunistic RCTs mitigates some of the difficulties of large scale experiments [Newman & Jaciw, in preparation].)

Scaling-up a program outside of the context of an effectiveness trial is more about the commitment of school and district leaders, state education department policies, instilling a commitment to change among teachers and administrators, and the development of collaborative supports within the school or professional learning community. It certainly helps, as with the RAISE project, that there is funding for training institutes, state coordinators, recruitment activities, as well as the program developers, but the growth in the size of the program is also largely dependent on processes within the schools and education agencies. These processes are under-theorized and as Sternberg et al. (2011) contend "little—arguably, almost nothing—is known about the factors that lead to successful scaling up" and that there has "not been a systematic review of the available knowledge, either at the level of theory or at the level of empirical evaluation of hypotheses and observations on the process of upscaling." The current study may be a contribution to this relatively new field. There are many important differences between the context of large-scale RCT and the scaling-up of programs outside of field trials but this is the topic of a separate analysis by our team.

DEVELOPING A THEORY OF ACTION FOR SCALE-UP

The focus of this study is to understand the processes involved in scaling up Reading Apprenticeship in different states and contexts, as well as the stages of transition that occur as ownership is transferred from the developers to local districts and schools. Given this focus, our theory of action builds upon Adelman and Taylor's (1997) four phases of scale-up and Coburn's (2003) four dimensions of scale-up.

Adelman and Taylor's (1997) model depicts four overlapping phases of scale-up. In the first stage, *Creating readiness*, efforts are directed toward disseminating program information, building interest, and negotiating policy frameworks for involvement. The second phase, *Initial implementation*, includes guiding the adaptation of the intervention by creating temporary mechanisms to facilitate implementation (e.g., mentors or coaches). The third phase, *Institutionalization*, ensures long term ownership and sustainability of the intervention which requires ongoing leadership to take responsibility for the intervention, and coordination mechanisms to keep the intervention running. The fourth phase, *Ongoing evolution*, is concerned with accountability and continually informing practices for improvement through formative and summative evaluation. Within each of these four phases are activities carried out by the scale-up staff, as well as collaborative efforts between scale-up staff, organizational leadership, and stakeholders.

Coburn (2003) proposed an expanded "conceptualization of scale consisting of four interrelated dimensions:" depth, spread, sustainability, and shift in reform ownership. Beyond just changes in classroom structure (e.g. materials, classroom organization), depth of reform-centered knowledge also includes changes in the teachers' underlying assumptions about pedagogical principles and expectations of students and how students learn. Spread pertains to increasing the number of schools or classrooms using a program, as well as the spread of reform-related norms, beliefs, and principles within a classroom, school, and district. This idea of spread includes an increase in the number of participants across sites (external spread), as well as within classrooms, schools, and districts (internal spread). Borrowing from the scale-up model presented at the 2015 conference of the National Center for Scaling Up Effective Schools (Cannata, Rutledge, Redding, & Nguyen, 2015) we use "scaling out" to refer to external spread and "scaling in" to refer to internal spread. Sustainability is the distribution, adoption, and maintenance

of an innovation over a long term. Coburn identifies some of the biggest challenges of sustainability as competing priorities in schools, changing demands (within the school and larger policy demands), and teacher and administrator turnover. Shift in reform ownership concerns the ultimate goal of reform efforts—to transfer the reform-centered knowledge, authority, and agency from the "external" providers to the "internal" actors (e.g., teachers, administrators, schools, and local and state education agencies) thereby sustaining the reform in ways that make a difference to students. This expanded conceptualization of scale moves away from the idea of replication toward conceptual, organizational, and philosophical changes that can be sustained over time.

RAISE Scale-up Logic Model

A traditional logic model, with inputs on the left, outputs or intermediate outcomes in the middle, and final outcomes on the right does not lend itself to representing this complex, multilevel, iterative scale-up process. Instead, we developed an interactive logic model that shows four stages of development from initial project development to the project goal of Reading Apprenticeship being broadly institutionalized (see Appendix A for comprehensive narrative description of each stage of the logic model and accompanying figures). The RAISE scale-up logic model consists of four stages.

Stage 1: Development activities
Stage 2: Increased ownership
Stage 3: Sustained ownership
Stage 4: Reading Apprenticeship broadly institutionalized

Stage 1 comprises the design and construction of the four development activities (i.e., Professional Development for Reading Apprenticeship facilitators and teachers; Instructional Support Resources; Recruitment and Retention; and Project Development and Coordination). The processes and materials for these activities, which we call "SLI's RAISE" are developed through the i3 grant funds. Additionally, this stage includes the uptake of these activities within the recruited and implementing schools and districts. This stage is similar to Adelman and Taylor's (1997) first two phases: Creating readiness and Initial implementation. These activities are not only designed to spread the enactment of RAISE activities in the participating schools, but they are also expected to instill participant buy-in and capacity to the extent that, in the ensuing stages, the developers are able to transfer responsibility for and ownership of Reading Apprenticeship to local districts and schools, as described in Coburn's model.

The development activities are hypothesized to lead to five intermediate outcomes: (1) increased participation in RAISE, (2) classroom fidelity of Reading Apprenticeship, (3) buy-in to the Reading Apprenticeship framework, (4) capacity to implement and disseminate Reading Apprenticeship practices, and (5) student achievement. Our first two intermediate outcomes—increased participation and classroom fidelity of Reading Apprenticeship—correspond to Coburn's (2003) first two dimensions of scale-up: spread and depth. Our second two intermediate outcomes—increased local capacity and buy-in—are expected to lead to increased local ownership of Reading Apprenticeship in later stages of the process.

These intermediate outcomes will also interact with each other. As buy-in and commitment to Reading Apprenticeship increase, we hypothesize that district school administrators, and teachers will dedicate the time and resources necessary to increase capacity to implement and disseminate Reading Apprenticeship

at the local level. As capacity and support builds, we expect districts and schools to increase the numbers of teachers implementing Reading Apprenticeship by scaling in (within RAISE schools) and scaling out (to new schools); that is, existing RAISE schools will send more teachers to RAISE training and spread the Reading Apprenticeship ideas to new districts and schools. We also expect classroom fidelity of Reading Apprenticeship to lead to increases in student achievement, as evidenced by improved standardized student test scores (Corrin, Somers, Kemple, Nelson, & Sepanik, 2008; Greenleaf et al., 2009; Greenleaf, Schneider, & Herman, 2005).

Stage 2 (Increased ownership) and Stage 3 (Sustained ownership) are hypothesized to result from the intermediate outcomes. These stages correspond to Coburn's "shift in reform ownership" dimension. Stages 2 through 4 are also similar to the third phase in Adelman and Taylor's model, institutionalizing new approaches. In Stage 2, we hypothesize that as the local level begins to take ownership of the development activities, these activities are adapted to meet their needs, which further reinforces the intermediate outcomes.

Stage 4 is RAISE's ultimate goal, Reading Apprenticeship broadly institutionalized as the model of academic literacy instruction, and where activities are fully implemented at the local level with limited support from SLI. Once the intermediate outcomes are realized, we hypothesize two end outcomes: local level policy shifts and Reading Apprenticeship spreading with depth beyond the original Local Education Agencies (LEAs) that were recruited to join the project (SLI, 2010). The model also depicts the influences and feedback loops that are active during this stage. Our final stage corresponds to Coburn's dimension of Sustainability, balancing the centralized, on-going research and development functionality of the developers with the uptake of reform ownership at the local level.

Differences between Theory of Action for RCT and Scale-up Study

The evaluation of RAISE encompassed an RCT and the scale-up study. Each had different goals and different theories of action. The RCT was conducted in 42 schools in Pennsylvania and California, with the goal of estimating the impact of Reading Apprenticeship on student achievement on general disciplinary literacy while understanding the conditions under which the impact was found and the mediators of the impact. The scale-up study was a formative evaluation of the scale-up process in four states (not including California) with the ultimate goal of helping the developers bring the innovation to scale.

The RCT and the scale-up study were designed around complementary theories of how Reading Apprenticeship works. The theory of action for the RCT is focused on changing teacher practices so as to support an apprenticeship process in the classroom and thereby improve student cognitive capacities and academic dispositions measured by an achievement test and attitude measures. The theory operates primarily at the teacher-classroom-student level. In contrast, the primary outcome for the scale-up study is the project's success in building a self-sustaining capacity to implement and maintain the improvements. For scale-up, the logic model operates at organizational levels at and above the classroom: the support structures at the teacher, school, district (LEA), and state levels. The theory sees the elements at all these levels as forming potentially positive feedback loops and indicates potential sources that block successful scale-up.

The overall goal of the scale-up study is to understand how schools and ultimately school systems build capacity to implement and spread Reading Apprenticeship practice and sustain these efforts. In our review of the literature in this area, we found that unified theory of scaling-up education reforms is in its early stages, and few empirical studies have investigated this process. This is one of the first empirical studies of a scale-up process across multiple states and contexts. Our goal is to investigate how the program becomes rooted across several different contexts under authentic conditions of implementation. From this, we can continue to develop hypotheses to guide the implementation and scale-up process and begin to build generalizations about the conditions for successful implementation and scale-up of Reading Apprenticeship in various settings. The results of this study will add to the research knowledge and literature on educational scale-up, as well as scale-up of literacy programs. In addition, this study has informed the development and elaboration of the RAISE scale-up logic models and theory, which may contribute to understanding in the field about how to study the scale-up of other innovations.

Chapter 2: Overview of the Scale-up

In this chapter, we focus on one of the scale-up outcomes: increased participation in RAISE. We begin with a description of the RAISE scale-up timeline and recruitment approach, then present a summary of the numbers of schools, teachers, and students served by the initiative, and end with maps documenting the spread of RAISE across each state. While we acknowledge that scale-up can be simply measured in terms of the increasing numbers of participants over time (Slavin, 2002), the findings presented in later chapters go much further and examine the school processes that may affect the growth (or loss) of participation.

RAISE SCALE-UP RECRUITMENT PROCESS

SLI's model for scale-up included participation from four consecutive cohorts of RAISE teachers and schools across four years (Table 2.1). Each year the State Coordinators would recruit new districts and schools (scale-out), and new teachers from existing schools (scale-in), to participate in RAISE. The State Coordinators worked closely with SLI to plan the RAISE professional development in their state, which was held in one central location.[2] Teachers were offered $1000 to attend the 10-day RAISE Institute with their school team (five days in the first summer, two days in the winter, and three days in the following summer). Principals were also encouraged to attend to build administrator support and knowledge of the initiative. Monthly RAISE team meetings during the school year, which were led by a trained teacher leader at each school, were expected to provide additional support for implementation.

TABLE 2.1. YEARS OF PARTICIPATION FOR READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION COHORTS

Cohort	2011-12	2012-13	2013-14	2014-15[a]
1	Year 1	Year 2	Year 3	Year 4
2		Year 1	Year 2	Year 3
3			Year 1	Year 2
4				Year 1

[a] There was limited data collection in the last year of the grant.

From an evaluation perspective, as the scale-up process proceeded across contexts, states, and years, this model allowed us to quantify changes in implementation over time within a given cohort, compare cohorts in their first, second, and third years of the initiative, and track schools that gained or lost participants.

The State Coordinators from each state were responsible for identifying and recruiting districts, schools, teacher leaders, and teachers to participate in RAISE and adopt the Reading Apprenticeship framework.

[2] With Cohort 2, Michigan scaled out to the northwestern part of the state. Due to the distance between sites, Michigan held two RAISE Institutes: one in the southeast and one in the northwest.

They, along with the Multi-State Coordinator, provided regional knowledge and management of their local sites. They did so through assessing the interest and need of districts and schools, building relationships with participants, addressing barriers or concerns to participation, and disseminating information. They were also essential in building relationships with LEA and State Education Agency administrators, and understanding the local context around literacy initiatives, objectives, and assessments, and how RAISE aligned with local needs.

For the first cohort of RAISE scale-up schools, the State Coordinators focused their recruitment efforts on schools that had prior experience with Reading Apprenticeship (through district or regional trainings), schools that they had personal relationships with, and/or schools that were closer in geographic proximity in order to build a solid foundation of participation. They sent out recruitment letters/flyers. Together with SLI staff, State Coordinators held a major informational "kick-off" meeting with local stakeholders in each of the four states. They also focused on building administrative support and recruiting "school teams". During the planning stages of the grant, SLI estimated that each school would send nine teachers to the RAISE Institutes (three in each content area). They hypothesized that it would be important to establish a core RAISE team at each school, to build a critical mass of Reading Apprenticeship implementers so they could collaborate with and support each other. Recruiting multiple teachers per school (and per content area) would also allow for a larger number of students to be reached, in multiple content areas and grades, which would deepen and engrain practices in the school. A larger team would also make it less likely that teacher turnover would threaten the sustainability of RAISE. In subsequent years, the State Coordinators focused on spreading the initiative to new schools and districts (scaling out) and creating "hubs" of Reading Apprenticeship practice throughout the state, but more importantly on building the capacity of school teams and districts who were already participating (scaling in).

COMPARING SCALE-UP AND RCT RECRUITMENT

While teachers in both the scale-up and RCT studies attended the same RAISE professional development and were provided with the same implementation expectations and support from SLI, the recruitment strategies differed (Fanscali et al., 2016). First, recruitment of the 42 schools in the RCT occurred only in Pennsylvania and California (no schools in California were recruited as "scale-up" schools). Additionally, to be eligible to participate in the RCT, schools had to meet the following criteria.

- Serve grades 9-12 in a single building
- Serve high proportions of students who were eligible for free and reduced-price lunch, were English Language Learners (ELLs), and had low prior achievement
- Have no prior experience with Reading Apprenticeship
- Have district approval to participate in the RCT and provide study data
- Have teachers willing to participate in the study data collection (which included nine online surveys, classroom observations, and administering a student survey and assessment in each study year)

Schools were ineligible to participate in the RCT if they had any of the following characteristics.

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

The first cohort of schools in the RCT included 17 RAISE and 15 control schools in Pennsylvania and California. The second cohort included five RAISE and five control schools, all in California. RAISE schools in the RCT were allowed to send additional teachers in later cohorts (scale-in), but only four schools (all in Pennsylvania) choose to do so.

SCALE-UP NUMBERS SERVED AND MAPS OF PARTICIPATION

In the tables and infographics below, we have provided detailed information regarding the number of schools and teachers that participated in RAISE as Scale-up schools across the four cohorts. By the end of the 2014-15 school year, 1719 teachers from 239 schools were reached by the RAISE initiative (Tables 2.2 and 2.3). Almost 50% of teachers and 55% of the RAISE Scale-up schools were located in Michigan.

TABLE 2.2. READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP TEACHERS REACHED

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Indiana	49	76	62	71	258
Michigan	208	250	209	194	861
Pennsylvania	68	167	83	63	382
Utah	72	66	32	49	219
Total	397	559	386	377	1719

TABLE 2.3. READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP SCHOOLS REACHED

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Indiana	7	7	3	4	21
Michigan	33	66	19	14	132
Pennsylvania	11	25	2	7	45
Utah	15	9	5	12	41
Total (scaling out)	66	107	29	37	239
Total (scaling in)		25	50	35	
Total (per cohort)	66	132	79	72	
Total (cumulative)	66	173	202	239	

Figure 2.1 shows the cumulative number of students, teachers, and schools served by the end of each year. The size of the state represents the relative scale (compared to the other states), in terms of the number of participants. Figure 2.2 represents the number of schools that participated in the RAISE professional development, and the number that scaled-in and scaled-out each year.

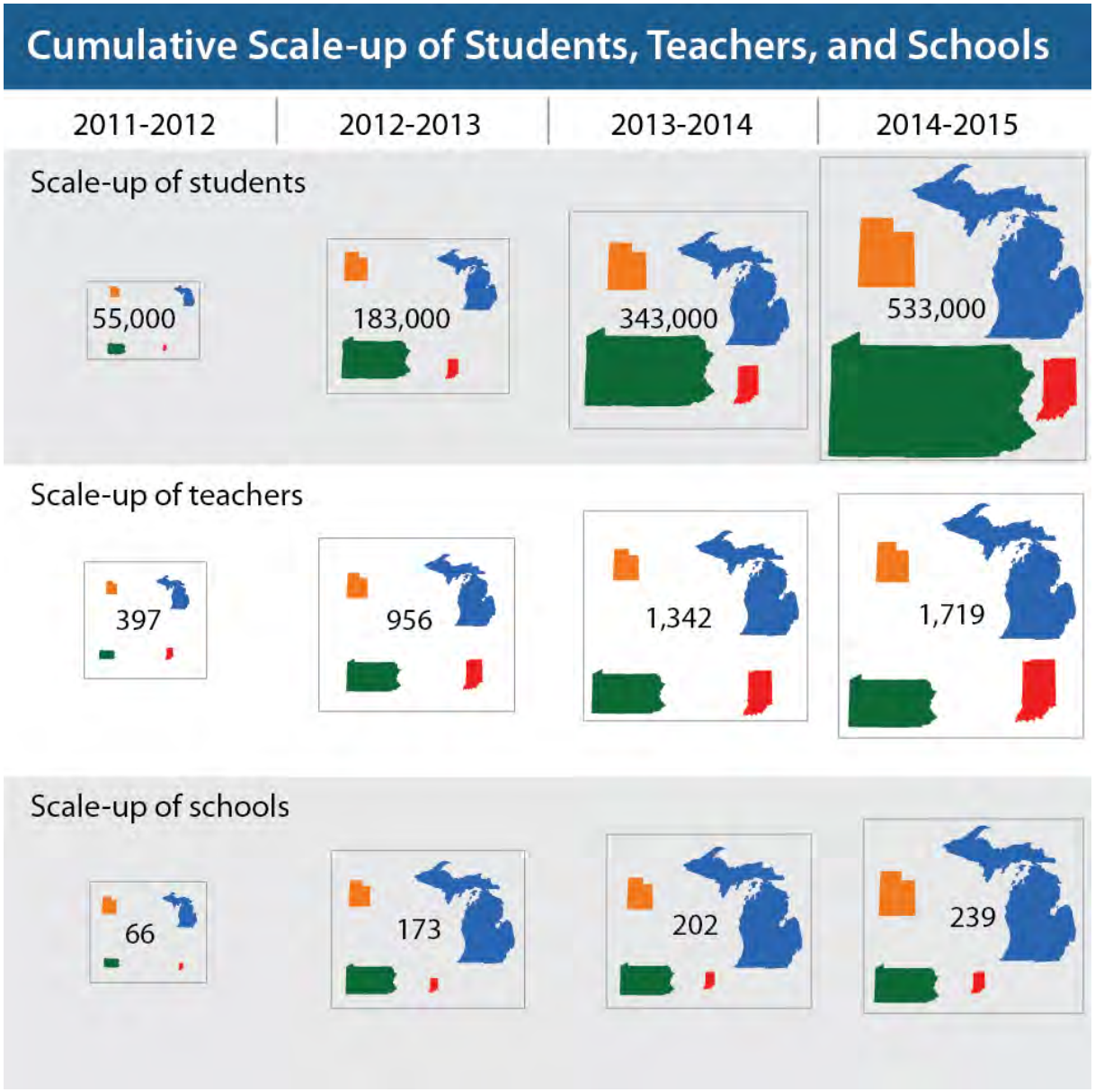


FIGURE 2.1. CUMULATIVE SCALE-UP OF STUDENTS, TEACHERS, AND SCHOOLS

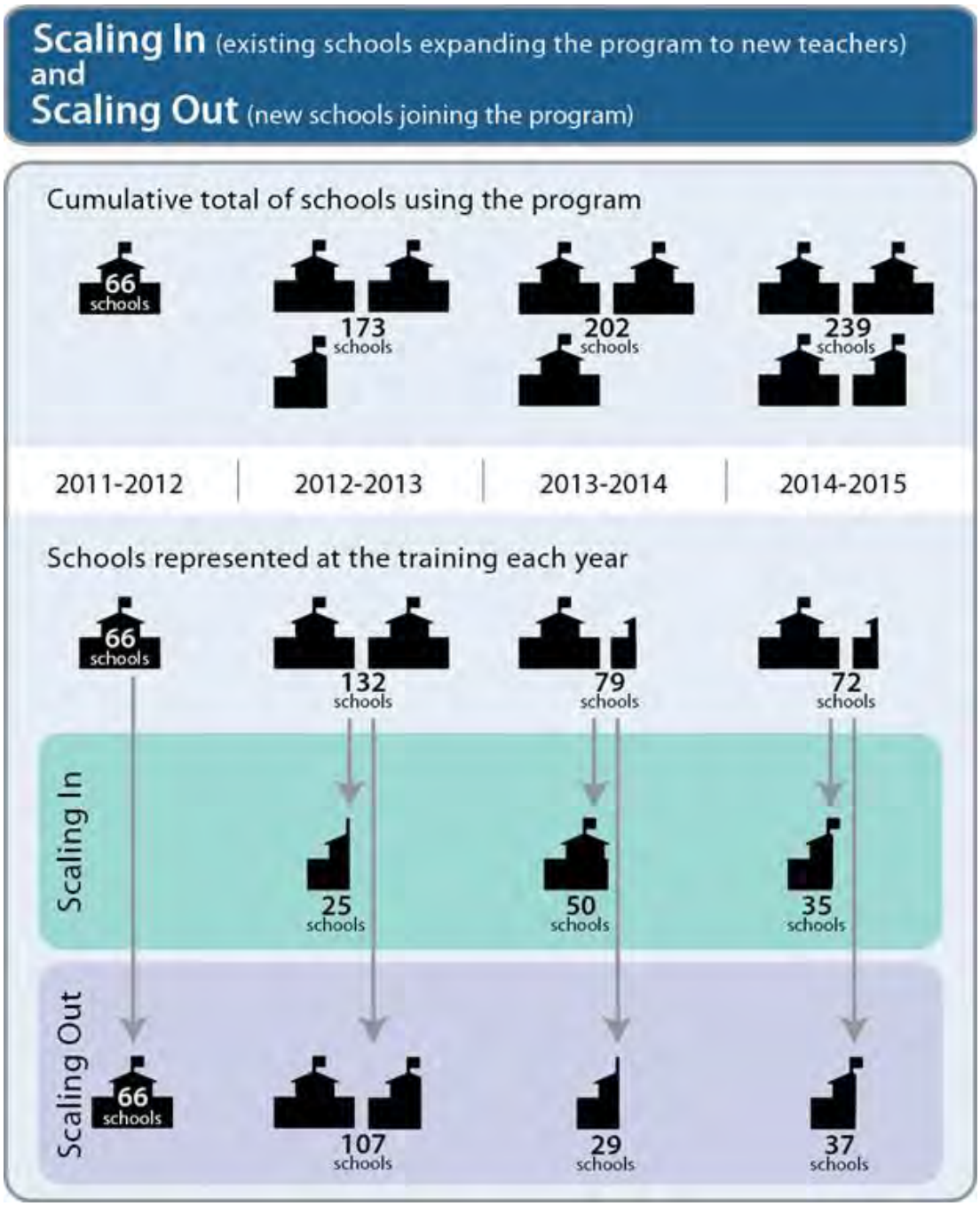


FIGURE 2.2. SCALING IN AND SCALING OUT

Across both the RCT and Scale-up sites, RAISE reached approximately 630,000 students and 1965 teachers in 274 schools. This provided us a powerful testbed for examining the scale-up processes of interest. In their 2010 i3 proposal, SLI projected the number of schools and teachers that would be participating in RAISE, by year and state. Table 2.4 shows the number of schools and teachers projected compared to the actual number of participating schools and teachers trained. For the first and second cohorts, SLI exceeded

the number of new schools trained; however, by the third and fourth cohort, the State Coordinators were more focused on scaling-in to existing RAISE trained schools, so the target numbers of schools were not met. The need to support teacher travel across larger geographical regions meant that fewer teachers could be served. Our estimates suggest that the projected number of students were served by the grant's end.

TABLE 2.4. COMPARISON OF PROJECTED NUMBERS TO ACTUAL PARTICIPATION

	SLI Projection		Actual Participation	
State	Estimated number of teachers	Estimated number of schools	Number of teachers trained	Number of new schools trained (scaling out)
Cohort 1	396	44	397	66
Cohort 2	666	74	559	107
Cohort 3	666	74	387	29
Cohort 4	666	74	377	37
Total (Scale-up sites only) [a]	2394	266	1720	239
Total (RCT + Scale-up sites)	2754	306	1965	274

[a] Cohort totals above includes Scale-up sites only.

Note. SLI stands for Strategic Literacy Initiative

For each year, we created maps of the Scale-up states identifying the districts and schools that were participating in RAISE. Districts with at least one school participating in RAISE are highlighted and the locations of the participating schools are marked with a black or orange dot. The initial purpose of the maps was for the SLI team and State Coordinators to identify "hubs" of participation in each state, to inform decisions about investment of further time and resources allocated to building capacity at the district or school level, and to help inform strategic recruitment for future cohorts. The maps also visually document the size of the scale-up, how the initiative spread across the states, and which schools and districts scaled in. Figure 2.3 shows the spread of the initiative by the 2014-15 school year, with all four cohorts of Scale-up district and schools.[3] The darker shades of blue represent districts that scaled-in and had multiple cohorts of teachers and/or schools participating in RAISE.

[3] Appendix B includes the maps of scale-up schools for each year of the initiative.

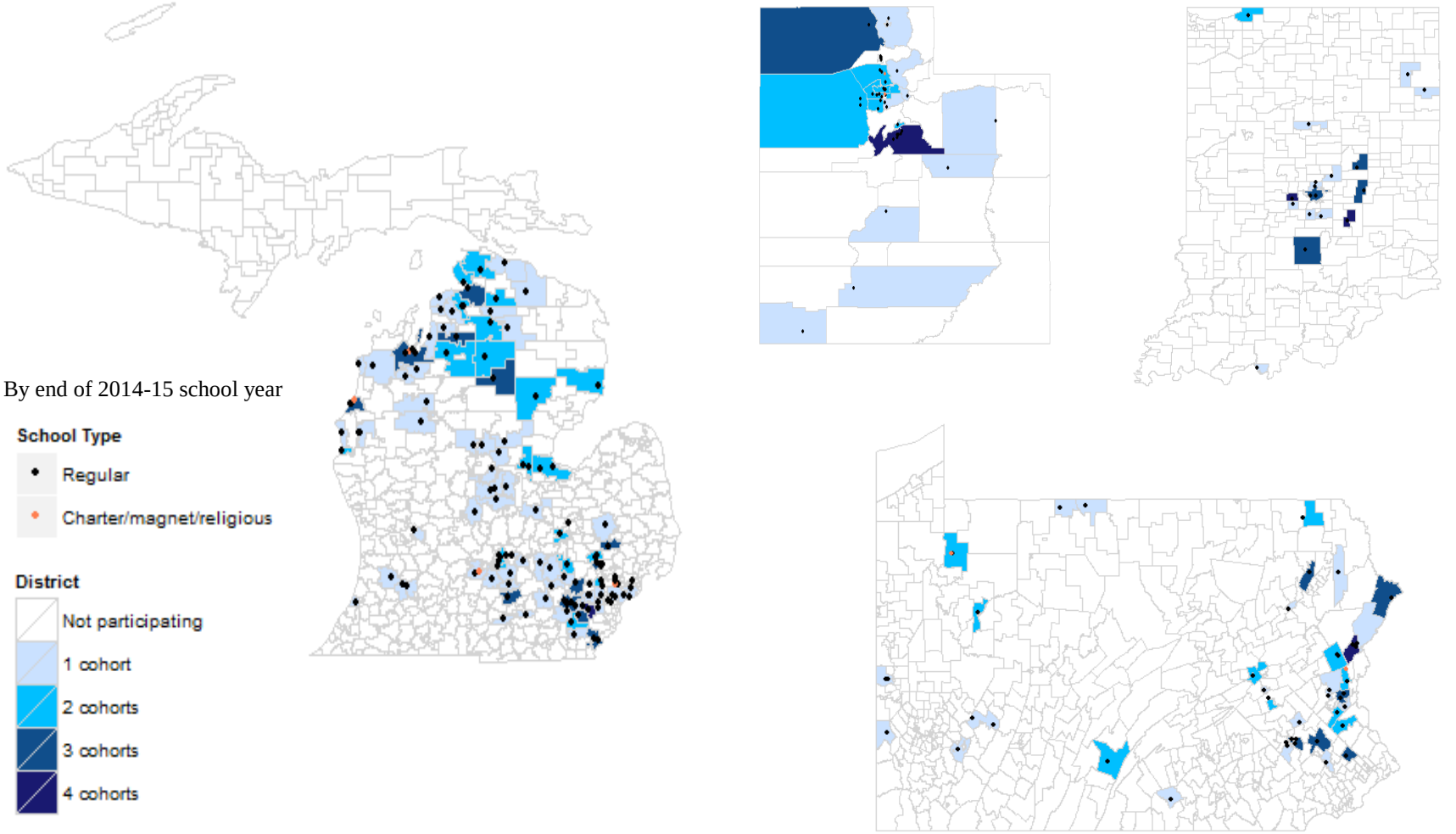


FIGURE 2.3. MAP OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP PARTICIPATION IN 2014-15

Chapter 3. Data Collection

In this chapter we present a description of the data collection and analytic approach of the RAISE Scale-up study. In addition to measuring the study's intermediate outcomes, these data allowed us investigate the transfer of responsibility for and ownership of the RAISE initiative from the Reading Apprenticeship developers to the local level, which is represented by movement through the stages of our logic model.

We collected multiple sources of qualitative and quantitate data. We have observed and documented key project activities; tracked the numbers of schools, teachers, and students served by the initiative; and surveyed participating teachers (three times a year during each year of implementation) and school administrators (annually). Through the surveys, we were able to measure general uptake of the RAISE activities, the extent to which they help districts and schools buy into the Reading Apprenticeship framework and build capacity, and how they take ownership of the initiative. We also conducted case studies of four schools in one state to gather a more in-depth understanding of how the scale-up process evolves, as well as to understand the contextual factors that are associated with the process. Data collection for the case studies included surveys, interviews, focus groups, informal classroom observations, and site visits with various stakeholders.

OBSERVATION AND DOCUMENTATION OF PROJECT ACTIVITIES

Throughout the study, researchers conducted informal observations of the RAISE professional development sessions and attended program team meetings and retreats. These qualitative data were used to gain an understanding of the Reading Apprenticeship framework and expectations for teacher and school implementation, and to inform the development of the study logic model, data collection activities, and analysis plans.

Researchers also attended key program team design meetings and meetings/retreats with the State Coordinators in order to understand program design intentions, and progress toward building capacity, engaging participants, and developing sustainability plans. During program team meetings, researchers also provided SLI and the State Coordinators with formative feedback and preliminary findings from study surveys about program implementation, the extent to which they were meeting project goals and recruitment targets, and to inform project decisions moving forward.

TRACKING PARTICIPATION

Researchers collected and entered all professional development attendance records from the 10 day RAISE Institute into a "participant tracker" database in order to track participation across states and subject areas. Researchers also tracked which schools and teachers agreed to complete study surveys and, less systematically, if/when schools or teachers were no longer participating in RAISE (either because they left the school or were no longer implementing Reading Apprenticeship). Data were updated during each year of the study as researchers obtained new information, which was primarily gathered through survey follow-up or direct communication with teachers or administrators. Data were linked across years to track the expansion and participation of states, districts, and schools.

Researchers collected school and demographic data for each participating school and district from the Common Core of Data (CCD), collected by the National Center for Education Statistics (NCES). Demographic characteristics included: local code (rural/suburban/urban, etc.), school size (student population, teacher FTEs, student to teacher ratio), grade range of school, student ethnicity, number of students eligible for free or reduced price lunch, school-level fiscal data (total salaries and other expenditures).

SURVEYS

During Years 1-3, researchers collected survey data from all RAISE teachers and administrators in the Scale-up schools who agreed to participate in the study and were still in a study school. Teachers were asked to complete three surveys per year (in the fall, winter, and spring) and administrators were asked to complete a survey in the spring of each year. The surveys were designed to collect information on the uptake of RAISE project activities, levels of buy-in and commitment to RAISE, and sustainability and contextual factors. Given the number of schools and teachers participating in RAISE, the project could not support additional stipends or incentives for teachers or administrators to complete surveys. As we have noted below, survey response rates declined across the study years.

Principal/School Administrator Surveys

Principal surveys were used to gather the school leadership perspective on the RAISE initiative. The administrator surveys were deployed in the spring of each year to consenting administrators who had teachers in either Cohort 1, 2, and/or 3 at their school. Table 3.1 includes the response rates to the administrator survey.

TABLE 3.1. PRINCIPAL SURVEY RESPONSE RATES, BY YEAR

	# of administrators who completed survey	Possible schools (with teachers who were part of Cohort 1, 2 and/or 3)	Response rate (out of total possible schools)
Year 1	39	66	59%
Year 2	82	173	47%
Year 3	73	202	36%

Specific domains measured in the surveys were guided by the logic model and included buy-in, commitment to Reading Apprenticeship, and sustainability of the initiative beyond the grant funding. The survey included the following domains.

Administrator Background

We collected the following administrator background data.

- Current position at school (e.g. principal vs. curriculum director)

- Years served as administrator overall
- Years served as administrator at current school
- Years served in any position at current school

Uptake of Development Activities

We asked questions regarding recruitment and retention processes to gauge the extent to which these efforts were successful. Specifically, we asked how the administrators heard about the RAISE initiative, why they choose to participate, and whom they contact with questions about RAISE.

While administrators are not required to attend the RAISE professional development or monthly team meetings, they are encouraged to do so in order to support their RAISE teachers. Therefore, we asked administrators if they participated in these activities. Additionally, we asked what types of support for Reading Apprenticeship implementation are provided to teachers by administrators at their school, and what kinds of discussions administrators have with their teachers about RAISE.

Finally, in order to gauge variability in resources/capacity of the leadership at each school involved in RAISE, we asked the role of the primary administrator who oversees RAISE (e.g. principal, literacy/curriculum director) and the administrator's level of involvement with the RAISE initiative.

Buy-in and Shift in Ownership

In order to gauge the level of buy-in of the school administrators, we asked about their level of commitment to RAISE and their agreement with the statement that Reading Apprenticeship is an appropriate framework for literacy instruction at the school and will increase student achievement.

An early indicator of "shift in reform ownership" is if the local level (i.e. participating district/LEA, schools, teachers) takes more responsibility for not only disseminating information about the initiative, but also recruiting additional schools and/or teachers to join the reform. Therefore, we asked the administrators several questions about if/why they had recommended RAISE to others. Additionally, in order for administrators to appropriately "use reform-centered ideas or structures in schools or district decision making," they must have a strong foundation of the reform-centered knowledge (Coburn, 2003). Therefore, we asked the administrators to rate their own level of understanding of the Reading Apprenticeship model.

Sustainability and Contextual Factors

In order to gain an understanding of specific sustainability issues, we asked administrators about challenges of sustaining RAISE in their school, to describe any district policy constraints that made the implementation of the RAISE initiative difficult, and if they believe RAISE would continue in their school without federal funding. We also asked about their knowledge, access, and likelihood of using several different supports to sustain RAISE in their school.

Sternberg et al. (2011) cite several contextual factors that are important for successful scale-up and sustainability, including a stable school/district working environment and administrators who encourage new practices/initiatives. Therefore, we asked administrators several questions about the stability of the school environment, including teacher and administrator retention rates and available

resources/data to inform decisions, and we asked how administrators generally feel about teachers implementing new instructional strategies.

Teacher Surveys

Teacher surveys were used to measure the uptake of RAISE activities, supports and barriers to implementation and capacity-building, and contextual factors. As mentioned above, teachers did not receive an additional incentive or stipend for participating in the study or completing study surveys. Following the deployment of each survey, researchers followed up via email, fax, and phone with all non-respondents. However, as noted in Table 3.2, teacher survey response rates declined over time.

TABLE 3.2. TEACHER SURVEY RESPONSE RATES, BY COHORT AND YEAR

Cohort	Sample	2011-12	2012-13	2013-14
Cohort 1	Schools with at least one teacher response (*n*=66)	92%	80-82%	71-73%
	Teachers (*n*=397)	66-75%	40-50%	29-34%
Cohort 2	Schools with at least one teacher response (*n*=132)		82-83%	69-72%
	Teachers (*n*=559)		62-68%	39-50%
Cohort 3	Schools with at least one teacher response (*n*=79)			66-68%
	Teachers (*n*=387)			39-46%

Note. We have provided the range of response rates across the three surveys each year.

The majority of the surveys included multiple choice or ordinal/interval scale questions lending to more efficient coding and analysis. The surveys included the following domains.

Teacher Background and Number of Students Taught per Subject

To help describe the context of implementation and/or to see if there are differences in our expected outcomes based on this measure, we asked teachers how many years of classroom teaching experience they have. Since there were several schools that had implemented Reading Apprenticeship prior to RAISE, we asked teachers how many hours of previous Reading Apprenticeship training they had received in order to examine differences in scale-up based on prior experience.

In order to track the number of students reached by RAISE, we asked the RAISE-trained teachers how many course sections and students they taught each year, in each of the focal subject areas.

Uptake of Development Activities

A majority of the survey questions centered on the development activities. Many of these questions were repeated across the three surveys in order to examine differences/changes in implementation during the school year. We asked questions about the uptake of the following development activities.

- Attendance at and preparedness and effectiveness of the RAISE Institutes
- Attendance at, helpfulness of, and activities that took place during the teacher leader meetings
- Attendance at, helpfulness of, and activities that took place during the monthly RAISE school team meetings
- Use and helpfulness of the Thinking Aloud website, which was designed to support online teacher collaboration and provide resources related to Reading Apprenticeship implementation
- Availability, types, and helpfulness of support for implementing Reading Apprenticeship in classrooms

We also asked teachers about their reasons for choosing to participate in RAISE and to rate the overall organization of the RAISE initiative. Additionally, we asked how often they used and how confident they are using Reading Apprenticeship pedagogical practices in their classroom, and if they had enough time to plan Reading Apprenticeship lessons. Finally, we asked a series of questions about the frequency and reasons for engaging in both formally and informally established collaboration with other teachers about RAISE implementation.

Building Capacity and Buy-in

In the first and third surveys, we asked teachers which activities were most effective in building their capacity to implement Reading Apprenticeship in their classroom. In order to gauge the level of teacher buy-in, we asked about their level of commitment to RAISE and their agreement with the statement that Reading Apprenticeship is an appropriate framework for literacy instruction at their school and will increase student achievement. We also asked teachers the extent to which they believed students improved in several academic and behavioral outcomes.

Shift in Ownership

The second survey focused on assessing the extent to which teachers were taking ownership of the RAISE initiative. Similar to what we asked administrators, we asked teachers to rate their own level of understanding of the Reading Apprenticeship model and if they had or would recommended RAISE to others. We also asked if they had or would consider taking on a RAISE-related teacher leadership position (e.g. teacher leader for school team, RAISE professional development facilitator). Additionally, we asked teachers about their level of responsibility/sense of agency for the success of RAISE at their school.

Sustainability and Contextual Factors

The third survey focused on sustainability and the contextual factors that may hinder or support successful scale-up. Specifically, we asked about the beneficial aspects of participating in RAISE; the challenges of implementing Reading Apprenticeship; how well RAISE aligned with the instructional goals, rigor, and needs of the students in their class/school; and teachers' plans to use the Reading Apprenticeship framework to inform instruction in their classroom in the next school year. As we did with the administrators, we asked the teachers to describe any school or district policy constraints that made the implementation of the RAISE initiative difficult, and if they believe RAISE would continue in their school without federal funding. We also asked teachers which supports they used for implementing Reading Apprenticeship following the professional development.

CASE STUDIES

The specific aim of the case studies was to capture a concrete picture of how contextual factors influence the support for implementation of Reading Apprenticeship as it continues to be scaled. The cases studies followed four schools in Michigan during the 2012-13 and 2013-14 school years. During the 2012-13 school year, we collected data from two district personnel (who were best positioned to speak about RAISE efforts and other literacy initiatives in their district), four head principals, four Reading Apprenticeship-trained instructional support staff, and 19 teachers. Participants were asked to participate in at least one of the following activities: teacher and administrator interviews, teacher focus groups, classroom observations, and/or open-ended surveys. During the 2013- 14 school year, the researcher collected interview data from 10 participants across the four schools who had participated in the research activities from the prior year (assuming they were still at the school). In total, there were three interviews for three of the case-study schools, and one interview for the fourth case-study school. All interviews were recorded and transcribed for analysis. Case study teachers were offered $250 per year for their participation in data collection activities.

ANALYTIC APPROACHES

Throughout the project, we have used a mixed methods approach, with both quantitative analyses and a qualitative strategy of inquiry. In the early stages of the study, we conducted teacher or administrator-level analysis of survey data and reported descriptive statistics on the uptake of the initiative, and early indications of enthusiasm and challenges. All subsequent analyses were conducted at the school-level in order to understand scale-up processes of the school team. Specific analytic methods are included in each of the following chapters.

Chapter 4: Early Indications of Enthusiasm and Initial Challenges

In the early stages of our study, we provided SLI with reports about initial RAISE implementation, levels of commitment, and challenges that teachers and administrators cited in surveys during their first year of implementation. This information was used for formative purposes and to provide insights into future analyses. In this chapter we provide a summary of those results, focusing first on Cohort 1 teacher survey data from the 2011-12 school year, using teacher-level descriptive statistics. Then, we present results from the 2011-12 Cohort 1 administrator survey, which provide important context and perspective from school leaders about the commitment to RAISE by the end of the first year of the initiative. Where appropriate, we make descriptive comparisons with Cohort 2 and 3 results in teachers' and administrators' first year of implementation. We present the findings of general participation and uptake of RAISE project activities; the extent to which teachers and school administrators began to report commitment, buy-in, and capacity to implement RAISE in their schools; and potential supports and barriers to sustainability.

Chapter 3 provides response rates for each of the surveys reported in this chapter. These results represent teacher and administrator self-reports from the first year of RAISE implementation. The data are from 59% of the Cohort 1 school administrators and 66-75% of the Cohort 1 teachers (depending on the survey). We do not know the implementation, commitment, or buy-in levels of those participants that did not consent to be part of the evaluation or complete the data collection activities.

INITIAL IMPRESSIONS FROM TEACHERS

Uptake of RAISE Activities and Supports for Reading Apprenticeship Implementation

To measure the effectiveness of the RAISE professional development institutes, we asked teachers the extent to which they agreed with the following statements after their participation in the Summer 5-Day Institute.

- Helped me collaborate with my colleagues to better understand the needs of my students
- Provided me with adequate resources and materials to implement what I learned in the professional development activities
- Led to changes in my classroom teaching practices

Figure 4.1 shows that 82% (n = 241) of the Cohort 1 teachers agreed or strongly agreed that the Summer 5-Day Institute helped them collaborate with their colleagues, 85% (n = 251) agreed or strongly agreed that they were provided with adequate resources and materials, and 90% (n = 265) agreed or strongly agreed that it led to changes in their teaching practices. These high levels of reported effectiveness were consistent with Cohort 2 and 3 teachers' reports about the RAISE Institute.

Results presented in this chapter are reported in more detail in Zacamy, Gray, Jaciw, and Newman, 2013.

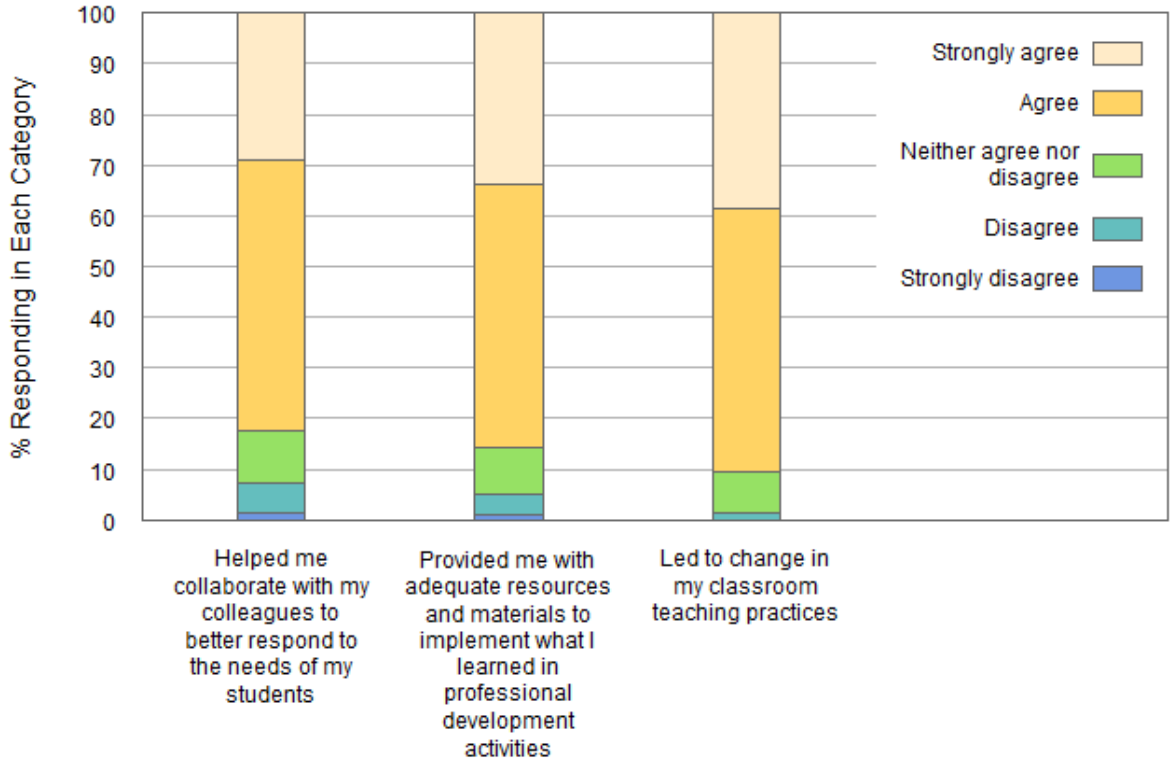


FIGURE 4.1. EFFECTIVENESS OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SUMMER 5-DAY INSTITUTE

Note. For this question, teachers were asked to select the one response option that they felt best answered the question.
n = 293

On each of the annual three surveys, we asked teachers if they had attended monthly RAISE school team meetings during a given period. In the first year of the study, 97% (*n* = 261) of the Cohort 1 teachers attended a monthly meeting between the beginning of the school year and January, 86% (*n* = 212) attended a monthly meeting in February and/or March, and 73% (*n* = 191) attended a monthly meeting in April and/or May. We saw a similar decline in attendance at monthly meetings during the first year of implementation for Cohort 2 and 3. The primary reasons for not attending the monthly meeting were that teachers had other obligations at the time the team meetings were offered or that the meetings were not offered at their school. For those who did attend meetings, we asked what activities took place. The following responses were most frequently selected.

- Sharing successful lessons (85% [*n* = 162])
- Problem-solving (67% [*n* = 127])
- Discussing problematic lessons (66% [*n* = 125])

On each survey, we also asked teachers if they had received any support for implementing Reading Apprenticeship in their classroom, outside of the RAISE monthly team meetings. On each of the three

surveys in their first year of implementation, roughly half of the teachers reported that they had received support for Reading Apprenticeship instruction. Resources, coaching/mentoring, and model lessons were the three most selected types of support received. Additionally, we asked teachers from whom they received support and the most selected responses were:

- Other RAISE teachers (65-75%, depending on survey)
- RAISE teacher leader (51-53% depending on survey)
- Principal or other administrator (17-24% depending on survey)

On the third survey, we asked teachers how often they used the Reading Apprenticeship pedagogical practices in their classroom, on average, during the 2011-2012 school year. Figure 4.2 shows that 67% (n = 174) of the teachers said they use these practices at least a few times a week, with 27% (n = 70) using them in each lesson. While the Reading Apprenticeship pedagogical practices are expected to be integrated throughout each lesson in which text and reading play a central role, it may take teachers several years to learn, become comfortable with, and fully incorporate new instructional strategies. Within the first year of RAISE (and prior to teachers receiving the full professional development), over two-thirds of the teachers reported using Reading Apprenticeship to inform their instruction on at least a weekly basis.

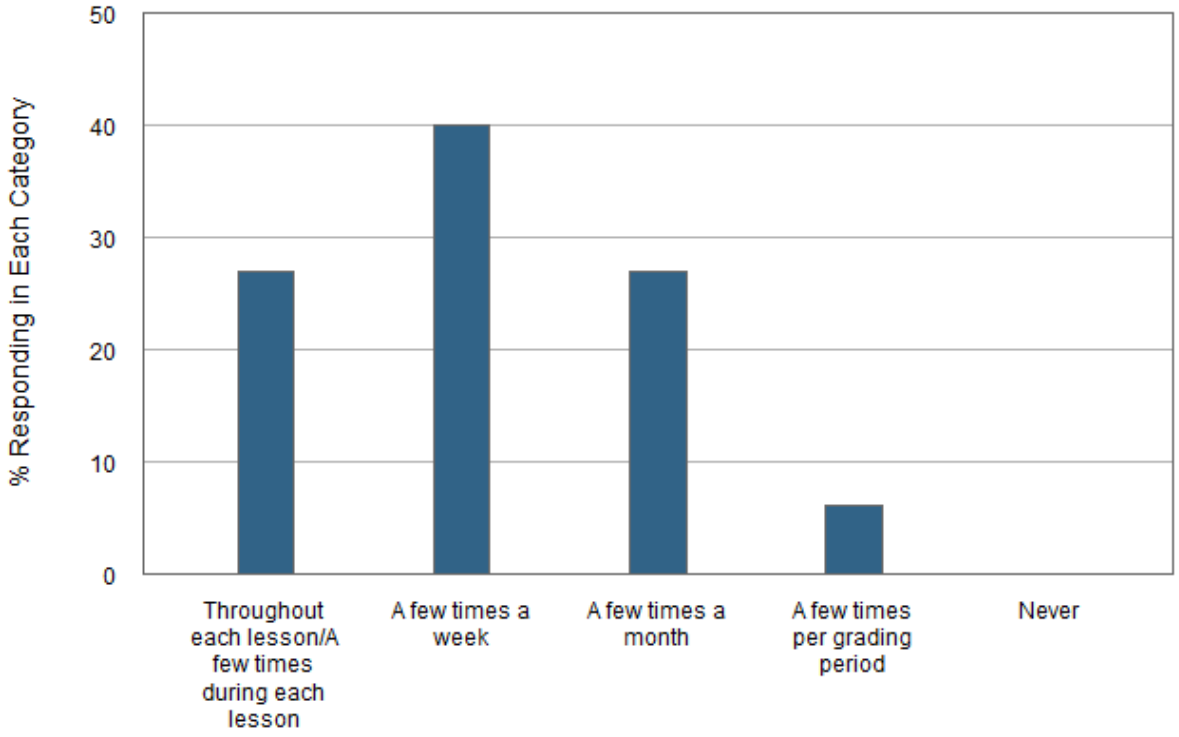


FIGURE 4.2. AVERAGE USE OF READING APPRENTICESHIP PEDAGOGICAL PRACTICES

Note. For this question, teachers were asked to select the one response option that they felt best answered the question.
n = 261

Building Capacity, Commitment and Buy-in

Several RAISE-related activities/resources are designed to help teachers build capacity to implement Reading Apprenticeship practices, including collaboration with teachers, attendance at the RAISE Institute, monthly team meetings, support from teacher leaders, support from administrators, and the *Thinking Aloud* website.[4] Therefore, on the third survey, we asked teachers which of these activities were most effective in building their capacity to implement Reading Apprenticeship in their classrooms. Figure 4.3 shows that a majority of the Cohort 1 teachers (60% [n = 165]) said that collaboration with other teachers was most effective at building their capacity. Teachers in Cohort 2 and 3 also reported that collaboration with other teachers was most effective at building their capacity. It is important to point out that less than 10% of the Cohort 1 teachers identified the monthly team meetings as the most effective activity, indicating that the collaboration may be occurring more informally or at other meetings/times. With the formal RAISE professional development opportunities ending after the first year of implementation, it is important for schools leaders to provide logistical supports for RAISE collaboration, such as having similar prep periods, common lunch times, and designating in-service days/times for Reading Apprenticeship collaboration. Teachers must also be committed to continuing to spend their time collaborating about RAISE, and develop social networks to deepen their knowledge and capacity to implement Reading Apprenticeship practices.

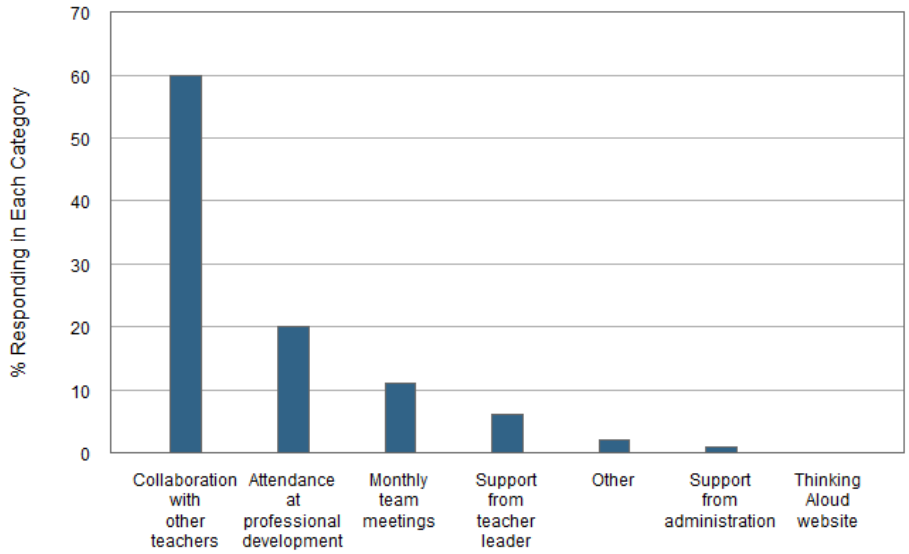


FIGURE 4.3. MOST EFFECTIVE ACTIVITIES AT BUILDING CAPACITY TO IMPLEMENT READING APPRENTICESHIP

Note. For this question, teachers were asked to select the one response option that they felt best answered the question. *n* = 259

[4] As noted in the initial description of the innovation for RAISE, program developers initially conceived and supported this site for teacher exchange, but withdrew support after seeing a very low rate of participation, even with significant staff support.

On the third survey, we asked teachers how committed they were to making Reading Apprenticeship work in their classrooms and in their schools. Figure 4.4 shows that 82% (n = 215) of the Cohort 1 teachers responded that they were either fully committed or fairly committed to making Reading Apprenticeship work in their classrooms, with 14% (n = 36) willing to give it a chance. Seventy percent (n = 176) of teachers responded that they were either fully committed or fairly committed to making Reading Apprenticeship work at their schools, with 24% (n = 63) willing to give it a chance. We found similar levels of commitment for Cohort 2 and 3 teachers in their first year of implementation.

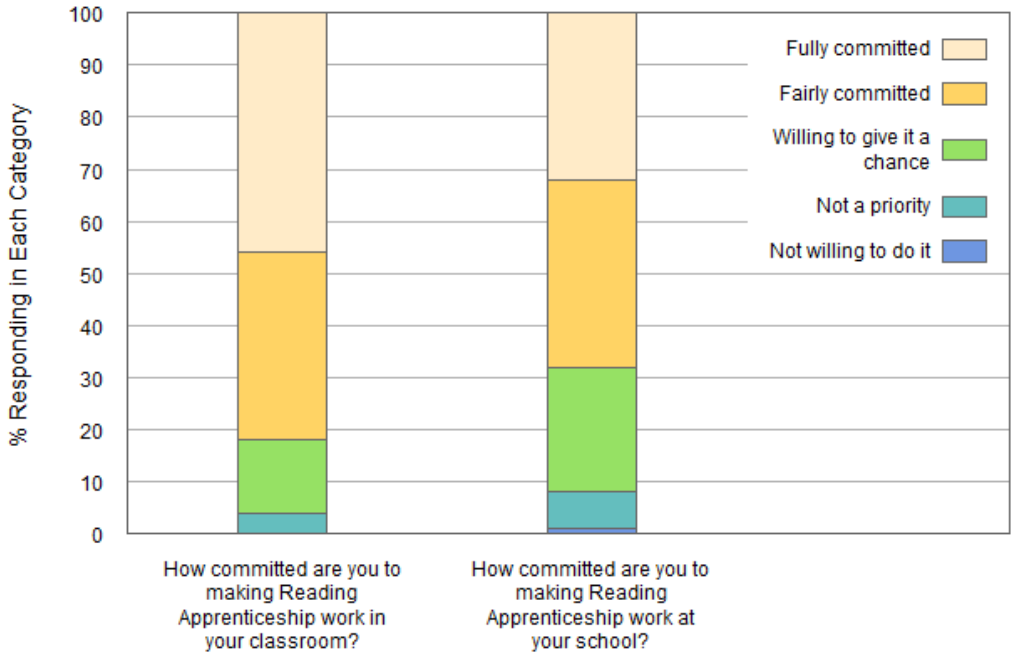


FIGURE 4.4. TEACHER COMMITMENT TO READING APPRENTICESHIP

Note. For this question, teachers were asked to select the one response option that they felt best answered the question.
n = 261 (classroom); n = 259 (school)

On the third survey, we also asked teachers the extent to which they agreed with the statements that Reading Apprenticeship is an appropriate strategy for literacy instruction and a means of improving student achievement. As shown in Figure 4.5, 92% (n = 239) of the Cohort 1 teachers said they agreed or strongly agreed with the statement that "Reading Apprenticeship is an appropriate framework for literacy instruction in my classroom," and 90% (n = 236) agreed or strongly agreed that "The implementation of Reading Apprenticeship will improve student achievement at my class/school/district." As with the levels of commitment, we found high levels of buy-in for Cohort 2 and 3 teachers in their first year of implementation.

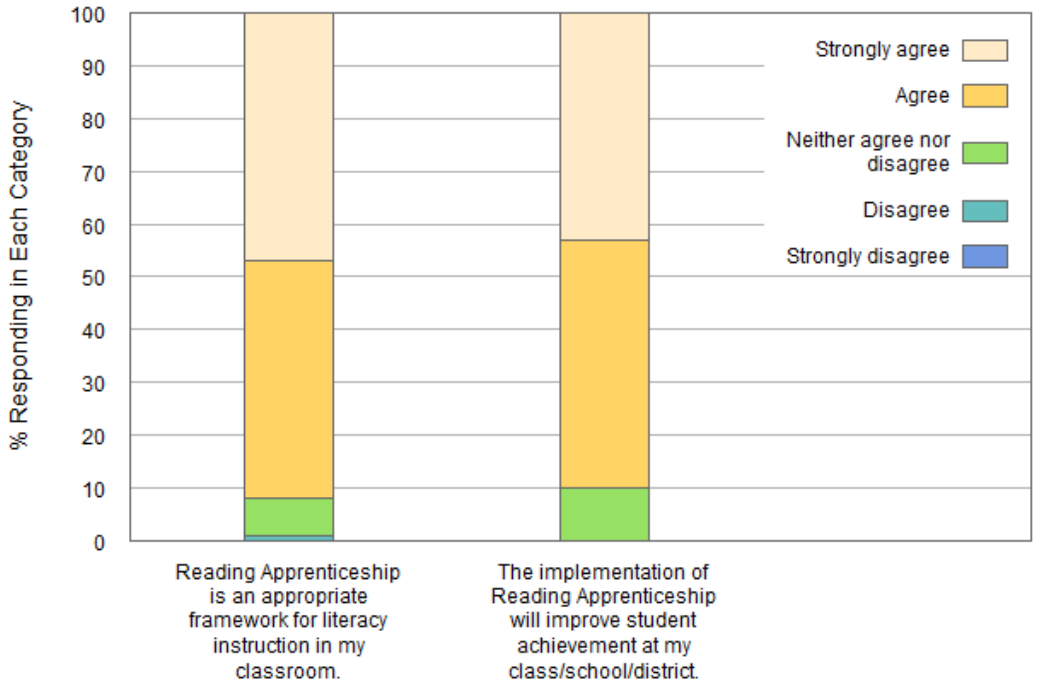


FIGURE 4.5. TEACHER AGREEMENT WITH BUY-IN STATEMENTS

Note. For this question, teachers were asked to select the one response option that they felt best answered the question.
n = 261

In response to a question on Survey 2, 88% (*n* = 245) of Cohort 1 teachers said that they strongly agreed or agreed with the following statement: "My school would benefit if more teachers participated in RAISE." Only three teachers (out of 277) said that they disagreed or strongly disagreed with that statement. This result suggests that participating teachers recognized the importance of building a team and having consistency in literacy instruction in order for RAISE to be implemented to its full potential.

Potential Support and Barriers to Sustainability

To gauge which factors may support successful scale-up of RAISE, we asked teachers which aspect of participating in RAISE they considered most beneficial. Table 4.1 shows the responses ordered by most to least selected. The following responses were most frequently selected.

- My students' literacy skills improved (22% [*n* = 57])
- My literacy instruction improved (19% [*n* = 50])

Both of these address measurable changes. While these changes were not directly measured in the RAISE Scale-up study, it is notable that teachers saw the value for such improvements. No teachers responded that there were no beneficial aspects of participating in RAISE.

TABLE 4.1. MOST BENEFICIAL ASPECT OF PARTICIPATING IN READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION (RAISE)

Benefits of participation	% of teachers who selected challenge
My students' literacy skills improved	22%
My literacy instruction improved	19%
Opportunities to collaborate with other RAISE trained teachers	14%
My students were more engaged	14%
My understanding of literacy (how readers make sense of text) improved	13%
The RAISE professional development institutes	6%
Reading Apprenticeship is aligned with existing materials and instructional approaches at my school	3%
Reading Apprenticeship is aligned with new or recently adopted standards	3%
My content area instruction improved	3%
My understanding of my content area improved	<1%
Other	2%
There are no beneficial aspects of participating in RAISE	0%

Note. For this question, teachers were asked to select the one response option that they felt best answered the question. Due to the rounding of decimals, percentages may not add up to 100%.

n = 260

To gauge which factors may hinder successful scale-up of RAISE, we asked teachers what challenges they faced in implementing Reading Apprenticeship during their first year of implementation. (Teachers were able to check all that apply.) Table 4.2 shows Cohort 1 teachers' responses ordered by most to least selected.

TABLE 4.2. CHALLENGES FACED IN IMPLEMENTING READING APPRENTICESHIP

Challenges in implementing Reading Apprenticeship	% of teachers who selected challenge
Competing priorities	54%
Student behavior	41%
Student ability	36%
Misalignment between Reading Apprenticeship and required curriculum	17%
Insufficient school administrator support	13%
Insufficient district support	13%
Insufficient understanding of how to implement Reading Apprenticeship in class	13%
Reading Apprenticeship is too much work	12%
Insufficient materials	11%
Insufficient parent support	10%
Insufficient training on Reading Apprenticeship	3%
Other	16%
I have not faced any challenges implementing Reading Apprenticeship	7%

Note. For this question, teachers were asked to select all response options that applied.
n = 261

In the surveys that were deployed to teachers in Cohorts 2 and 3, we added the following two additional challenges as response options to this question: Reading Apprenticeship slowed down the pace of my instruction and Insufficient time to collaborate. The three most selected responses for Cohort 2 and 3 teachers were:

- Reading Apprenticeship slowed down the pace of my instruction (48% [*n* = 168] of Cohort 2 teachers; 28% [*n* = 55] of Cohort 3 teachers)
- Competing priorities (44% [*n* = 153] of Cohort 2 teachers; 30% [*n* = 59] of Cohort 3 teachers)
- Insufficient time to collaborate (42% [*n* = 147] of Cohort 2 teachers; 26% [*n* = 51] of Cohort 3 teachers)

Competing school and district priorities has been well documented in the literature as a primary challenge to sustainability (Coburn, 2003), and teachers in all three cohorts cited competing priorities as

a primary challenge to sustaining Reading Apprenticeship. They also cited several classroom factors as hindrances, such as pacing of instruction, student behavior, and student abilities.[5] To a lesser extent, teachers reported organizational and time constraints as challenges to implementing Reading Apprenticeship. These results reflect the real challenges in implementing an instructional framework like Reading Apprenticeship but can be viewed in relation to the level of commitment and buy-in expressed by teachers. Additionally, in spite of the reported challenges, 91% (n = 238) of the Cohort 1 teachers reported that they planned to continue to use the Reading Apprenticeship framework to inform their instruction during the next school year (2012-2013).

INITIAL IMPRESSIONS FROM SCHOOL ADMINISTRATORS

Reasons for Participation

While there are several reasons why districts and schools may choose to join RAISE, we asked administrators to select the primary factors that led to their school's participation. As shown in Table 4.3, here are the most selected options by Cohort 1 administrators.

- There is prior research showing that Reading Apprenticeship is effective at improving student achievement (54% [n = 21]).
- The pedagogy corresponds to the literacy practices advocated by my school (45% [n = 18]).
- It was highly recommended to me (other than by teachers) (41% [n = 16]).

Administrators in Cohort 2 and 3 schools selected the same top three reasons for participation. This result suggests that SLI and the State Coordinators continue to focus on the strong research base behind Reading Apprenticeship and alignment with current practices during their recruitment efforts.

TABLE 4.3. PRIMARY FACTOR(S) THAT LED TO SCHOOL PARTICIPATION IN READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION

Reasons for participation	% of administrators who selected challenge
There is prior research showing that Reading Apprenticeship is effective at improving student achievement	54%
The pedagogy corresponds to the literacy practices advocated by my school	46%
It was highly recommended to me (other than teachers)	41%
Our teachers wanted to participate	31%
It is free professional development for our teachers	21%
It is aligned to the Common Core Standards	13%

[5] The Cohort 2 and 3 teachers who selected student ability were asked to describe which student abilities made sustaining Reading Apprenticeship a challenge. Three general themes emerged in their responses: Student motivation (e.g. engagement in school, in general), varied (reading) abilities in the class, students with "very" low reading/comprehension skills.

TABLE 4.3. PRIMARY FACTOR(S) THAT LED TO SCHOOL PARTICIPATION IN READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION

Reasons for participation	% of administrators who selected challenge
It was required by the district	8%
Other	18%
I am not aware of the reason(s) our school decided to participate	3%

Note. While administrators were able to select all response options that applied, we asked them to select no more than three of the options.

n = 39

Buy-in and Increased Ownership

We asked administrators the same questions about levels of commitment and buy-in to RAISE as we asked their teachers. Ninety-five percent (*n* = 37) of the Cohort 1 administrators responded that they were either fully committed or fairly committed to making RAISE work at their schools. Only 5% (*n* = 2) responded that they were willing to give RAISE a chance. None of the administrators selected "Not a priority," "Not willing to do it," or "I don't know enough about Reading Apprenticeship or the RAISE initiative to respond." It is important to point out that this survey was deployed in May, giving the administrators a year of RAISE implementation in their schools to establish their commitment level.

As shown in Figure 4.6, 67% (*n* = 26) of administrators said they strongly agreed with the statement that "Reading Apprenticeship is an appropriate framework for literacy instruction at my school" and 74% (*n* = 29) strongly agreed that "The implementation of Reading Apprenticeship will improve student achievement at my school." These high levels of commitment and buy-in were echoed by the Cohort 2 and 3 administrators in their first year of implementation.

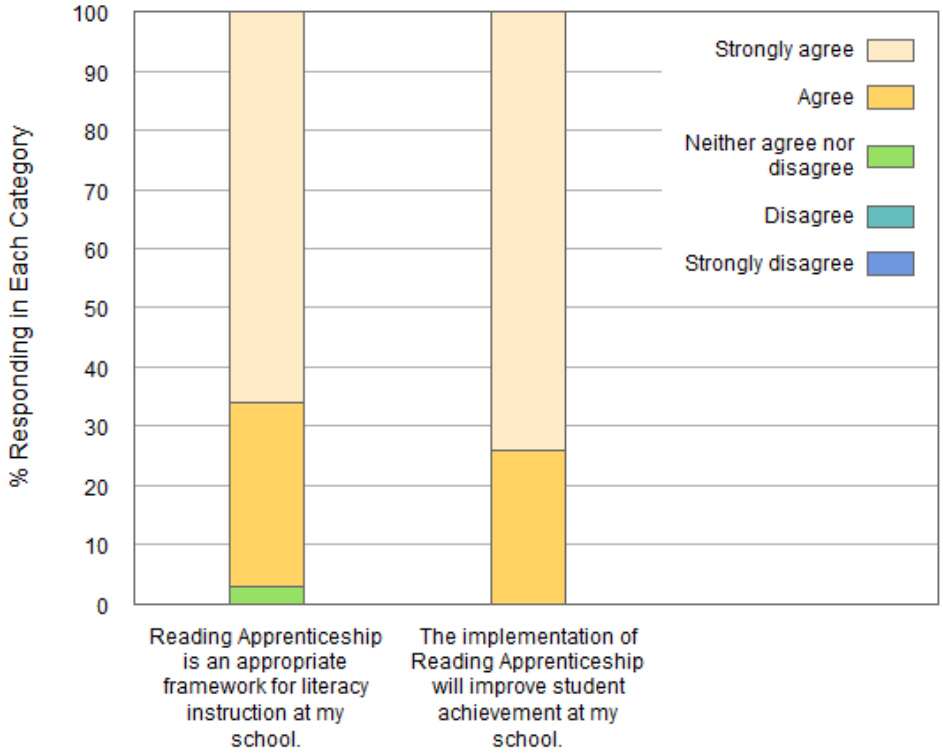


FIGURE 4.6. ADMINISTRATOR AGREEMENT WITH BUY-IN STATEMENTS

Note. For this question, administrators were asked to select the one response option that they felt best answered the question. *n* = 39

Recommending RAISE to Others

One expected indicator of a "shift in ownership" and a factor contributing to spread is that the recruitment and retention processes are transferred to the local level, with principals and teachers recruiting or recommending that other schools/teachers join RAISE in subsequent years. Eight-five percent (*n* = 33) of the Cohort 1 administrators responded that they had recommended joining RAISE to other teachers at their school, and two responded that they had not because all teachers in the appropriate subject areas at their schools were already participating in RAISE, indicating that this transfer of responsibility was already occurring. We asked the four administrators who responded that they had not recommended RAISE to the non-participating teachers at their school to indicate why, and all selected "I plan to, but haven't done it yet." Additionally, two administrators responded that someone else at their schools is talking to them about joining RAISE, and one administrator selected "I don't think there is enough instructional support available." Additionally, 67% (*n* = 26) of the administrators said that they had recommended joining RAISE to other school personnel (e.g. administrators, instructional coaches, or teachers from other schools).

Potential Support and Barriers to Sustainability

We asked several questions related to the specific contextual factors that may hinder or support successful scale-up and sustainability. Table 4.4 shows how the Cohort 1 administrators responded to the question asking them what they thought the biggest challenges to sustaining RAISE in their school long term would be. The following three response options were most selected.

- Competing initiatives (56% [*n* = 22])
- Budget constraints (33% [*n* = 13])
- Misalignment between RAISE and teacher preferences (67% [*n* = 26])[6]

Additionally, 23% (*n* = 9) of the administrators responded that they did not think there would be any challenges to sustaining RAISE in their school long term. Cohort 2 and 3 administrators responded in similar ways, with the same three primary challenges.

TABLE 4.4. CHALLENGES TO SUSTAINING READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION (RAISE) IN SCHOOL LONG TERM

Challenges to sustaining RAISE	% of administrators who selected challenge
Competing initiatives	56%
Budget constraints	33%
Misalignment between RAISE and teacher preferences	23%
Insufficient district support	13%
Teacher turnover	8%
Administrative turnover	3%
Other	7%
I don't think there would be any challenges to sustaining RAISE in our school long term	23%

Note. While administrators were able to select all response options that applied, we asked them to select no more than three of the options.

n = 39

With the ongoing concern of funding and budget constraints, we asked if the administrators thought that RAISE would continue in their schools without the i3 federal funding. Across the four states, 38% (*n* = 15) of the Cohort 1 administrators said "Yes," 26% (*n* = 10) said "No," and 36% (*n* = 14) said "I don't know."

[6] There were three additional answer options to this question that are not shown in this table. No administrators selected "Reading Apprenticeship is too difficult for our students," "Misalignment between RAISE and district literacy policies," or "I don't know enough about the RAISE initiative to respond."

POSITIVE IMPRESSIONS FROM THE FIRST YEAR OF IMPLEMENTATION

Findings from teacher and administrator survey data from the first year of the study (2011-12), suggested high levels of buy-in and commitment from the initial cohort of RAISE Scale-up participants. Teachers were integrating Reading Apprenticeship into their classroom instruction and collaborating with one another to improve their practices. While attendance at the RAISE monthly meetings declined during the first year, teachers cited collaboration and support from other RAISE teachers as the most effective means in building their capacity to implement Reading Apprenticeship in their classrooms. Administrators were looking to RAISE as a means to improve student achievement and almost all who had responded to the surveys had recommended joining RAISE to other teachers in their schools. We found similar trends for Cohort 2 and 3 teachers and administrators.

While initial impressions of RAISE were positive, slightly more than half of the teachers and administrators identified competing initiatives and priorities as a primary challenge to implementing and sustaining RAISE long term. It is worth contextualizing the RAISE initiative within a period of significant change in state and national curriculum standards and associated assessments. Based on reported challenges to implementation, we encouraged SLI and the State Coordinators to continue to work with the local level stakeholders to make productive connections between RAISE goals and existing and new initiatives. That is, to provide support to demonstrate how adopting RAISE could be a beneficial mechanism through which they could meet state and district level mandated requirements, rather than feeling overwhelmed with transitions and "one more initiative." We continued to investigate whether the high levels of implementation and commitment continued over time or if we began to see more variability in the scale-up process for certain schools.

Chapter 5: Implementation and Commitment Over Three Years

While the initial look at the survey data considered survey responses of individual teachers and principals, in the subsequent analyses we report in this chapter, we have taken the school as the unit rather than individual participant. At this stage of the work, we considered the processes of interest to be related to the cohesion (or not) of the community developing within the school. In this chapter we examine the relationships between implementation of key RAISE activities and commitment to the initiative over time. For formative purposes, this could give us an indication of RAISE activities most associated with intermediate scale-up outcomes related to change in the sense of ownership. This analysis primarily focuses on Cohort 1 schools across three year of implementation; however, we have also provided comparisons of the same measures across Cohorts 1-3.

HYPOTHESES AND METHODS

Following Cohort 1 into their second and third year of implementation (2012-13 and 2013-14), we examined the trends and relationship between the uptake of the RAISE initiative and sustainability. This followed the RAISE Scale-up logic model where we hypothesized that in early stages as teachers deepen Reading Apprenticeship practice and strengthen support ties over time, we would see an increase in buy-in and capacity to implement and disseminate Reading Apprenticeship. Also, we expected that "ownership" of the initiative would begin to be transferred to the local level, which would support sustainability as formal supports from the developers were withdrawn. Therefore, we identified key indicators of participants' uptake of RAISE activities, including if they received support for Reading Apprenticeship instruction, their attendance at the RAISE monthly school meetings, and the extent to which they used Reading Apprenticeship practices. We also identified early indicators of scale-up "success", including teachers' buy-in and commitment levels to Reading Apprenticeship in their classroom and school, and if teachers plan to continue using the framework in the next year. We framed these analyses in terms of the following hypotheses:

1: Teachers in schools with higher levels of collaboration and support will have higher levels of buy-in, commitment, and sustainability of RAISE.

2: Teachers in schools where Reading Apprenticeship practices are used more frequently will have higher levels of buy-in, commitment, and sustainability of RAISE.

To measure these hypotheses, we first identified key indicators of participation in RAISE activities and scale-up outcomes, and mapped these to the RAISE Scale-up logic model. Most of the indicators were measured through three years of teacher survey data, for a total of up to nine survey occasions. The number of RAISE teachers at each district and school was tracked in a participant database each year. The hypotheses described above reflect the relationships between the indicators of participation in RAISE activities and the scale-up outcomes that are listed in Table 5.1.

A version of this chapter was presented at the 2015 annual meeting of the American Educational Research Association in Chicago, IL (Zacamy, Newman, Lin, & Jaciw, 2015)

TABLE 5.1. INDICATORS OF PARTICIPATION IN READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION (RAISE) ACTIVITIES AND SCALE-UP OUTCOMES

Indicators of participation in RAISE activities and indicators of scale-up outcomes	Logic model component	Year 1 data source	Year 2 data source	Year 3 data source
Indicators of participation in RAISE activities				
Receipt of support for Reading Apprenticeship implementation	Instructional support resource	Survey 1-3	Survey 4-6	Survey 7-9
Attendance at monthly team meetings	Instructional support resource	Survey 1-3	Survey 4-6	Survey 7-9
Average Use of Reading Apprenticeship pedagogical practices	Professional development[a]	Survey 3	Survey 6	Survey 9
Scale-up outcome				
Level of *buy-in* of the Reading Apprenticeship Framework	Buy-in	Survey 1 & 3	Survey 4 & 6	Survey 7 & 9
Level of *commitment* to Reading Apprenticeship	Buy-in	Survey 1 & 3	Survey 4 & 6	Survey 7 & 9
***Sustaining* Reading Apprenticeship practices**	Sustainability	Survey 3	Survey 6	Survey 9

Note. For Cohort 1 Year 1= 2011-12 school year, Year 2 = 2012-13 school year, Year 3 = 2013-14 school year

[a] Average use of Reading Apprenticeship pedagogical practices is an indicator that teachers are using the practice and strategies they learned during the RAISE Institute.

Next, we conducted three types of analyses. All of the analyses were carried out at the school level, using school averages of the indicators of participation in RAISE activities and of the outcome variables:

(1) We estimated the change in school averages of the indicators of participation in RAISE between Year 1 and Year 3.

(2) We estimated the change in school averages of indicators of scale-up outcomes between Year 1 and Year 3.

(3) We measured the association between changes in the indicators of participation in RAISE and changes in the scale-up outcomes. The analyses allow us to draw some hypotheses about the how changes in indicators are related to changes in outcomes. This allows us to test the theory that increases or intensity of the uptake of RAISE activities positively reinforce or increase levels of buy-in, commitment and sustainability Additionally, if we do not find relationships between these measures at this stage, we can focus on other factors (of implementation or contextual) that potentially influence these changes.

As noted in Chapter 3, survey response rates declined across the three years. We continued to send surveys to teachers each year, unless they left the participating school or no longer wanted to participate. But teachers were not offered an incentive to complete the surveys. The decrease in responses primarily happens at the teacher level, although by the third year thirteen schools were lost.

We examined for the possibility of response bias with the concern that teachers who continued to complete surveys were more engaged in RAISE than those who stopped. If this were the case, the survey results would be biased in a positive direction. In general, we found that the commitment levels were lower for teachers who stopped completing surveys during the first year (compared to those who continued to respond), but the responses for those who stopped completing surveys in the second year were similar to those who continued completing surveys until the end. This suggests that the reason for discontinuing participation in the surveys in the later years (during Year 2 or Year 3) may have been more about survey fatigue or teachers leaving study schools, rather than disengagement with RAISE. It is important to keep this in mind when interpreting the results. Additionally, due to the correlational nature of all of the analyses, we cannot infer causality from the results.

FINDINGS OF LONGITUDINAL ANALYSIS

The results in this section are from the first cohort of RAISE teachers and schools from their first through third year of implementation (AY 2011-12, 2012-13, 2013-14). We report trends in changes in participation in RAISE activities and scale-up outcomes across the three years, as well as whether the relationships between these changes in participation are related to changes in the outcomes.

Changes over Time in Indicators of Participation in RAISE Activities

In the following sections, we provide a series of graphs that illustrate school level average responses to each of the survey questions, across Year 1 through Year 3. Each graph shows the survey occasions/time points on the x-axis and the school level responses on the y-axis. The blue dots represent school averages at each response level, and the size of the dots are proportionate to the number of schools at each point (i.e. the bigger the dot, the more schools are represented). We have also indicated the overall sample mean and median with a purple and green circle, respectively.

Receipt of Support for Instruction

In each survey, we asked teachers to report if they had received support for implementing Reading Apprenticeship, other than through the RAISE monthly meetings, during the prior four weeks of instruction (teachers were asked to exclude support from the monthly meetings in their response). As shown in Figure 5.1, we found that during Year 1 there was a decline in teachers reporting that they received support; this decline is statistically significant ($p < .001$). Looking at the mean and median reported levels, however, we see an increase from the end of the first year to the beginning of second year. This resurgence could be due to either the timing of the school year or participation in the last three days of RAISE professional development, which occurred in the summer prior to Year 2 and focused on designing instruction with Reading Apprenticeship integration. This trend, however, does not continue for the rest of Year 2 or Year 3, when there was no further professional development. There are also more schools in Year 2 and 3 where either all teachers reported receiving support or no teachers reported receiving support.

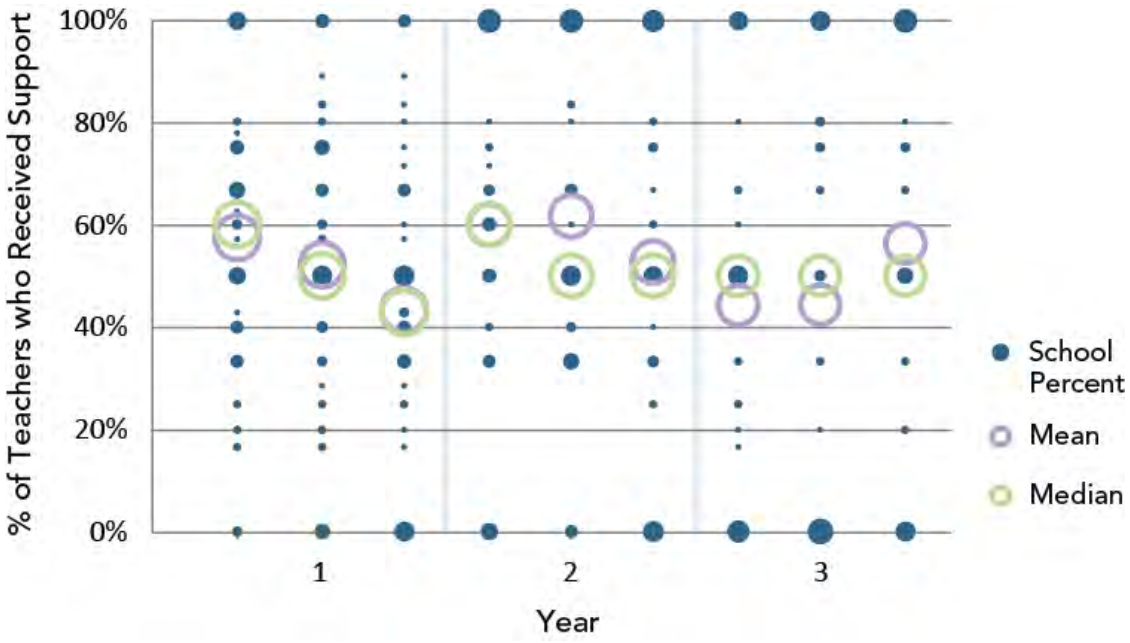


FIGURE 5.1. SUPPORT RECEIVED FOR READING APPRENTICESHIP INSTRUCTION

n = 61 schools in Year 1; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

Attendance at Monthly Meetings

Also on each survey, teachers reported if they had attended a monthly meeting between the prior and current surveys (or between the beginning of the school year and the current survey for survey 1, 4, and 7). As shown in Figure 5.2, we found a significant decrease in the attendance at monthly meetings within each year and across the three years. While in Year 1, nearly all teachers (96%) reported that they attended a monthly meeting in the first survey, by the end of the second and third year, the average dropped to 35% and 21%, respectively. Additionally, most schools (73%) had no teachers reporting that they attended a monthly meeting. This reduction in the average attendance at monthly meetings is statistically significant ($p < .001$). This finding suggests a possible decrease in participation or uptake of the program. It is also possible, however, that teachers were finding other ways of collaborating and supporting their Reading Apprenticeship implementation.

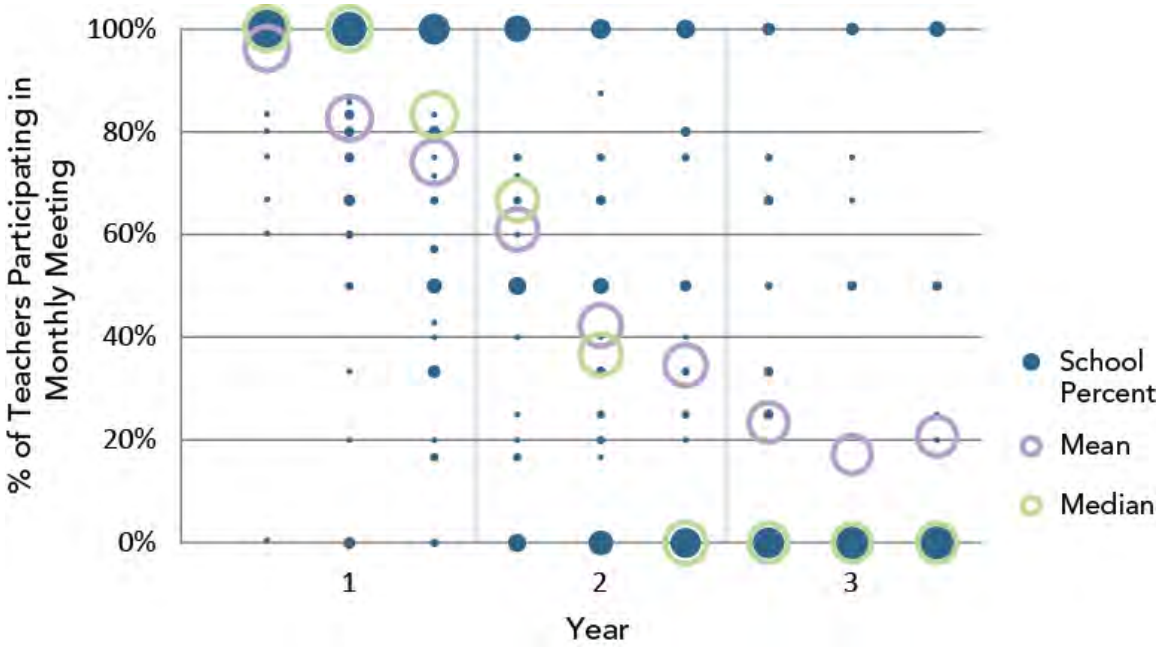


FIGURE 5.2. ATTENDANCE AT MONTHLY MEETINGS

n = 61 schools in Year 1; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

Use of Reading Apprenticeship Practices

While the Reading Apprenticeship pedagogical practices were expected to be integrated throughout instruction, it may take teachers several years to learn, become comfortable with, and fully incorporate new instructional strategies. At the end of each year, we asked teachers how often they used the Reading Apprenticeship pedagogical practices in their classroom, on average, during the school year. Teachers responded on the following scale.

- Throughout each lesson (5)
- A few times during each lesson (4)
- A few times per week (3)
- A few times per month (2)
- A few times per grading period (1)
- Never (0)

At the end of the first year, a majority of teachers reported that they were implementing Reading Apprenticeship practices on at least a weekly basis. While it might be reasonable to expect that the use of the Reading Apprenticeship practices would increase in the second and third year of implementation as teachers are becoming more familiar with the framework, we found that the average reported use is consistent across the years (Figure 5.3). What is interesting here, however, is that the variation in the response increases. In Year 1, we see many of the responses clustered between a 2 and 4 (a few times per month and a few times during each lesson, respectively). By the third year, we see a few schools (*n* = 3),

where the average is "throughout each lesson", and one school reported that they are no longer using Reading Apprenticeship.

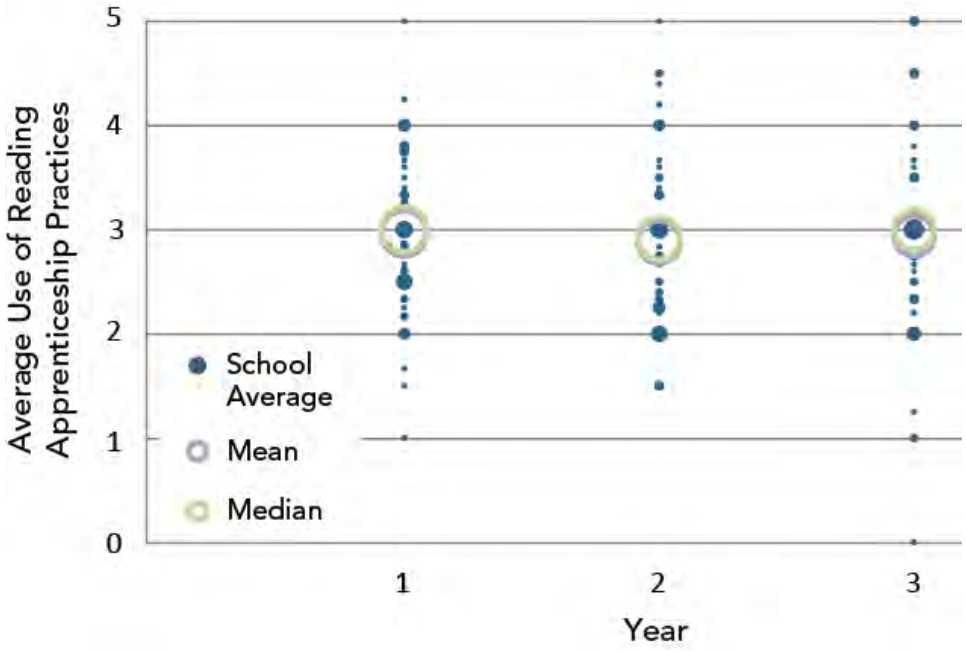


FIGURE 5.3. AVERAGE USE OF READING APPRENTICESHIP PRACTICES

n = 61 schools in Year 1; *n* = 54 schools in Year 2 *n* = 48 schools in Year 3

Changes in Indicators of Scale-up Outcomes

Buy-in to Reading Apprenticeship Framework

As explained in our RAISE Scale-up logic model, we defined buy-in as the belief that Reading Apprenticeship is an appropriate strategy for literacy instruction, and a means of improving student achievement. Therefore, we asked teachers at the beginning and end of each school year to rate their levels of agreement with those statements (5 = Strongly agree; 0 = Strongly disagree). Cohort 1 teachers reported high levels of buy-in, as reflected in a large majority of the teachers agreeing or strongly agreeing with those statements. While teachers' buy-in levels remain high, we did find a slight decrease over time in the school mean levels of agreement with Reading Apprenticeship being an appropriate strategy for classroom instruction (between survey 1 and 9) (p = .048) and as a means of improving student achievement (p = .048) (Figures 5.4 and 5.5).

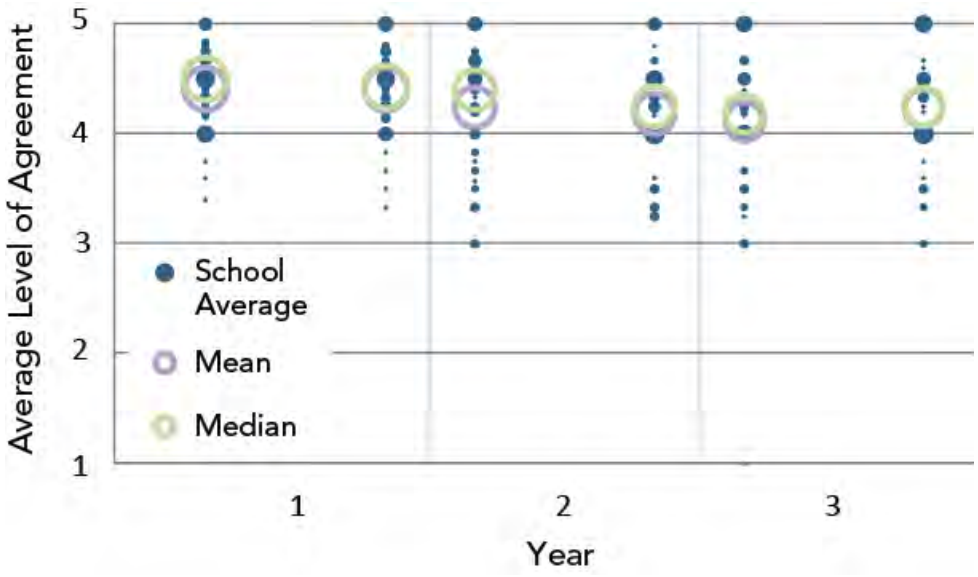


FIGURE 5.4. BUY-IN OF READING APPRENTICESHIP AS APPROPRIATE LITERACY INSTRUCTIONAL STRATEGY FOR CLASSROOM

n = 61 schools in Year 1 ; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

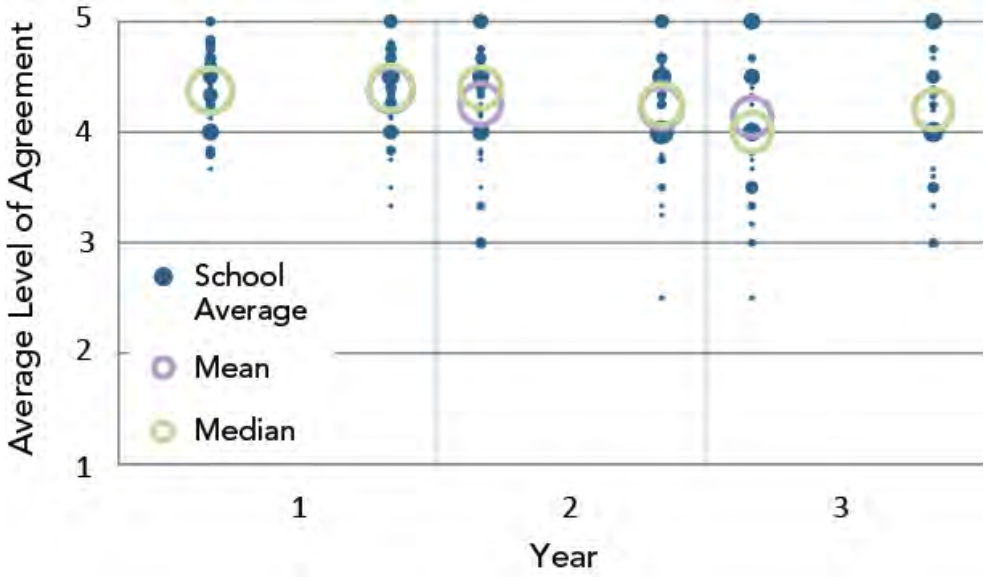


FIGURE 5.5. BUY-IN OF READING APPRENTICESHIP AS MEANS TO IMPROVE STUDENT ACHIEVEMENT

n = 61 schools in Year 1; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

Commitment to Reading Apprenticeship

At the beginning and end of each year, we also asked teachers to report their level of commitment to making Reading Apprenticeship work in their classroom and in their school (5 = Fully; 4 = Fairly; 3 = Willing to give it a chance; 2 = Not a priority; 1 = Not willing to do it).

As we found with levels of buy-in, Cohort 1 schools reported high levels of commitment, with a majority being fully or fairly committed to making Reading Apprenticeship work (Figures 5.7 and 5.8). There is, however, a statistically significant decrease in commitment levels between the first and third year. In Year 1 the decrease in average reported commitment in classrooms and schools was not statistically significant. In Year 2, the picture changed, with a decrease in both average reported commitment in classrooms ($p < .001$) and in schools ($p < .001$). This drop-off in reported commitment in Year 2 was greater with respect to schools than classes ($p < .001$). During Year 3, there is a slight decrease in commitment levels between the beginning and end of the year; this decrease, however, is not statistically significant. Between the first and the third year, we also found that on average teachers reported being more committed to making Reading Apprenticeship work in their classes than in their schools ($p < .001$).

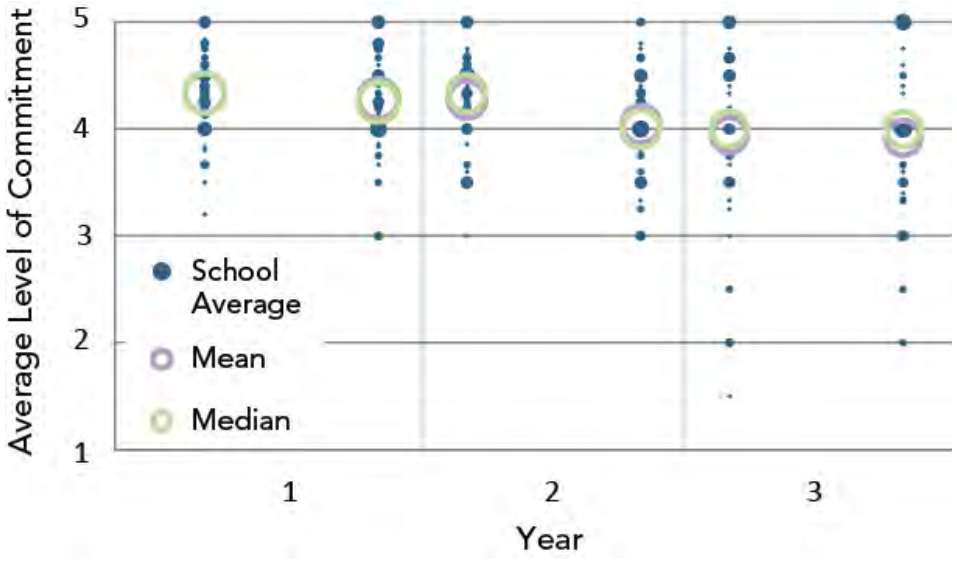


FIGURE 5.6. COMMITMENT TO READING APPRENTICESHIP IN CLASSROOM

n = 61 schools in Year 1 ; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

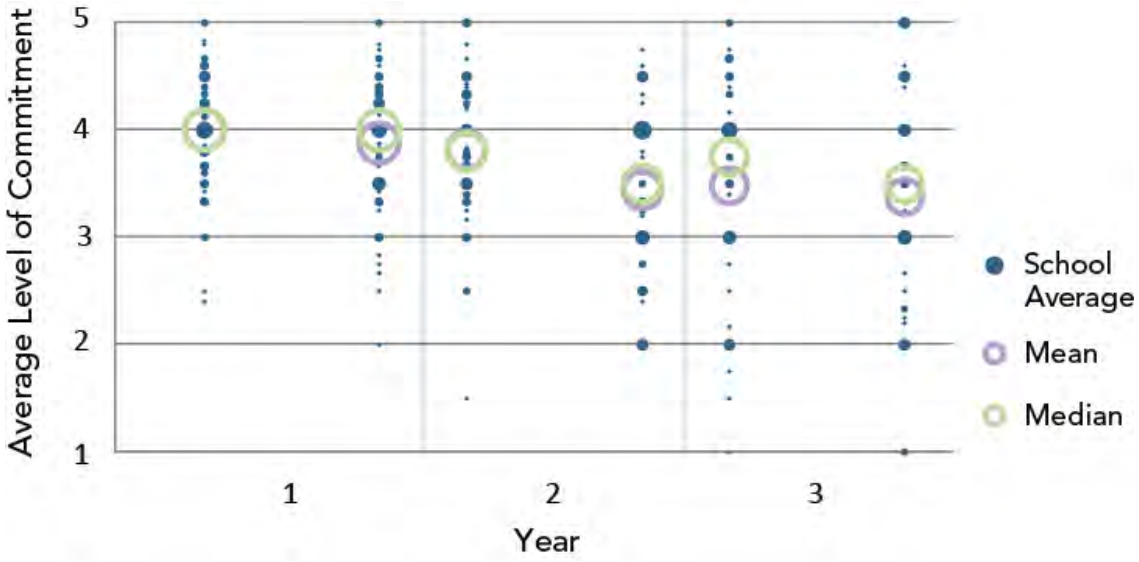


FIGURE 5.7. COMMITMENT TO READING APPRENTICESHIP AT SCHOOL

n = 61 schools in Year 1; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

Continued Use of Reading Apprenticeship Practices

One measure of sustainability is teachers' report of whether they planned to continue using the Reading Apprenticeship framework to inform their instruction in the following school year. At the end of Year 1, on average 91% of teachers in each school said they would continue to use Reading Apprenticeship (Figure 5.8). By the end of Year 2 and Year 3, on average 85% and 82% of the teachers in each school said they would continue using Reading Apprenticeship practices in their third and fourth year of implementation, respectively. The reduction in the percentage of teachers reporting that they would continue implementation at the end of the first and third year is not statistically significant (p =.064). The median percent of teachers in each school responding that they plan to continue using Reading Apprenticeship stayed constant across the three years.

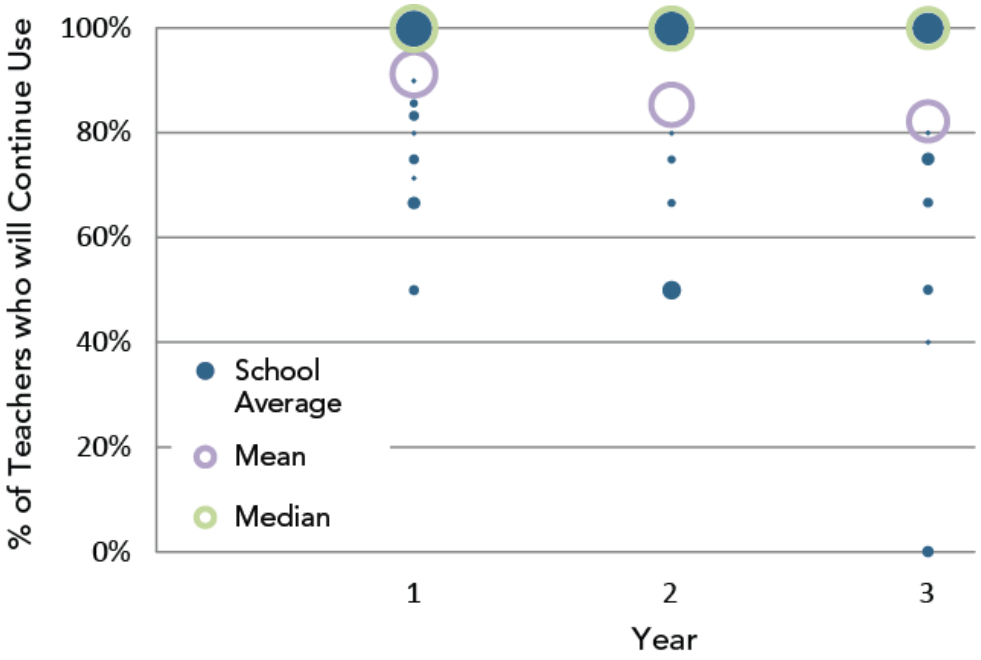


FIGURE 5.8. CONTINUE USE OF READING APPRENTICESHIP

n = 61 schools in Year 1 ; *n* = 54 schools in Year 2; *n* = 48 schools in Year 3

Relationship between Indicators of Participation in RAISE Activities and Indicators of Scale-up Outcomes

In addition to assessing the change in indicators of participation in RAISE activities and scale-up outcomes over the three years, we examined if there was a relationship between the change in the indicators of the uptake of RAISE from Year 1 to Year 3 and the change in the indicators of scale-up outcomes over the same time period.[7] We cannot infer causality from these results; however, they allow us to consider how certain activities potentially influence outcomes. We found that the following relationships were statistically significant (Table 5.2).

- Change in receiving support for Reading Apprenticeship instruction is positively associated with change in self-reported levels of commitment to making Reading Apprenticeship work within the school
- Change in attendance at monthly meetings is positively associated with change in self-reported levels of commitment to making Reading Apprenticeship work within the school
- Change in the average use of Reading Apprenticeship is positively associated with all reported outcomes

[7] Because we did not measure several of the indicators of participation in RAISE activities across all nine survey occasions, we assess the change in the indicators between the first and last surveys in which we asked the question; the change in the outcome score is limited to the same interval.

TABLE 5.2. RELATIONSHIP BETWEEN CHANGES IN INDICATORS OF PARTICIPATION IN READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION ACTIVITIES AND SCALE-UP OUTCOMES

	Receiving support for instruction	Attendance at monthly meetings	Average use of Reading Apprenticeship
Buy-in as appropriate literacy strategy	0.15	0.05	0.35*
Buy-in as means to improve student achievement	0.20	0.09	0.33*
Commitment in classroom	0.29	0.17	0.55***
Commitment in school	0.35*	0.31*	0.46**
Continued use of Reading Apprenticeship next year	0.14	0.09	0.61***

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Figure 5.9 shows the positive and significant relationship between the average use of Reading Apprenticeship practices and commitment to making Reading Apprenticeship work at the school. Each of the 48 blue dots represents a Cohort 1 school and indicates its change in average responses. While the changes in these variables are related, we also note that there is a lot of variability in the responses. The schools represented in the upper right quadrant are schools whose responses to both questions increased over the three years, and the schools represented in the bottom left quadrant are schools whose responses decreased to both questions over the three years.

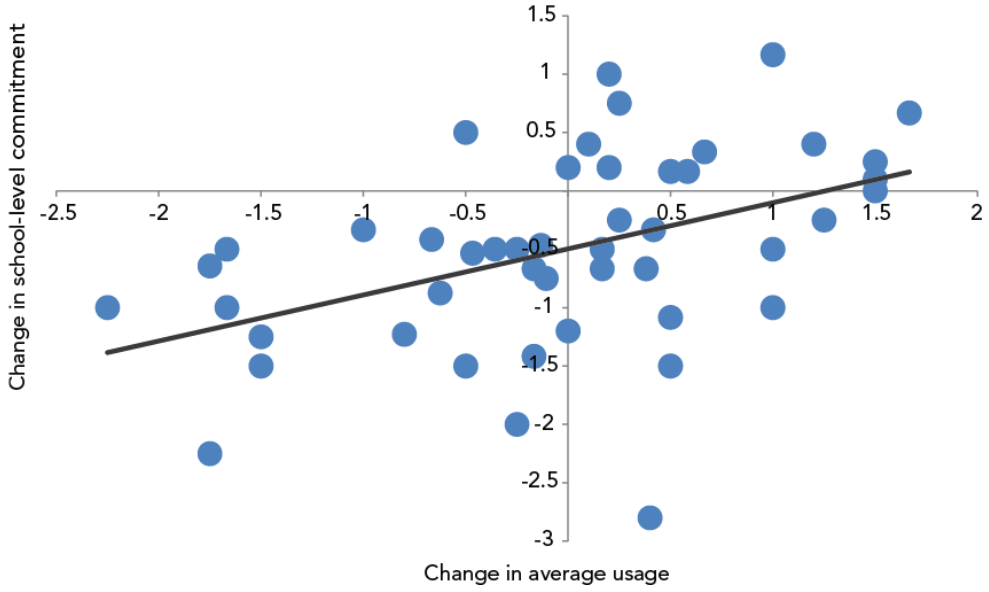


FIGURE 5.9. RELATIONSHIP BETWEEN CHANGE IN AVERAGE USAGE OF READING APPRENTICESHIP PRACTICES AND SCHOOL-LEVEL COMMITMENT

Note. The correlation coefficient is .46. $p < .01$

SIMILAR TRENDS ACROSS COHORTS

While the focus of this chapter was on Cohort 1 schools, we also explored whether there were similar trends over time in the uptake of RAISE activities and indicators of scale-up "success" for schools participating in Cohorts 2 and 3.[8] In general we found similar trends across the three cohorts. Figures 5.10 and 5.11 show the comparison in the trends over time for Cohorts 1-3 in their attendance at monthly meetings and average use of Reading Apprenticeship practices.

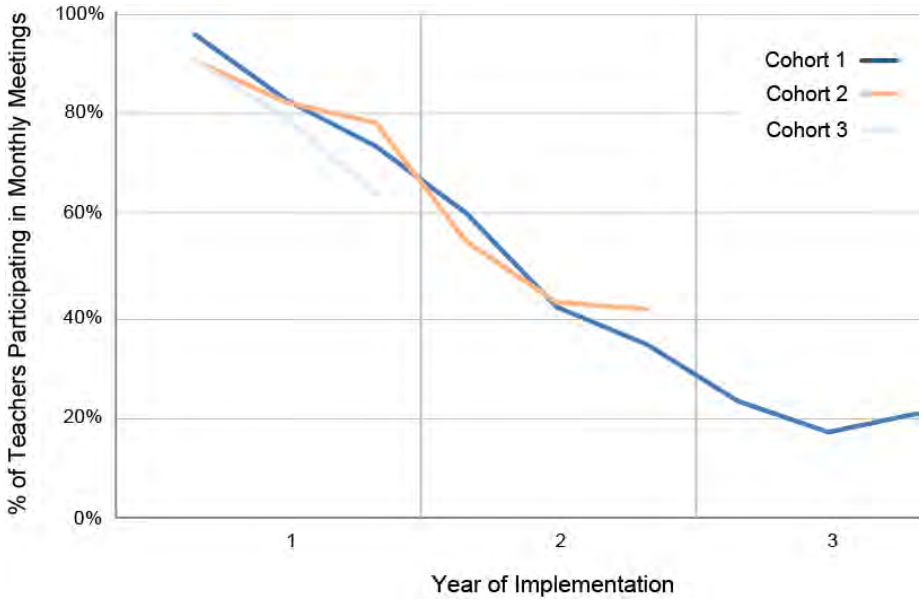


FIGURE 5.10. ATTENDANCE AT MONTHLY MEETINGS ACROSS COHORTS

Cohort 1: *n* = 61 schools in Survey 1-3 ; *n* = 54 schools in Surveys 4-6; *n* = 48 schools in Surveys 7-9
Cohort 2: *n* = 108 schools in Survey 1-3 ; *n* = 94 schools in Surveys 4-6
Cohort 3: *n* = 54 schools in Survey 1-3

[8] Cohorts 2 and 3 have fewer survey occasions than Cohort 1. Cohort 2 has two years of survey data. Their first year of implementation (surveys 1-3) was AY 2012-13 and their second year of implementation (surveys 4-6) was AY 2013-14. Cohort 3 has one year of survey data. Their first year of implementation (surveys 1-3) was AY 2013-14.

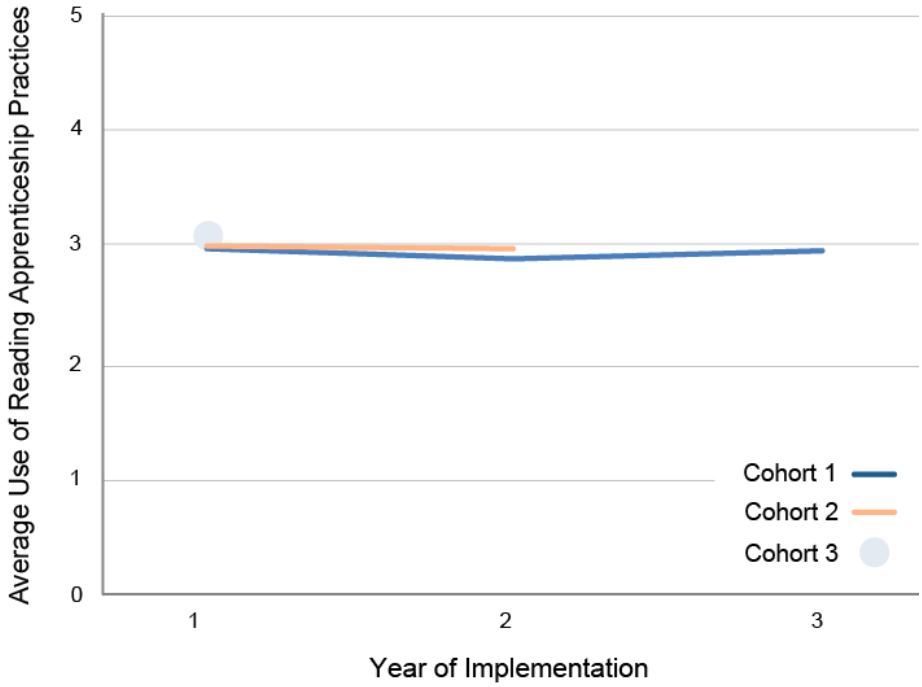


FIGURE 5.11. AVERAGE USE OF READING APPRENTICESHIP PRACTICES ACROSS COHORTS

Cohort 1: *n* = 61 schools in Survey 1-3 ; *n* = 54 schools in Surveys 4-6; *n* = 48 schools in Surveys 7-9

Cohort 2: *n* = 108 schools in Survey 1-3 ; *n* = 91 schools in Surveys 4-6

Cohort 3: *n* = 51 schools in Survey 1-3

Y-axis scale: 5 = Throughout each lesson; 4 = A few times during each lesson; 3 = A few times per week; 2 = A few times per month; 1 = A few times per grading period; 0 = Never

Figures 5.12 and 5.13 show the comparison in the trends over for Cohort 1-3 in their commitment to making Reading Apprenticeship work at the school level and continued use of Reading Apprenticeship practices.

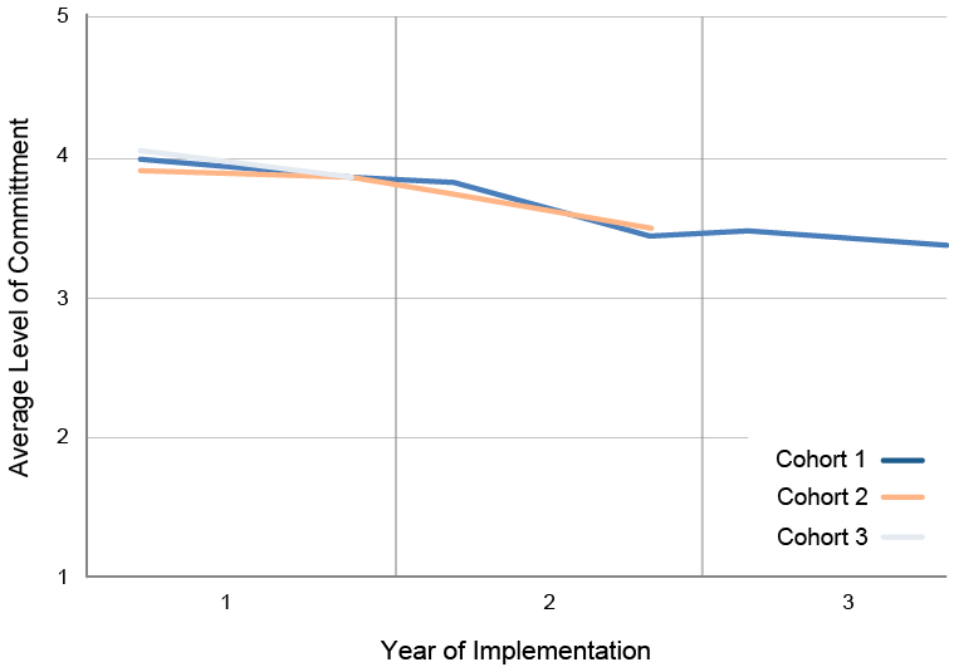


FIGURE 5.12. COMMITMENT TO READING APPRENTICESHIP AT SCHOOL ACROSS COHORTS

Cohort 1: *n* = 61 schools in Survey 1-3 ; *n* = 54 schools in Surveys 4-6; *n* = 48 schools in Surveys 7-9

Cohort 2: *n* = 109 schools in Survey 1-3 ; *n* = 95 schools in Surveys 4-6

Cohort 3: *n* = 54 schools in Survey 1-3

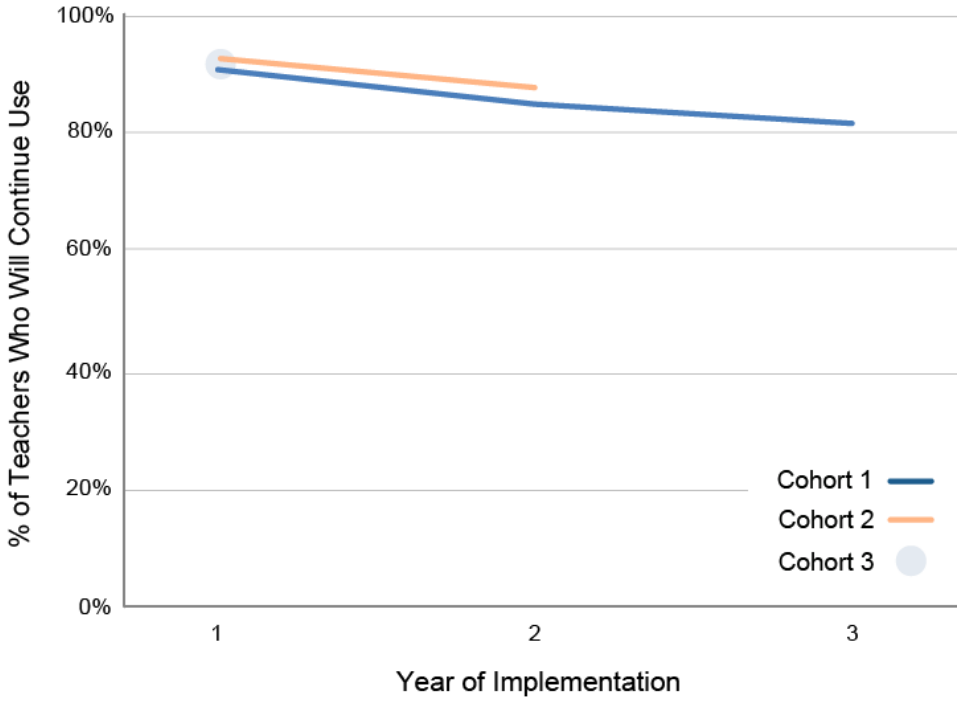


FIGURE 5.13. CONTINUED USE OF READING APPRENTICESHIP ACROSS COHORTS

Cohort 1: *n* = 61 schools in Survey 1-3; *n* = 54 schools in Surveys 4-6; *n* = 48 schools in Surveys 7-9

Cohort 2: *n* = 108 schools in Survey 1-3; *n* = 91 schools in Surveys 4-6

Cohort 3: *n* = 52 schools in Survey 1-3

IMPORTANCE OF SCHOOL LEVEL COLLABORATION

In general, we found that the uptake of RAISE activities and commitment levels were not as strong in the second or third year as the first year. While reported levels of participation in RAISE activities, buy-in, and commitment were high in the first year of implementation, on average this enthusiasm decreased or leveled off by the third year. One of the most obvious decreases over time was teachers' attendance at the RAISE monthly team meetings, which was considered an important mechanism for collaboration and support. In spite of the decrease in participation, reported levels of buy-in remained high. The statistically significant positive relationship between the change in commitment to making Reading Apprenticeship work at the school and change in participation in each of the RAISE activities for Cohort 1, suggests that school level engagement or collaboration is an important process related to continued use of the program.

We also found a greater spread in responses by the third year. In Figure 5.3, for example, while the average amount of practices reported remains about the same, we can notice what looks like a greater spread. Figure 5.9 shows a horizontal spread. The possibility that over time some schools may increase and others decrease the amount they use Reading Apprenticeship led us to a concern that there may be very different processes at work. Our first approach to looking at these processes was the case studies of four schools, which we report in the next chapter. We continue with a quantitative analysis comparing characteristics of schools that gained or lost participants over time.

Chapter 6. Case Studies of Local Challenges and Adjustments

In our quantitative analysis presented in Chapter 5, we saw that there was growing variability among schools in their uptake of Reading Apprenticeship and in their contexts in terms of policy, resources, leadership, and challenges at the school, district and state levels. Our aim in conducting case studies in four of the Michigan schools was to get concrete pictures of how these contextual factors influence school-level scale-up processes. The cases highlight the challenges schools face in this process, and the local-level problem-solving that happens as they navigate those challenges. We conducted the case studies across two years of data collection.

SAMPLE AND METHODS

The case study sample includes four comprehensive high schools in two districts in Michigan. One district, Sage, included two schools with prior exposure to Reading Apprenticeship (that is, prior to RAISE), and the other district, Debutte, had two schools that were first exposed to Reading Apprenticeship through the RAISE initiative. In the fall of 2012, all RAISE participants from the selected schools were asked to participate in the longitudinal case studies. This included the head principal; RAISE teacher leaders; teachers across ELA, history, and biology content areas; and a Reading Apprenticeship-trained instructional support staff member. We should note here that the last category of "instructional support staff" was not a built-in part of the RAISE intervention, but rather an additional support position present in all four of the case study schools. These roles included: two literacy coordinators (a position created within the Debutte district), one Assistant Principal who had previously received training and taught with Reading Apprenticeship, and one teacher leader who was also a Reading Apprenticeship facilitator for SLI. The teachers in all four schools were in the first cohort of RAISE participants and attended the RAISE professional development during the 2011-12 school year.

During the 2012-13 school year (year 1 of the case study data collection), we collected interview data from two district personnel (who were best positioned to speak about RAISE efforts and other literacy initiatives in their district), four head principals, and four Reading Apprenticeship-trained instructional support staff. Additionally, we collected survey data from 19 teachers, and we collected interview, classroom observation, and focus group data from a sample of those teachers. During the 2013-14 school year (year 2 of the case study data collection), we collected interview data from 10 participants across the four schools who had participated in the research activities from the prior year (assuming they were at the school then). Table 6.1 contains information on the participants cited in this chapter.

A version of this chapter was presented in Hegseth, Zacamy, and Newman, 2015.

TABLE 6.1. READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION (RAISE) CASE STUDY PARTICIPANT INFORMATION

Case study school	Context	Participant	Position
Debutte School District: New to Reading Apprenticeship through the RAISE initiative			
Pershing High School	High level of trust between teachers and administrators Labeled a Focus School [a] in AY 2013/14	Jane	Literacy Coordinator
		Floyd	Principal
Carnegie High School	Multiple administration changes Labeled a Focus School in AY 2012/13 and a Priority School in AY 2013/14	Ryan	Principal
		Lucy [b]	Literacy Coordinator
		Terry	RAISE Teacher Leader- ELA
Sage School District: Had exposure to Reading Apprenticeship prior to the RAISE initiative			
Churchill High School	Often acts as an "exemplar" Reading Apprenticeship school [b] Labeled a Focus School in AY 2013/14	June	Reading Apprenticeship-trained Facilitator and RAISE Teacher Leader - ELA
		Serena	Co-RAISE Teacher Leader- ELA
		Brett	History Teacher
Twin Lakes High School	Administration change and labeled Focus School in AY 2013/14	Jasper [c]	Principal
		Shannon	Assistant Principal
		Lindsay	Building Coach and RAISE Teacher Leader- Science

Note. AY = Academic Year

[a] According to the Michigan Department of Education, "Focus Schools consist of the ten percent of schools on the Top-to-Bottom list with the largest achievement gaps between its top 30 percent of students and its bottom 30 percent, based on average scale score. In addition to being required by USED for ESEA Flexibility, identifying Focus Schools is a critical step toward Michigan achieving its overriding goal of closing the achievement gap within schools and reducing the achievement gap statewide." (MDE, 2016). Priority schools "are those schools in the bottom 5% of a complete "top-to-bottom" list of schools... This placement is often indicative of minimal student outcomes in a number of subject areas over time, poor achievement coupled with dramatic declines in student performance or substantial achievement gaps, or all of these factors. As a result, Priority school districts need to develop a reform/redesign plan for the school that focuses on rapid turnaround as a way to dramatically improve student achievement." (MDE, n.d.)

[b] Based on strong Reading Apprenticeship implementation, this school was often observed by and visited by teachers both within the district and county, as well as from out of state.

[c] This participant was newly hired at the school for AY 2013-2014.

As described in Chapter 3, data collection included surveys, interviews, focus groups, and classroom observations. These data were qualitatively coded using the NVivo coding software. The researcher coded all data in three passes. First, the researcher reviewed the data and coded for themes that are central to the larger RAISE Scale-up study's logic model and research questions (e.g., mentions of collaboration, resources, training, etc.). During the second pass, the researcher coded for mentions of time (e.g., this year, last year, before or after the RAISE initiative, etc.) and stakeholders (e.g., references to one's principal, teacher leader, district superintendent, etc.), in order to analyze changes across time, and the influence of various stakeholders in each context. In the third and final pass, the researcher coded more interpretively for 1) discussion of factors that influenced the implementation and sustainability of RAISE, 2) participant attitudes toward RAISE and the Reading Apprenticeship framework, and 3) any instances where participants made comparisons across various stakeholders or across time. Codes were developed both internally and externally. The researcher developed some codes based on existing scale-up theories and literature. In addition to these external codes, themes within the data were identified and internal codes were developed in order to capture prevalent, emerging patterns. Codes from 2012-13 were largely maintained to analyze data from the 2013-14 school year. (See Hegseth, Zacamy, & Newman, 2014 and Hegseth, Zacamy, & Newman, 2015 for detail about participant sample selection, data collection, and analysis methods.)

RAISE SCALE-UP CASE STUDY FINDINGS

To document the prevailing themes within these cases, we first discuss each school individually, with a description highlighting the major changes—across the two school years—that have implications for implementing and sustaining RAISE. We then present four key challenges that the case study participants encountered during the two years of data collection, and ways in which certain contextual factors further complicate the ability to navigate those challenges.

Pershing High School

Pershing High School is one of the two case study schools in Debutte School District: the district that did not have exposure to Reading Apprenticeship prior to the RAISE initiative. Approximately 11% of the Pershing teacher population attended the RAISE professional development in the first or second year of the initiative, and approximately 5% of the total teaching population attended the Cohort 1 Institute. There were twice as many RAISE-trained ELA teachers as there were RAISE-trained biology or history teachers. Pershing is an urban school, with 45% of the student population eligible to receive free or reduced-price lunch. Approximately 88% of the Pershing student population is White (many of these students are of Arab descent), 7% Black, 3% Hispanic, 1% Asian, and less than 1% American Indian/Alaskan Native. According to 2012-13 NCES data, 39% of the student population in Debutte School District is ELLs. Due to the relatively high number of economically disadvantaged students who attend Pershing, the school receives Title I funding to help with their endeavors to improve academic

achievement.[9] When conducting research at Pershing during the 2012-13 school year, we interacted with the head principal, the literacy coordinator, and five teachers (one science, one history, and three ELA teachers). During the 2013-14 school year, the researcher was only able to interview the school's literacy coordinator, Jane.

Findings from the first case studies report revealed Pershing High School to be a context characterized by relational trust, which previous research claims is essential to educators' ability to experiment with and sustain improvements (Bryk & Schneider, 2002; Meier, 2002). Prior to RAISE, a great deal of trust existed between the principal and the literacy coordinator, Jane, as well as between Jane and teachers with whom she worked. This set of trusting relationships enabled the Principal to promote a high degree of autonomy and agency for staff leading the RAISE initiative at the school. In his 2013 interview, Principal Floyd exhibits his trust in Jane when stating:

> "She is phenomenal, and not only creative in her own way, but she does the research to find the best ways, she convinces a big district of 20,000 kids in three big high schools of ways that we're going to improve. ...So when she comes to me with a little bit of excitement and says, 'this needs to happen, we need to train so and so, and I got three teachers that are excited or five or whatever, can we start this?,' I'll take the money out of my own wallet to help her out, because I know that she's not just trying something. She knows that it'll work for us" (Floyd, interview, April 25, 2013).

As Pershing's literacy coordinator, Jane was in charge of curriculum, textbook selection, facilitating and coordinating building and district-level professional developments on literacy, as well as the evaluation of Pershing's ELA teachers. Jane's role as evaluator and trainer—though much appreciated by teachers and administrators—proved to complicate implementation efforts at Pershing and in Debutte (further discussion to follow).

The degree of trust and autonomy that Principal Floyd gave Jane greatly facilitated her work, especially considering the magnitude and scope of her responsibilities. During the second-year interviews, Jane validated Principal Floyd's trust in her when discussing the resources that support the classroom implementation of Reading Apprenticeship.

> "We have a Title 1 budget, smaller than it was last year, but still I have funds to support the [Reading Apprenticeship] teachers, which are really discretionary in terms of what I want, what the teachers need. So we've purchased [materials to support Reading Apprenticeship implementation]…We also have some money set aside in addition to what the district is giving us, for follow up training, or follow up conversations. I suppose it's more like, within the building, getting subs, and having teachers have conversations about what's going on in their classroom...our principal has been very supportive of that, it's pretty much whatever I need, which is nice" (Jane, interview, March 13, 2014).

---

[9] The researcher made every attempt to protect the identity of case study participants and schools, which is why background information on each school is reported in an approximate and concise manner. Unless otherwise stated, student demographic information is based on 2012-13 NCES data. Pseudonyms are used for all participants.

Though the high degree of trust in Reading Apprenticeship leaders remained constant at Pershing during the second year of data collection, there were many changes with varied implications for the spreading and sustaining of RAISE. At the start of the 2013-14 school year, Pershing was named a Focus School, a term used in Michigan for schools that have a 30% or larger gap in achievement between the high and low achieving students at the school. Unlike Carnegie (the other case study school in Debutte School District), which was named a Focus School during the 2012-13 school year, Pershing did not receive any Focus School funding to help with remedying disparities in achievement due to increased budget cuts across the district. Additionally, because Carnegie was labeled a Priority School (which has harsher implications than the Focus School status) at the same time that Pershing became Focus, Carnegie was able to recruit some of Pershing's teachers who were deemed "highly effective" by local evaluation criteria. It can be argued that Focus School status may have supported the uptake of and buy-in to RAISE at Pershing, as teachers and administrators were actively seeking out ways to improve teaching practices and student achievement. This was evident when administrators selected Reading Apprenticeship as one of the three initiatives that were integral to Pershing's School Improvement Plan. However, while Jane pointed out that administrators' awareness of the need for transformed teaching made them more open to supporting Reading Apprenticeship, she also voiced concern that the Focus School label increased the pressure for Reading Apprenticeship to quickly generate positive test results.

> "I think principals are noting that stand and lecture doesn't get there, and that there has been a lot of acceptance of Reading Apprenticeship, mainly because there's so much choice and you can start off small and build up, but I'm not sure if we have real data results yet. So we have acceptance, it looks good, but on the other hand, [this year,] two of the three high schools fell into this Priority/Focus status, too. Not a lot of people brought that up yet, but they will sooner or later if our scores don't raise. We can show them other kinds of scores, but that's common assessment or lexile scores, which are fine, except it's the Michigan Merit Exam that everybody cares about" (Jane, interview, March 13, 2014).

Though Jane worried that continued participation in Reading Apprenticeship in the Debutte district would not be supported unless state test scores improved, the initiative did seem to have spread throughout Pershing High School across time. The efforts of Jane and other Debutte literacy coordinators demonstrated a transfer of ownership of the initiative to the local level.

While RAISE teachers had a difficult time coordinating time to meet for their RAISE team meetings due to conflicts in prep periods during the 2012-13 school year, prep periods were coordinated for each content area during 2013-14. This allowed Jane to meet with teachers in varied content areas during their common prep periods to talk about RAISE. In 2013-14, Jane helped to facilitate an SLI-run Reading Apprenticeship training in the spring for Debutte teachers. She planned to hire substitute teachers one time each semester to pull out teachers for lesson study and RAISE conversations, but was only able to do so during the first semester because of scheduling conflicts and a record number of snow day disruptions.

In 2013-14, Jane was only responsible for supporting Pershing (as opposed to the previous year when she was also required to support the district's middle schools) and she continued to co-lead district Reading Apprenticeship trainings with the other Debutte literacy coordinators. They led three cohorts

through district trainings, with 25 Debutte district teachers in each cohort. These trainings occurred once a month for half a day, and then, the literacy coordinators further supported their teachers at their school site. While these Reading Apprenticeship trainings were not as in-depth as those at the RAISE Institute, they helped to augment the number of teachers who were exposed to the Reading Apprenticeship framework at Pershing, thus increasing external spread beyond the two high schools involved in Reading Apprenticeship through the RAISE grant.

One benefit of this spread and transfer of ownership seemed to be a collaborative, school-wide effort to encourage and institutionalize time for "extensive reading" and to build reader identity, both core features of the Reading Apprenticeship framework. During the 2012-13 school year, Jane coordinated an effort to have teachers bring in books so that there could be classroom libraries for Sustained Silent Reading (SSR). In 2013-14, Jane used some funds to continue to build up classroom libraries. As she reports:

> "no longer are we having students being scolded in the other classrooms for bringing out a book to read when they're done with their work, because that actually was happening a few years ago. So, I think everyone knows now that kids are doing the SSR reading, they're reading independently, and if they don't have anything else in class to do, you can say 'get your reading book out' and they will. So, that's been very different" (Jane, interview, March 13, 2014).

While spread of the Reading Apprenticeship work through RAISE was occurring at Pershing, a concern surfaced in interviews with Jane, as well as in a focus group with Pershing and Carnegie teachers. The district Reading Apprenticeship trainings—shorter and less in-depth than the more intensive RAISE training—seemed to result in teachers only understanding Reading Apprenticeship as a list of strategies, rather than a framework for teaching. Related to this issue, Jane also reported having a hard time bringing the RAISE-trained teachers and the district-trained teachers together to collaborate, due to their different training experiences and resulting approaches to Reading Apprenticeship.

This issue of different levels of depth of Reading Apprenticeship professional development, and the relationship to districts' needs for "getting the reforms out broadly," speaks broadly to implementation of complex reforms requiring deep transformation in teachers' beliefs, identity, knowledge, and practice. We cite this as one of the four major challenges that our case studies serve to illuminate.

In summarizing data from Pershing informants on supports and barriers to implementing and sustaining Reading Apprenticeship over the two years of the case study, the primary supports participants reported in the first year revolved around resources, like the literacy coordinator and money for supplies. According to first-year participants, the initiative was also starting to spread throughout the school and district due to the district trainings. The primary barriers cited revolved around (1) the lack of a common prep period to collaborate, (2) resistance from teachers and students to Reading Apprenticeship implementation and routines (such as Talking to the Text), and (3) too many competing initiatives and tension between district-trained (with just five half-days of training) versus RAISE-trained (sending teachers to the 10 days of intensive professional development of the RAISE Institute) teachers' levels of understanding. In year 2, similar types of resources were available to support the initiative, even in the face of budget cuts, and Reading Apprenticeship continued to spread

throughout the school and district. Though time and resources were available for collaboration, and Jane and some teachers were helping to spread and support the initiative, there were concerns tied to the Focus School status, resistance from certain content areas (specifically, social studies), diminishing resources, and a lack of improved state test scores to validate the sustained use of Reading Apprenticeship amidst a high-pressure climate relying heavily on such metrics.

Carnegie High School

Carnegie High School is the second case study school in Debutte School District. Carnegie was similar to Pershing in that the school had a small percentage of teachers who were RAISE-trained, though also like Pershing, Carnegie had literacy coordinators who had trained increasing numbers of teachers in Reading Apprenticeship at the district level. Approximately 9% of the Carnegie teacher population attended a Cohort 1 or 2 RAISE Institute, with less than 5% of the total teaching population attending the Cohort 1 Institute. There were over twice the number of RAISE-trained ELA teachers than there were RAISE-trained science teachers, and no RAISE-trained history teachers. Carnegie High School is an urban school, with 59% of the student population eligible to receive free or reduced-price lunch. Approximately 86% of the Carnegie student population is White, 9% Black, 4% Hispanic, and less than 1% Asian or American Indian/Alaskan Native, respectively. As was stated for Pershing, Carnegie is in a district that contains many ELLs and students of Arab descent. The head principal, Principal Ryan, and some other administrators were new to the school at the beginning of the 2012-13 school year, which is also when Carnegie was named a Focus School. At the start of the 2013-14 school year, Carnegie was given the more severe label of Priority School. While this designation has many negative implications, it did result in Carnegie receiving additional funding, some of which was used to hire an SLI RAISE facilitator to coach teachers. As in the case of Debutte's ability to fund literacy coordinators, this coach was also an element *added to* the basic grant-funded RAISE intervention. With the Priority label, Carnegie was also permitted to recruit "highly effective" teachers from other schools in the district in order to generate improvement. Additionally (and unrelated to the Priority School designation), there was a change in literacy coordinators from the first to the second year of the data collection. When conducting research at Carnegie during the 2012-13 school year, we interacted with the head principal, the literacy coordinator, and three teachers (one biology and two ELA teachers, as there were no RAISE-trained Carnegie teachers in history). During the 2013-14 school year, we conducted in-person interviews with the new literacy coordinator, Lucy, and teacher leader, Terry, as well as a Skype interview with Principal Ryan.

Findings from the first-year case studies report revealed Carnegie as a school in a relatively fragile condition, largely due to its Focus School label and the fact that the teachers felt too many demands from competing initiatives. Despite this, Carnegie stakeholders seemed optimistic about the future of both the school and Reading Apprenticeship. In the second year of data collection, the school continued to struggle through a variety of changes and outside pressures, as well as a new literacy coordinator and the ramifications of the Priority School designation. However, Principal Ryan's vision for the school and optimism about the future remained constant. In his first year interview, he discussed how he viewed this new designation as something that could be turned "from a negative to a positive."

> "Carnegie was deemed a Priority School, and that kind of changed everything for us. This being my second year as principal in the building, it provided a tremendous opportunity that then took away all excuses, all whatever there would be to think that it wasn't urgent to do as much as we could to raise student achievement. And [Reading Apprenticeship] was obviously one of the things that we committed to as an administrative team and now as a staff and I think it's had a tremendous amount of growth. I've seen it in the classroom, so that Priority designation provided us the opportunity as far as having total buy in, and also having some money to purchase resources that we used to obviously support [Reading Apprenticeship] Reading and science were the two areas that we needed to improve the most, and that's what we committed to, and we saw [Reading Apprenticeship] as critical piece as anything to doing that" (Ryan, Skype interview, April 28, 2014).

Another seeming consistency across the two years of data collection was that, when compared to other case study schools, there was less teacher autonomy over when and how Reading Apprenticeship collaboration took place at Carnegie. When asked about who the leaders were with regards to RAISE at her school, teacher leader Terry mused:

> "I don't know if I view any of us as leaders, I'm supposed to be the RAISE teacher leader, you would have no idea that I was. You would not know, other than having a RAISE classroom, it's not like I'm getting up at meetings and sharing things and talking about it, it's really not that I want to. I don't want the pressure" (Terry, interview, March 13, 2014).

Lucy, the literacy coordinator who was hired in the 2nd year of the study, confirmed that the school had an administrator-driven approach to collaboration when replying to a question about who was designated to lead Carnegie's Reading Apprenticeship professional learning community.

> "[whispering] I designated myself. And then the writing went to [another teacher] who's actually a classroom teacher. I couldn't do both, and so, I think I just wanted to control it a little bit right off the bat, get it up and going, because it's across content areas, and I thought I had a better perspective on that than Terry with more of a language arts... I'm not sure, we'll see" (Lucy, interview, March 13, 2014).

While Lucy did not attend the RAISE Institute, she had been trained in the Reading Apprenticeship framework while she was a Title I resource teacher at a different school, and she reported receiving a lot of support from the other literacy coordinators in Debutte district.

The implications of the changes at Carnegie High School during the second year of data collection were varied. The change from Focus to Priority designation alone created a host of ambiguous effects on the spreading and sustaining of RAISE. Seeming benefits of the designation were that Carnegie received additional funding, which enabled them to do things such as build up classroom libraries for SSR and hire an SLI coach to come in and do a lesson study with a group of teachers. Carnegie was also able to recruit a new wave of "highly effective" teachers, some of whom had already been Reading Apprenticeship-trained at Pershing, and some of whom Terry described as young, energetic, and willing to try a new framework such as Reading Apprenticeship. Though these benefits and the added sense of solidarity were salient, the Priority School status did result in a host of obligations and pressures, such as

frequent observations from outsiders, and an extended school day with support classes that teachers needed to plan for and teach. Terry's interview presented the complications of this new condition. When asked about whether the school's literacy coordinator, Lucy, was leading RAISE meetings, Terry stated:

> "Yeah, I think we've got [RAISE monthly meetings], but I can't remember the last time we were able to do that. It doesn't mean we didn't do it, there's so much going on... like we have the SIP [school improvement plan] committees coming to observe us and they're all from central office, so what do they want to see? And then we get a list of what they want to see. They want to see pair work, and this on the board and student work and Word Walls. Well then we're running around making sure everything there is good, and then Peggy Black's going to come in and observe, and she wants to see co-teaching and duh duh duh da da, so then we're doing that. And then Lucy's going to observe and she needs to see the writing process, or some part of it. So we're so busy, there's just so many observations now and we have to structure our schedule around it. So as far as like, 'you're on the RAISE committee you're on the SIOP committee,' yes, there are those things, but I would say they're taking second place" (Terry, interview, March 13, 2014).

Terry did not harbor any resentment toward administration with regards to the different obligations, and the fact that the administration drove collaborative activities. Her interview also reflected on how overall attitudes and practices had changed across time, since the initial introduction of RAISE to a small group of teachers.

> "Carnegie right now is a climate of people who realize that… whatever they tell you to do, just do it. That's our climate right now. 'You spent last year bitching about all the meetings, it got worse, they could fire you if you don't- they want you to do A, B, and C, why are you going to complain? Embrace it. Do it.' So I think the fact that we are a Priority School, we are changing in spite of all this weight on us. ...we still have to get that observation part down more, but I think that in spite of our busyness, it's kind of made us try a few things that we wouldn't have tried before" (Terry, interview, March 13, 2014).

Though there were added observations and obligations linked to the Priority label, Principal Ryan remained consistent with his views of the top initiatives needed for Carnegie's efforts toward improvement: Reading Apprenticeship, Formative Assessment, and 6+1 Writing. While there continued to be a disconnect between Principal Ryan's more focused vision for the school and the many conflicting priorities that Terry felt, there seemed to be more understanding across stakeholders in the 2nd year of the case study.

> "I don't think that our administration is against RAISE, I think that they're for it, they just don't have the time. Sometimes, in a time of crisis, you forgive a lot of people. We're not critical of each other at Carnegie this year. We're not saying, 'I really need time to meet with these people, and you could provide it, you're just not.' That's not even an option, they're so inundated. You know that thing in RAISE about believe everyone has their best intentions? That's where we're at Carnegie. When we say 'why would Principal Ryan do this?' Now we say 'you know what? It's probably because of this" (Terry, interview, March 13, 2014).

Additionally, Reading Apprenticeship seemed to have spread across staff and students at Carnegie with the continuation of district trainings, the coaching from the SLI consultant, and the energies of the new literacy coordinator. Both Terry and Principal Ryan spoke about some of those changes in a positive light.

> "One of our science teachers actually, he teaches a language arts intervention, and he had his students create metacognitive logs based on their readings. And he submitted them to us and it was called 'The 20 Project,' and it meant students read SSR+ for 20 minutes a day. And he shared all of them that his students had written from that class, and...many of the students spoke of how they used to love reading as a child, and about third or fourth grade, there's this shift that takes place where all of the reading now is academic. It's no longer high- interest, and that's either as a result of time, or lack of resources, and they began to hate reading, most of them. And now, through Silent Sustained Reading Plus, and this extended time, the intervention class, they have come back to love reading again, and they're reading to their brothers and sisters. There were some very powerful, emotional things that were written that really told us we were on the right track, and obviously with the improved fluency, comprehension, and love for reading, we know that that will spill over into, obviously social studies, language arts, math, science, so we think that's going to have a tremendous impact on us for years to come. And it was as a result of the SSR+ model" (Ryan, Skype interview, April 28, 2014).

> "[Lucy's] been really instrumental. Sometimes administrators can make you feel like there's a weight on your shoulders and sometimes they can make you feel like 'we're going to help you with the weight and we'll get through.' And that's how she makes me feel, and we feel like she's a breath of fresh air, so we're very thankful" (Terry, interview, March 13, 2014).

Analyzing changes in supports and barriers across the two years of data, it became clear that the supports that were most commonly cited across participants revolved around spread of the RAISE initiative across staff and students, and the 10-day Institute. The most-often cited barriers pertained to insufficient time for collaboration (and the existing collaboration was all administrator-driven), district trainings for non-RAISE staff that delivered a much less robust version of Reading Apprenticeship and a district context replete with competing initiatives, and not enough spread of RAISE across content areas at Carnegie. In the second year of data collection, the primary themes discussed—in both negative and positive ways—pertained to Carnegie's context and climate as a Priority School and the change in literacy coordinators, as well as resources. Though there continued to be insufficient time for collaboration, collaboration continued to be administrator-driven, and Carnegie was still immersed in a context full of competing obligations, participants demonstrated more understanding and tolerance for those difficulties, given the larger constraints that the school was facing from the Priority label.

Churchill High School

Churchill High School is one of two case study schools in Sage School District, the district that had exposure to Reading Apprenticeship prior to the RAISE initiative. Unlike the two Debutte case study schools, Churchill had approximately 90% of its teaching staff trained in Reading Apprenticeship at the time of data collection. In addition to the district-provided Reading Apprenticeship training that

teachers received, which occurred prior to the RAISE initiative, approximately 16% of the Churchill teaching population attended a Cohort 1, 2, or 3 RAISE 10-day Institute. An equal number of teachers were RAISE-trained in science and history, and more than twice that number were RAISE-trained in ELA. Churchill is the only case study school that sent teachers to the Cohort 3 Institute. Churchill is an urban school, and approximately 26% of its student population is eligible for free or reduced-price lunch. Forty-nine percent of Churchill students are White, 18% Asian, 20% Black, 6% Hispanic, and less than 1% American Indian/Alaskan Native. Both Churchill and Twin Lakes are in a district where only 4% of the students are ELLs, which presents the schools with a very different student population than in Debutte School District. However, similar to Pershing and Carnegie, Churchill receives Title 1 funding. The head principal, Principal Dwight, has been Churchill's principal for over 15 years. June, the teacher leader who has been a Reading Apprenticeship professional development facilitator at the school and district since Sage began implementing Reading Apprenticeship (prior to RAISE), has been in her role for over six years, and has been a classroom teacher for over 15 years. When conducting research at Churchill during the 2012-13 school year, we interacted with the head principal, the Reading Apprenticeship facilitator and teacher leader, and four other classroom teachers (one biology, one history, and two ELA teachers, one of whom also assumed a Reading Apprenticeship teacher leader role at Churchill). During the 2013-14 school year, we conducted follow-up interviews with the Reading Apprenticeship facilitator, the co-teacher leader, and the history teacher.

Churchill High School emerged as a school that embodied leadership with respect to Reading Apprenticeship and RAISE at a high level of fidelity to the model. The school continually received visitors from other districts, counties, and states, as it is widely acknowledged as an exemplar Reading Apprenticeship school. These visitors tour the school, converse with Churchill teachers and administrators about Reading Apprenticeship, and observe teachers implementing Reading Apprenticeship lessons. Some of the teachers at Churchill have been Reading Apprenticeship facilitators for SLI (both before and during RAISE), in addition to their leadership responsibilities at the school. When asked to discuss the relationship between the school's Reading Apprenticeship implementation and its visibility to the outside community, June responded with:

> "When we do have outside visitors, I send out an email to the staff, and just say 'we have visitors coming in, to see what great work you are all doing, so please welcome them.' So even if they're not coming into their classroom they realize that 'wow, Churchill is being acknowledged for its great work.' Not only locally, but we've had people come from out of state, two trips from Chicago and just different places. So it's kind of like a PR thing, I also have a background in marketing, so I tend to think more that way oftentimes about promoting. And when we had the group coming in for the teacher leader meeting, I crafted an email for Principal Dwight to send to our superintendent and the assistant superintendent, and invite them. Unfortunately they weren't able to come, but I wanted them to know what great work Churchill is doing. And when the new superintendent took over in the fall, I had received a new Reading for Understanding book from [Reading Apprenticeship state coordinator] Sid Mason, and then we also had a video of [Reading Apprenticeship] at Churchill...so I said to Principal Dwight, 'Principal Dwight, would you like to give this to [the superintendent]?' he goes, 'Oh yeah.' So he gave her the book and the video, so it's just the little things like that, it's kind of self- promotion, but we don't want

> it to die. I put eight years into this program and it's working, so we don't want to change, we want to continue to work on that" (June, interview, March 14, 2014).

Churchill's status as a Reading Apprenticeship exemplar school and their fidelity to the framework have been constant for many years. Additionally, there were a few other consistencies present across the two years of data collection at Churchill, even while neighboring schools and the surrounding district were fraught with change. One such consistency was that students' reading scores seemed to keep improving, and Churchill's administrators kept attributing this to Reading Apprenticeship. So, even though Churchill was labeled a Focus School at the beginning of the 2013-14 school year due to its overly large gap between the highest and lowest achievers, there is still evidence from the Michigan Merit Exam and other metrics that create ongoing enthusiasm for Reading Apprenticeship. As the school's co-teacher leader, Serena, and literacy coordinator, June, described,

> "And then, with our scores, overall, as a school, reading is top...even at the beginning of the school year, we sat down at a staff meeting, and they had us look at science and social studies, I think it was the ACT, the MME, and we were looking at those things, and I don't want to say it was me, but as a school, that was high. And they even said 'well, we've had Reading Apprenticeship going on, obviously you see the results, you see the SRS scores, you see the ACT scores in this area,' so it made me feel very proud...we see it as a school" (Serena (co-teacher leader), interview, March 14, 2014).

> "And it isn't that our bottom 30% is so low, it's our top 30% is so high, so [AP] Steven was showing some of the data at one of the opening sessions of the school year, and he was just crowing about the fact that, even when you looked at the data for the bottom 30%, the growth area that our bottom 30% was achieving was in reading. And he attributed that to Reading Apprenticeship" (June, interview, March 14, 2014).

One participant, Social Studies Teacher Brett, did report that when the data are disaggregated, minority groups are not achieving at the same rate as other student populations. So while, as June said, across students there appeared to be growth with the bottom 30%, Brett's point was that when the data were disaggregated by race, minority students were not achieving at the same rate as non-minority students. He was the only participant to mention this disparity, while the other participants focused on how the test scores were largely viewed to be in favor of Reading Apprenticeship. June reported the direct benefits of these achievement data in that she believed her administrators used these test scores and anecdotal data to argue for June's release period, which she devoted to performing Reading Apprenticeship facilitator tasks. June was the only person to maintain this release period during the 2013-14 school year; all other Sage schools had to cut the district designated release time for teacher leaders to work on Reading Apprenticeship implementation, collaboration, coaching, and trainings.

Other consistencies across time included steady administrative support and a related reliable stream of resources devoted to Reading Apprenticeship. Less positive consistencies across time included June reporting, for the second year in a row, that she was still running up against resistance from math teachers. Though math was not part of the RAISE initiative, June recognized the need for the math

department's buy-in when generating building-wide support for and implementation of Reading Apprenticeship.

> "math is kind of a different culture, and I haven't been able to bridge that. And that bothers me, because so many kids struggle with math, and I struggle with math myself, so I don't feel like I could even make some valid suggestions about that. But that's the area where I've not been able to bridge and so, I'm still hopeful that it will happen"(June, interview, March 14, 2014).

Additional consistencies seen across the years of data collection were that teacher Brett continued to discuss how state and district requirements and administrative duties prohibited him from using his time to collaborate with others, and Co-Teacher Leader Serena continued to discuss a desire for an ever-increasing amount of teachers to be sent to the RAISE professional development. Thus, key themes and contextual factors at Churchill remained constant across time, even when faced with a Focus School designation, and while immersed in a district plagued by budget cuts and full of middle and high schools where—according to some Churchill and Twin Lakes participants—support for Reading Apprenticeship was dwindling. The primary notable change across time that was cited pertains to June being able to obtain the time and resources needed for small group collaboration, which she was missing in her first-year interview.

In the analysis of first-year data, prevalent supports cited by participants had to do with a supportive school context and resources, such as strong leaders, and a district-supported release time for one such leader to facilitate school-wide Reading Apprenticeship activities and support teachers. Barriers cited pertained to collaboration (which, depending on the participant, was reported to be insufficient in that there was not enough time, it was not content-specific, or it happened in large groups instead of small); resistance from content groups like math or from teachers who thought they knew Reading Apprenticeship because of their pre-RAISE exposure to it (Hegseth, Zacamy, & Newman, 2014); and dwindling resources (not enough funding for the substitute teachers who facilitated in-school planning time, or to continue having two teachers with a release period). In 2013-14, there was continued talk about how strong leaders and time for collaboration supported Reading Apprenticeship implementation. Additionally, there was discussion by all those interviewed surrounding how improved test scores supported the continued use of Reading Apprenticeship. With regards to barriers toward implementing and sustaining RAISE, June and others continued to discuss resistance within the school and added discussions about lack of spread in the Sage middle schools. Brett continued to voice his distress over a lack of time for collaboration due to district or state obligations and conditions. Despite the barriers reported in the final year of data collection when the data were considered in their entirety, it seemed that Churchill's supportive school context and strong leadership had only improved and become reinvigorated with time.

Twin Lakes High School

Twin Lakes High School is the other case study school in Sage School District, the more experienced district with regards to Reading Apprenticeship. Like Churchill, Twin Lakes had around 90% of its teaching staff trained in Reading Apprenticeship, an additional group of teachers who participated in the RAISE 10-day Institute, and a teacher leader, Lindsey, who had a district-supported release period

devoted to her coaching/Reading Apprenticeship leadership responsibilities during the 2012-13 school year. Twin Lakes is unique in that 20% of its teaching population had attended the RAISE Institute as part of Cohort 1 or 2, 11% of Twin Lakes teachers attended the Cohort 1 training, and approximately equal numbers of ELA and history teachers were RAISE-trained, whereas more than double that number of teachers was RAISE-trained in science. Twin Lakes High School is a suburban school, and approximately 16% of its student population is eligible for free or reduced-price lunch. Sixty percent of Twin Lake students are White, 17% Black, 5% Hispanic, 10% Asian, and less than 1% American Indian/Alaskan Native.[10] As stated previously, Churchill and Twin Lakes are in a district where only 4% of the students are ELLs. Twin Lakes is also different from the other case study schools in that it is fairly new: opened and founded within the past six years. Principal Amara was Twin Lake's head principal from its inception until the start of the 2013-14 school year, when she moved out of state, and Principal Jasper assumed her position. Shannon is one of Twin Lake's assistant principals (AP) and was formerly a Reading Apprenticeship facilitator and teacher leader in a nearby district where she taught. Thus she brought significant background as an administrator supporting Reading Apprenticeship at Twin Lakes. When conducting research at Twin Lakes during the 2012-13 school year, we interacted with the head principal, Amara; AP Shannon; the Reading Apprenticeship facilitator and teacher leader, Lindsey; and four other classroom teachers (two biology, one history, and one ELA teacher). During the 2013-14 school year, the researcher conducted interviews with the new principal, Jasper, as well as AP Shannon and Lindsey.

Twin Lakes was unique among the case studies for several reasons. First, it was the only case study school that had a teacher leader from a non-ELA content area. Lindsey taught science, which many participants saw as being a benefit in terms of convincing teachers outside of ELA to buy in to Reading Apprenticeship's usefulness and applicability across disciplines. The school also set itself apart in terms of its ability to see connections between Reading Apprenticeship and other initiatives or external conditions (like the Common Core State Standards), thereby keeping it relevant and revisited across time. This was best exemplified when RAISE teachers asked to have time for literacy collaboration in the monthly small learning communities (SLCs), thereby institutionalizing collaboration for both Reading Apprenticeship and a district writing initiative, the Washtenaw Writing Collaborative. This tactic of framing Reading Apprenticeship in such a way as to keep it relevant throughout the school can be seen in the following excerpts from interviews with Lindsey and AP Shannon, who spent the 2013-14 school year attempting to connect Reading Apprenticeship to a new priority put forth by Sage, (and further supported by the new principal): data teams.

> "We have our SLC meetings…what we did last year is we put literacy as part of that meeting agenda, because typically we didn't have a very full agenda for those meetings, and we didn't want to lose that time… that's changed this year… the SLC meeting time is now focused on data teams, which to myself and Jasper and Shannon, is a good opportunity to look at literacy.

[10] Eight percent of the students at Twin Lakes identify with two or more races/ethnicities.

> As a scenario… I might bring some student work samples to my data team and say 'ok this is the problem I'm having, I want my students to be more metacognitive, here's some student work, what are you seeing?' Share those with your peers, 'what are you seeing? And how might I address that?'… there's a discussion of possible interventions that we could do, 'take that back to your classroom, give it a try, collect another sample of student work or data of some kind, bring that back and talk about how it went, if you see evidence of it, your kids are going deeper.'…. So I think that's how data teams and the Reading Apprenticeship protocol [for examining student work] is a little different, it's just more formalized, that you're collecting that data throughout the year, regarding some sort of skill that you want your students to get better at over the course of the year, whereas before it was just kind of like 'yeah, I'm seeing a difference'" (Lindsey, interview, March 12, 2014).

> "We created a list of what things we could do to move forward [with connecting Reading Apprenticeship to data teams]… Lindsey [and I] talked about kind of back-dooring it, not coming out and saying 'this is data teams work' or 'this is literacy work,' but doing some of the work, and afterwards talking about this is how it connects to our literacy work, and this is how it connects to our data teams work...because I feel like people, you know how sometimes you just hear words so much, and they shut down and they turn off, and then the negativity comes.

> … it really is the same work, but it has to do with how you frame it and how the expectations are presented. And so I think that was our first step was that student work protocol and moving forward with that, but just being more intentional about thinking about those connections prior to the rollout or prior to the introduction of it, and then making those connections more visible, instead of expecting [teachers] to make those connections themselves" (Shannon, interview, March 13, 2014).

Apart from Shannon and Lindsey's aligned approach to sustaining Reading Apprenticeship by adapting and connecting it to other priorities, there was a great deal of change at Twin Lakes from the 2012-13 school year to the second year of data collection. Two changes, which were unanimously reported as being prohibitive to Reading Apprenticeship implementation, revolve around turnover of Reading Apprenticeship trained staff and district budget cuts, resulting in Lindsey losing her release period for additional Reading Apprenticeship support.

> "We lost a lot of the staff that was trained initially, so with staff turnover, the dynamic of the staff has really changed, where when we were in the crux of it when I first got here, a large majority of the staff had been trained and so as staff have moved on and gotten other jobs, and you've gotten new staff in, and you've grown your staff, not all of them have been trained. So that's a different dynamic, we've got work to do" (Shannon, March 13, 2014).

> "Essentially with [budget cuts], there was no way for us to continue to give Lindsey that release period…. So, we've seen the impact of that on how that looks as far as the follow through and what's been happening. We recently met at the ISD for kind of the administrator meeting, and we were talking about, she's been able to do some things through PD opportunities and some staff

> meeting times, but not nearly kind of the impact she was able to have in the past" (Shannon, interview, March 13, 2014).
>
> "We're losing the FTE for the coach to be able to do the support, and that there were concerns as to how that might affect the implementation, because it had been successful so far in this building" (Jasper, interview, March 13, 2014).

Not only did Lindsey not have time to support teachers in their implementation of Reading Apprenticeship, all participants interviewed discussed a lack of interest from teachers who had not yet attended the RAISE professional development. Given this set of changes, the two administrators, Jasper and Shannon, and teacher leader Lindsey discussed whether or not to make training mandatory for certain teachers. They were tempted to do so even though they realized that a key reason Reading Apprenticeship had been successful in Sage School District was because it spread voluntarily, from the teachers up.

Like Churchill, Twin Lakes was designated a Focus School at the start of the 2013-14 school year, yet because of district-level budget cuts, the school did not receive any Focus School funding to help address its achievement gap. Though there has been a history of support for Reading Apprenticeship at the district level, Twin Lakes stakeholders (like their Churchill counterparts) noticed some diminished enthusiasm and effort in the district's other high schools and middle schools, and attributed that to a lack of administrative support for Reading Apprenticeship. It remains to be seen whether the new superintendent will be of assistance in sustaining this longstanding initiative. It also remains to be seen how the new principal and resulting new set of priorities will influence Reading Apprenticeship implementation at Twin Lakes. Already, the data team's priority has interfered with the formerly institutionalized SLC time devoted to literacy collaboration. However, Shannon and Lindsey were both optimistic that Principal Jasper "gets" Reading Apprenticeship more than former Principal Amara did, due to his previous experience as both a science teacher and a principal at a project-based tech school, as compared to Amara's background in business teaching. They were optimistic that Principal Jasper will work with them to make connections between Reading Apprenticeship and data teams explicit to the rest of the staff in the coming years.

> "Jasper is coming from a different district, but he's very familiar with [Reading Apprenticeship] and is totally seeing the connection between [Reading Apprenticeship] and some of the other initiatives that we're having in the district, or in our school, and that's such a breath of fresh air because that's something I don't think Amara ever really got. And so he's really supportive of that, and I think that he and Shannon will both really advocate for having some release time for next year" (Lindsey, interview, March 12, 2014).
>
> "Jasper is totally on board with it, he gets it… So I don't feel some of the pushback we had previously, 'cause Amara, I don't think she had a real good understanding of the work. I think he does, and he gets the instructional aspect of it, and the importance of it, and I think he's just struggling with how to fit it in and how to make it work, and then, as well, struggle with the lack of resources available to dedicate to that" (Shannon, interview, March 13, 2014).

The implications of all of these school and district-level changes and conditions seem to be that only a small number of teachers on the Twin Lakes team had what they needed to sustain their Reading Apprenticeship implementation. Shannon and Lindsey both discussed the fact that only those who: 1) were present at the founding of the school (when Reading Apprenticeship was embedded in the curriculum), or 2) attended the district-facilitated literacy teacher leader meetings possessed the knowledge and were getting the support necessary for ongoing implementation of Reading Apprenticeship. The rest of the staff was not receiving as much coaching from Lindsey, they were not receiving as much collaboration time during SLC meetings, and they seemed resistant to opportunities for training. This, plus the loss of staff that had previously been trained in Reading Apprenticeship, resulted in much institutional memory loss, and year 2 participants reported that only those in leadership positions still see the ways in which Reading Apprenticeship connects to and is needed across different disciplines and priorities.

As with other schools in the case study, Twin Lakes interviewees reported changes with regards to the cited supports and barriers to Reading Apprenticeship implementation. In 2012-13, supports revolved around institutionalized time for collaboration, strong administrative and teacher leaders (outside of ELA) who had ample opportunities to lead, resources (time for collaboration and for Lindsey to provide additional support for her colleagues' implementation), and a collaborative school culture aligned to the RAISE initiative. Barriers included a lack of certain supplies and funding for summer planning time; some resistance from teachers, students, and the principal; and the competing priority of high-stakes testing. During the 2013-14 school year, supports pertaining to strong leaders remained constant, but discussion of barriers became much more heightened. Again, these barriers primarily related to decreases in the resources provided by the district to support Reading Apprenticeship, as well as a context full of competing priorities and new staff members' resistance to becoming RAISE-trained. An additional barrier that was not present at Twin Lakes in year 1, and was certainly lamented in year 2 of data collection, was the decrease in time for collaboration.

In year 1 of data collection, Twin Lakes' supports for implementing and sustaining Reading Apprenticeship far outweighed any barriers discussed. All of that changed in year 2, even though supportive leaders with aligned goals and visions remained constant (and reportedly improved because of the new head principal).

CHALLENGES

With the above cases in mind, we will describe four key challenges that surfaced over the two years of data collection.

The first two challenges pertain to the comparatively inexperienced Debutte district, and the ways in which Debutte schools—Pershing and Carnegie—and participants balanced depth and spread (as defined by Coburn [2003]) during their first few years of RAISE implementation. The final two challenges are ones that the more experienced Sage schools—Churchill and Twin Lakes—and school leaders faced. These challenges revolve around how each school tried to sustain RAISE across time.

Challenge 1: Being a RAISE Teacher Leader While a Reading Apprenticeship Novice

The RAISE program defined the responsibilities of "teacher leader" as convening monthly RAISE team meetings and being one of the teachers working on deepening their understanding and practices of the Reading Apprenticeship instructional approach. The idea of "leading from practice" is a core idea underlying SLI's conception of the teacher leader role envisioned in the RAISE intervention. However, teacher leaders were never expected to be coaches in Reading Apprenticeship. In addition, beginning in 2012-2013, teacher leaders were supported in the politics and strategies of leadership (to advocate for attendance at the meetings, to work with more resistant colleagues, to seek support from administrators), and in deepening their own understanding and implementation of Reading Apprenticeship.

While the teacher leader role as defined by RAISE presented challenges enough, many teacher leaders expected even more from themselves. For example, Carnegie High School's RAISE teacher leader, Terry, did not yet claim for herself a strong, well-defined RAISE leadership role. She did not yet consider herself (or any other teacher in her school) to be a RAISE leader in the early stages of implementation and did not personally want that pressure. When asked what advice she would give to someone in her shoes at a school that is just beginning to implement RAISE, Terry stated:

> "I always like the advice that was given to me, and I probably have said it on almost every interview we've had, and that's don't try to do everything at once. Just pick a couple things out. And that's been a comfort to me because then I don't get overwhelmed… It's a kind of thing that builds on itself, so just relax, I'm not going to be [a more experienced Churchill ELA teacher] Serena today (Terry, interview, March 13, 2014).

In her interview, Terry discussed how impressed and invigorated she was when she visited Churchill and observed one of Serena's RAISE lessons. While Serena's lesson gave Terry ideas, Terry acknowledged that she is "not yet Serena" in terms of having the same level of knowledge, understanding, and implementation of Reading Apprenticeship, and she was not yet expected to be. Leader or not, she found comfort in the advice that she should start small, and build up her own Reading Apprenticeship implementation in her classroom, prior to taking on more public, far-reaching RAISE responsibilities.

Comparing Co-Teacher Leader Serena from a school with significant past experience with Reading Apprenticeship and Teacher Leader Terry helped highlight the predicament Terry described. Serena had more time with Reading Apprenticeship. She was exposed to Reading Apprenticeship prior to the RAISE initiative, and she attended the RAISE Institute with Terry. Serena also completed the RAISE facilitation training, which greatly deepened her experience with both implementing the Reading Apprenticeship framework, and training others to do so. Finally, Serena was in a district and a school with multiple people who have modeled, and still model, how to be a RAISE leader. June is her co-teacher leader, and there were Reading Apprenticeship leaders prior to Serena who set a precedent for what that role and those responsibilities entailed. Serena discussed some of her supports when asked about how changes over the 2013-14 school year influenced her own implementation:

> "It makes me think back to when I first was introduced to [Reading Apprenticeship], that overwhelming feeling, where do I start? I want to do a good job, I can't do it all, and being told 'you only have to start with one thing.' And it helps me feel stronger about what I do, it's like some things even now click in, even though I've been doing it for awhile, but because I've had that opportunity to take baby steps, I feel more confident. Like when I had people come visit this year, it wasn't just everyone starts Talking to the Text, and in one period, [the visiting teachers] saw so many things being done that I naturally do. And by having this opportunity, having the opportunities to have the small meetings, the one-on-ones, the guests come in, I feel more confident now by being able to take those baby steps and so, even this year, with the small bit I've done, it's made me conscious of how I am as a [Reading Apprenticeship] teacher, and where are my strengths and where are my weaknesses? … it's allowing me to practice more, having more people come in, having these conversations, having these questions. So I feel like it's making me even better as an individual, just having more and more practice, with being a [Reading Apprenticeship] teacher, and as it grows I grow" (Serena (co-teacher leader), interview, March 14, 2014).
>
> Serena goes on to add, "now that I'm actually a trainer, the people that I work with, the people who have been doing it, the people who are leaders above me, that's been helpful with how to work with the [new] teachers …so the leaders who are part of the [RAISE facilitation training] program for sure, and our teacher leader meeting leaders, keep me going" (Serena (co-teacher leader), interview, March 14, 2014).

Like Terry, Serena emphasized the need to take time and small steps with Reading Apprenticeship implementation. SLI also acknowledged—and reiterated to teachers during the RAISE Institute—that the kind of deep pedagogical changes that RAISE is trying to impact will not happen overnight. Teachers are expected to implement changes in their classroom slowly, and build expertise over time. Serena cited time (to go slowly), additional facilitation training, multiple collaboration and leadership opportunities, and the examples and guidance of other leaders both at Churchill and in the broader Reading Apprenticeship network, as important supports for her own implementation and leadership. Terry had very few, if any, of these added supports at Carnegie. Yet she, like Serena, assumes the title of RAISE teacher leader, and any corresponding pressures or sense of responsibility. Herein lies Terry's challenge:

How might Terry navigate the need to slowly develop her own knowledge and capacity surrounding Reading Apprenticeship implementation, while also continuing to convene and facilitate her RAISE team colleagues, who are also "just beginning"?

If Terry ignores her leadership responsibilities to work solely on her own implementation, then there is risk that both teachers and administration will shift their focus and interest away from the nascent initiative, because there is no strong teacher voice advocating for it. If Terry continues to let Literacy Coordinator Lucy take the lead, then RAISE is being advocated by someone in an evaluative position, as opposed to someone who is speaking directly from the classroom. While the support of administrators is important in the RAISE Scale-up model, it is hypothesized that having a strong teacher leader who is a

classroom teacher and can support the "bottom up" spread will likely be more powerful. Therefore, a second question highlights Terry's and Carnegie's challenge moving forward:

How might Terry collaborate with Literacy Coordinator Lucy in such a way that Lucy takes more of a broad school advocacy role, while Terry speaks for RAISE from the position of a classroom teacher, thereby supporting a powerful "bottom up" spread of Reading Apprenticeship practices?

Challenge 2: Attempting to Both Build Capacity and Scale Up

Terry's challenge can be seen a microcosm for broader trends happening in Debutte School District, implicating both Pershing and Carnegie. That is, Debutte was simultaneously trying to build and deepen capacity to implement Reading Apprenticeship with fidelity by sending small groups of teachers to RAISE trainings, while more broadly spreading Reading Apprenticeship by means of district trainings for those teachers who had not participated in the in-depth RAISE training.

Again, comparing the beginning stages of Reading Apprenticeship implementation in Debutte to that of the more experienced Sage district is helpful for illuminating Debutte's challenge. Churchill's Reading Apprenticeship building leader, June, recounted the way in which she was advised to start Reading Apprenticeship years ago, prior to the RAISE initiative and the i3 grant.

> "We began in 2006, and our leader was an assistant principal named [Sally]. She was trained years before 2006, so she understood it, and she was our mentor, and when we became trainers, she said, 'This is how it's going to become embedded in our school, you will not piecemeal it, you will not, even if a teacher comes up to you and says- "why don't you just give me one of those strategies to use?"' she said, 'you will not do that. Every teacher will go through the three days of training because it's important that they understand the framework, that this is not a group of strategies that you use. This is how you do business in your class.' And we agreed totally, and so we were very small and very methodical in how we began to train our teachers, and the first year, it was just [a former Churchill Reading Apprenticeship leader] and I. We spent the first year just doing it in our own classrooms and people began to ask us, 'what are you doing?' and so we'd talk about it, but we just said 'you have to go through the training.' It's like this secret club, and my other colleague in English asked me to come in to her classroom and model a Reading Apprenticeship lesson with her students, and she sat there and she watched and she was sold. And so then the next year, I think we had four or five teachers trained during the summer, and then we met with them in small meetings, every month, our principal paid for subs and we did two and a half hour meetings with that small group. Then we got a few more teachers and we met with them in small groups. Now, our biggest group was about four years ago, and we did all 9th and 10th grade teachers that had not been trained, so then that enlarged the group, so we still want to continue to do small group meetings, because we felt those were the most effective and to bring in the teachers that are buying in to the practice, and maybe a couple of outliers, bring them in too, try to convert them a little more, but keep that momentum going. I'm very proud of Churchill, when [State Coordinator] Sid Mason has visitors, he calls, 'when can we come to Churchill?' That, to me, shows us that we did it the right way. We didn't just go through and say 'you all will be trained you all will use it.' We did it, in a very methodical process. And we've

> supported them in a very strategic way, and I think that's why it's still alive at Churchill" (June, interview, February 12, 2013).

In describing the beginning stages of Reading Apprenticeship at Churchill, June evidenced a few key components to ensuring both fidelity and ongoing momentum that can be used to inform the process of scaling up through RAISE. First, as discussed by Terry and Serena in the section above, June talked about the need to start small, and for teacher leaders to focus first on their own Reading Apprenticeship classroom implementation. Then, she stressed the need for interested teachers to attend a formal training, instead of just gleaning isolated strategies from June and other teacher leaders. Finally, June mentioned supporting other teachers' implementation by means of small groups, which potentially limited how quickly Reading Apprenticeship could spread at Churchill.

Debutte School District took a different approach to spreading Reading Apprenticeship through the RAISE initiative. While a small percentage of teachers did attend the RAISE Institute, Reading Apprenticeship as a framework for instruction was largely being spread by means of district trainings, which were facilitated by literacy coordinators at the district's three comprehensive high schools. Teachers trained at the district attended half-day trainings instead of full days, and people who were not implementing Reading Apprenticeship in a classroom trained them.

In both of their year 2 interviews, Literacy Coordinators Jane and Lucy discussed the complications inherent in their providing these district trainings to teachers, rather than investing the time in sending teachers through the RAISE Institute. Both were Reading Apprenticeship-trained, and had had some opportunities to implement Reading Apprenticeship in the classroom. However, Jane admitted the merits of teachers learning from leaders who were consistently implementing Reading Apprenticeship in a classroom, like Teacher Leader Lauren. Similarly, Lucy admitted that she relied on Jane and the other literacy coordinator to largely support her when they conducted trainings, as these women had deeper understanding of Reading Apprenticeship than she did. Jane also admitted her difficulties in bringing RAISE-trained and district-trained teachers together to collaborate. Jane and teachers at Pershing and Carnegie sensed different levels of understanding of the Reading Apprenticeship framework, which they felt were contingent on whether teachers attended the RAISE Institute. Finally, as previously mentioned, Jane and Lucy evaluated some of the teachers at their school. They provided different support than what a RAISE leader in a non-evaluative position could provide.

Despite these complications, Debutte School District had many viable reasons for training teachers at the district. Due to the Focus and Priority labels, these schools were feeling pressure to show fast improvement. They were also experiencing—like schools across the nation—cuts in budgets and resources. Given dwindling resources, it seemed sensible to build capacity internally by giving literacy coordinators the task of Reading Apprenticeship training and support, instead of relying on external trainings that were only "free" for a limited time during the i3 grant or on already overtaxed teacher leaders. In this high-pressure climate with limited time and resources, it seemed promising to adopt this train-the-trainer strategy, to spread Reading Apprenticeship as quickly as possible. The potential tradeoff, however, was the depth of practice afforded through the RAISE Institute.

How can Debutte schools, facing the pressure of the Focus/Priority labels and afforded limited time and money, spread Reading Apprenticeship in a way that promises rapid buy-in and growth, while not compromising depth of practice, or genuine capacity and improvement?

Debutte needed to act fast enough to assure both internal and external stakeholders that they were taking their Focus and Priority labels and their improvement plans, seriously. Yet they could not act so fast as to spread a version of Reading Apprenticeship that risks being shallow, ineffective, or incomplete.

Challenge 3: Sustaining Support with Diminishing Resources

During the two years of the case study data collection, the Sage schools did not seem to acutely struggle with the tension of building capacity of RAISE leaders and spreading Reading Apprenticeship to a broader base of teachers. This district has a history with Reading Apprenticeship, and though this at times meant that teachers resisted RAISE because they thought they already understood the framework, this history also meant that there was enough time to build depth of practice and a genuine groundswell of teacher interest in Reading Apprenticeship.

Churchill and Twin Lakes did, however, grapple with challenges of their own. These challenges related to navigating the expansion and contraction of implementation across time, in the face of contextual changes and potential dwindling interest in a not-so-new initiative. Both schools dealt with a contraction, Churchill in year 1 of data collection and Twin Lakes in year 2. Though these issues were different from those of Debutte, they also prove challenging in terms of finding a solution that did not, in turn, cause a new set of challenges.

June cited many supports that she and Churchill High School experienced as they implemented Reading Apprenticeship. Across the two years of data collection, June and other teachers repeatedly stated that Churchill administrators were supportive. There had been capable and inspiring leaders at Churchill since the beginning stages of implementation, and the school had a history with both Reading Apprenticeship implementation and with demonstrating its capability to other schools and districts.

Despite these supports, June and Serena alluded to one challenge during the first year of data collection. Though they did receive supports for Reading Apprenticeship implementation, those resources were subject to increasing budgets cuts; they went from having two release periods to work on school-level Reading Apprenticeship implementation to having only one (for June) in 2012-13. Furthermore, because of limited funding for collaboration, June and Serena both noticed a diminished amount of small group collaboration during the 2012-13 school year. This was especially concerning, given that, throughout the two years of data collection, they cited small group work as a necessary support for ongoing Reading Apprenticeship implementation and growth.

> "I think the small groups really make a difference because you feel like you can be yourself, it's not like a staff meeting, and I really think even though it's a small thing, it made a difference in the past, and some of the people that are in RAISE, it's because of that one-on-one, small group comfort. And that they feel like 'wow, I appreciate this because of what you guys have done with it' and it makes people more comfortable. So I feel like those smaller meetings really can make a

> bigger impact in the long run, that's what I found here at Churchill… it creates those personal relationships" (Serena, interview, March 14, 2014).

Constrained by diminishing resources and release time, the leaders made the decision to turn their attention away from providing small group support, and toward providing broader school-level support. Serena described the way in which resources were allocated during the 2012-13 academic year:

> "We'll try to have times where we can meet maybe after school, it's not easy always doing it during the day, just because of scheduling, I would say last year probably a little bit more … activity with [RAISE] meetings than this year. With this year we've had [RAISE activities] more blended into staff meetings… we have Common Core coming up, so our administration has had us guide those meetings…" (Serena, interview, April 24th, 2013).

In her year 1 interview, June also mentioned the support of having administrators who allotted her and Serena time in staff meetings and professional development to discuss Reading Apprenticeship. Additionally, she described a building-wide Reading Apprenticeship activity:

> "Last fall, my retired partner, my principal paid her, as a consultant, to work with me, and we did an election day Read-a-thon for Reading Apprenticeship in the entire building. And our focus was not on who's running for president, our focus was on the nine proposals. We had nine proposals on the ballot, so we designed an entire day of reading the proposals, we did a pretest of the students' knowledge of what they understood, what they knew about the proposals, first hour. We did a posttest last hour, to assess their understanding of the nine proposals, and then we divided, each class period into two proposals, and the three local proposals was the fifth hour, and we designed Reading Apprenticeship lessons, for every single class, and distributed that to every teacher in the building, so that the entire day was devoted to Reading Apprenticeship practices around the election proposals. And it was tremendous, I mean, you talk about building a professional learning community, that was the entire building that was doing the same kind of a lesson, and we gave them Reading Apprenticeship lesson plans" (June, interview, February 12, 2013).

June's excitement about this building-wide event was evident in her recounting of the Reading Apprenticeship day. Serena was similarly positive about the RAISE team's opportunities to discuss Common Core and Reading Apprenticeship in staff meetings. Both leaders viewed such opportunities as supports. Those professional learning events and staff meetings allowed the leaders to spread awareness of Reading Apprenticeship, and stoke many teachers' interest.

However, throughout their interviews, it was clear that neither June nor Serena viewed whole-school activities as sufficient when supporting their school's ongoing work with Reading Apprenticeship. These whole-school activities were a support, but one that needed to be supplemented by small group activities to deepen practice. In her 2014 interview, when reflecting on the 2012-13 school year, June mentioned how she had been "relegated to whole-building kinds of things" and remembers that she was "really beginning to worry about … losing momentum, in the building, and I felt like we needed to get back to some of those smaller groups meetings" (June, interview, March 14, 2014). Churchill's case highlighted the following questions.

How can leaders balance providing school-wide and small group supports to scale and sustain an initiative in a climate constrained by diminishing resources of time and money? Furthermore, how might teachers navigate such a balance when administrators, the district, or the state determine the majority of resource allocation decisions?

The year 2 case studies report (Hegseth, Zacamy, & Newman, 2015) described the ways in which June navigated this challenge, and secured resources for small group collaboration during the 2013-14 school year. Despite June's and Churchill's success, the challenge remains. Diminishing resources inevitably limit choices, and so a balancing act must be negotiated during and across school years. Churchill teachers largely had to rely on whole-school supports in 2012-13, but June—with the help of district and school support—sensed when and how to change to small group supports the following year. Knowing what supports were needed when, and sensing when those supports must (or could) change, proved to be an ongoing challenge that merits continual reexamination.

Challenge 4: Mandating Participation to Ensure a Sustained Effort

The Twin Lakes case introduced the fourth challenge. Twin Lakes seemed to be on firm footing in the first year of data collection. Assistant Principal Shannon and Teacher Leader Lindsey were knowledgeable about Reading Apprenticeship, the supports it requires, and how to connect it to external demands and internal school culture and competing initiatives. Further, Twin Lakes was the only case study school to successfully institutionalize time for RAISE small group collaboration by folding it into SLC meetings.

As strong and steady as Twin Lakes seemed in 2012-13, the second year of data collection brought a new (yet supportive) principal, a loss of the release period for Teacher Leader Lindsey, and a great deal of teacher turnover. Twin Lakes quickly went from a well-supported environment that was infused with Reading Apprenticeship expertise, to an environment experiencing diminishing resources and decreasing knowledge of, and buy-in to, Reading Apprenticeship.

Both Lindsey and Shannon demonstrated their awareness of a key and purposeful strategy for effectively and authentically spreading Reading Apprenticeship:

> "The reality is that it is different when it comes from your peers…I think that's why Reading Apprenticeship in [this county] has seen so much success, is because the way it was rolled out. It was not a top-down rollout or approach at all, it was very much a groundswell of teachers telling teachers, and 'this is what it looks like' and 'here's the data I got in my classroom by implementing these things,' and those kinds of things. … it just can't be top-down, this is what we're doing, and everybody's doing it. And as an administrator now I struggle with that because, seeing both sides of it, but I do struggle with that, because you want to say, 'no I need you to do this' …but if your teachers aren't in a place to receive that, then it's not going to matter" (Shannon, interview, March 13, 2014).
>
> "I think that's one of the reasons why Reading Apprenticeship has been very successful is kind of the buzz that it generates among staff, and when we have the opportunities to actually sit down and talk about what we're doing in our classroom, that's where that sort of thing comes up. And I

> think that's something that we relied on at Twin Lakes for a long time, to be the motivation for people to go to trainings, and then, I don't know, I just feel like we hit a plateau or something. We've also had some staff changeover, where there's the district transfer process that has happened every single year and has brought new teachers into our building, so we've lost a few really motivated, really gung-ho Reading Apprenticeship teachers" (Lindsey, interview, March 12, 2014).

And both participants demonstrated that they were uncertain about how to reconcile their knowledge of how Reading Apprenticeship *should be* spread, in light of Twin Lake's current "plateau" of enthusiasm. When the buzz or groundswell has died down, due to teacher turnover and other contextual factors, what strategies should be used to spread?

> "We're kind of in a position where we're asking ourselves should we make people go [to the Institute]? Because there are people that I think would really benefit from it, and I think would be really receptive to going, as well. I put together the application this year to send a group, and I only had two people that wanted to go. And I know it's a huge commitment…but it's such great PD. I had two people that wanted to go, I sent out multiple requests, several times to the same people, and nobody, that was really frustrating, so I've been consulting with Principal Jasper and Shannon about how to address that, how do we get people involved again?" (Lindsey, interview, March 12, 2014).

Researcher: Lindsey mentioned to me that she's trying to pull together an application for training?

> Shannon: "Yeah, and people were not really interested, which was frustrating because I'm torn on what to do. Because, I know the philosophy is that you're not supposed to require people to do it, because that's been what WestEd has said the whole way through. You don't require, this needs to voluntary, this needs to be something people want to do. But at the same time, I'm at the point where I feel like, no I need to require some people to go, because they're not going to go otherwise, and then we're going to lose momentum … so I'm having a bit of an internal struggle with that, as to how to move forward with keeping that momentum going and keeping the initiative alive" (Shannon, interview, March 13, 2014).

Put simply, Twin Lakes leaders struggled with whether to mandate participation to ensure a sustained effort:

When faced with diminishing enthusiasm, should schools compromise a key "bottom up" recruitment strategy and mandate participation in training, in order to spread the initiative and try to assure its survival?

It is unknown what Twin Lakes eventually decided regarding whether to make participation in the RAISE Institute mandatory for select teachers. If they did not make training mandatory, they risk continuing to implement in an environment that shows diminishing buy-in and knowledge of Reading Apprenticeship. If, however, they did mandate training to continue to spread Reading Apprenticeship through RAISE, they are departing from what has worked in the past to engage teachers. As Shannon said, "you want to say, 'no I need you to do this' …but if your teachers aren't in a place to receive that, then it's not going to matter" (Shannon, interview, March 13, 2014).

LOCAL PROBLEM-SOLVING

The case studies illustrated the wide range of complex problems faced by participants in scaling up Reading Apprenticeship, as well as the local problem-solving efforts of teachers and administrators. We can map these case study observations back to the interest in creating the conditions for scaling through empowering a local team.

While the focus of our study has been on school-level processes during scale-up, it is clear that substantial challenges arise from mandates and decisions made at the district and state level, such as the Focus/Priority designation. Other challenges simply arise from the depth of change and corresponding strength of commitment that integrating Reading Apprenticeship into instruction calls for. We saw, in each case, the school personnel puzzling through challenges and making choices to optimize the reach of the program—or at least preserve it in place.

With Challenge 1, *Being a RAISE Teacher Leader While a Reading Apprenticeship Novice,* we were reminded of the bootstrapping involved in the startup of a new program and the importance of having mentors available. Also, we saw the complexity of simultaneously working on personal expertise—getting the program to work in one's own classroom—while trying to support and maintain the community of RAISE teachers in the school.

With Challenge 2, *Attempting to Both Build Capacity and Scale Up,* we saw clearly the tensions inherent in prioritizing depth of practice and spread concurrently. This was perhaps the most salient issue for the field as we think about the difficulties involved in scaling up new and transformative teaching practices. Reading Apprenticeship requires an investment in terms of staff time and bringing in external expertise for professional learning. If RAISE schools and leaders just focused on building depth with a small group of teachers, they may be sacrificing the spread needed to maintain administrative support and become rooted in the school.

With Challenge 3, *Sustaining Support with Diminishing Resources,* we discussed a school's response to resource cuts that were outside of the school's control. Experienced leaders with fewer resources for small group meetings had to move to whole school activities and presentations for a time. While these may be useful for spread, they were considered less useful for depth of practice.

With Challenge 4, *Mandating Participation to Ensure a Sustained Effort,* we saw local RAISE leaders struggling with the pros and cons of a more organic voluntary approach versus a mandatory approach to sustaining an intervention. While school leaders considered making Reading Apprenticeship training mandatory as a way to maintain its place in the school, they understood the difficulty in requiring a program that has grown best from more grass-roots support.

Many of the challenges presented in these cases resulted from constraints coming from outside the schools. Our study of scale-up maintained a focus on the school-level and did not systematically collect data at the district or state levels. As we return, in this report, to the quantitative analysis of teacher and principal surveys from the larger sample of Scale-up schools, we have to keep in mind the variety of specific challenges and responses that to some degree were different in each of the schools in the RAISE project. While technically we might say that these external factors created noise for the quantitative analysis, we do not want to dismiss their importance. Each of the dots in our graphs represents a large

number of practitioners responding strategically to their often unique local challenges. This chapter provides a glimpse of the concrete, working environment of the broader group of teachers and principals answering our surveys.

Chapter 7: Teacher Teams as Predictors of Scale-up Within Schools

While the case studies of four of the schools provided a window into the processes of local (school and district) problem-solving, we returned to our survey data and participant tracker to see if more general patterns could be found in school-level processes. We were particularly interested in processes that may account for sustaining and growth of participation in a school contrasted with cases where interest trailed off.

We also drew on the results of the analysis of the first cohort of RAISE teachers, which raised questions about the mechanisms for sustaining Reading Apprenticeship implementation in RAISE. We framed this step as an investigation of what contextual conditions or malleable factors may affect scale-up.

A challenge, however, was to identify a quantitative outcome measure of scale, that is, purely in terms of the growth in the number of teachers, schools, or districts. The RAISE recruitment approach—which included both internal spread (within participating schools, districts, and states) and external spread (to new schools or districts)—provided us with an opportunity to measure the growth (or loss) of participation over time. We focused on changes in the number of teachers within schools since we found most of the changes occurred at that level. That is, there were few changes in number of schools in a district or number of districts within a state. This unit was also consistent with our prior analyses conducted at the school level and became a useful quantitative scale-up outcome measure.

QUANTITATIVE ANALYSIS OF SCALE-UP OUTCOMES: "GAIN AND LOSS"

Through our tracking of participation in RAISE training over the three years, we were able to categorize which schools ended up with fewer or with more teachers than were enrolled in the first year training opportunity; that is, which schools gained or lost participants.

Figure 7.1 shows how the 167 schools included in the analysis fell out. This divides the schools into those that lost participating teachers between the initial training and the end of the third year, those that gained participants, and those that stayed the same. In the graph the gray part of the bars, which we label "churn," represents the teachers who were trained but lost to the project (either no longer implementing the program or leaving the school). The "gainer" and "loser" sets of schools started with around the same number of teachers in their first training cohort. In almost one half of the participating schools (77), the number of participating teachers did not change. In 56 schools, RAISE gained participants. In 34 schools, teachers were lost.

A version of this chapter was presented at the annual conference of The National Center on Scaling Up Effective Schools in Nashville, TN (Zacamy, Newman, Lazarev, & Lin, 2015).

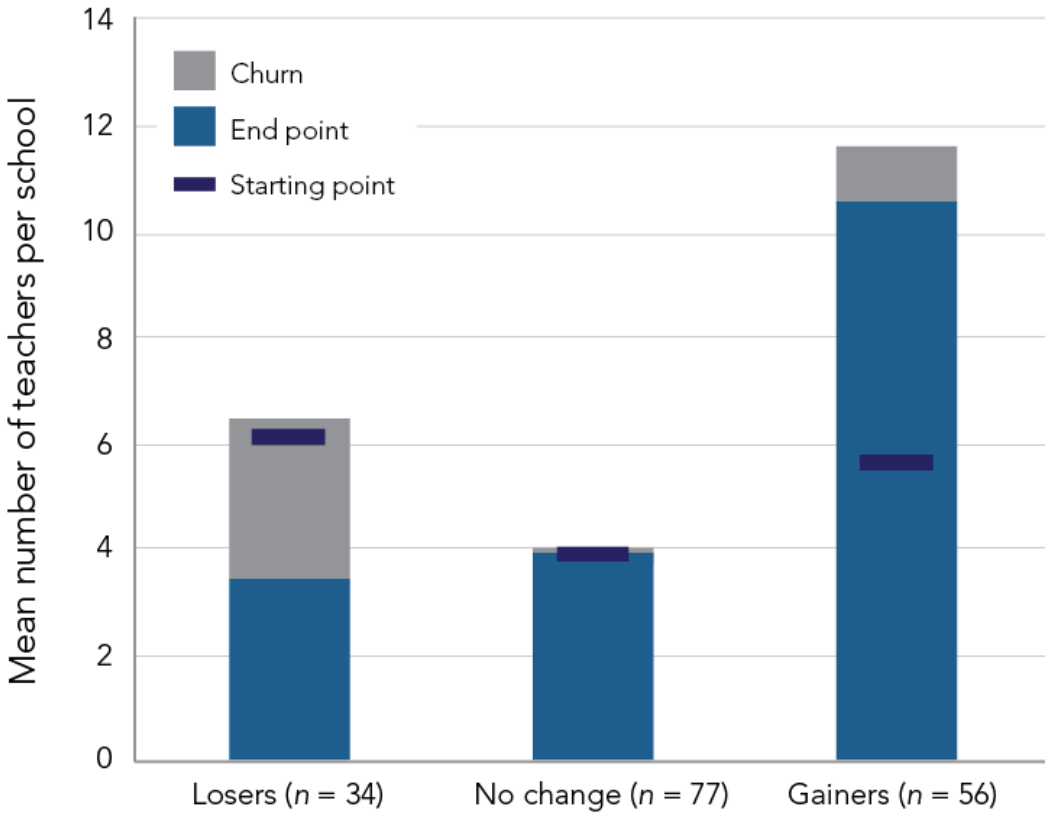


FIGURE 7.1. SCHOOLS THAT LOST OR GAINED TEACHERS DURING THEIR TWO OR THREE YEARS IN THE PROJECT

The difference between the number of participating teachers at the end of the study period and the number of teachers enrolled in the initial training—we call this the GL ("Gain/Loss") metric—is an indicator of a positive or negative scale-up process within the school. We, of course, recognize that this is a rough estimate because, for example, teachers may leave the school for many reasons, and teachers may learn Reading Apprenticeship from fellow teachers without being counted in one of the project training sessions. Nevertheless, the substantial variation among schools suggested that GL may be a productive measure with which to move beyond analysis of survey responses by themselves. Using data from the teacher and administrator surveys and school/district demographic characteristics from the first year of implementation, we would be able to identify contextual and malleable factors that may predict this metric (gain or loss in the number of participating teachers). Our goal is to establish factors that are associated with positive gains, which we consider indicative of a schools' supporting a process with potential for program scale-up.

We used the "participant tracker" as the primary data source for calculating the GL metric. The participant tracker was updated with information as researchers received it. The method for uncovering teachers or schools that were no longer participating in RAISE was primarily survey follow-up or other direct communication with teachers or administrators. Given the number of schools and teachers involved in this initiative, we did not have the resources to track "attrition" as closely as we would have

in a more structured study (e.g., a randomized control trial). Additionally, teachers were not provided with an additional stipend or incentive to participate in the RAISE Scale-up study surveys and response rates declined over the three years. The tracker served the important function of tracking which district, schools, and teachers were participating, but it was not initially designed as a formal data collection tool for research purposes. It does, however, allow us to understand the processes of "attrition" or expansion beyond what is possible with only the survey data. We note that there is a possibility of underestimating both growth and "attrition" or loss using this data source.

The GL metric was calculated for the schools that began RAISE implementation in Cohort 1 or 2 and tracked participation through spring 2014. We coded if the school gained, lost, or had no change in number of participants from their initial training point to the end of that school year. (For schools that started in Cohort 1, we coded the GL metric using data from 2011-2014 and for Cohort 2, 2012-2014.)[11]

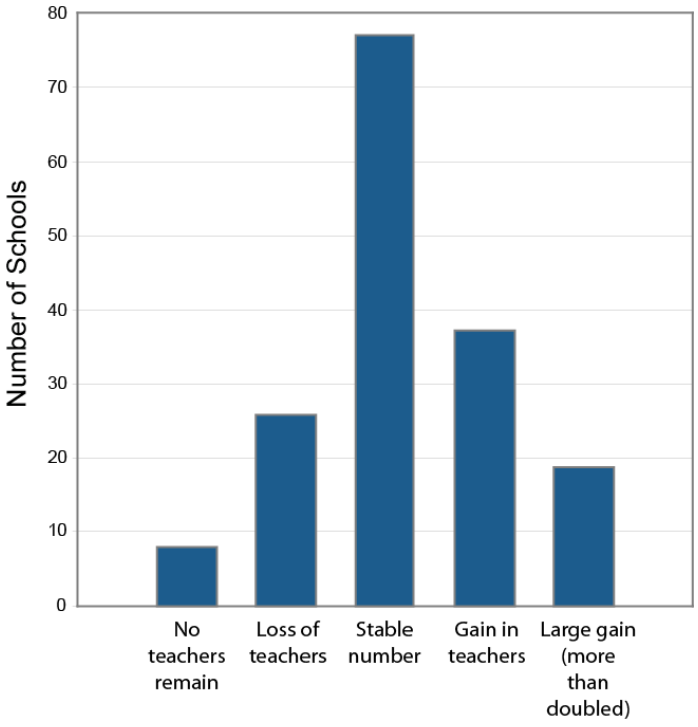


FIGURE 7.2. DISTRIBUTION OF SCHOOLS BY GAIN/LOSS IN THE NUMBER OF TEACHERS PARTICIPATING IN RA

Figure 7.2 shows the distribution of the GL for the 167 schools in the sample described above divided into five GL categories. We see that this metric varies widely, in relation to the initial enrollment, from no teachers remaining (8 schools where none of the teachers reported participating in RAISE by the end of the study) to schools that gained more than twice the number of teachers they started with. There is a positive dynamic overall indicated by the two taller "gainer" bars on the right. A majority of the schools experienced no net change, however, we know that most such schools belong to the second cohort and therefore had less time to change. Our initial observation is that schools vary in the likelihood that the innovation will pick up, rather than lose implementers and that substantial variation warrants a quantitative analysis of the potential predictors of these outcomes.

Potential Predictors of GL

Potential predictors of the GL outcome can be divided into three categories.

[11] Schools joining RAISE for the first time in the third or fourth cohort were excluded from this analysis since there would not be a long enough period of time to measure changes in number of participants. Teachers in the third cohort who joined from schools in the first two cohorts were included in the count.

1. Contextual variables. These are fixed/preexisting conditions such as school size, and student demographic characteristics. Including these variables would allow researchers and program developers to understand which contexts may be more fertile than others for RAISE taking hold.
2. Malleable conditions resulting from the RAISE development activities. These can be considered indicators of program implementation, and reflect the activities of the RAISE project, such as attendance at monthly RAISE team meetings. Results for these may help program developers with resource allocation, if we can identify which malleable factors will predict longer term growth of the initiative.
3. Intermediate scale-up outcomes. These are indicators of buy-in, commitment, ownership, etc. that result from implementation and training in fertile contexts.

Table 7.1 lists the potential predictors and the data sources drawn on for analysis.

TABLE 7.1. POTENTIAL PREDICTORS OF GL

Potential Predictor	Data Source
Contextual Factors	
School size (number of enrolled students)	NCES
Student-teacher ratio	NCES
% of students enrolled in free or reduced priced lunch	NCES
% of students in each ethnicity category	NCES
Level of agreement with the statement that the principal is involved in district decisions about educational issues (1-5 scale)	Principal survey
Level of agreement with the statement that the school receives instructional resources commensurate to meet needs of students (1-5 scale)	Principal survey
Level of agreement with the statement that district provides school with sufficient data/information to make informed decisions (1-5 scale)	Principal survey
Years served as principal	Principal survey
Principal turnover (# of head principals the school had in past five years)	Principal survey
Teacher turnover (level of agreement with the statement that teachers tend to stay at school for more than five years; 1-5 scale)	Principal survey
Principal retention (level of agreement with statement that the principal's immediate plan is to stay at their school; 1-5 scale)	Principal survey
Teacher retention (how current retention rate compares to last five years; 1-5 scale)	Principal survey
Years of teaching experience	Teacher survey
Rating of school leadership (averaged rating about shared vision about education, atmosphere of trust and respect, comfort raising issues and concerns, school leadership support, and that teachers are held to high professional standards for delivering instruction; 1-5 scale)	Teacher survey

TABLE 7.1. POTENTIAL PREDICTORS OF GL

Potential Predictor	Data Source
Malleable Factors	
Principal attendance at training (principals were not required to attend)	Principal Survey
If participating in RAISE was required	Teacher Survey
Attendance at RAISE monthly school meetings (# of meetings attended)	Teacher Survey
Level of agreement with the statement that teacher has enough time to plan Reading Apprenticeship lessons (1-5 scale)	Teacher Survey
Use of Reading Apprenticeship practices (0-5 scale)	Teacher Survey
Intermediate Outcomes	
If principal recommended joining RAISE to other teachers or personnel	Principal Survey
If principal thinks RAISE would continue in their school without federal funding	Principal Survey
Level of responsibility for the success of Reading Apprenticeship at their school (1-4 scale)	Teacher survey
Level of commitment to making Reading Apprenticeship work in school (1-5 scale)	Teacher Survey
If principal completed survey (indicator of commitment)	Principal Survey

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.

GL Analytical Model

Our analytical approach to modeling the GL outcome takes into account that GL depends potentially on the factors listed above and on the enrollment in the initial training. Enrollment in the initial training may depend on the same contextual factors as GL but not on the intermediate outcomes. We used a conventional approach for this type of situation involving endogenously determined variables: estimating a two-stage regression (2SLS). In the first-stage equation, initial enrollment is regressed against contextual variables—school characteristics—and data on the teachers' prior acquaintance with Reading Apprenticeship, which can be considered independent of the current RAISE implementation. Predicted values of initial enrollment are included in the second-stage equation—the equation of our primary interest—which regresses the GL metric on all three types of covariates. In addition, we included in both equations two school-size factor variables (log of 9th grade enrollment and the schoolwide student-teacher ratio) to account for the natural limits to program growth set by the school size.[12]

[12] We found that school size factors affect the initial enrollment but do not affect the GL metric directly.

Our analytical sample included 130 schools—all schools with both NCES data and teacher survey data available (78% of the total). In a large proportion (60%) of the schools in the analytical sample, principals did not respond to the survey, which was expected to be an important source of GL predictor variables. Consequently we estimated two types of our two-stage model: Model 1 without principal survey responses, using the full analytical sample; and Model 2, using a smaller sample, with principal survey variables included. Since most survey items were not statistically significant in the estimated Model 2, we estimate two additional "shrunk" versions of Model 2: one (Model 2s) including only those teacher survey items that are estimated with a *p* value of .25 or lower, and another (Model 2sm) with the number of covariates further reduced by removing the least significant terms in Model 2s (*p* value greater than .5).

We also ran the same analysis with a dataset in which both cohorts were limited to two years of GL data. This gives a shorter term perspective but makes the two cohorts equivalent. This alternative revealed only minor changes to the results.

WHAT WE LEARNED FROM ANALYSIS OF GL PREDICTORS

Our analyses produced several interesting results, some that point to key mechanisms for spread and others that raise additional questions. These are presented in Tables 7.2 and 7.3, which are based on analytical model 2. Model 2's results are consistent with the alternative models'. See Table C1 in Appendix C for the full set of results.

Contextual Factors

Table 7.2 shows the results for contextual variables, which as a group had only marginal significance as the predictors of the GL. The size of the school, unsurprisingly, is positively associated with GL, but the relationship is weak (*p* values in the range of .09-.14). The association with school ethnic composition represented by the percentages of the three major ethnic groups in the student population is weak: all three ethnic group variables are significant only in one of the models (Model 2sm), and the differences are not substantial.

TABLE 7.2. CONTEXTUAL FACTORS AS PREDICTORS OF GL

Factor	Results for model 2	
	Estimate	*p* value
Contextual Factors		
School size (number of 9th grade students enrolled)	3.00	.09
Student-teacher ratio	-3.18	.43
% of students enrolled in free or reduced priced lunch	-2.10	.57
% of students in each ethnicity category		
% Hispanic	20.61	.14
% Black	24.23	.05
% White	19.90	.07

TABLE 7.2. CONTEXTUAL FACTORS AS PREDICTORS OF GL

Factor	Results for model 2	
	Estimate	*p* value
Level of agreement with the statement that the principal is involved in district decisions about educational issues	- 0.83	.23
Level of agreement with the statement that the school receives instructional resources commensurate to meet needs of students	- 0.40	.48
Level of agreement with the statement that district provides school with sufficient data/information to make informed decisions	0.42	.46
Years served as principal	- 0.15	.03
Principal turnover (# of head principals the school had in past five years)	- 0.15	.74
Teacher turnover (level of agreement with the statement that teachers tend to stay at school for more than five years)	- 0.48	.58
Principal retention (level of agreement with statement that the principal's immediate plan is to stay at their school)	- 0.77	.05
Teacher retention (how current retention rate compares to last five years)	- 0.13	.80
Years of teaching experience	- 0.03	.83
Rating of school leadership	1.01	.27

Other school context variables or school characteristics—including socioeconomic status, principal leadership, teacher turnover at the school (principal level of agreement that teachers tend to stay at the school for more than 5 years), or changes to teacher retention (as reported by the principals)—are not significantly related to the GL. District contextual factors—such as resource allocation, principal involvement in district decisions, and the district providing schools with data to inform decisions—also have no significant relationships to the GL. This suggests that Reading Apprenticeship is equally scalable across all types of communities covered by this study.

We did find a significant negative relationship between two principal survey variables and the GL outcome: years of experience serving as a school administrator and principals' plans to stay at the school (considered a proxy for the principal future/ possible turnover). Surprising as they are at first glance, these relationships can be explained by a greater enthusiasm for Reading Apprenticeship (resulting in higher GL numbers) among younger principals who are characterized by lower years of experience and lower commitment to staying at the same school for a longer period of time.

Malleable Factors and Intermediate Outcomes

In Table 7.3, we find results that suggest establishing a community of practice early on (measured through data on attendance at monthly meetings and commitment to making it work in the school) supports longer-term spread and sustainability. We found that teachers' level of commitment to making Reading Apprenticeship work in their school and their attendance at the RAISE monthly team meetings in their first year of implementation predict whether the initiative grows in their school. We also found that principal involvement, as indirectly indicated by whether they completed a survey for the study, is a strong predictor of teachers joining the initiative (positive association with the GL in Model 1).

TABLE 7.3. MALLEABLE FACTORS AND INTERMEDIATE OUTCOMES AS PREDICTORS OF GL

Factor	Model 2	
	Estimate	*p* value
Malleable Factors		
Principal attendance at training	-0.23	.57
If participating in RAISE was required	-1.75	.25
Attendance at RAISE monthly school meetings (no. of meetings attended)	0.60	.02
Level of agreement with the statement that teacher has enough time to plan Reading Apprenticeship lessons	-0.16	.83
Use of Reading Apprenticeship practices	0.00	.99
Intermediate Outcomes		
If principal recommended joining RAISE to other teachers or personnel	-0.48	.66
If principal thinks RAISE would continue in their school without federal funding	0.48	.70
Level of responsibility for the success of Reading Apprenticeship at their school	0.95	.22
Level of commitment to making Reading Apprenticeship work in school	2.21	.03
If principal completed survey (from model 1)	1.84	.01

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.

We did not find that malleable factors related to implementation activities were related to the GL. The finding that it made no difference if teachers had enough time to incorporate Reading Apprenticeship into lesson plans, or how frequently teachers used Reading Apprenticeship practices in their lessons, may suggest that, in the initial phases of implementation, the extent of actual use of Reading Apprenticeship in class is less important than the teacher buy-in. Additionally, findings that it did not matter for the GL if the principal recommended Reading Apprenticeship to others in their school or if

they believe Reading Apprenticeship will continue without federal funding present Reading Apprenticeship as a self-sustaining teacher-driven reform that does not require a lot of administrative input or financial support, which makes it a potentially viable option for schools with limited resources.

IMPORTANCE OF TEACHER COLLABORATION

Our approach to scale-up is to consider the intermediate processes that may be kicked off by the program training and direct support but take hold in a school among teachers and the principal. We take an increase or decrease in number of teachers sent to training as a measurable indicator that a positive or negative process is taking place. Instead of just measuring the total number of teachers "reached" by Reading Apprenticeship's professional training, we also considered the loss of teachers who had gone to training but left the school or discontinued the program. Scaling-up in this view is a net gain of teachers where success requires that the challenges, such as competing priorities, are outweighed by positive experience with the program. Our Gain/Loss (GL) metric is useful because it points in both directions so it accounts for the dynamic tension between the scaling up of a program and the loss of momentum, which can be simultaneously present.

We considered this measure as a proxy for processes within the school that support or detract from the program's sustainability. We then examined not only the survey results from the first year of the school's participation, but also the contextual variables that characterized the school in order to find predictors of the GL metric. Our findings provide the program developers with hypotheses as to conditions for success of their program, as well as suggestions for focus of the intervention. Insofar as the survey questions map to the categories that researchers have pointed to, these results can be seen as providing support or raising questions about processes, at least as applied to Reading Apprenticeship.

As a broad conclusion from this study, we see that a community of practice matters. A predictor of additional teachers joining the team in future years is the extent to which teachers are committed to making the Reading Apprenticeship work at the school. This is a better predictor than the extent that teachers implemented Reading Apprenticeship practices in their own classroom early on. We also see that attendance at the monthly meetings (school mean for teacher responses) in the first year predicts new teachers joining in subsequent years. This is interesting because we also saw a precipitous decline in attendance during the year, and the decline continues into subsequent years. It may be that meetings are essential in establishing the community but don't provide a useful form of support once established.

The role of the principal in promoting the processes detected by GL is interesting. The survey response rate for principals in the first year was 54%, but that in itself was a strong predictor for later gains. A level of interest is important but there was no relation to the principal attending the training. A principal directly suggesting that teachers join Reading Apprenticeship was also unrelated to gains, suggesting that the teacher team is what is important. There was an indication that less experienced principals—and those unsure about their tenure at the school—predict gains, which may suggest that schools with established veteran principals are not as fertile ground for new programs like Reading Apprenticeship.

We found that school size has little effect, if any, on the scale-up. School characteristics such as percent free or reduced-price lunch and percent minority were not strong predictors of GL. This suggests that Reading Apprenticeship can succeed across diverse communities.

We chose to conduct this analysis at the school level, since that was the focus of most of our data collection. The choice of the school was also related to the nature of the program, which saw the school community as critical for successful implementation. Our larger theory of action recognizes the important processes occurring at the district and state levels, especially when it comes to institutionalization and ensuring continued funding. The GL method can be applied with data from district administrators where a scale-up project encompasses a large number of districts. In the current project, few districts added or lost schools.

The Scale-up study didn't consider student achievement as a variable. Working in four states, each with different sequences of high-school student assessments obtaining achievement data was not feasible. The RCT that ran in parallel to the scale-up work used a researcher-administered test, a strategy that was not feasible with the much larger scale-up school sample. School-level economic-status data (Free and Reduced-Price Lunch) is a usual correlate of test scores and we found that poor schools were just as likely to see growth of participation as middle-class schools. A natural question is whether schools that were successful in scaling-in were also more likely to see achievement gains in the use of Reading Apprenticeship than schools where enthusiasm for the program apparently waned. We turn now to address the question of whether the impact estimates from the RCT can be applied to the Scale-up schools.

Chapter 8. Can Randomized Control Predict Impact in Scale-up Schools

This chapter examines the strategy inherent in the i3 program, which calls for rigorous tests of the effectiveness of each innovation in which the program invests. The idea behind a large-scale RCT is that the program developer is unable to provide extraordinary support to the schools so the estimated treatment effect can be generalized to schools that use the program under more ordinary circumstances. Usually, this is framed as an issue of external validity where the similarity of the schools in the RCT is compared with those in a target population in terms of demographics and other pre-existing characteristics (Cronbach, 1982; Shadish, Cook, & Campbell, 2002; Tipton, 2013). We are asking a somewhat different question: are the supports and processes in the schools, which we have examined closely in this report, comparable between the RCT treatment group and the RAISE Scale-up schools?

We took advantage of an unusual situation in which we were able to collect the same data from teachers and principals in the treatment group of an RCT and from teachers and principals involved in an implementation of the same program outside of the RCT but in very comparable schools. We drew on results reported by Fancsali, et al. (2015) and additional analyses for the RAISE RCT. The comparison of these contexts provided a demonstration of the differences in participant engagement and an approach to estimating the likely impact of the program when implemented outside of an RCT. We could not directly compare student growth between RCT and Scale-up schools because of the inherent differences in the study design and unavailability of a uniform outcome measure in the latter (like the ETS assessment administered by the researchers in RCT schools). We did, however, have a rich set of teacher and principal survey questions, which were asked of both groups, as a basis for our analysis. We were also able to analyze the student achievement data from the RCT at the teacher level to produce value-added (VAM) scores so as to identify, from teacher surveys, the characteristics associated with effective teaching in the Reading Apprenticeship context.

We addressed three questions.

1. Based on survey questions asked of both RCT and Scale-up teachers, what are the differences between RCT and Scale-up schools in the internal school processes and resulting teacher beliefs and motivations?
2. For the teacher characteristics found in the RCT treatment schools to be associated with the teacher's contributions to student achievement (VAM scores), do we find that these characteristics also differentiate between the RCT and Scale-up schools? If the characteristics differentiating the two contexts are associated with student achievement outcomes, this could provide indirect evidence of whether Scale-up schools may have more, less, or the same level of program effectiveness as found in the RCT.
3. For the teacher characteristics found (in Chapter 7) in the Scale-up schools to be associated with gains in participating teachers, do we find that these characteristics also differentiate between the RCT and Scale-up schools? If the characteristics differentiating the two contexts are associated with scaling-in, this could provide indirect evidence of whether RCT schools may have more, less, or the same level of the hypothesized scaling-in processes associated with teacher teams we identified in Chapter 7.

A version of this chapter was presented at the annual conference of the Society for Research in Educational Effectiveness in Washington D.C. (Newman, Lazarev, & Zacamy, 2016).

ECOLOGICAL VALIDITY

We introduce a term, *ecological validity*, as a way to frame our comparison of the RCT and Scale-up schools. We interpret this term in a specific way: an experiment's ecological validity can be threatened if the controls or processes—put in place to assure internal or external validity—result in a deviation from how the program would be implemented outside of an experiment. Gibson (1966) provides the classic example in the study of visual perception where he compares the laboratory apparatus in which the subject is stationary with how the senses are used in a natural setting where the subject is moving his eyes and body. In the same spirit, Cole and his colleagues (Cole, Hood, & McDermott, 1978; Newman, Griffin, & Cole, 1989) questioned the ecological validity of laboratory-based cognitive psychology as a study of the thinking and problem solving occurring in classrooms.

Shadish, Cook, and Campbell (2002) consider ecological validity as a kind of external validity, but it may be more useful to view it more specifically in terms of the effect of the experimental design itself. Nobody will argue that an RCT is not superior to the comparison of two intact groups in terms of internal validity. But, it is possible that the RCT imposes constraints and involves processes that do not occur outside of the experiment. Since these constraints may be salient to the practitioner and stakeholders of the evaluation, it is important to recognize them in the explanation of rigorous design, otherwise, the value of the resulting evidence may be lost in translation.

The analysis reported here addresses the question: In the context of projects funded by programs such as i3, where a rigorous evaluation accompanies the scale-up of the same innovative program, is it possible to provide systematic evidence that an RCT is under- or over-estimating the impact that the program has when implemented outside of the trial? A common approach to scale-up research is to consider the problems of implementation when a program is evaluated in a large-scale trial. We are asking a different question: do the processes involved in recruiting, training, supporting, randomizing, and measuring for the RCT have a positive or negative impact on the outcomes? Can we identify mechanisms by which those impacts may be occurring?

COMPARING THE SCALE-UP SCHOOLS TO THOSE IN THE RCT

We can compare teachers and schools in the RCT's treatment group to other participants implementing the same program under ordinary conditions of recruitment and program implementation. Where an RCT makes use of a state administered test as a student outcome measure, the same measure can be obtained from the Scale-up sites and a quasi-experiment could be considered. In our case, the RCT used a researcher-administered outcome measure, whereas the Scale-up schools studied were spread over four states, making a direct comparison of outcomes impossible. Where both groups are surveyed, as in the current study, differences in levels of implementation or engagement can provide a contrast with the implication of potential differential impacts. This study takes several additional steps in identifying productive mediating processes in the RCT and measuring the presence of those in the comparable Scale-up schools. Our goal is to approximate the potential positive or negative impacts of the ecological invalidity of our RCT. Our goal is also to illustrate an approach to improving the validity of rigorous tests of program effectiveness.

Since the evaluation of SLI's RAISE project consisted of an RCT with 42 schools in Pennsylvania and California (Fancsali, et al., 2015) and this formative scale-up study of an additional 239 schools in four states, including Pennsylvania, we could compare schools in the treatment group of this RCT with schools implementing the same program outside of the constraints of randomized control. We put a special emphasis on the schools in Pennsylvania, where teachers in the RCT and Scale-up schools were trained in the same summer institutes and shared other project supports such as the same State Coordinator. In Pennsylvania, we analyzed data from the 11 treatment schools participating in an RCT and the 31 schools implementing the same intervention in the scale-up group. Table 8.1 provides a comparison of characteristics of the scale-up and RCT schools in the whole sample and in the Pennsylvania subsample.

TABLE 8.1. CHARACTERISTICS OF THE SCALE-UP AND RCT SCHOOLS IN THE WHOLE SAMPLE AND IN THE PENNSYLVANIA SUBSAMPLE

Variable	RCT	SU	*p* level, t-test
All schools			
School size (total students)	1349	1040	0.11
Student-teacher ratio (school total)	18.2	19.2	0.45
Percent free and reduced lunch (grades 9-11)	38.3	38.3	0.99
Percent minority students (grades 9-11)	48.9	23.2	0.00
Average years of experience (participating teachers)	10.7	11.4	0.52
Pennsylvania only			
School size (total students)	987	1196	0.29
Student-teacher ratio (school total)	13.9	15	0.1
Percent free and reduced lunch (grades 9-11)	29.1	24.8	0.57
Percent minority students (grades 9-11)	18	17.7	0.96
Average years of experience (participating teachers)	10.2	10.5	0.82

As described in detail in Chapter 2, recruiting for RCT and scale-up differed substantially. Constraints on recruiting in the RCT, compared to the Scale-up schools, introduces differences in selection that make a comparison challenging. In the RCT, the target number of schools and teachers was determined through a power analysis, while on the scale-up side, the target number of schools and teachers was set in the proposal in terms of the number that would be reached in the project. For Scale-up schools, recruitment was conducted over four years with four cohorts of schools and teachers, while recruiting for the RCT was completed in the first two years so as to allow teachers in the treatment schools to be in the program for two years. In Scale-up, new schools were added in each cohort and in many cases, new teachers were trained in schools that had joined in the prior cohort. A constraint on the RCT schools is also notable: teachers from a participating school could not have participated in Reading Apprenticeship

training prior to the project, whereas a number of Scale-up schools included some where teachers may have already been trained in Reading Apprenticeship and had been implementing the practices for some years.

DESIGN, DATA COLLECTION AND ANALYSIS

The analysis we report here is a comparison of schools from the RCT and scale-up contexts where the outcomes are teacher and principal self-reported surveys. We do not have a baseline for the outcome measure but have documented commonalities between the two groups. In order to increase comparability, all schools were from the same state and all teachers were trained in the same summer workshops. We proceeded with the following steps.

1. Perform descriptive comparison of the two contexts, and determine if there are significant differences between the treatment group of the RCT and schools participating in Scale-up.
2. Estimate a linear regression model of student outcomes (at the teacher level) in the RCT to determine if any of the differentiating characteristics identified in the first step are associated with the student outcomes. This should lead to a narrower set of characteristics relevant to the questions for this analysis.
3. Further explore associations between the relevant characteristics identified in the second step and other available indicators that may illuminate the mechanisms driving the differences between the two contexts (if any).

Data

The data used in this study included the following sets.

1. *Fidelity of implementation metrics used in the RCT to characterize the fidelity of the treatment condition to the goals of the program developers*. Variables of interest included teacher attendance of monthly meetings and total days of training. These were based on tracking participation of teachers and their survey responses.
2. *Teacher characteristics*. In this study, these are limited to years of experience and subject taught.
3. *Self-reported teacher survey responses*. A large number of the same questions were asked in both contexts. An overlapping subset of survey items asked in the spring in both contexts was used in the analysis (see Appendix D for list of questions). Most items were Likert-scaled. Two groups of items—Level of Preparation and Level of Confidence in using Reading Apprenticeship—were each aggregated into a single variable by averaging the scores. One item—Support for literacy instruction—was based on "check all that applies" and encoded as total count of checked types of support.

A large number of survey questions did not overlap between the two contexts: the RCT put greater focus on classroom practices, while Scale-up included many additional questions about administrator participation.

Analysis

The first step in the analysis involved performing a t-test for the differences between teachers in the two groups of schools and identifying characteristics with significant differences between the two groups of schools (RCT and Scale-up).

The second step required constructing a teacher-level student outcome measure – teacher value-added. We used a conventional approach, in which value-added was calculated as teacher fixed effects, T, in the student-level regression of student outcomes (posttest) on pretest and student characteristics:

$Y_t = \alpha Y_{t-1} + T + \beta X + \varepsilon$.

A study-administered student assessment developed by ETS was used for both pretest and posttest.

The main analysis in the second step involved performing linear regression analysis, with the teacher value added as the outcome and a variety of teacher characteristics as covariates:

$T = \beta Z + \varepsilon$. We include only teacher outcomes and teacher characteristics in Z in this model.

The third step involved analyzing linear regression models with relevant teacher outcomes identified in step 2 as the left-hand side variables and metrics of fidelity of implementation and teacher characteristics as covariates.

DIFFERENCES BETWEEN RCT AND SCALE-UP SCHOOLS

Comparative descriptive statistics for teachers in RCT and Scale-up schools revealed a number of differences suggesting better teacher outcomes in the Scale-up group (Table 8.2). The pattern of differences is similar in Pennsylvania and across all schools.

TABLE 8.2. DIFFERENCES BETWEEN RCT AND SCALE-UP SCHOOLS

Variable	RCT (treatment)	Scale-up	*p* value (t test)
All schools			
Confidence in using Reading Apprenticeship practices	3.54	3.62	.25
Level of preparation	3.54	3.71	.01
Total days of training	9.03	9.15	.52
Years of teaching experience	10.80	11.53	.36
Level of commitment to the success of RAISE in school	2.56	4.22	< .01
Level of understanding of Reading Apprenticeship	3.30	3.42	.05
Level of effectiveness of Reading Apprenticeship in improving student achievement	3.72	4.08	< .01
Level of alignment of Reading Apprenticeship with class goals	2.20	2.64	< .01

TABLE 8.2. DIFFERENCES BETWEEN RCT AND SCALE-UP SCHOOLS

Variable	RCT (treatment)	Scale-up	*p* value (t test)
Level of alignment of Reading Apprenticeship with class standards	2.20	2.55	< .01
Support for literacy instruction	.94	.80	.21
Number of monthly meetings attended	5.24	5.78	.05
Pennsylvania only			
Confidence in using Reading Apprenticeship	3.65	3.65	.98
Level of preparation	3.62	3.67	.57
Total days of training	9.27	8.93	.16
Years of teaching experience	10.09	10.91	.42
Level of commitment to the success of RAISE in school	2.43	4.14	< .01
Level of understanding of Reading Apprenticeship	3.28	3.45	.04
Level of effectiveness of Reading Apprenticeship in improving student achievement	3.57	3.97	< .01
Level of alignment of Reading Apprenticeship with class goals	2.20	2.64	< .01
Level of alignment of Reading Apprenticeship with class standards	2.20	2.55	< .01
Support for literacy instruction	.94	.80	.21
Number of monthly meetings attended	5.80	6.14	.33

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.

The striking difference was the level of commitment to the success of RAISE in the school. Note that this question was not about the teacher's personal success but about helping other teachers to be successful. Differences in the level of understanding of Reading Apprenticeship and teacher perception of effectiveness of Reading Apprenticeship in improving student achievement are also substantial. At the same time, there are no significant differences in the fidelity of implementation, level of preparation, or teacher confidence in using (various aspects of) Reading Apprenticeship.

RELATION TO STUDENT ACHIEVEMENT

The strong positive results on several variables in the Scale-up schools we analyzed may suggest that those schools might also result in better student outcomes. However our analysis of the correlates of student outcomes (teacher VAM) shows that only teachers' confidence in their ability to implement the program in the classroom has a significant positive association with student outcomes (Table 8.3).

TABLE 8.3. CORRELATES OF STUDENT GROWTH (COEFFICIENT ESTIMATES FROM LINEAR REGRESSION MODEL OF TEACHER VALUE-ADDED.)

Variable	Estimate	*p* value
Confidence in using Reading Apprenticeship	0.20	.01
Level of commitment to the success of RAISE in school	- 0.09	.18
Level of understanding of Reading Apprenticeship	0.05	.36
Level of alignment of Reading Apprenticeship with class goals	0.01	.95
Level of alignment of Reading Apprenticeship with class standards	0.02	.87
Level of effectiveness of Reading Apprenticeship in improving student achievement	- 0.02	.81
Years of teaching experience	0.00	.28
Support for literacy instruction	- 0.01	.58
Subject: Biology	0.02	.74
Constant	- 0.50	.17

Note. R-squared: 0.13. Sample: all RCT treatment schools. RAISE stands for Reading Apprenticeship Improving Secondary Education.

The level of confidence is responsible for about 2/3 of explained variation in the VAM scores of the RCT treatment teachers and is a substantial correlate of student outcomes. Based on the results in Table 8.2 and the analysis of teacher VAM distribution, one Likert scale-point difference in level of confidence is associated with 31 percentiles of teacher VAM distribution.

Exploration of Level of Confidence

Although we didn't find evidence of a difference in the level of confidence between RCT and Scale-up teachers, the possible strong role for this indicator as a proximate cause of student outcomes suggests the need to take the next step and explore possible associations between teacher confidence and other characteristics. One can hypothesize a list of factors that may lead to higher level of confidence: formal training, better understanding in general that comes with experience and prior acquaintance with the program, and instructional support. We set up a linear regression model with level of confidence as the outcome to test these hypotheses using available data. We used all available teacher data for this analysis including a binary context indicator (one for Scale-up, zero for RCT) in order to identify possible differences that are not detectable via the simple difference testing.

Results of this analysis (presented in Table 8.4) show that level of preparation, level of understanding, and instructional support are all significantly associated with teacher confidence. The first of these variables is more substantial quantitatively than instructional support. There is no significant difference between Scale-up and RCT schools.

TABLE 8.4. REGRESSION MODEL OF CONFIDENCE (SCALE-UP AND RCT TOGETHER)

Variable	Estimate	*p* value
Level of preparation	0.44	< .01
Scale-up (or RCT) sample	- 0.04	.52
Level of understanding of Reading Apprenticeship	0.23	< .01
Support for literacy instruction	0.05	.04
Years of teaching experience	0.00	.73
Subject: History	- 0.12	.04
Subject: Biology	- 0.33	< .01
Constant	1.30	< .01
Note. R-squared: 0.34.		

Preparation and Understanding of Reading Apprenticeship

Level of preparation and level of understanding are essentially intermediate teacher outcomes. We conclude our exploration with regression analyses of these variables. The results in Tables 8.5 and 8.6 show that only level of understanding is significantly different between teachers in Scale-up and RCT schools, but this difference is not quantitatively substantial. This difference has no potential to substantially affect teacher confidence and ultimately student outcomes.

TABLE 8.5. FACTORS ASSOCIATED WITH LEVEL OF PREPARATION

Variable	Estimate	*p* value
Total days of training	0.01	.60
Number of monthly meetings attended	- 0.01	.22
Scale-up (or RCT) sample	0.04	.60
Years of teaching experience	0.00	.67
Support for literacy instruction	0.07	< .01
Subject History	- 0.30	< .01
Subject Biology	- 0.31	< .01
Constant	3.76	< .01
Note. R-squared: 0.08		

TABLE 8.6. FACTORS ASSOCIATED WITH LEVEL OF UNDERSTANDING OF READING APPRENTICESHIP

Variable	Estimate	*p* value
Total days of training	0.00	.94
Number of monthly meetings attended	0.02	.09
Scale-up (or RCT) sample	0.16	.03
Years of teaching experience	0.00	.61
Support for literacy instruction	0.03	.11
Subject History	- 0.19	< .01
Subject Biology	- 0.17	< .01
Constant	3.24	< .01
Note. R-squared: 0.03		

Results from this analysis—related to student achievement—show that although substantial differences in teacher outcomes are observed between Scale-up and RCT, there is no evidence that these differences would affect differential student outcomes in the two contexts. Closer examination of classroom practices that mediate student achievement, identified in the RCT analysis (Fancsali, et al., 2015), were not used in our analysis because the relevant survey questions were not asked of the Scale-up teachers. As a result, while we identified areas of potential threat to ecological validity, we had only one indirect indicator of teacher effectiveness to work with (confidence in using the program).

DIFFERENCES BETWEEN SCALE-UP AND RCT RELATED TO SCALING-IN

We can look in the other direction and consider whether RCT schools had characteristics of schools that gained new participants uncovered in Chapter 7. The strongest predictor of scaling-in was the level of commitment to making Reading Apprenticeship work at the school and attendance at monthly meetings during the initial year. As we saw in Table 8.1, both of these characteristics were found more in Scale-up than RCT. We noted that these characteristics are major features of the RAISE scale-up model and a reason to consider the approach to have been successful in promoting scaling-in. So, while the Scale-up and RCT treatment schools were equivalent in the teacher characteristic associated with student achievement, the Scale-up schools on average—that is, including both schools that gained and those that lost participants—had more of the characteristics associated with the program taking hold. With the substantial differences between the settings, our findings raise the question of how to attribute the impact estimate from the RCT to the schools adopting the program outside the constraints of an experiment.

Chapter 9. Conclusion

Our study of the scale-up of Reading Apprenticeship was both a formative effort providing SLI with feedback as the basis for continuous improvement and research into school-level processes associated with the success of scaling up. These strands came together in the finding that establishment of a teacher team with school-level commitment to success was the strongest predictor of gains in participation, or "scaling-in", within the school. This report contributes practical recommendations to program developers and approaches to the scientific study of scaling up.

CONTRIBUTION TO CONTINUOUS IMPROVEMENT

RAISE works. When we examine the innovation that SLI was testing in bringing Reading Apprenticeship to scale, we find that creating a school-level teacher team was the strongest correlate of later gains in teacher participation at that school. As we outlined, the idea of having teachers from multiple disciplines engaged in the program and the support provided by school-based teacher leaders followed the theory that a site-based teacher team would be an essential part of bringing Reading Apprenticeship to scale. Rather than working—as in earlier RCTs—with small groups of intervention teachers and providing them with more responsive support, in this situation the SLI program developers were working at a level of scale involving hundreds of schools. This, in combination with their theory of action that promotes site-based team work to extend professional development impact, meant that they built into the intervention a structure that empowered local teams.

Where the internal team and commitment to school-wide implementation was found, the program was more likely to take hold and scale-in. The problem is that not all schools increased participation of their teachers. We found that 20% of the schools lost participants, and some lost all their participants. We do not have an explanation for why there was this difference in scaling-in. We were unable to find a relationship to pre-existing conditions such as student economic status, minority status, or school size. Our case studies of four of the schools indicated a wide variety of conditions at the district level and above that could influence the likelihood that Reading Apprenticeship would take hold. These were not fully accounted for in our quantitative analyses and so represent sources of noise through which the signal of teacher teams nevertheless showed through.

Our research can only provide working hypotheses that future implementation of Reading Apprenticeship can make use of. Importantly, we have no reason to suspect that the program works less well in more disadvantaged settings. And at the school level, it was equally likely to take hold in the four states in which it was tried. Our findings suggest that a focus on building teacher teams with a commitment to the program in the school (not just their classrooms) in the initial year may be the best way to get Reading Apprenticeship to scale-in. This includes assuring time for site-based meetings and collaborative planning. Continuing these meetings once the team has solidified may not be a critical element, and the precipitous drop in meeting participation we observed over time did not appear to be a problem for the sustainability of the program.

Short of conducting an RCT on implementation supports (e.g., randomly giving teams strong support in half the schools and not in the others) we cannot state that the RAISE approach caused the program to

take hold. More troubling as researchers, we could not identify the conditions moderating the presumed effect such that, under the same treatment condition, some schools gained participants and others lost participants. Our study did not delve into the organizational levels above the school where policy choices, pressures to comply with regulations, competing programs, union disputes, and community pressures may make the ground less fertile for growth of the program.

CONTRIBUTION TO THE THEORY OF SCALE-UP

Our study of comparable schools implementing the same program in the context of an RCT versus in a context not designed with research controls highlights characteristics that are often not attended to in rigorous effectiveness research but are pertinent to understanding the effectiveness and scalability of the program. In the current study, a direct comparison of school achievement was not possible since the outcome measure used in the RCT was researcher-administered. However, we are able to look at teacher characteristics associated with higher student achievement in the RCT and we can look at teacher (and principal) characteristics associated with school-level gains in program participation in Scale-up. We found that RCT schools had lower levels of program-related characteristics associated with successful scale-up. Scale-up and treatment schools had similar levels of program-related teacher characteristics that were associated with greater student achievement in the RCT treatment schools. But schools that were successful in growing participation internally were not more likely to have the teacher characteristics associated with achievement gains. This suggests that program effectiveness in a controlled study may not be indicative of the program's prospects of wider adoption and, in fact, such studies may not create the best conditions for future program sustainability.

In the definition of scaling suggested by Robert Slavin (Slavin, 2002) the metric involves the multiplication of the growth in numbers and the impact, as measured, for example, in an RCT. This product (of scale x impact) is how at the end of a large project, the overall success can be described. Scientific research, however, aims to measure the potential of a program when implemented in other settings. The impact estimate from an RCT is not just a measure of what happened in the sample, but provides a guide to what one might expect elsewhere under similar conditions. Our approach to scalability is similar. We are looking for ways to assess the scalability of a program, which is more important than counting the students, teachers, and schools that were reached in a program.

The question "Does it work?" takes on a different meaning. From the point of view of an administrator, both impact and scalability may play a part in deciding whether a program is suitable for a school, district or state. Scalability can be primary because if the program cannot take hold in a district, its potential for impact has no value. On the other hand a program with no research track record or one that shows it is no more effective than others may still seem promising enough to try it out in a local pilot if is otherwise a good fit. Considerations such as available time, compatibility of the program with policy directives from the state or district, or cost, must be addressed. In reviewing three decades of research on this topic, Honig & Coburn (2008) find that, once central office district administrators have gathered evidence either through reading scientific reviews or conducting local program evaluations, these are never used "directly" to make decisions. They find rather that scientific evidence is incorporated into "working knowledge" that is practical and "mediates" between information sources

and decisions. This "working knowledge" can include a wide range of other kinds of information and evidence, including teacher feedback, implementation issues, past experience, or what the neighboring district superintendent said about it. For example, in many cases a decision to move forward was made before the evaluation was complete or even started; thus the evidence from it is used (or ignored) to support that decision or to maintain enthusiasm. In other words, Honig & Coburn (2008) argue that there is a strong element of advocacy in how evidence is filtered and used.

Our study found more positive use of evidence by school administrators, although still balanced by questions of feasibility of implementation. Our case studies illustrated instances where the positive findings from prior research on Reading Apprenticeship were used as a reason to sustain the program in order to reach a school or district's goals related to their designation as a Priority or Focus school. Our surveys of principals found that over half considered evidence of Reading Apprenticeship's effectiveness in choosing to join the project. This reason was second only to "The pedagogy corresponds to the literacy practices advocated by my school" as a reason. Clearly, then there was a mixture of reasons, but the frequent consideration of evidence is consistent with the policies that were incorporated in the i3 program, which encourages the use of evidence in educational decisions. That is, i3 makes it a requirement of programs like Reading Apprenticeship that they have a level of evidence as a condition of funding and conduct a rigorous evaluation as a part of its project. This provides an example, at least, of the role of evidence in decision making that is less cynical than the administrator citing evidence to support a decision post hoc, or what Donald Campbell (1969) described as the "trapped administrator"—a school administrator who has made a bold decision to invest considerable resources in an unproven program. When the evaluators are called in, the only politically acceptable outcome is to show that the program worked.

From our study of the scale-up processes we see that the "scale x impact" view of scale has two potential weaknesses. The first, as we note above, is that we need to know not just how much impact a program had but what is the potential impact of a program if an administrator were to decide to implement it in his or her district. For this, we need a measure of likely scalability under the local conditions in the district. Scale-up research that will be useful to the decision-maker must be predictive of sustaining and scaling a program. Our study has taken a small step in suggesting a simple metric (GL) of "scaling-in" that may provide a useful approach for studies that look at the predictors of scalability.

The second potential weakness of "scale x impact" view is that the conditions required to measure impact may be incompatible with the processes leading to scale. Starting with the recruitment of schools into an RCT compared to recruitment into a program outside of an experiment, we see additional constraints on the candidates for the RCT. The greater commitment among teachers outside of the RCT to making the program work in the school suggests one possible outcome of this difference. Initial recruitment and further "scale-in" growth within the Scale-up schools—where internal teams attracted new recruits—make the two contexts not comparable. Whatever the processes driving growth and attrition in the scale-up context, they result in a population with greater commitment to the schoolwide success of the program. A quasi-experiment comparing the RCT treatment schools with the Scale-up schools in terms of impact on student achievement (if a common outcome measure were available)

would face uncontrollable selection bias. We do not see how the impact estimate generated by the RCT can be applied to the Scale-up schools.

We see the "Does it work?" question applied to scaling to be more primary to practitioner decisions than the question about impact. The question of whether a program is scalable may not be readily amenable to the RCT design. Our own study took the form of a formative experiment (Newman, 1990) in which we collaborated with SLI to track and provide feedback on the adjustments SLI made to maximize the scaling in and scaling out of RAISE. In that sense the research shares goals with recent work on continuous improvement (Bryk, Gomez, Grunow, & LeMahieu, 2015; Cohen, Peurach, Glazer, Gates, & Goldin, 2013.) This is very different from conducting a large-scale RCT to measure the impact of a program at scale—a design that would not address the question our study tackled. We have attempted to identify school-level processes of an innovation that are associated with growth in the number of teachers served. These processes go beyond the innovator's recruiting, training and direct support of the schools, which in an RCT design are characterized as the "treatment" and are described as the first stage in the logic model we presented initially. As processes that cause the innovation to grow and be sustained begin to take hold, we see a divergence between the scale-up and RCT contexts. Putting greater program evaluation efforts into understanding these processes may be productive for projects funded by programs like i3 where the program is scaled-up and researchers can examine the processes that support program growth.

References

Adelman, H. S., & Taylor, L. (1997). Toward a scale-up model for replicating new approaches to schooling. *Journal of Educational and Psychological Consultation*, *8*(2), 197-230.

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York: Russell Sage Foundation.

Bryk, A. S., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning to Improve: How America's Schools Can Get Better at Getting Better.* Cambridge, MA: Harvard Education Publishing.

Campbell, D.T. (1969). Reforms as experiments. *American Psychologist*, 24: 409-429.

Cannata, M., Rutledge, S., Redding, C., and Nguyen, T. (2015) *With Scale in Mind: A Continuous Improvement Model for Implementation*. Presentation at the conference of the NCSU, Nashville, TN

Coburn, C. (2003). Rethinking scale: Moving beyond the numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3-12

Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K. E., & Goldin, S. (2013). *Improvement by design: The promise of better schools*. Chicago, IL: University of Chicago Press.

Cole, M., Hood, L., & McDermott, R. P. (1978). *Ecological Niche-Picking: Ecological Invalidity as an Axiom of Experimental Cognitive Psychology*. Unpublished manuscript. Retrieved September 27, 2015 from http://www.academia.edu/11407068/Ecological_Niche_Picking_Ecological_Invalidity_as_an_Axiom_of_Cognitive_Psychology

Corrin, W., Somers, M. A., Kemple, J., Nelson, E., & Sepanik, S. (2008). *The Enhanced Reading Opportunities Study: Findings from the Second Year of Implementation* (NCEE 2009-4036). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco, CA: Jossey-Bass.

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.1). Palo Alto, CA: Empirical Education Inc.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Greenleaf, C., Hanson, T., Herman, J. Litman, C., Madden, S., Rosen, R., Kim-Boscardin, C., Schneider, S. & Silver, D. (2009). *Integrating Literacy and Science Instruction In High School Biology: Impact On Teacher Practice, Student Engagement, and Student Achievement*. Final Report to the National Science Foundation Grant #0440379.

Greenleaf, C., Schneider, S., & Herman, J. (2005). *An Efficacy Study of Reading Apprenticeship Professional Development for High School History and Science Teaching and Learning*. Teacher Quality Research Reading/Writing Grants, U. S. Department of Education Institute of Education Sciences, Grant # R305M050031, Greenleaf PI.

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., ... & Jones, B. (2011a). Integrating literacy and science in biology: 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

Hegseth, W., Zacamy, J., & Newman, D. (2015). *Case Studies of the Scaling And Sustaining of Reading Apprenticeship In Four Michigan Secondary Schools: Report 2*. (Empirical Education Rep. No. Empirical_ i3SU-7019-CS2-Y3-O.1). Palo Alto, CA: Empirical Education Inc.

Hegseth, W., Zacamy, J., & Newman, D. (2014). *Case Studies of the Scaling and Sustaining of Reading Apprenticeship In Four Michigan Secondary Schools*. (Empirical Education Rep. No. Empirical_ i3SU-7019-CS1-Y2-O.1). Palo Alto, CA: Empirical Education Inc.

Honig, M. I., & Coburn, C. (2008). Evidence-Based Decision Making in School District Central Offices. *Educational Policy, 22*(4), 578-608.

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.

McDonald, S., Keesler, V., Kauffman, N., & Schneider, B. (2006). Scaling up exemplary interventions. *Educational Researcher*, *35*(3), 15–24.

Meier, D. (2002). *In Schools We Trust: Creating Communities of Learning in an Era of Testing and Standardization*. Boston: Beacon Press.

Michigan Department of Education (MDE). (2016). *Focus Schools*. Retrievable from http://www.michigan.gov/mde/0,4615,7-140-22709_62253---,00.html

Michigan Department of Education (MDE). (n.d.). *Frequently Asked Questions about Michigan's Priority Schools*. Retrievable from http://www.michigan.gov/documents/mde/Priority_FAQ_427729_7.pdf?20130802150528

Newman, D. (1990). Opportunities for research on the organizational impact of school computers. *Educational Researcher*, 19, 8-13.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. New York: Cambridge University Press.

Newman, D., & Jaciw, A. P. (in preparation). The Unvarnished Report of Conducting an RCT of a Statewide STEM Program in the Process of Scaling Up. In R. Maynard, L.V. Hedges, & S. Wachter (Eds.), *Teaching Cases on Designing and Fielding Experimental Design Evaluations in Education* (provisional title). Available at SSRN: http://ssrn.com/abstract=2519677

Newman, D., Lazarev, V. & Zacamy, Z. (2016) *Comparing a program implemented under the constraints of an RCT and in the wild.* Paper presented at the annual conference of the Society for Research in Educational Effectiveness, Washington DC.

Sternberg, R. J., Birney, D., Jarvin, L., Kirlik, A., Stemler, S., & Grigorenko, E. L. (2011). Scaling up educational interventions. In R. J. Sternberg & M. Constas (Eds.), *Translating educational theory and research into practice*. Mahwah, NJ: Erbaum.

Schneider, B., & McDonald, S. (2007). Introduction. In B. Schneider & S. McDonald (Eds.), *Scale-up in Education*, Volume 1, Ideas in Principle, pp 1-15.

Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). *Reading for Understanding: How Reading Apprenticeship Improves Disciplinary Learning in Secondary and College Classrooms*. Hoboken, NJ: Wiley Jossey-Bass. Retrievable from http://readingapprenticeship.org/book/reading-for-understanding/

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Slavin, R. E. (2002, November). Remarks offered as a member of the Panel on Scaling at the Interagency Education Research Initiative (IERI) Annual Principal Investigators Meeting. Washington, DC. (as cited in: McDonald, S., Keesler, V., Kauffman, N., & Schneider, B. (2006). Scaling up exemplary interventions. *Educational Researcher*, *35*(3), 15–24.)

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. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Strategic Literacy Initiative (SLI). (2010). *Scaling Up Content-Area Academic Literacy in High School English Language Arts, Science and History Classes for High Needs Students* (pp. 1-37, Rep.). San Francisco, CA: WestEd.

Tipton, E. (2013). Improving generalizations from experiments using propensity score subclassification: Assumptions, properties, and contexts. *Journal of Educational and Behavioral Statistics, 38*, 239-266.

Zacamy, J., Newman, D., Lin, L, & Jaciw, A.P. (2015). *RAISE Scale-up: Implementation and Sustainability in Year 3.* Paper presented at the annual meeting of the American Educational Research Association in Chicago, IL. Retrieved on April 19, 2015 from http://empiricaleducation.com/pdfs/RAISEp.pdf

Zacamy, J., Jaciw, A. P., Lin, L., & Newman, D. (2014). *Year 2 Interim Report of RAISE Scale-up Study*. (Empirical Education Rep. No. Empirical_i3SU-7019-IR1-Y2-O.1). Palo Alto, CA: Empirical Education Inc.

Zacamy, J., & Newman, D. (2014, April). Formative Evaluation on the Process of Scaling Up Reading Apprenticeship. In J.A. Hamilton (Chair), *Evaluation of Reading Apprenticeship Improving Secondary Education*. Symposium conducted at the annual meeting of the American Education Research Association, Philadelphia, PA.

Zacamy, J., Gray, S., Jaciw, A., & Newman, D. (2013). *Year 1 Interim Report of Reading Apprenticeship/RAISE Scale-up*. (Empirical Education Rep. No. Empirical_i3SU-7019-IR1-Y1-O.1). Palo Alto, CA: Empirical Education Inc.

Zacamy, J., Newman, D., Lazarev, V., & Lin, L. (2015). *School Processes That Can Drive Scaling-up of an Innovation, or Contribute to its Abandonment*. Paper presented at the national conference of The National Center on Scaling Up Effective Schools.

Appendix A: Detailed RAISE Scale-up Logic Model

In this appendix, we provide a comprehensive narrative description of each stage of the RAISE Scale-up logic model guiding our study. We also present the accompanying logic model figures. As described in the Study Overview section of this report, the arrows in the logic model figures represent relationships or interactions between different components of the process. They change color and directionality through the different stages of the model.

STAGE 1: DEVELOPMENT ACTIVITIES AND INTERMEDIATE OUTCOMES

The Stage 1 diagram (Figure A1) consists of two concentric circles. The inner green circle, which represents the funding and management of SLI's RAISE, contains the four key development activities. The outer blue circle contains the intermediate outcomes, which are the result of direct uptake of the development activities by the districts, schools, and teachers.

Development Activities

The program developers provide schools and districts with the resources, information, and skills to participate in RAISE. Here we describe the four activities.

1. Project development and coordination
2. Recruitment and retention
3. Professional development for Reading Apprenticeship facilitators and teachers
4. Instructional support resources

The Project development and coordination and Recruitment and retention activities are similar to Adelman & Taylor's (2007) Creating readiness stage which refers to developing interest and dissemination of information, and creating agreements and policies for implementation. The Professional development and Instructional support resources activities align with Adelman & Taylor's Initial implementation stage, which involves supporting and guiding the adaptation and employment of the intervention in new contexts by creating temporary mechanisms to facilitate implementation (e.g., mentors or coaches).

Project Development and Coordination

The SLI co-directors are responsible for overall project leadership and guidance in management of the scale-up process. They maintain project budgets, make key decisions, and guide the process during each phase. SLI secures funds to supplement the i3 grant through partnerships with private sector organizations for materials, resources, salaries, and stipends for project development. In addition, the SLI co-directors lend their expertise in the Reading Apprenticeship method, the Reading Apprenticeship philosophy, and orientation to instruction to lead the core intellectual work. SLI administrative staff supply general project coordination (e.g., reserving space for trainings, communicating with teachers/administrators). This core group is similar to what Adelman & Taylor (2007) call the "change team." They are responsible for developing and following through with the "big picture" process of scale-up through developing linkages of resources across sites, resolving large-scale problems systematically, and ensuring effective diffusion. Furthermore, as part of RAISE scale-up, the evaluation

team collects quantitative and qualitative data on the scale-up process and provide formative feedback to the SLI co-directors to inform practice.

Recruitment and Retention

The State Coordinators (SCs) are responsible for identifying and recruiting districts, schools, teacher leaders, and teachers to participate in Reading Apprenticeship professional development and adopt the Reading Apprenticeship framework. The State Coordinators from each state, as well as the Multi-State Coordinator, join the "change team" and provide regional knowledge and management of their local sites. State Coordinators are responsible for recruitment and site management through assessing the interest and need of districts and schools, building relationships with participants, addressing barriers or concerns to participation, and disseminating information. The co-directors and support staff work with the SCs, district contacts, and school administrators to identify and recruit teacher leaders. Teacher leaders are recruited from among RAISE-trained teachers at each school.

Retention of schools and districts involves frequent and ongoing communication between schools/districts and State Coordinators. Retention of teachers includes ongoing support and professional development, as well as a ladder of movement in which outstanding Reading Apprenticeship teachers are identified and ask to be trained as teacher leaders, and potentially as RAISE Institute facilitators. Model Reading Apprenticeship classrooms are also identified as exemplars for training and professional development purposes.

Professional Development

Professional development is the primary vehicle for bringing Reading Apprenticeship principles and pedagogy into districts, schools, and classrooms. The professional development team at SLI consists of subject area leads and support staff who were responsible for updating existing Reading Apprenticeship professional development modules and implementing the plan for the RAISE professional development.[13] This team also identify, recruit, and train a group of Reading Apprenticeship facilitators who conduct the RAISE Institutes. In addition, the professional development team develops the training modules and materials for the facilitator and teacher trainings.

Facilitator professional development. The professional development team selects the facilitation team from a group of Reading Apprenticeship certified consultants and previously trained Reading Apprenticeship teachers and coaches. The facilitation team attends a two-day intensive training and collaborate through an online resource website to deepen their understanding of the Reading Apprenticeship model and framework and content-specific Reading Apprenticeship training modules. They also work in facilitation teams to plan which team member is responsible for implementing each module at the upcoming RAISE Institutes.

Teacher professional development. The RAISE Institutes consists of 65 hours (10 days across two years) of training on the Reading Apprenticeship model and philosophy as follows.

[13] The professional development team works in consultation with the SLI co-directors.

a) Five full days of training in the first summer prior to implementation focusing on the foundation of Reading Apprenticeship
b) Two full days of training during the first year of implementation focusing on formative assessment, differentiation, and planning for implementation
c) Three full days of training in the summer following the first year of implementation focusing on formative assessment and planning for implementation

The goals for professional development are fivefold.

a) Articulate and define the Reading Apprenticeship model and framework (social, cognitive, knowledge building, and personal dimensions)
b) Define, model, explore, and practice Reading Apprenticeship instructional strategies that foster metacognitive inquiry, collaboration that facilitates metacognitive inquiry and conversations; and students' use of reading comprehension strategies
c) Describe the teachers' role in an Reading Apprenticeship classroom including formative assessment and differentiation of instruction
d) Teach discipline-specific reading comprehension strategies and instructional practices
e) Plan for implementation

A key aspect of the professional development is working to change teachers' perspectives from seeing themselves only as teachers to seeing themselves also as learners. As learners, teachers continually improve their practices, learn from the experiences of other Reading Apprenticeship teachers and teacher leaders, and approach the implementation of Reading Apprenticeship as a learning process, similar to those of their students. SLI intends to accomplish this through inquiry-based, collaborative discussion of metacognitive processes, and a lot of professional reading and small-group discussion.

Instructional Support Resources

Instructional support resources are also available in four forms: (a) meetings for teacher leaders, (b) monthly on-site support meetings for teachers led by teacher leaders, (c) administrator online course, and (d) *Thinking Aloud* website.

Monthly webinars/in-person meetings for teacher leaders. In addition to attending the RAISE Institute, teacher leaders receive further support focused on the following.

a) Articulating the Reading Apprenticeship model and framework
b) Methods for providing on-site support to teachers
c) Tools and resources for teachers

During the first year of the project, the SLI staff presents the teacher meeting agendas to the teacher leaders, but in the following years, the SLI staff works more collaboratively with the teacher leaders to prepare and review the teacher meeting agendas. (In the 2011-12 school year, support was provided through monthly webinars. Starting in the 2012-13 school year, these webinars were replaced with three day-long, in-person meetings with all teacher leaders in the state. The goals of the webinars and in-person meetings were the same.)

Monthly on-site support meetings for teachers. The teacher leaders took what they discuss and learn during the teacher leader meetings and facilitate one monthly on-site meeting with their school's RAISE teachers. These meetings are similarly structured during the first year of implementation and the agenda for meetings is prepared by State Coordinators and SLI for continuity across schools and districts. During the meetings, the teacher leaders provides support to teachers, helps them problem solve, and provides tools to facilitate implementation. These meetings are designed to foster a professional community among the Reading Apprenticeship teachers through teacher collaboration and learning. Activities include sharing of practices, reviewing student work, using Reading Apprenticeship protocols to guide discussion and reflection about practices, reviewing videos of practice, and reading and discussing professional articles.

Administrator online course. Administrators also have the opportunity to participate in an online course about Reading Apprenticeship so they can support Reading Apprenticeship instruction in their school classrooms. The course is developed in collaboration with SLI and the State Coordinators and is designed to prepare administrators to articulate the Reading Apprenticeship model and framework, recognize Reading Apprenticeship practices, provide an infrastructure for supporting teachers (e.g., space for monthly meetings, supplies and materials, allowing for time for collaboration), and provide tools and resources for teachers (e.g., model lessons, rubrics for practice, protocols for collecting and reviewing student work). The course does not focus on evaluating teachers. While the course is optional, administrators are encouraged to attend.

Thinking Aloud website. Additional resources for facilitators, administrators, teachers, and teacher leaders are provided through an online portal, called *Thinking Aloud* (developed in years 1-2 of the initiative). The *Thinking Aloud* website provides the means for educators to support one another, share ideas, ask questions, discuss strategies, and build a stronger professional network of the Reading Apprenticeship community.

Intermediate Outcomes

Here we hypothesize *how* the Stage 1 development activities lead to the five intermediate outcomes, as depicted by the green arrows in our logic model.

Buy-in to the Reading Apprenticeship Framework

We define buy-in as commitment to Reading Apprenticeship as an appropriate strategy for literacy instruction and as a means of improving student achievement. Our model contains four green arrows leading from the four development activities to buy-in. Project coordination includes communication with teachers/administrators that is: (1) intended and designed specifically to increase staff buy-in and (2) the channel through which schools and districts will get the support and materials to implement and expand Reading Apprenticeship. Recruitment and retention also leads to increased buy-in: recruitment offers teachers and schools the chance to participate, and retention offers incentives for participants to continue use, as well as to evolve in their practice. The professional development and instructional support resources are designed to convince staff at all levels of the district, from teachers to administrators, that Reading Apprenticeship is an appropriate and effective method for teaching literacy instruction and improving student achievement.

Increased Capacity to Implement and Disseminate Reading Apprenticeship Practices

Our model contains green arrows leading from three development activities (project development and coordination, professional development, and instructional support resources) to increased capacity. Project development and coordination, as well as recruitment and retention activities are expected to directly lead to the increased capacity of states, districts, and schools to implement Reading Apprenticeship through allocation of funding and dissemination of information. In addition, as a result of participation in the Reading Apprenticeship professional development activities and as a result of receiving instructional support, teachers, teacher leaders, and principals are expected to have increased capacity to implement and disseminate Reading Apprenticeship practices. As teachers, teacher leaders, and administrators become well versed in Reading Apprenticeship, we hypothesize that they put into place and maintain structural supports (e.g. meeting space for teachers, time for collaboration) and create and sustain resources (e.g., materials and tools for teachers).

Increased Participation in RA

A key outcome in most scale-up work is to spread ideas and interventions to larger and more diverse populations (Schneider & McDonald, 2007). This intermediate outcome corresponds to Coburn's dimension of spread, which she describes as the spread of reform-related norms, beliefs, and principles within a classroom, school, and district. In our logic model, this outcome relates to both spread from within, as well as outward expansion to more districts, schools, and classrooms. There are three development activities from our logic model (project development and coordination, active recruitment, and professional development) that are hypothesized to increase the number of teachers, schools, and districts using the Reading Apprenticeship framework. Specifically, project development and coordination help with funding and building of local partnerships, which allows for more schools to implement Reading Apprenticeship. Active recruitment and retention also results in more involvement from teachers, schools, and districts. By the end of the grant period, SLI's goal is to have trained 2,800 teachers and 240 teacher leaders, and impact 410,000 students (SLI, 2010). The professional development is the primary method of disseminating Reading Apprenticeship norms, beliefs, and principles.

Classroom Fidelity of Reading Apprenticeship

The goal of the Reading Apprenticeship professional development is to transform academic literacy teaching. In this logic model, we operationalize this goal as classroom fidelity of the Reading Apprenticeship framework. This outcome corresponds to Coburn's dimension of depth, which is defined by changes in teachers' beliefs, norms of social interaction, and pedagogical principles enacted in the curriculum. At the classroom level, fidelity is characterized by increased numbers and varieties of texts, collaborative activities and assignments for students, use of metacognitive inquiry, and instruction promoting equity. Our model contains two arrows leading from two development activities (professional development and instructional support resources) to classroom fidelity of Reading Apprenticeship. Professional development provides teachers with the skills to implement Reading Apprenticeship with fidelity and continually improve on their practices, and the instructional supports further improve teachers' understanding of Reading Apprenticeship practices. Furthermore, it is hypothesized that use of instructional supports leads to changes in teachers' and administrators' beliefs about literacy instruction, as well as provide a forum for collaboration and support, thus resulting in higher classroom fidelity.

Increased Student Achievement

The fifth intermediate outcome in this process is student achievement. Reading Apprenticeship has been shown to have positive effects on student achievement in previous studies (Corrin et al., 2008; Greenleaf et al., 2009; & Greenleaf, Schneider, & Herman, 2005). While there are no direct links between the development activities and this outcome, it is a critical intermediate outcome in this process.

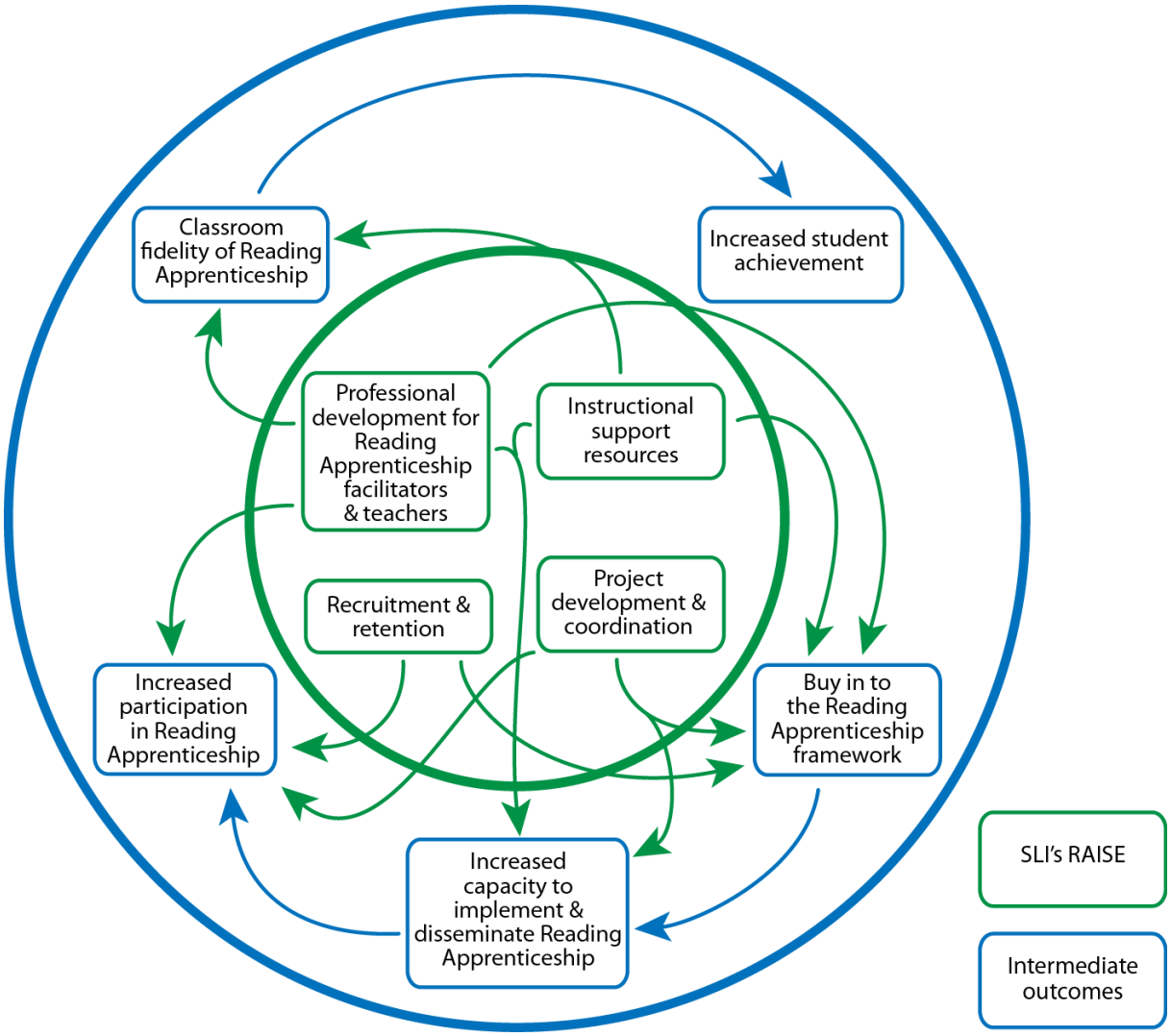


FIGURE A1. STAGE 1: DEVELOPMENT ACTIVITIES

Note. SLI stands for Strategic Learning Initiative. RAISE stands for Reading Apprenticeship Improving Secondary Education.

STAGE 2: INCREASED OWNERSHIP

At Stage 2 of our model (Figure A2), ownership of RAISE begins to transition from the developers to the districts, schools, and teachers, and a dynamic "cycle of improvement" develops. This stage, together with Stage 3, corresponds to Coburn's dimension—Shift in reform ownership—which refers to a transfer

in ownership from the "external" providers to the "internal" actors. Adelman and Taylor describe ensuring long-term ownership and sustainability of the intervention, which requires (a) ongoing (local) leadership to take responsibility for the intervention, and (b) maintenance of planning, implementation, and coordination mechanisms to keep the intervention running. They state that "institutionalizing new approaches entails ensuring that the organization assumes long-term ownership and that a blueprint exists for countering forces that erode progress" (Adelman & Taylor, 2007, p. 220). Here we describe how the initial development activities become a shared responsibility between the SLI team and the local organizations (in this case, schools and districts).

Project Development and Coordination

The SLI co-directors continue to be responsible for overall management of the scale-up process, as well as secure funds to supplement the i3 grant. Schools and districts also begin to examine local funding sources that can be dedicated to continuing and expanding Reading Apprenticeship. External formative evaluations are ongoing, but the local level also begins to develop tools to be able to evaluate their implementation and needs for future self-assessment. Local actors take more responsibility for organizing the dissemination of information about the overall pedagogical principles of Reading Apprenticeship in general, and specifically about the RAISE project development, professional development, and support opportunities that are available to their local schools and teachers.

Recruitment and Retention

The SCs continue to identify, recruit, and retain districts, schools, teacher leaders, and teachers to participate in the Reading Apprenticeship professional development and adopt the Reading Apprenticeship framework in their schools. Local district and school administrators work closely with the SCs to identify and recruit additional teachers and schools from existing RAISE schools and districts (i.e. scaling-in) to join the scale-up efforts. Districts and schools also play an active role in reaching out to neighboring schools and districts to share their experience with Reading Apprenticeship and invite them to join (i.e. scaling-out). Retention of Reading Apprenticeship teachers, teacher leaders, and schools becomes increasingly complicated as more actors are involved. The SCs depend more on local administrators to support retention efforts and alert them to issues that may jeopardize retention.

Professional Development

Professional development for new teachers continues to include 65 hours of professional development (RAISE Institutes) on the Reading Apprenticeship model and philosophy. As veteran Reading Apprenticeship teachers and teacher leaders increase their depth of understanding of the Reading Apprenticeship model, they play an important role in supporting newly trained Reading Apprenticeship teachers during the training and at their local sites. There are also increased opportunities for RAISE-trained teachers to apply for and join the professional development facilitation team.

Instructional Support Resources

The monthly meetings continue to occur, however, there is more leeway and flexibility for teacher leaders to prepare their own agendas and respond to specific school needs. Furthermore, the *Thinking Aloud* website is monitored by the SLI team, but at the local level, teachers and administrators use the website to develop networks with Reading Apprenticeship teams in other states.

Cycle of Continuous Improvement

The four development activities from Stage 1 (project development and coordination, recruitment and retention, professional development, and instructional support resources) become shared responsibilities between the developers and the local actors. Each of these activities is adapted to local contexts and needs and is planned with the idea of sustaining Reading Apprenticeship locally. The intermediate outcomes are established and reinforced, and are beginning to become independent from the resources, funding, and involvement of the SLI team. This cycle of improvement is characterized by continuous interactions and feedback loops between the development activities and intermediate outcomes.

As the cycle developed, not only do we expect a higher measure of each of the intermediate outcomes as the process evolved (i.e. increase in participants, more capacity to implement, deeper classroom fidelity, higher student achievement, more buy-in), but also that, as they increased, they are reinforced and supported from within (the classroom, school, district, state) rather than by the developer (i.e. the transfer of ownership). Here we describe each of these arrows in the cycle in relationship to the intermediate outcomes.

Buy-in of Reading Apprenticeship Framework

Our model contains one purple arrow leading from increased student achievement to buy-in. As student achievement increased, we hypothesize that teachers, schools, districts and states become more committed to implementing and expanding Reading Apprenticeship. That is, the results feed back into the uptake or buy-in of Reading Apprenticeship. Furthermore, our model depicts one purple arrow leading from buy-in to instructional support resources. We hypothesize that as teachers, schools, and districts took ownership of Reading Apprenticeship, teachers, teacher leaders, and administrators use the instructional support resources to supplement and inform their practices, as well as to develop networks with other Reading Apprenticeship professionals. Teachers, teacher leaders, and administrators provide feedback to their State Coordinators and the SLI team about how these resources are used and whether additional instructional supports are needed at their local level.

Increased Capacity to Implement and Disseminate Reading Apprenticeship Practices

The purple arrow leading from increased capacity to instructional support resources shows teachers and administrators taking ownership of the instructional supports, such as the monthly school team meetings and web portal, and adapting these supports to fit their local contexts. Our model also depicts one purple arrow leading from increased capacity to professional development. As schools, districts, and states built capacity to support the implementation of Reading Apprenticeship, we hypothesize that local actors play a more active role in the professional development by providing feedback to inform the professional development of teachers and teacher leaders and becoming trained Reading Apprenticeship facilitators. Furthermore, as districts and states begin to develop their own professional development to support the sustainability of Reading Apprenticeship, additional feedback is provided to improve the overall RAISE project. The local level actors also build the capacity to take more ownership of project coordination, and recruitment and retention activities, as represented by the two purple arrows leading from this intermediate outcome to those development activities.

Increased Participation in Reading Apprenticeship

There is a purple arrow leading from professional development to more teachers, schools, and districts using Reading Apprenticeship. As the development of teachers, teacher leaders, and administrators is increasingly supported at the local level, more students are impacted by Reading Apprenticeship.

Classroom Fidelity of Reading Apprenticeship

Our model contains two purple arrows leading from two development activities to classroom fidelity. These two purple arrows are the same as the green arrows described in Stage 1. As these development activities become increasingly shared between the SLI team and local actors, support and guidance to address challenges and issues with implementation in schools occurs more from the local level. Furthermore, through the web portal resources, Reading Apprenticeship teachers and teacher leaders become linked with a wider network of professionals engaged in Reading Apprenticeship. Through building this support network, teachers, teacher leaders and administrators strengthened their commitment. Within this process, schools build capacity, improve performance, and maintain fidelity to the Reading Apprenticeship model.

Student Achievement

In our model, one purple arrow from student achievement leads to buy-in. As participating states, districts, and schools receive information regarding effects on student achievement, their support for Reading Apprenticeship increases. As support continues to build, more resources are put towards Reading Apprenticeship professional development, development of teacher leaders, and ownership over the tools and systems once provided by the developers.

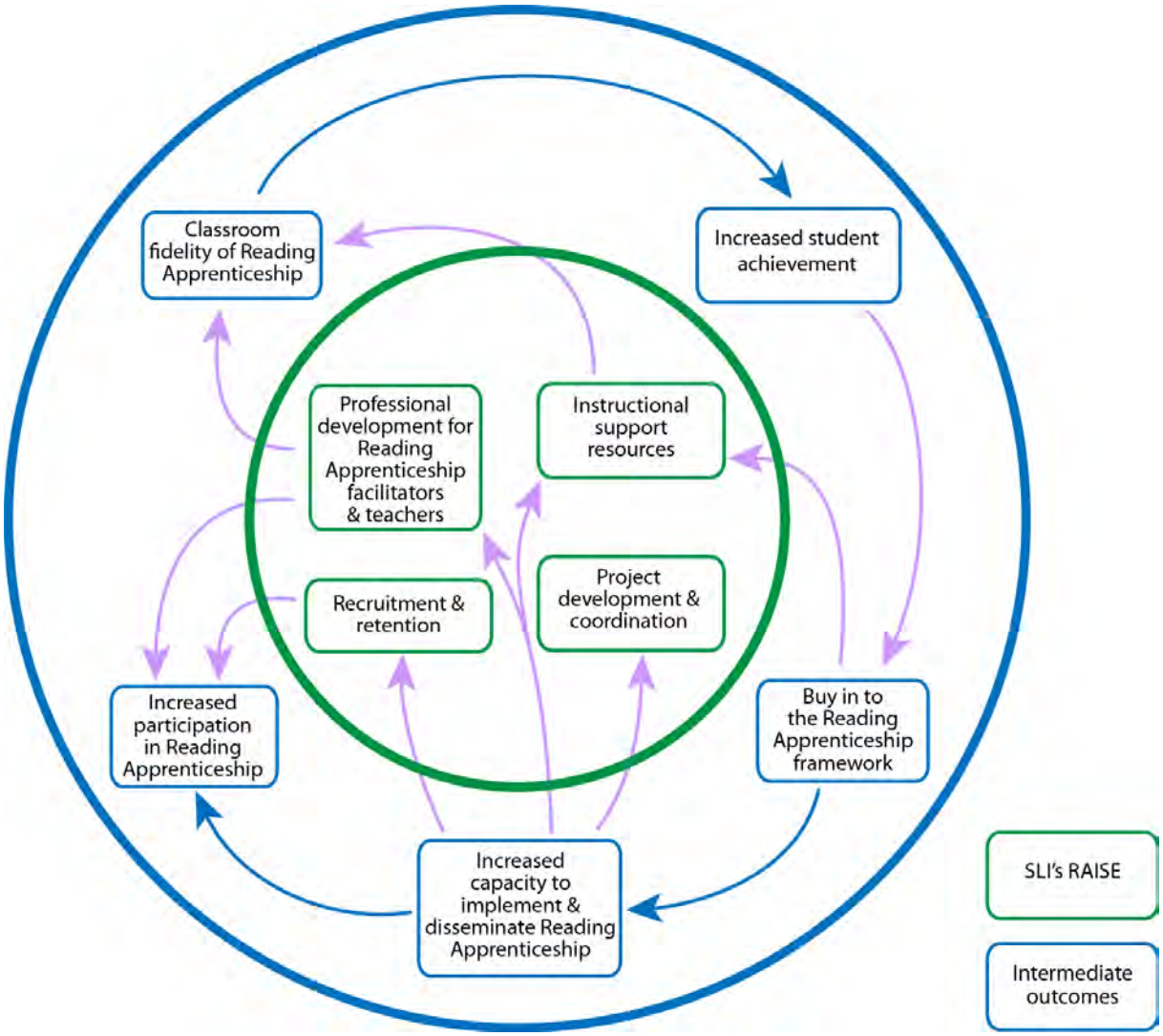


FIGURE A2. STAGE 2: INCREASED OWNERSHIP

Note. SLI stands for Strategic Learning Initiative. RAISE stands for Reading Apprenticeship Improving Secondary Education.

STAGE 3: SUSTAINED OWNERSHIP

The third stage in our logic model (Figure A3), Sustained ownership, involves a withdrawal of resources and support from the SLI team and a transfer of more responsibility and ownership of the activities to sustain Reading Apprenticeship to the local schools and districts. In this stage, the green outlines around the development activities begin to fade, signifying the diminishing presence of the SLI team and sustained ownership of the RAISE project goals at the local level. Furthermore, the schools and districts take responsibility for the intermediate outcomes and the interactions among them, thus the blue arrows are also replaced by purple arrows, signifying that the cycle is sustained at the local level.

Responsibilities for recruitment and retention, professional development, and instructional support resources are transferred to the local level. Project coordination is also transferred to the local level.

In this stage, we expect that Reading Apprenticeship has been fully implemented in a large number of schools and districts and that many teachers, teacher leaders, and administrators are involved. While the developers are minimally involved in the project coordination, we hypothesize that states or districts either seek external funding or allocate internal resources for implementing and sustaining Reading Apprenticeship in schools. Furthermore, states, in collaboration with school districts, will recruit and train new and replacement teachers on an as-needed basis, as well as continue to provide incentives for teachers and teacher leaders who are doing exceptionally well to serve as models for others, or be trained at a higher level. Professional development opportunities and instructional support resources are offered by states and districts. The *Thinking Aloud* website portal will continues to be used to create and maintain social networks for Reading Apprenticeship professionals. Schools and districts will begin to shift their academic policies in support of broadly implementing Reading Apprenticeship long term. Districts develop evaluation plans for identifying needs, strengths, and areas of change for self-assessment. This stage is similar to the fourth and last phase of Adelman & Taylor's model (2007), Ongoing evolution, and is concerned with accountability in outcomes, as well as in continually evolving practice for improvement through formative and summative evaluation.

The cycle of improvement continues in this stage. The purple arrows depicted in Stage 3 are the same as the purple arrows in Stage 2. However, these relationships between activities and intermediate outcomes have strengthened over time, and continue to evolve as ownership of the RAISE reform efforts is more thoroughly transferred to the local level.

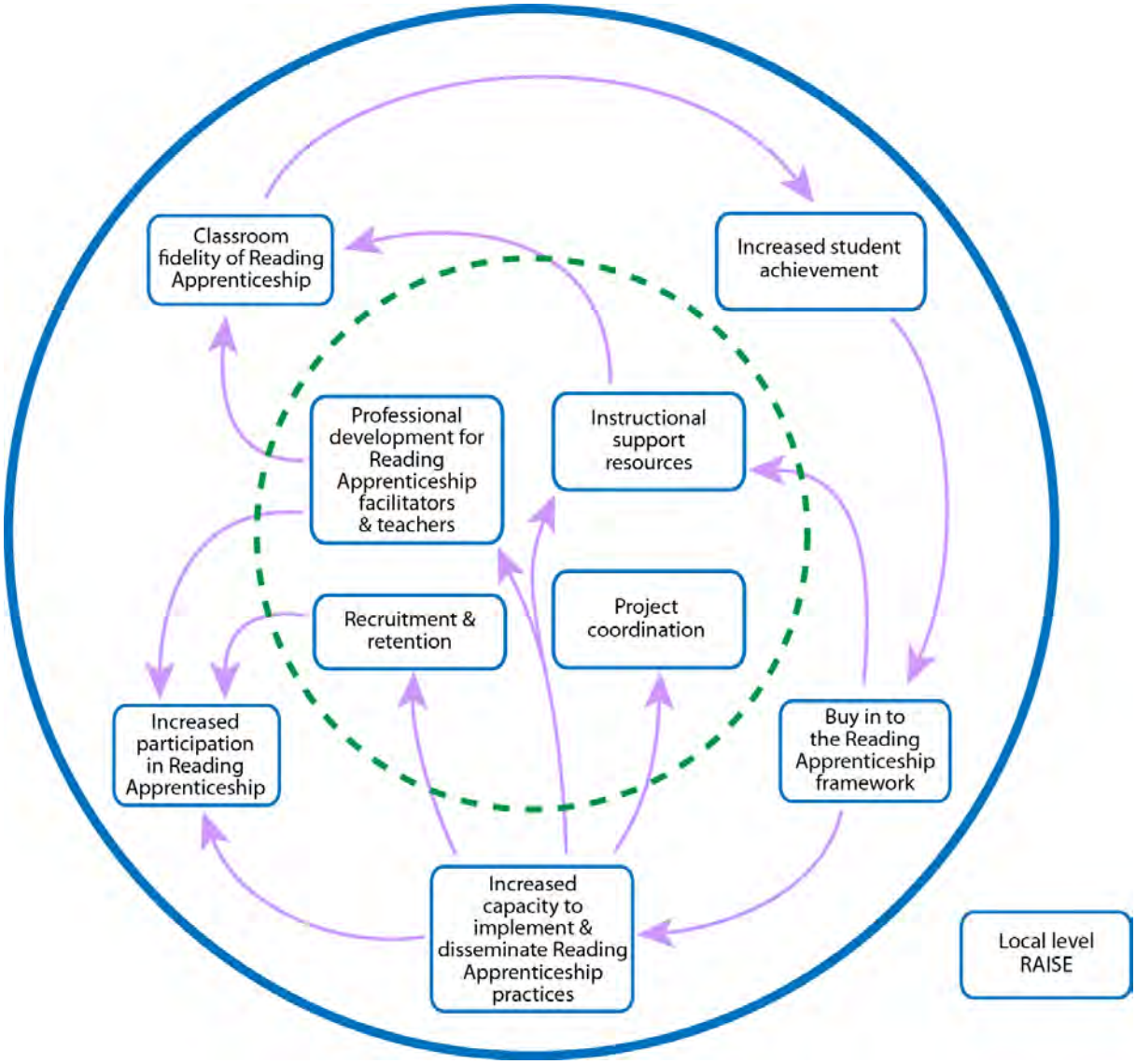


FIGURE A3. STAGE 3: SUSTAINED OWNERSHIP

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.

STAGE 4: READING APPRENTICESHIP BROADLY INSTITUTIONALIZED

This last stage retains the arrows and boxes depicted in Stage 3, and the cycle of improvement is ongoing; however, in this last stage (Figure A4), all activities are implemented at the local level and are built to sustain Reading Apprenticeship, as well as to help other LEAs develop similar capacity. This stage corresponds to Coburn's Sustainability dimension, which is described as the distribution, adoption, and maintenance of an innovation long-term.

By Stage 4, Reading Apprenticeship has become a norm and standard in the originally recruited LEAs; there is solid commitment and support at all levels built into the system. In addition, all of the

intermediate outcomes are realized, which is hypothesized to lead to two end outcomes: 1) Reading Apprenticeship becoming institutionalized as the LEAs' model of academic literacy and 2) LEAs demonstrating capacity to scale Reading Apprenticeship with fidelity broadly in the regions (SLI, 2010). Specifically, there are three black arrows leading from Classroom fidelity of Reading Apprenticeship, Increased capacity to implement and disseminate Reading Apprenticeship, and Increased participation in Reading Apprenticeship to Demonstrated capacity to scale Reading Apprenticeship with fidelity broadly in the regions. We expect that in this final stage, an increase in local units implementing Reading Apprenticeship with fidelity contributes to an increase in participation broadly in the region. Additionally, there are two black arrows depicted in the logic model leading from increased student achievement and Buy-in to Reading Apprenticeship becomes institutionalized in the LEAs' model of academic literacy. As depicted in the logic model, increase in student achievement and continued support and commitment (buy-in) for Reading Apprenticeship leads to policy shifts at the school, LEA, and state level where Reading Apprenticeship is "institutionalized" as the local model of academic literacy instruction. Our model also consists of black arrows leading from Reading Apprenticeship becomes institutionalized in the LEAs' model of academic literacy to Demonstrated capacity to scale Reading Apprenticeship with fidelity broadly in the regions and vice versa. Policy shifts that support Reading Apprenticeship institutionalization result in an increase in units that implement Reading Apprenticeship. The increase in units further reinforces institutionalization and policy at the school, district, and state levels.

Logic Model Updates

In the early development of the RAISE Scale-up logic model, we focused on the literature that described the "shift in reform ownership" as the primary dimension for scale-up. One of the key areas of investigation in our study was how the developers create conditions and built capacity to shift the ownership to the local level. However, there is another component of the process that we realized must be accounted for in the logic model driving this study: balancing the centralized, on-going research and development functionality of the developers with the uptake of reform ownership at the local level.

Since the inception of Reading Apprenticeship, SLI has followed a "design research" model in which they have maintained a dialogic exchange with the field. At each stage of implementation, SLI has included a research component, and revised and improved Reading Apprenticeship based on that research. While the core theory and pedagogy behind Reading Apprenticeship has remained constant, the R&D team continued to develop new resources and supports to deepen the Reading Apprenticeship professional development experience and practice in the field. In this scale-up process, the local level is expected to adapt these resources to their contextual needs, and SLI continues to improve and revise these components as they learn from the field.

Since the process of generative scale-up continues through the RAISE project, it has, therefore, been built into our logic model. In the original version of the logic model, as ownership strengthened at the local level, we hypothesized that the presence of the developers would diminish, until it completely disappeared. We revised the logic model so that the developers' presence fades, but remains as they interact, build relationships, and improve the program based on what they learn from the field.

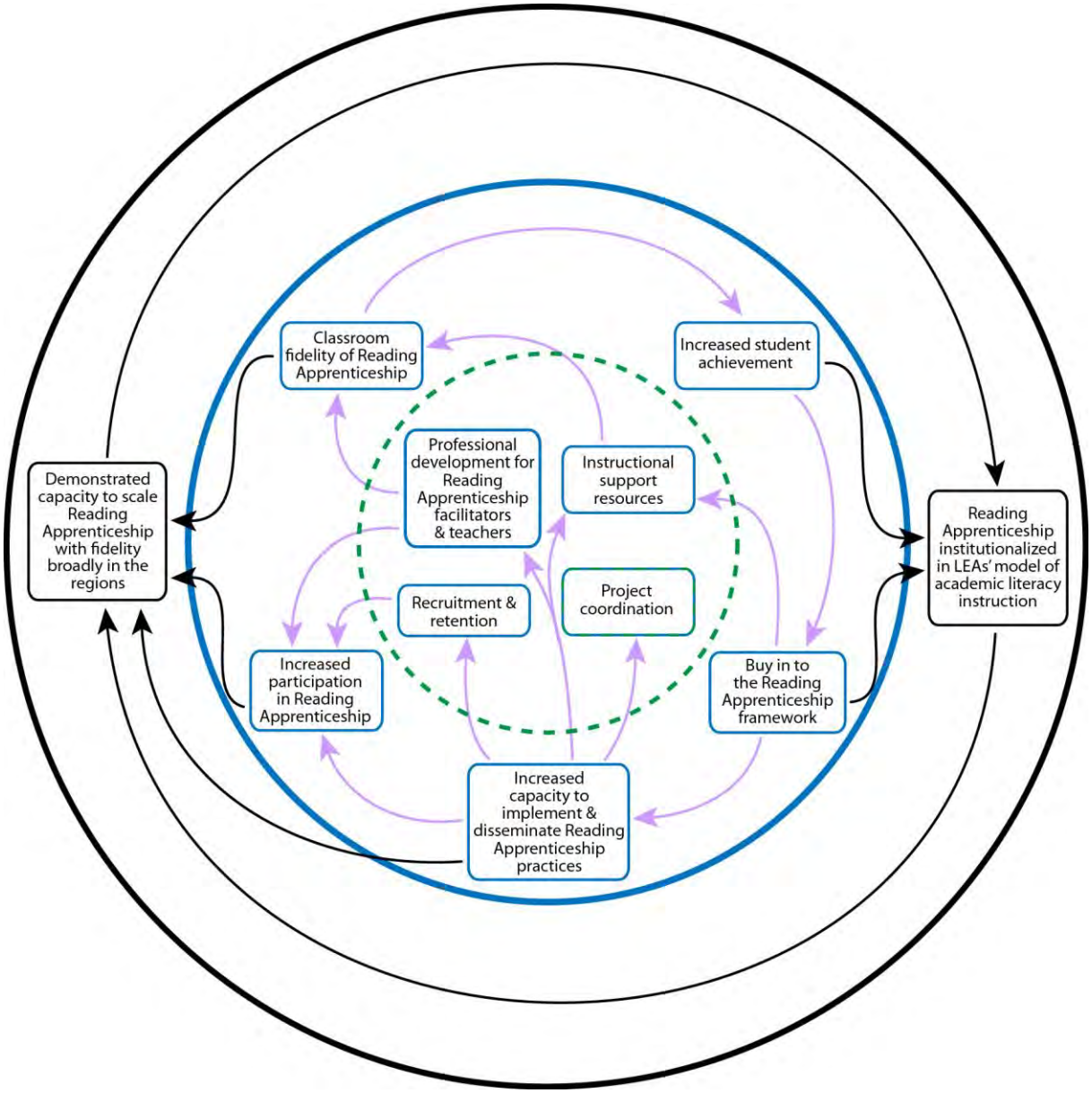


FIGURE A4. STAGE 4: READING APPRENTICESHIP BROADLY INSTITUTIONALIZED

Note. LEA stands for local education agency.

Appendix B: State Maps of RAISE Scale-up, By Year

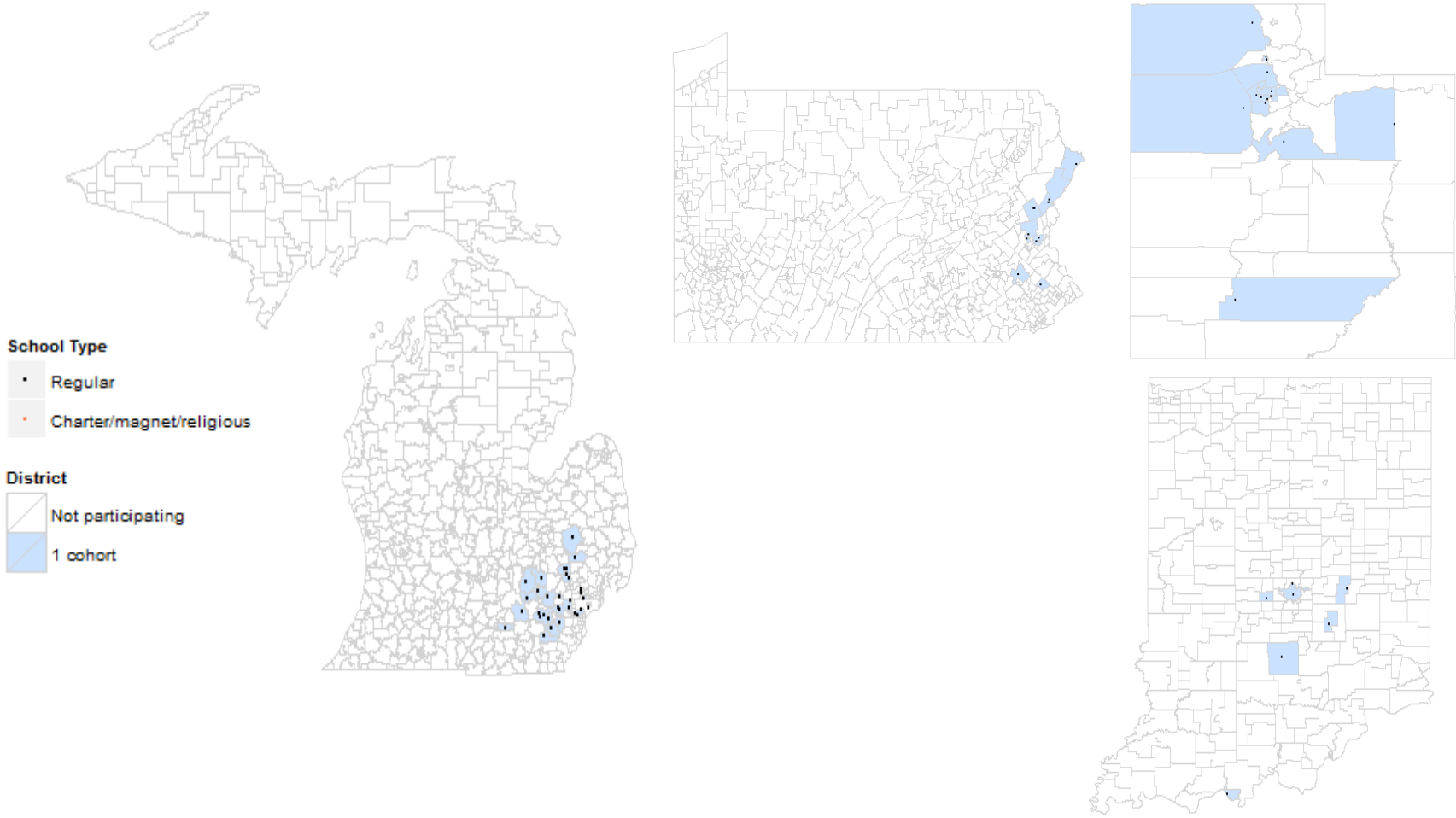


FIGURE B.1. MAP OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP PARTICIPATION IN 2011-12

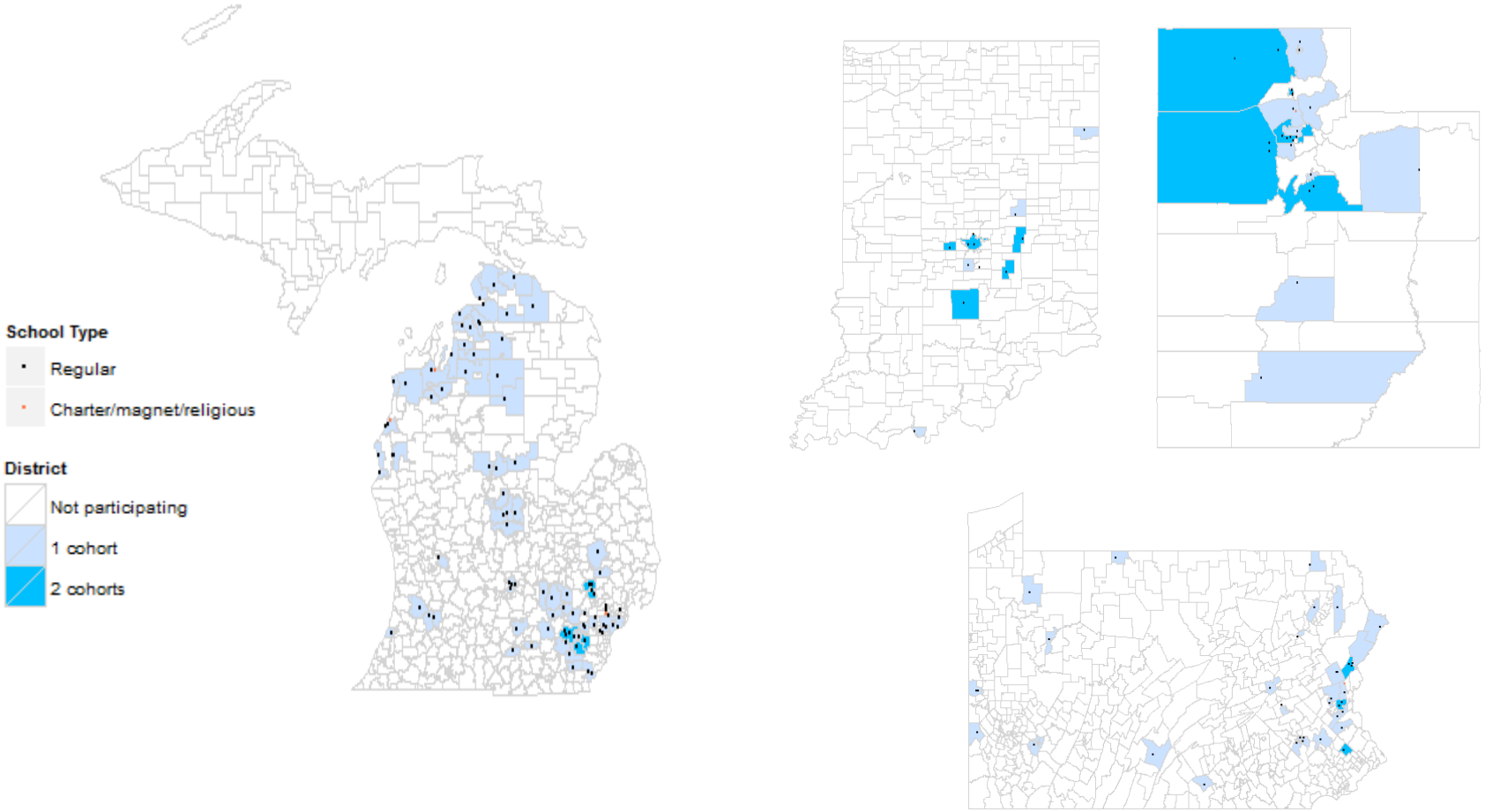


FIGURE B.2. MAP OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP PARTICIPATION IN 2012-13

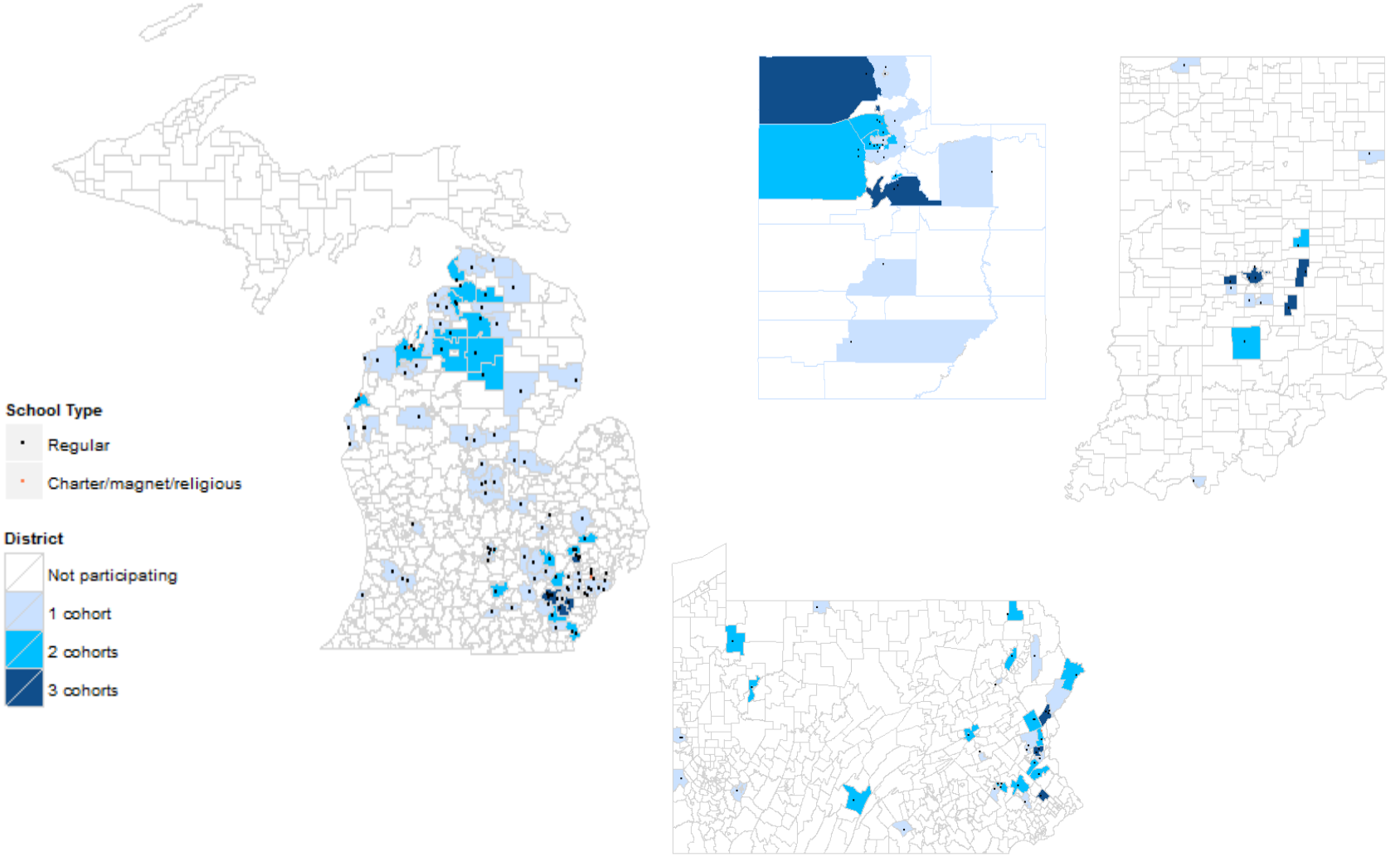


FIGURE B.3. MAP OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP PARTICIPATION IN 2013-14

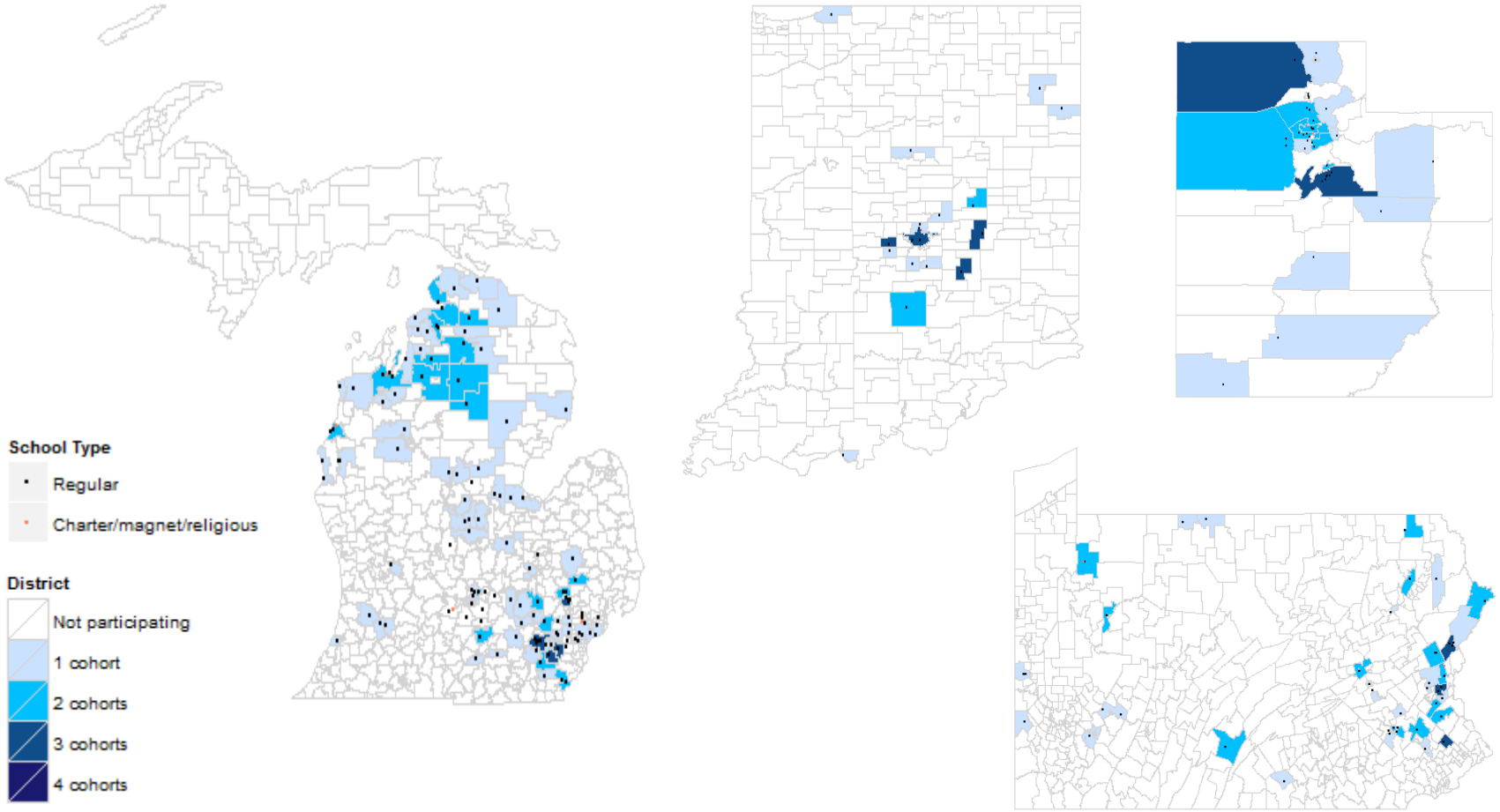


FIGURE B.4. MAP OF READING APPRENTICESHIP IMPROVING SECONDARY EDUCATION SCALE-UP PARTICIPATION IN 2014-15

Appendix C: Results from GL Analysis

TABLE C1. RESULTS FROM GL ANALYSIS

Factor	Model 1		Model 2		Model 2s		Model 2sm	
	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value
Contextual Factors								
School size (no. of 9th grade students enrolled)	2.07	.14	3.00	.09	3.00	.09	2.49	.10
Student-teacher ratio	-2.90	.30	-3.18	.43	-3.18	.43	-2.74	.47
% of students enrolled in free or reduced priced lunch	-1.77	.39	-2.10	.57	-2.10	.57		
% of students in each ethnicity category								
% Hispanic	10.29	.27	20.61	.14	20.61	.14	23.37	.03
% Black	15.76	.07	24.23	.05	24.23	.05	22.03	.02
% White	12.13	.11	19.90	.07	19.90	.07	21.80	.02
Level of agreement w/ statement that the principal is involved in district decisions about educational issues			-0.83	.23	-0.28	.67		
Level of agreement w/ statement that the school receives enough instructional resources to meet needs of students			-0.40	.48	-0.09	.88		
Level of agreement with the statement that district provides school with sufficient data/info to make informed decisions			0.42	.46	0.09	.88		
Years served as principal			-0.15	.03	-0.12	.08	-0.13	.03
Principal turnover (no. of principals at school in past five years)			-0.15	.74	-0.05	.92		
Teacher turnover (level of agreement w/ statement that teachers tend to stay at school for more than five years)			-0.48	.58	0.05	.95		
Principal retention (level of agreement with statement that the principal's immediate plan is to stay at their school)			-0.77	.05	-0.62	.11	-0.48	.18
Teacher retention (current retention rate compared to last five years)			-0.13	.80	-0.39	.46	-0.51	.31
Years of teaching experience	-0.08	.35	-0.03	.83				
Rating of school leadership	0.57	.24	1.01	.27				

TABLE C1. RESULTS FROM GL ANALYSIS

Factor	Model 1		Model 2		Model 2s		Model 2sm	
	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value
Malleable Factors								
Principal attendance at training			-0.23	.57	-0.30	.35	-0.26	.40
If participating in RAISE was required	-1.26	.22	-1.75	.25				
Attendance at RAISE monthly school meetings (no. attended)	0.45	.01	0.60	.02	0.66	.01	0.66	<.01
Level of agreement with the statement that teacher has enough time to plan Reading Apprenticeship lessons	-0.49	.24	-0.16	.83				
Use of Reading Apprenticeship practices	-0.12	.77	0.00	.99				
Intermediate Outcomes								
If principal recommended joining RAISE to other teachers or personnel			-0.48	.66	-0.94	.40	-0.56	.59
If principal thinks RAISE would continue in their school without federal funding			0.48	.70	1.29	.32	1.29	.25
Level of responsibility for the success of Reading Apprenticeship at their school	0.49	.28	0.95	.22				
Level of commitment to making Reading Apprenticeship work in school	0.95	.07	2.21	.03	2.59	<.01	2.00	<.01
If principal completed survey	1.84	.01						
Constant	-14.56	.18	-22.96	.23	-22.96	.23	-26.89	.06
R2	0.31		0.56		0.43		0.39	
(Adj. R2)	(0.21)		(0.32)		(0.25)		(0.27)	
Sample size (school)	130		76		81		85	

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.

Appendix D: Survey Questions Used in Chapter 8

TABLE D1. SURVEY QUESTIONS IN COMMON BETWEEN SCALE-UP AND RCT

Question: Please rate your level of confidence in your ability to do the following.	
	Answer options
Provide opportunities for reading a variety of texts of different types/genres	Not at all confident Less than moderately confident Moderately confident More than moderately confident Very confident
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 or writing activities in groups (small groups or whole class), (i.e. setting norms, creating safety, providing prompts that promote collaboration, and providing guidance/feedback)	
Question: Thinking about the RAISE PD overall, how well do you feel it prepared you to use the following components in your classroom?	
	Answer options
Providing students opportunities for reading a variety of texts of different types/genres	Not at all prepared Less than moderately prepared Moderately prepared More than moderately prepared Very prepared
Teaching students to analyze their own thinking about reading texts	
Structuring lessons so that students are held accountable for reading (e.g., 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)	
Providing explicit instruction around reading comprehension strategies (e.g., setting a reading purpose, previewing text, chunking, visualizing)	
Modeling/demonstrating reading comprehension strategies (e.g., setting a reading purpose, previewing text, chunking, visualizing)	
Supporting students in working on reading or writing activities collaboratively, (i.e. setting norms, creating safety, providing prompts that promote collaboration, and providing guidance/feedback on student participation)	
Giving 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)	
Facilitating students' active engagement in learning through the use of inquiry-based instructional methods	
Asking students to pose questions and problems about course readings	
Employing 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	

TABLE D1. SURVEY QUESTIONS IN COMMON BETWEEN SCALE-UP AND RCT

Question: Which activities took place during the last monthly meeting you attended?		
Answer options (Check all that apply)		
Reviewing student work	Discussing problematic lessons	Problem solving
Sharing successful lessons	Exploring Reading Apprenticeship tools/protocols	Other
Question: What were the reason(s) you did not attend the RAISE school team meetings?		
Answer options (Check all that apply)		
I prefer not to respond	Meetings were not offered	I had other obligation(s) at that time
I was not interested		Other
Question: How helpful were these activities in improving Reading Apprenticeship implementation in your classroom?		
Answer options (Choose one)		
Not at all helpful	Less than moderately helpful	Moderately helpful
More than moderately helpful		Very helpful
Question: Thinking back on the last four weeks of instruction, what kind of support for implementing Reading Apprenticeship in your classroom have you received? Please do not consider any activities from RAISE monthly meetings in your response.		
Answer options (Check all that apply)		
I did not receive support for literacy instruction	Coaching/mentoring	Model lessons
Observation/feedback	Resources	Classroom management help
Political Support (e.g. someone "backed you up" in a conflict over your implementation of literacy instruction)	A change in school/district policy that was relevant to literacy instruction	Other
Question: From whom did you receive support?		
Answer options (Check all that apply)		
Principal or other school administrator	Teacher leader	Department head
Other teachers	Program representative	Other
Question: In general, how helpful was this support for improving Reading Apprenticeship implementation in your classroom?		
Answer options (Choose one)		
Not at all helpful	Less than moderately helpful	Moderately helpful
More than moderately helpful		Very helpful
Question: What kinds of discussions have you had with other RAISE trained teachers about Reading Apprenticeship this school year? Please exclude discussions during your RAISE team monthly meetings and the RAISE professional development institutes.		
Answer options (Check all that apply)		
Reviewing student work	Discussing problematic lessons	Discussing what helps students learn the best
Sharing successful lessons	Discussing resources for literacy instruction	Other
I did not have these types of discussions with other teachers about literacy instruction		

TABLE D1. SURVEY QUESTIONS IN COMMON BETWEEN SCALE-UP AND RCT

Question: What challenges have you faced in implementing Reading Apprenticeship?		
Answer options (Check all that apply)		
Lack of materials	Lack of parent support	Lack of administrative support
Competing priorities	Too much work to implement	Not enough training on RA
Student behavior	Student ability	None
Lack of understanding of how to implement Reading Apprenticeship in my class		Other
Question: How committed are you to making Reading Apprenticeship work in your classroom?		
Answer options (Choose one)		
Not a priority	Willing to give it a chance	Fully committed
Question: How would you describe your own level of understanding of the Reading Apprenticeship framework?		
Answer options (Choose one)		
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)		
Question: Thinking back over your experience, how well do you think Reading Apprenticeship was aligned with the goals of your classroom?		
Answer options (Choose one)		
Not well aligned	Somewhat well aligned	Very well aligned
Question: Thinking back over your experience, how well do you think Reading Apprenticeship was aligned with the content standards of your classroom?		
Answer options (Choose one)		
Not well aligned	Somewhat well aligned	Very well aligned
Question: To what extent do you agree with the following statement? The implementation of Reading Apprenticeship will improve student achievement in my classroom.		
Answer options (Choose one)		
Not at all effective	Less than moderately effective	Moderately effective
More than moderately effective		Very effective

Note. RAISE stands for Reading Apprenticeship Improving Secondary Education.