THE NATIONAL CENTER ON SCALING UP EFFECTIVE SCHOOLS

School Processes That Can Drive Scaling-up of an Innovation, or Contribute to its Abandonment

Jenna Zacamy | Denis Newman | Valeriy Lazarev | Li Lin

Conference Paper
October 2015

VANDERBILT PEABODY COLLEGE

EDC Learning transforms lives. | Florida State University 1851 | Georgia State University | The University of North Carolina at Chapel Hill | Wisconsin University of Wisconsin-Madison

The National Center on Scaling Up Effective Schools (NCSU) is a national research and development center that focuses on identifying the combination of essential components and the programs, practices, processes and policies that make some high schools in large urban districts particularly effective with low income students, minority students, and English language learners. The Center's goal is to develop, implement, and test new processes that other districts will be able to use to scale up effective practices within the context of their own goals and unique circumstances. Led by Vanderbilt University's Peabody College, our partners include The University of North Carolina at Chapel Hill, Florida State University, the University of Wisconsin-Madison, Georgia State University, the University of California at Riverside, and the Education Development Center.

This paper was presented at NCSU's second national conference, Using Continuous Improvement to Integrating Design, Implementation, and Scale Up. The conference was held on October 7-9, 2015 in Nashville, TN. The authors are:

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

The research reported here was supported by an Investing in Innovation (i3) Validation grant (#U39B100255) awarded to WestEd. The opinions expressed are those of the authors and do not necessarily represent the views of WestEd, the sponsor, or the National Center on Scaling Up Effective Schools.

Table of Contents

Introduction

This paper reports findings from a multi-year study of the scale-up of Reading Apprenticeship (RA), an approach to improve academic literacy by helping teachers provide the support students need to be successful readers in the content areas. WestEd's Strategic Literacy Initiative (SLI), began developing the program 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, WestEd 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.[1,2] 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,[3] but also a parallel effort to study the schools outside of the RCT that were to receive the same professional development and other supports in implementing RA. This scale-up study was intended to provide formative feedback to the SLI developers to help them achieve their goal to build local education agency capacity to disseminate, support, and sustain academic literacy improvement in high school subject areas within their regions. Whereas, the i3 and similar funding can "prime the pump," the project must build the capacity to disseminate, support, and sustain the innovation. Ultimately, adoption at the state-level, driven by local adoption and evidence of success, will keep it going.

This formative evaluation of RAISE implementation collected and analyzed data on the number of trainings, the reach of the program, and the program elements that were taken up or not by participants. In this study, we followed 239 schools in four states (Utah, Michigan, Indiana, and Pennsylvania) as they participated in the expansion of the program. Schools and teachers were added to the project each year. We surveyed the teachers and principals and participated in many of the project meetings and training events. We tracked participation in the scale-up through sign-in sheets at each of the training events and maintained a spreadsheet[4] of all the schools and teachers recording new additions each year and, less systematically, as teachers left the school or stopped participating in the program. Approximately 1720 teachers received training in the scale-up study side of the overall i3 project.

To support inquiry into the scale-up process, we developed an unconventional spiraling logic model (Zacamy, Newman, Lin, & Jaciw, 2015) described below, 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 and changes in educator attitudes.

---

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

[2] Through RAISE, WestEd served approximately 280 schools, 1950 teachers and 409,500 students in five states (California, Indiana, Michigan, Pennsylvania, Utah).

[3] The RCT is being reported separately in a final report as well as reports several exploratory analyses currently in preparation.

[4] This "participant tracker" document, while not a formal data collection device, became the basis for the analysis we present here.

Our first set of studies, reviewed below (and reported in Zacamy, Jaciw, Lin, & Newman, 2014; Zacamy & Newman, 2014; Zacamy et al., 2015), examined survey responses to gauge adoption of RA, buy in by teachers and school administrators along with measures of participation in program activities. These analyses provided useful insights and suggestions about malleable factors that could be focus of future improvements. The surveys, however, did not directly address the actual scaling-up of RA, that is, the numerical increase in number of teachers and schools using RA. This is an issue addressed in our logic model, which hypothesized that "as capacity and support builds, we expect districts and schools to increase the numbers of teachers implementing RA; that is, schools will send more teachers to RAISE training and spread the RA ideas to other districts and schools." We were also sympathetic to the idea that scale-up can be measured in terms of the increasing numbers of participants over time (Slavin, 2002). With these considerations in mind, we harvested the data from the participant tracker and created a dataset of schools and teachers with which we could measure the growth and loss of RAISE participation within states, districts, and schools. These increases or decreases in number of participants becomes an outcome that can be considered the goal of the investments in professional development that are hypothesized to promote intermediate outcomes such as local level ownership and commitment.

The current paper reports our analysis linking growth in numbers of participants to contextual variables, e.g., school size, percent free and reduced price lunch (FRPL), to program events, e.g., monthly meetings, and to teacher and principal level of commitment.

Before going into these findings, we need to describe the program that SLI was scaling up.

Reading Apprenticeship Framework and Logic Model

What Reading Apprenticeship Is and Prior Research

Reading Apprenticeship (RA) 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 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 RA 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 RA intervention is inquiry-based professional development intended to transform teachers' understanding of their role in adolescent literacy development. This PD engages teachers in:

- 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 RA 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 RA instruction (Greenleaf et al., 2011a,b; Kemple et al., 2008; Somers et al., 2010).

Innovations for the i3 Implementation at Scale

The i3 validation project, that this scale-up study followed, proposed to provide RA professional development and supports for implementation for approximately 2800 teachers in 306 schools, across 5 states. The project included several innovations to scale to this level, including:

1. **Teacher leaders** recruited for each school team from the participating team of teachers. Teachers were expected to participate in at least monthly on-site teacher meetings facilitated by the teacher leaders.
2. **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 WestEd office. These activities included:
   - communicating and coordinating project activities at the state and local level
   - convening and facilitating cross-site Teacher Leader meetings
   - working directly 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 the central WestEd office staff and other state coordinators
3. **85 Professional Development Facilitators** trained and apprenticed to deliver the revised, discipline-based 10-day RA professional development series. As part of this, SLI developed materials, protocols, and assessments to support facilitator development and "certified" facilitators to use the professional development materials for site based trainings (i.e. outside of the RAISE 10-day Institute).
4. **RAISE administrator programs and materials.** SLI developed an online administrator course and provided opportunities during the teacher professional development sessions for RAISE administrators to share their ideas, needs, and perspectives on their school teams' implementation.

Scale-up Research vs. Scaling-up

In this 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 in to 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 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 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 small 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.

The focus of this study is to understand the processes involved in scaling up RA 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). 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, 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.

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 RA 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 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 RA 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 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* school systems build capacity to implement and disseminate RA 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 also 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 scale-up process and begin to build generalizations about the conditions for successful scale-up of RA 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.

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 RA being broadly institutionalized.[5] 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 "WestEd'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 RA 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 RA, (3) buy-in to the RA framework, (4) capacity to implement and disseminate RA practices, and (5) student achievement. Our first two intermediate outcomes—increased participation and classroom fidelity of RA—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 RA in later stages of the process.

These intermediate outcomes will also interact with each other. As buy-in and commitment to RA increase, we hypothesize that districts, schools, and teachers will dedicate the time and resources necessary to increase capacity to implement and disseminate RA at the local level. As capacity and support builds, we expect districts and schools to increase the numbers of teachers implementing RA; that is, schools will send more teachers to RAISE training and spread the RA ideas to other districts and schools. We also expect classroom fidelity of RA 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

[5] See Zacamy, Newman, Lin, and Jaciw (2015) for comprehensive narrative description of each stage of the logic model and accompanying figures.

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, RA 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 RA spreading with depth beyond the original 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, but acknowledges balancing the centralized, on-going research and development functionality of the developers with the uptake of reform ownership at the local level.

Study of RAISE Scale-up: Questions and Methods

As shown in Table 1, SLI's model for scale-up included participation from four consecutive cohorts of RAISE teachers and schools.

TABLE 1. YEARS OF PARTICIPATION FOR RAISE 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.

Cohort 1 included 65 schools and 365 teachers across the four states. Cohorts 2, 3, and 4 included teachers from existing RAISE schools (i.e., schools where teachers participating in previous cohorts) as well as from new schools. By the 4th Cohort, 239 schools and approximately 1720 teachers have participated in the scale-up sites. As the scale-up process proceeded across contexts, states, and years, we were able 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 have gained and lost participants.

The scale-up study is guided by research questions investigating the spread of RA the scale-up process, and contextual factors that affect scale-up. In addition to measuring the study's intermediate outcomes, these questions investigate the transfer of responsibility for and ownership of the RAISE initiative from the RA developers to the local level, which is represented by movement through the stages of our logic model.[6]

To address these questions, we have observed and documented key project activities; tracked the numbers of schools, teachers, and students served by this initiative; and surveyed participating teachers (three times a year during each year of implementation) and school administrators

[6] For a more detailed description of the research questions and rationale, see *Year 1 Interim Report of Reading Apprenticeship/RAISE Scale-up.*

(annually).[7] Through the surveys, we were able to measure general uptake of the RAISE project 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.[8]

We have not measured classroom fidelity of RA implementation or the effect of RA on student achievement in this study since a concurrent RCT has investigated these outcomes.[9] We have prepared several reports addressing a number of these questions. The focus of the current report is on connecting quantitative evidence of scaling-up and diminishing of participation with indicators of processes within schools that predict this growth or loss.

Findings

Early Indications of Enthusiasm and Initial Challenges

Findings from teacher and administrator survey data from the first year of the study (2011-12), suggest considerably high levels of buy-in and commitment from this initial cohort of RAISE scale-up participants. Across the four states,

- 95% of the *administrators* responded that they were either fully committed or fairly committed to making RA work at their schools.
- 82% of the *teachers* responded that they were either fully committed or fairly committed to making RA work in their classrooms.
- 97% of the *administrators* said that they believe student learning at their school will improve if more teachers join RAISE.
- 92% of *teachers* said that they strongly agreed or agreed with the statement that "Reading Apprenticeship is an appropriate framework for literacy instruction in my classroom."

Additionally, teachers reported high ratings of effectiveness of the RAISE professional development, with 90% of these teachers strongly agreeing or agreeing that the Summer 5-Day Institute led to changes in their teaching practices. The results also suggest high levels of implementation, with 67% of the teachers reporting using the RA pedagogical practices at least a few times a week, and 27% using them in each lesson in the first year. While half of the teachers reported that implementing RA was moderately challenging, 91% said that they planned to use the RA framework to inform their instruction during the next school year (2012-2013 school year). We found similar trends for the first year of implementation for the second and third cohort of RA teachers.

---

[7] We have 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 included surveys, interviews, focus groups, and site visits with various stakeholders. Results from the case studies are not reported in this paper; see *Case Studies of the Scaling and Sustaining of Reading Apprenticeship in Four Michigan Secondary Schools: Report 2* for the year 2 case study report.

[8] See Appendix A for a detailed description of the data collection.

[9] The final reports on the concurrently conducted RCT as well as additional reports on exploratory analyses are currently in preparation. Schools from the RCT are not included in the scale-up study although a comparison of schools in the RCT and scale-up is on-going.

Administrators identified competing initiatives (56%) and budget constraints (33%) as the two primary challenges to sustaining RAISE long-term. Additionally, when asked if they thought RAISE would continue in their schools without the i3 federal funding, over half responded "No" or "I don't know." Over half of the teachers identified competing priorities as a primary challenge that they faced in implementing RA during the 2011-2012 school year (Zacamy, Gray, Jaciw, & Newman, 2013; Zacamy et al., 2014).

Implementation and Commitment Over Three Years

While the initial look at the surveys considered individual responses, in most of the subsequent analyses we report, we have taken the school as the unit rather than individual teacher or participant. This was because we considered the processes of interest to most likely be related to the cohesion (or not) of the community developing within the school.

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 initiative and sustainability. This followed the RAISE scale-up logic model where we hypothesized that in early stages as teachers deepen RA practice and strengthen support ties over time, we will see an increase in buy-in and capacity to implement and disseminate RA. Also, we expected that "ownership" of the initiative will begin to be transferred to the local level, which will support sustainability as formal supports from the developers are withdrawn. Therefore, we identified key indicators of participants' uptake of RAISE activities, including if they received support for RA instruction, their attendance at the RAISE monthly school meetings, and the extent to which they used RA practices. We also identified early indicators of scale-up "success", including teachers' buy-in and commitment levels to RA in their classroom and school, and if teachers plan to continue using the framework in the next year.

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, this enthusiasm decreased or leveled off by the third year. Figures 1 and 2 illustrate school level average responses to each of the survey occasions, across Year 1 through Year 3. The graphs show the survey occasions 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 is proportionate to the number of schools at each point. We have also indicated the overall sample mean and median with a purple and green circle, respectively.

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. As shown in Figure 1, 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, by the end of the third year of implementation 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 ($p < .001$).

In spite of the decrease in participation, reported levels of buy-in remained high. At the beginning and end of each year, we asked teachers to report their level of commitment to making RA work in their school. As shown in Figure 2, Cohort 1 schools reported high levels of commitment, with a majority being fully or fairly committed to making RA work. There is, however, a statistically significant decrease in commitment levels between the first and third year.

We then examined whether decreases in the uptake of RAISE activities were related to changes in the indicators of success. We found a statistically significant positive relationship between the change in commitment to making RA work at the school and change in participation in each of the RAISE activities, suggesting that school level engagement or collaboration is an important process related to continued use of the program.

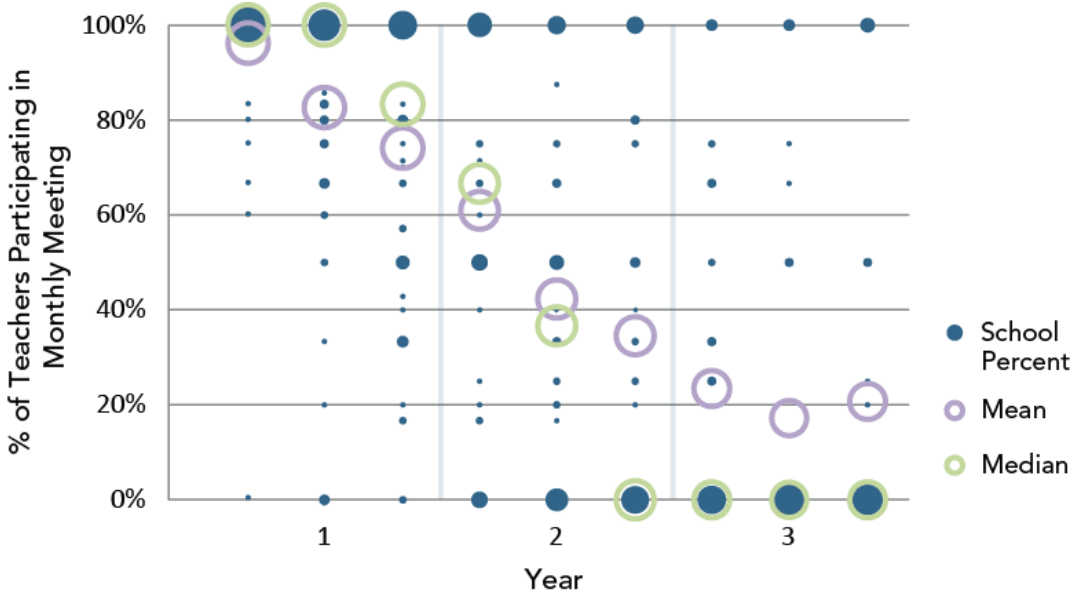

FIGURE 1. COHORT 1 ATTENDANCE AT MONTHLY MEETINGS OVER THREE YEARS

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

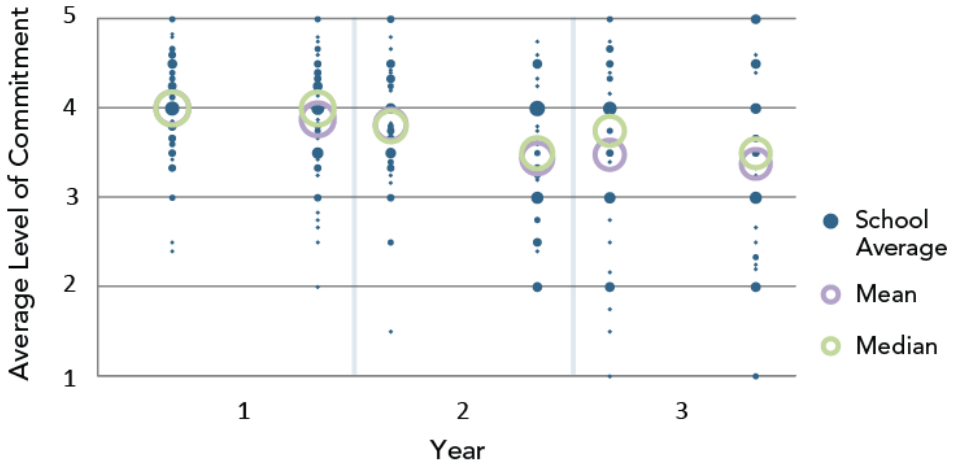

FIGURE 2. COHORT 1 COMMITMENT TO READING APPRENTICESHIP AT SCHOOL OVER THREE YEARS

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

We also found a greater spread in responses by the third year with schools reporting at both high and low levels, whereas in the first year most responses were generally clustered together. This variation further supports the hypothesis that contextual factors at the school or district levels may support or hinder the scale-up process and should be explored. While we noted a decline in the survey response rates across the three years and recognize that this may introduce some response bias, we did not find major differences in commitment levels for teachers who stayed in the sample until the end of the study and those that left in the second year. If there was a response bias (in the positive direction), then the decrease in the participation of RAISE activities and commitment levels would be underreported, and our conclusions would be further substantiated (Zacamy, Newman, Lin, & Jaciw, 2015).

Quantitative Analysis of Scale-Up Outcomes: "Gain and Loss"

The results of the analysis of the first cohort of RAISE teachers raised questions about the sustainability and about the mechanisms for sustaining Reading Apprenticeship implementation instantiated in RAISE, and led us to further investigate what contextual conditions or malleable factors may affect scale-up. A challenge, however, was to identify an objective outcome measure of scale.[10] The RAISE recruitment approach, which included both internal (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.

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 3 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.

[10] While our surveys of teachers and administrators provided useful data on their implementation and commitment, as well as reported challenges to sustainability, the decline in response rates by the third year introduced potential response bias of self-reported measures.

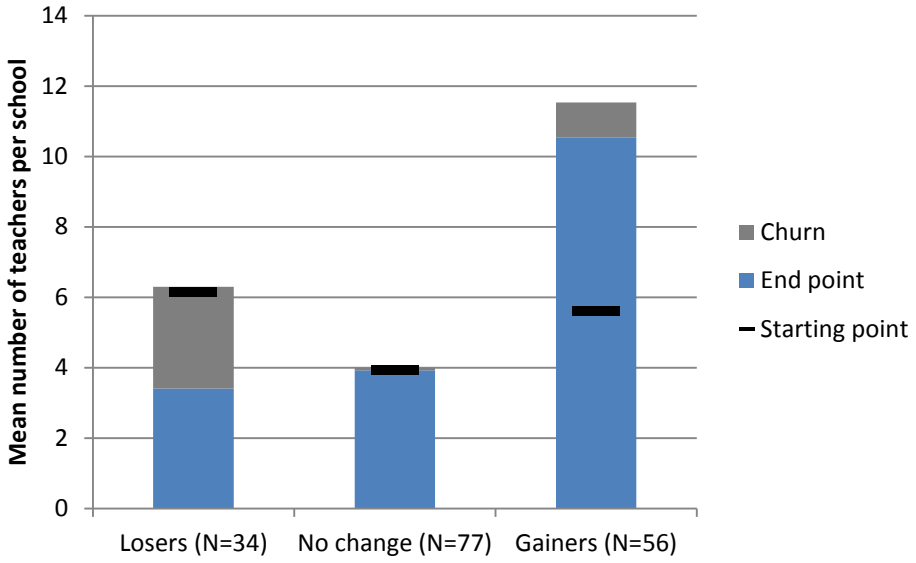

FIGURE 3: SCHOOLS THAT LOST OR GAINED TEACHERS DURING THEIR TWO OR THREE YEARS IN THE PROJECT

The difference between the number of participating teachers at the end of the study period and the number of teacher enrolled in the initial training—we term this the GL ("Gain/Loss") metric—is an indicator of a positive or negative scale-up process within the school. 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. This database was used to track district, school, and teacher level attendance at the RAISE Institute (using the attendance logs), which schools and teachers agreed to complete study surveys, and if/when schools or teachers were no longer participating in RAISE (either because they left the school or were no longer implementing RA). Data were linked across years to track the expansion and participation of states, districts, and schools.[11]

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 scale-up study

[11] We did not track if teachers attended RA professional development outside of the 10-day RAISE Institute (i.e. through district or school site based trainings).

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 "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.)[12]

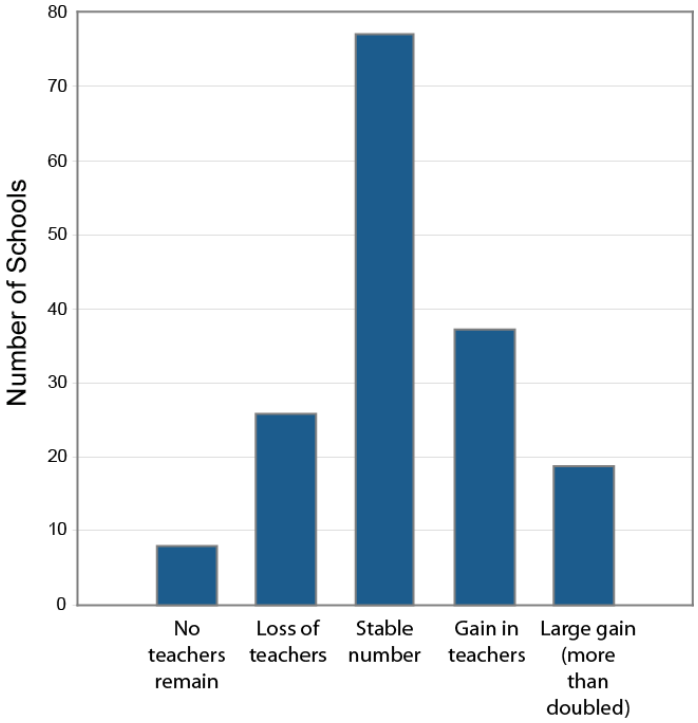

FIGURE 4. DISTRIBUTION OF SCHOOLS BY GAIN/LOSS IN THE NUMBER OF TEACHERS PARTICIPATING IN RA

Figure 4 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 dynamics 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:

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.

[12] 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.

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 2 lists the potential predictors and their sources.

TABLE 2. 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
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

TABLE 2. POTENTIAL PREDICTORS OF GL

Potential Predictor	Data Source
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

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 RA, 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.[13]

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.

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

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. See Table B1 in Appendix B for the full set of results.

First, there are a set of results that suggest establishing a community of practice early on supports longer-term spread and sustainability. We found that teachers' level of commitment to making RA 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).

We found that contextual variables 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.

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 RA 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 RA (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.

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 RA into lesson plans, or how frequently teachers used RA practices in their lessons, may suggest that, in the initial phases of implementation, the extent of actual use of RA in class is less important than the teacher buy-in. Additionally, findings that it did not matter for the GL if the principal recommended RA to others in their school or if they believe RA will continue without federal funding present RA as a self-sustaining teacher-driven program that does not require a lot of administrative input or financial support, which makes it a potentially viable option for schools with limited resources.

Discussion

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 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 RA.

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 RA work at the school. This is a better predictor than the extent that teachers implemented RA practices in their own classroom. 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 RA 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 RA.

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 RA 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.

This research has taken a step toward integrating the quantitative view with the process view of scale-up. Our work will continue, especially in ways that will assist the program developers in improving the growth and sustainability of their program.

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.

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

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.

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.

Kemple, J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., and 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.

Newman, D. and Jaciw, Andrew 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

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.

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/

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. *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.

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, Z., 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, Z., 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.

Appendix A: Detailed Data Collection Activities

Across the five year study, researchers collected multiple sources of data for qualitative and quantitative analysis including professional development records; principal and teacher surveys; interview/focus groups with teachers, instructional support staff, and administrators; and site visits. Data from informal interviews, emails, and discussions have also been used to inform data collection and reporting.

Professional Development Observations and Attendance Records

Throughout the study, researchers conducted observations of a sample of the RAISE professional development in order to gain a strong understanding of the Reading Apprenticeship framework, expectations for teacher and school implementation, and how the training agendas are designed to build capacity and engage participants in the RAISE initiative. Researchers also used components of the training to inform survey design. We collected artifacts (e.g., handouts, agendas, resource materials) from observed sessions and 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. Data were linked across years to track the expansion and participation of states, districts, and schools.

District and School Demographic Characteristics

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.

Principal/School Administrator Surveys

During Years 1-3, researchers conducted annual surveys of principals and/or school administrators in order to gather the school leadership perspective on the RAISE initiative. Using the data from the participant database and professional development attendance records, researchers requested consent to participate in the surveys from all RAISE schools and teachers. Given the number of schools and teachers participating in RAISE, the project could not support additional stipends or incentives for teachers to complete surveys. Following the deployment of each survey, researchers followed up via email, fax, and phone with all non-respondents.

Specific domains measured in the surveys were guided by the logic model and included buy-in, commitment to RAISE, and sustainability of the initiative beyond the grant funding. The administrator surveys were deployed in May of 2012, 2013, and 2014 to consenting administrators who had teachers in either Cohort 1, 2, and/or 3 at their school. 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 RA 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 RA 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 RA 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 question 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

During Years 1-3, all consented RAISE teachers currently in their schools received three surveys per year (fall, winter, spring). A 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 RA prior to RAISE, we asked

teachers how many hours of previous RA 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 site
- Availability, types, and helpfulness of support for implementing RA 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 RA pedagogical practices in their classroom, and if they had enough time to plan RA 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 RA 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 RA is appropriate framework for literacy instruction at 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 RA 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, CIT). 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 RA, how well RAISE aligned with the instructional goals, rigor, and needs of the students in their class/school, and teachers' plans to use the RA 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 RA following the professional development.

Appendix B: Results from GL Analysis

TABLE B1. 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 (number 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 with the statement that the principal is involved in district decisions about educational issues			-0.83	.23	-0.28	.67		
Level of agreement with the statement that the school receives instructional resources commensurate to meet needs of students			-0.40	.48	-0.09	.88		
Level of agreement with the statement that district provides school with sufficient data/information 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 (# of head principals the school had in past five years)			-0.15	.74	-0.05	.92		

TABLE B1. 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
Teacher turnover (level of agreement with the 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 (how current retention rate compares 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				
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 (# of meetings 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

TABLE B1. 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
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	