4

Teaching Reading for Understanding: Summarizing the Curriculum and Instruction Work of the Five Core Reading for Understanding Teams

Gina Biancarosa, *University of Oregon*
Peter Afflerbach, *University of Maryland*
P. David Pearson, *University of California, Berkeley*

CONTENTS

INTRODUCTION

In examining the teaching and learning of reading comprehension, the five teams (excluding the Educational Testing Service [ETS] because of its exclusive emphasis on assessment—see Chapter 3) in the Reading for Understanding (RfU) consortium pursued different but complementary goals regarding the related processes, components, and uses of comprehension. The RfU teams designed instruction that addressed different aspects of comprehension development, from emphases on the key antecedents of decoding and listening comprehension, to explicit strategies for making and monitoring meaning, to activities that require students to put the fruits of their comprehension to work for some other purpose, to collaborations and conversations that promote rich talk about text, where the goal is developing or refining many kinds of knowledge,

including insights about the human condition, knowledge that describes and explains how the natural and social worlds work, and even metaknowledge about the nature of language, knowledge, and understanding.

The portfolio is expansive and complex, culminating in well-designed and implemented randomized controlled trials (RCTs) that incorporated a wide range of independent variables, often targeting those malleable factors discussed extensively in Chapter 2. All of the interventions emanated from a theoretical base about the nature and development of reading comprehension (but not always the same theoretical base). They detailed explicit models (theories of action) of how particular facets of the reading comprehension puzzle can be shaped in instructional settings to elicit changes in performance. The details of the actual interventions were, in general, as well informed by the wisdom of practice as by the theories on which they were built; teachers were involved as co-designers or critics along the way, often in extensive design research efforts.

The RfU teams focused on a range of outcomes. Outcomes ranged from discrete component skills, often representing near transfer of instructional targets, to complex comprehension, writing, and editing tasks, representing far transfer of instructional targets. Measures of these outcomes ranged from curriculum aligned to curriculum independent. Finally, they included researcher-developed measures, measures developed by the primary RfU assessment teams, and otherwise commercially available measures.[1] They measured teaching as well as learning, always documenting what actually occurred in the intervention classrooms and, often, in the business-as-usual (BAU) control groups. In contrast to many prior efforts in pedagogical research, these were statistically well-powered efforts, with samples sufficiently large and well defined to detect even small effects.[2] In short, there was every reason to believe, going into the RCT phase of the RfU initiative, that if there were effective interventions to be found, they would be found in this initiative.

As a reminder, the focus in this chapter is to summarize the efforts and key findings from each of the five RfU teams before shifting the focus, in Chapter 5, to a panoramic analysis and synthesis of findings as well as pedagogical themes, practices, and insights across the teams. Given the vast scope of the RfU endeavor, we first unpack in some detail what each team learned in its efforts so that readers might appreciate the breadth, depth, and nuance of the RfU instructional portfolio. As we move to Chapter 5, we assess the impact of their commonalities and distinctions. Our reasoning was that if we could tell the story and reveal the essence and core of each team's effort, we would set the stage for a more meaningful cross-team synthesis.

This approach is necessary given the differences in how each RfU team approached its work. For example, two of the teams, the Language and Reading Research Consortium (LARRC) and Reading, Evidence, and Argumentation in Disciplinary Instruction (READI), were focused and integrated across the 5-year cycle of work; they had what

[1] To assist the reader who wants to pursue a deeper examination of the specific measures used within and across the five teams, Appendix 4-1 provides a compendium of all the measures used in the RCTs reported in this chapter. Chapter 3 provides a more extensive review of many of these measures in its appendix as well.

[2] Appendix 4-2 summarizes the demographic information, by team and RCT, of the students involved in the RCTs.

we came to call a "long runway" leading from their initial conceptualization and design work to their culminating efficacy studies. As a contrast, another team, the Florida Center for Reading Research (FCRR), rapidly developed a diverse portfolio with at least eight "variations" on its curriculum and instruction (C&I) theme—a collection of comprehension tools for teachers (and students). The Catalyzing Comprehension through Discussion and Debate (CCDD) and Promoting Adolescents' Comprehension of Text (PACT), each with at least two major strands of parallel research, landed somewhere in the middle. Despite the diversity of approaches to the work, each team was required by the final (fifth) year of the RfU to conduct an efficacy trial or RCT on at least one significant pedagogical intervention. Given the fact that each team met this requirement of conducting one or more major efficacy trials, we decided to summarize the efficacy trials of each team and work our way back into the development efforts that led up to the trials.

In this chapter, we begin with a rationale for the curriculum and instruction portfolio for the entire RfU consortium. This is followed by the briefest of overviews of the work of each team, just to provide a sense of the range of curriculum and instruction efforts across teams, before turning to the heart of the chapter: a more elaborate account, in order of the grade levels targeted, of the work of each team—LARRC, FCRR, CCDD, PACT, and READI.

THE RATIONALE FOR THE PEDAGOGICAL EMPHASIS IN THE RFU

Making progress in understanding all of the facets of reading comprehension—its nature, development, pedagogy, and assessment—was important to the designers of the RfU initiative. In fact, progress in each of those areas is contingent on progress in the others. Instructional improvement in the absence of strong linkages to theories of its development is likely to live a short life; and it is impossible to evaluate the impact of instruction without indices (good assessments) of development over time.

Instruction as First Among Equals

Improving instruction was the soul of the RfU initiative, as well it should have been—and should be—because it is the lack of progress in achievement, presumably attributable to a lack of successful pedagogical tools, that each and every RfU team set out to change. First and foremost, the crucial piece of evidence motivating this unusual and substantial investment in such a specific program of research (approximately $120 million over more than 5 years) was that too many students from grades 4–12 score below par on national (NAEP, 2019a) and international (e.g., PISA, 2018) assessments of reading comprehension achievement. Not only have scores been too low, but they have reflected little or no year-to-year progress in reading comprehension performance over the past two and a half decades (NAEP, 2019b), with particularly stable scores at the secondary level. A third concern is that these flat trends exist in the face of increasing expectations both within school and in the postsecondary worlds of work and tertiary education (NGA & CCSSO, 2010). Whether employed or pursuing a degree, students must read increasingly complex texts and perform increasingly complex reading-related tasks. Ironically, advances in the digital delivery and portrayal of information,

even in still and dynamic images, have only increased the range of "texts" that students must master and information that students must process to be competent in school and the workplace. It appears that many students are not up to the task. This shortcoming has been brought into sharper relief than ever in light of the widespread acceleration in new standards over the past decade, most prominently represented by the Common Core State Standards (NGA & CCSSO, 2010) as well as many state standards (e.g., Texas Essential Knowledge and Skills). It may well be the case that the traditional comprehension curricula that led us into the second decade of the 21st century are simply not up to the demands of today's literacy standards.

Two movements in particular highlight these shortcomings: disciplinary literacy and deeper learning. Disciplinary literacy is grounded in the increasing realization that while generic reading skills and practices represent a good start, they will not suffice in specific disciplines of the academy—literature, mathematics, the arts, the sciences, and the social sciences (NGA & CCSSO, 2010; Shanahan & Shanahan, 2008). Instead, learning in the disciplines requires discipline-specific reading strategies and mastery over discipline-specific discourses used to frame reasoning, explanation, and argumentation (Wineberg, 2001). A second movement, most commonly identified with the label of deeper learning (R. Anderson, personal communication, September 17, 2019; Goldman, Snow, & Vaughn, 2015; NRC, 2014), suggests that comprehension, at least simple comprehension of the text, is not enough; readers must go beyond comprehension to synthesize, analyze, critique, and apply what they learn while reading in the service of other goals or products— evaluating arguments or explanations within and across texts, working across sources to construct new arguments, and using information to solve important problems (in the spirit of project-based learning, for example).

Entering the RfU era, the field was informed by substantial research-based knowledge of reading comprehension. From the 1970s to the 1990s, we had, as documented in Chapter 2, gained increased understanding of how comprehension was orchestrated by readers as a process with many constituent parts (Anderson, Hiebert, Scott, & Wilkinson, 1985; Pressley & Afflerbach, 1995). We were, with the help of sociocultural perspectives (Freebody & Luke, 1990; Gee, 2000; Purcell-Gates, Perry, & Briseño, 2011), gaining knowledge of the contexts in which comprehension may be best taught, or learned, and used. Yet, this research and theory did not seem to matter much in relation to improving many students' comprehension performance. That was the context in which the RfU initiative was initiated.

The Pedagogical Charge

To address these issues and concerns across the pre-kindergarten (pre-K) through grade 12 continuum of reading comprehension development, the Institute of Education Sciences initiated the RfU grant program, providing a bold rationale and focus:

> Although the nation has invested billions of dollars in teaching children to read, many American students continue to struggle in reading. The latest data from the National Assessment of Educational Progress show that 1 out of 3 fourth-graders and 1 out of 4 eighth-graders cannot read at the basic level. That is, when reading grade appropriate material, these students do not understand what they read. It is difficult to imagine that

> students who cannot understand what they read will be successful in school or gain the skills necessary to succeed in the 21st century workforce. (IES, 2009, p. 5)

It was essentially a realization that while the history of teaching reading comprehension had been marked with some successes, it was also marked with failure to reach all students so that they might realize their potential as learners, workers, citizens, and individuals. The RfU teams were asked to change this pattern of performance that falls short of expectations, and it is to their work that we turn our attention.

Previewing the Curriculum and Instruction Portfolio of Work

We preview the entire range of activity across the five teams as a way of appreciating the breadth, as well as the interrelatedness, of activity carried out across the entire initiative. Then, on to a deeper analysis of the work of each team.

LARRC, one of two "early" (pre-K through grade 5) teams, created Let's Know! (LK), a 25-week multicomponent, supplemental curriculum for pre-K through grade 3 intended to help develop and improve children's language skills in anticipation of improving reading comprehension. LK was designed to improve both lower- and higher-level language skills—vocabulary, comprehension monitoring, and text-structure knowledge—as well as general language comprehension.

FCRR, the second "early" team, focused on assessing the value added of several component interventions, most focusing on one or more linguistic or cognitive skills, both proximal (did students improve on the specific component taught?) and distal (did the learning transfer to more general measures of language comprehension, literacy skill(s), or knowledge?). They were especially interested to learn whether the interventions were effective for children with weaker entry-level language and decoding skills.

CCDD implemented a program comprising two interventions as part of their RfU work: Word Generation (WG) and the Strategic Adolescent Reading Intervention (STARI). WG was designed for students in grades 4–8 to emphasize motivation, vocabulary, background knowledge, content-specific demands of text, and complex lines of argument to foster development of students' academic language, perspective-taking ability, and deep reading comprehension through the demands of discussion, debate, and writing. STARI was an omnibus, multicomponent program that addressed "fluency, word study, and comprehension, aiming to move struggling students two grade levels ahead in 1 year," as well as students' motivation and engagement (LaRusso, Donovan, & Snow, 2016, p. 14).

PACT investigated the role of cognitive processes, motivation, and intervention components to improve reading comprehension. PACT researchers developed two major multicomponent interventions: PACT, with a focus on reading comprehension and knowledge acquisition within middle and high school history classes, and Comprehension Circuit Training (CCT), which incorporated word identification, vocabulary enhancement, and comprehension and metacognition strategy development within middle school English language arts (ELA) classrooms. A major component of both PACT and CCT was team-based learning (TBL), a collaborative structure for promoting student-to-student support of learning.

Researchers within the READI team examined the development of students' disciplinary knowledge by focusing on higher-level reading comprehension strategies and evidence-based argumentation (EBA) to support adolescent learners in grades 6–12. The fundamental READI goal was to expand students' abilities to move beyond basic reading comprehension, to think critically about text, and to construct arguments from insights gleaned from the close reading of multiple text sources within the disciplines of history, science, and literature. READI researchers identified core constructs in the disciplines and centered instruction around them. READI also focused on students' development of discipline-specific epistemic orientations (understanding the nature, sources, and limitations of knowledge), which was regarded as key to suitable framing of reading tasks, successful comprehension, and transfer to new situations. Finally, READI emphasized the development of teacher learning as a key mediator of student learning.

EXAMINING THE RFU TEAM PORTFOLIOS

Language and Reading Research Consortium

Overview

LARRC, one of two "early" (pre-K through grade 5) teams, enacted a continuous line of inquiry with a singular focus to develop its pedagogical portfolio. Over the 5 years, LARRC scholars created, refined, tested, and fully evaluated LK—a 25-week supplemental curriculum for pre-K through grade 3 designed to develop and improve children's lower- and higher-level language and comprehension skills. These included vocabulary, comprehension monitoring, text structure, story grammar knowledge, and general language comprehension. The logic of the curriculum was that the cumulative effect of improvement in component skills would serve as a path to improved reading comprehension. Results from an RCT in which variations of the LK curriculum were compared to a BAU control revealed consistent, large, statistically significant effects favoring the LK curriculum on intervention-aligned measures of the vocabulary taught in the program and comprehension monitoring (see Table 4-1 for a summary of all effect sizes). Relative to BAU, minimal effects were found for understanding orally presented narrative and expository texts.

Developing the Let's Know! Curriculum

The LK curriculum was developed using the Curriculum Research Framework (Clements, 2007), which involved an iterative process of curriculum development encompassing three goals: (1) establishing foundations for curriculum, (2) building a student learning model, and (3) evaluating the effectiveness of curriculum. As the LK curriculum was created, researchers conducted pilot tests for implementation, feasibility, and efficacy, with formative and summative assessments included in the design and refinement process. Development of the LK curriculum was paralleled by a comprehensive design study (LARRC, 2016) in which researchers worked hand in glove with teachers and other school personnel to make certain that LK was well situated in the

contexts of schooling, that is, relevant to and supportive of existing curricula, classroom practices, and participating student and teacher needs.

Following the development of LARRC's LK curriculum, related inquiry assessed the influence of the curriculum on teaching—whether LK increased the quantity and quality of instruction (Pratt & Logan, 2014). Researchers used a single class observation to examine the impacts of LK on teachers' use of 18 language-focused comprehension supports and general classroom quality. The classroom observations were analyzed using the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008) and Snippets coding protocols (Pianta, Mashburn, Downer, Hamre, & Justice, 2008). Snippets allowed for examination of teachers' use of the language-focused comprehension supports prominently featured in LK lessons. Researchers determined that teachers working with the innovative LK curriculum exhibited significantly greater use of language-focused comprehension supports than did teachers in the comparison group. In addition, teachers using LK exhibited significantly higher classroom quality indicators, as indexed by the CLASS observation protocol. In short, the team concluded that LK had a positive influence on teacher behaviors.

LARRC researchers then examined the influence of differential "doses" (varying levels of LK vocabulary instruction) on students' vocabulary and comprehension development (LARRC, Arthur, & Davis, 2016). Researchers compared a single-dose version of the curriculum that they eventually dubbed LK-Broad (the normal LK vocabulary curriculum—LK^B), a double-dose version that they dubbed LK-Deep (LK vocabulary curriculum with each lesson repeated to double time on task—LK^D), and BAU vocabulary instruction using a quasi-experimental design. Measures focused on students' pretest and posttest vocabulary knowledge of words occurring within LK, as well as target vocabulary measures that assessed increases in students' knowledge for words taught in specific units and lessons. Vocabulary was assessed with the oral prompt, "Tell me what (vocabulary word) means." Coders used a detailed scoring rubric to assign two points for a correct definition, one point for partially correct responses, and zero points for an incorrect definition. Researchers determined that there were no statistically significant differences in students' vocabulary achievement when comparing LK^B to LK^D; however, effect size estimates for double-dose treatment (LK^D) were consistently greater than for the single-dose condition (LK^B). When analyzed as a single condition, the two variations of LK (LK^B and LK^D) produced superior mastery of taught vocabulary compared to BAU. When examined by grade level, results were consistently significant, positive, and large. The researchers speculated that the "dosing differences" received by students in LK^B and LK^D, in effect, may not have been so different. Qualitative data revealed that teachers in the single-dose condition unexpectedly provided students with learning opportunities related to new vocabulary words, frequently put the unit words on word walls, and may have referred to them outside of the LK lessons. While the firewall between treatment groups was not firm, the researchers concluded that "robust" vocabulary instruction at either the single- or double-dose intensities had positive effects on children's learning of targeted words.

Again, employing a quasi-experimental design, LARRC researchers (Johanson & Arthur, 2016) examined further the impact of these two conceptually different variations of LK—LK^D and LK^B—on a range of pre-kindergartners' more proximal component

skills (taught vocabulary, comprehension monitoring, and text-structure knowledge). LK^B included five different lesson types—grammar, vocabulary, inferencing, comprehension monitoring, and text-structure knowledge—whereas LK^D included only three of the lesson types present in LK^B—vocabulary, inferencing, and comprehension monitoring—but with additional practice time and opportunities. As with LARRC, Arthur, and Davis (2016), vocabulary was assessed by prompting students with, "Tell me what (vocabulary word) means," and scoring responses on the three-point scale. Comprehension monitoring was assessed as children listened to passages, identified inconsistencies in the passages, and then identified strategies that could correct the inconsistencies. The text-structure assessment required students to listen to two passages and then respond to multiple-choice items for which they selected the best main ideas and appropriate titles for the passages. Furthermore, researchers used a Listening Comprehension Measure, adapted from the Qualitative Reading Inventory, Fifth Edition (QRI-5; Leslie & Caldwell, 2011), as a distal measure. LARRC, Johanson, and Arthur (2016) hypothesized that children who were exposed to either LK^B or LK^D would significantly outperform children receiving BAU on measures of these skills. Both LK^B and LK^D students outperformed BAU students but the two levels of LK did not differ from one another.[3] Results on the proximal measures of comprehension monitoring and text-structure knowledge did not yield any significant effects.

Summarizing the research on the way to the RCT. To summarize to this point in the LARRC trajectory, LARRC research conducted in anticipation of the RCT began by enacting the Curriculum Research Framework to guide a systematic approach to curriculum development that focused on language comprehension for children in pre-K through grade 5. The collaborative development work was informed by the prior research on vocabulary and knowledge acquisition and guided by the experience of working teams of varied stakeholders, most notably classroom teachers, as they refined the curriculum in design studies and pilot studies. Following the development of LK, ensuing studies (LARRC et al., 2016; LARRC, Pratt, & Logan, 2014) focused on the curriculum's effect on teacher behaviors, student learning in relation to instruction (i.e., the development of the "component language skills" of vocabulary, comprehension monitoring, and text structure), and overall language comprehension. Perhaps the most apt summary is that the results supported the conclusion that students' vocabulary skills and comprehension monitoring, but not their overall listening comprehension, improved for both LK^D and LK^B compared to BAU.

The Randomized Controlled Trial

The combined curriculum development, design studies, and examinations of curriculum efficacy led LARRC researchers (LARRC, Jiang, & Davis, 2017) to conduct a culminating RCT to investigate the influence of the LK curriculum on students' comprehension and comprehension-related skills (comprehension monitoring, understanding

[3] Note that effect sizes for the proximal measures were reported using a rate ratio, which is an effect size often reported for negative binomial regression analyses, as in this case for measures representing counts. The effects can be interpreted as a score that is X times as large as the comparison condition.

narrative and expository text through inferencing, and text-structure knowledge) and vocabulary. Overall, the results of the RCT indicated consistent, large, statistically significant effects of the LK curriculum on comprehension monitoring and vocabulary measures relative to the BAU condition. Minimal effects were found for making inferences and using text structure, such as compare and contrast, to support comprehension of expository texts, and for sequencing events to support narrative comprehension.

Methods. The RCT was conducted with a cohort of 766 students enrolled in 132 classrooms in 61 schools in 6 states. Pre-kindergarteners numbered 167, with 155 students in kindergarten, 139 in grade 1, 155 in grade 2, and 150 in grade 3. Fifty-three percent of the students were female, and students averaged 6.5 years of age at the start of the academic year. Eighty-six percent of students were White, 8 percent were Black, 4 percent were Asian, and 2 percent were of other races; 12 percent were Hispanic or Latino. Six percent of participating students had individualized education programs. Nine percent of students had family incomes less than $25,000, 24 percent of students came from families with incomes of $25,001 to $50,000, 13 percent of students had family incomes of $50,001 to $75,000, and 45 percent of students had family incomes greater than $75,000. The mothers of half of the students held a bachelor's degree or higher, and 20 percent of students received free or reduced-price lunch. Teachers averaged 42.2 years of age and close to 14 years of teaching experience in pre-K through grade 3. The teacher population was 94 percent White, 3 percent Hispanic or Latino, and 2 percent Black. The average K–3 class size was 21 students. Pre-K classrooms averaged 17 students; 22 percent of pre-K classrooms were sponsored by Head Start.

Classrooms were randomly assigned to one of three conditions: LK^B, LK^D, and BAU. As detailed earlier, LK^B and LK^D differed in the use of practice lessons, text mapping, and Read to Know. In LK^D, text mapping and Read to Know lessons were replaced with lessons on integration and Words to Know, which provided additional practice. In LK^B, students learned text mapping, which focused on texts and grammatical structure, and Read to Know lessons in LK^B encouraged students to independently apply comprehension-related skills during reading. Both versions provided the same total number of lessons and weekly minutes of instruction.

Random assignments were blocked by school site and by grade. The BAU control classrooms received typical language arts instruction. In both LK conditions, teachers implemented four units over 25 weeks during the academic year. There were three 7-week units and one 4-week unit. Weekly instruction consisted of four 30-minute lessons, for a total of 120 instructional minutes each week. Each unit was themed (e.g., animals or folktales) and instruction focused on a specific type of text structure (e.g., compare-contrast and cause-effect). As well, instruction focused on new vocabulary words (including semantic relations among words), inference making, comprehension monitoring, story grammar, and main idea.

Students were assessed at multiple points during the study. At the end of each of the four units, teachers administered standardized curriculum-aligned measures (CAMs) to assess students' achievement in relation to the LK target strategies and skills. CAMs served as proximal measures of students' learning outcomes in comprehension monitoring, understanding text, and vocabulary. The comprehension monitoring CAM measured a student's ability to identify information in orally presented

passages that did not make sense and to apply comprehension-monitoring strategies. The understanding text CAM for narrative text required students to listen to, retell, and then answer questions (predominantly curriculum-aligned inference and text-structure questions) about the narrative. For expository texts, students responded to main idea and detail questions. Vocabulary assessment and scoring was the same as in the earlier studies. Students were also assessed with standardized measures that aligned with CAMs, including the Expressive Vocabulary Test, and the Test of Narrative Recall. In addition, researchers used questionnaires to obtain demographic and classroom information.

LARRC researchers used chi-square tests with categorical data and analyses of variance with continuous data to determine the initial equivalence of groups across conditions based on demographic variables. Following, analyses were conducted to determine the impact of LK^B and LK^D on CAMs for comprehension monitoring (find the inconsistent statement and tell how to fix it), narrative text listening comprehension (both recall and answering questions), and the inclusion of story grammar elements, expository text (answering main idea and detail questions), and vocabulary. Given the range of grades investigated (pre-K to grade 3) there were both floor and ceiling effects for some CAMs, and researchers used multilevel-censored normal response models to account for non-normal distributions.

Results. Table 4-1 summarizes all of the effect sizes for the LARRC RCT, with and without key covariates. In analyses that did not account for covariates, students in LK^B classrooms outperformed BAU students on only two measures: comprehension

TABLE 4-1 LARRC Effect Size Summary by Intervention, Assessed Construct, and Grade Level With and Without Covariates

	Listening Comprehension				
Grade	Narrative	Expository	Comprehension Monitoring	Target Vocabulary	Story Grammar Understanding
Let's Know!-Broad					
Pre-K	–0.07 / –0.09	0.73 / 0.69	0.78 / 0.67	**1.55** / **1.38**	0.28 / 0.31
K	0.44 / 0.26	**0.47** / **0.43**	**1.73** / **1.63**	**2.49** / **2.38**	0.03 / –0.09
1	0.35 / 0.32	0.40 / 0.37	**1.25** / **1.17**	**2.67** / **2.43**	0.01 / **–0.33**
2	–0.20 / –0.02	–0.06 / –0.04	**0.79** / **0.87**	**1.52** / **1.58**	–0.13 / 0.10
3	0.46 / 0.37	–0.29 / –0.24	**0.95** / **0.89**	**2.15** / **2.16**	0.35 / 0.20
Let's Know!-Deep					
Pre-K	–0.13 / –0.07	**0.84** / **0.80**	**0.97** / **0.96**	**1.95** / **1.88**	0.10 / 0.17
K	0.11 / 0.07	0.25 / 0.32	**1.63** / **1.66**	**3.45** / **3.48**	0.07 / 0.13
1	–0.01 / 0.03	–0.04 / –0.04	**1.16** / **1.17**	**2.45** / **2.36**	0.25 / 0.14
2	–0.25 / –0.33	0.01 / –0.01	**1.27** / **1.28**	**3.16** / **3.04**	–0.02 / 0.01
3	**1.31** / **1.24**	0.22 / 0.24	**1.06** / **1.08**	**2.98** / **2.80**	0.34 / 0.20

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and contrasts with business as usual. Effects with covariates follow the slash; covariates included all pretest measures, parent education, gender, age, race, and school/site. All measures were researcher developed and aligned to the LARRC intervention.

monitoring in most grades in kindergarten through grade 3 and taught vocabulary in all grades. Students in LK^D classrooms also outperformed BAU students on comprehension monitoring and taught vocabulary in all grades. For proximal comprehension questions, only grade 3 students in LK^D treatment significantly outperformed the BAU group. For the story grammar portion of understanding narrative text, no significant differences were found between either LK group and the BAU group, but for understanding expository text, pre-K students in LK^D outperformed students in BAU. Finally, students in LK^D outperformed students in LK^B in three instances: understanding of expository text in grade 3 and taught vocabulary in kindergarten and grade 2.

When the entire set of covariates (all pretest measures, parent education, gender, age, race, and site/school) were included in analysis, results were remarkably similar. All previously observed significant effects were again observed with very little change in effect sizes. In some cases, effects were modestly stronger and in others modestly weaker, but none of the differences could be considered relevant for practical purposes. However, beyond the previously mentioned effects, four new significant effects were observed. Controlling for covariates, grade 3 LK^D students also outperformed LK^B students on taught vocabulary. Also controlling for covariates, kindergarten and grade 1 LK^D students outperformed LK^B students on the one measure that showed no previous effects: story grammar. Story grammar surfaced as the single significant negative effect when grade 1 LK^B students were compared to BAU; additionally, for story grammar, LK^D students did not differ from BAU.

Follow-Up to the Randomized Controlled Trial

The LARRC team conducted a second RCT with an entirely new cohort of pre-K to grade 3 students. While parallel results for the second cohort are not yet available, LARRC published results for the two cohorts combined with a focus on distal outcomes, namely, reading comprehension, in grades 1–3 (LARRC, Jiang, & Logan, 2019). This study examined not only direct effects of LK on reading comprehension, but also whether effects on reading comprehension were mediated by the language outcomes targeted by LK. In addition, because differences between the two versions of LK (i.e., LK^D and LK^B) were not substantial in the first RCT, the two LK conditions were combined for this study. Thus, the treatment group represents two academic years of children in grades 1–3 in one of two LK conditions, and the comparison was again BAU instruction.

The study included 997 students in grades 1–3 in 184 classrooms, 62 percent of which were in suburban locations, 25 percent were in urban locations, and the remaining 13 percent were in rural locations. Depending on grade level, 29 to 43 percent were from racial or ethnic minority backgrounds, more than 90 percent spoke English as their primary language at home, and 48 percent had mothers who had earned an associate or higher degree.

Implementation of LK was consistent with the first cohort RCT described previously. Students took the same CAMs previously described with the exception of the story grammar task. In addition, at the beginning and end of the school year, students also took the Gates-MacGinitie Reading Test (GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and an adaptation of the QRI-5 (Leslie & Caldwell, 2011). Students also took the Test of Narrative Retell: School-Age (TNR) as a pretest (Petersen & Spencer, 2012).

The effects of LK were estimated using a multilevel multivariate regression that yielded direct effects for students and for classrooms on all CAMs simultaneously once child-level covariates were controlled. Covariates included student demographics, including age, and pretest measures on which conditions significantly differed initially, which included the vocabulary CAM in all grades and the TNR in grade 3. A second multilevel, multivariate regression included indirect effects of LK on reading comprehension, which was parameterized as a latent variable based on GMRT and QRI-5 results, via CAMs.

As summarized in Table 4-2, direct effects of LK on CAMs replicated those of the RCT. Large, significant effects were found in all three grades for LK vocabulary, and moderate to large significant effects were found for comprehension monitoring, with largely null effects on the listening comprehension measures. What this study adds to the picture, however, is that vocabulary was a mediator for large and significant effects on reading comprehension.

Summary

Both versions of the LK curriculum contributed to consistent and reliable gains on some indices of students' reading development, notably vocabulary and comprehension monitoring, but not on others, namely, listening comprehension (answering questions) about narrative and expository texts and discerning the structure of narratives (the story grammar measure). The follow-up study added to this picture by demonstrating that vocabulary learning mediated large, significant, indirect effects on reading comprehension. Although the follow-up study did not estimate direct effects for reading comprehension, the mediating effect of vocabulary learning is important in that it demonstrates that learning taught vocabulary in LK translated into impressive gains in reading comprehension.

TABLE 4-2 LARRC Effect Size Summary for Direct and Indirect Effects of LK by Assessed Construct and Grade Level for the Follow-Up Study

	Direct Effects			
	Listening Comprehension		Comprehension Monitoring	
Grade	Narrative	Expository		Target Vocabulary
1	0.09	0.12	**1.24**	**2.23**
2	–0.18	0.03	**0.71**	**1.98**
3	**0.33**	–0.05	**0.55**	**2.14**
	Indirect Effects on Reading Comprehension			
1	0.06	–0.01	–0.09	**2.26**
2	–0.24	0.01	0.14	**1.89**
3	0.48	–0.04	–0.12	**1.89**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and contrasts with business as usual. Direct measures were researcher developed and aligned to the LARRC intervention. The indirect reading comprehension latent variable was based on scores on GMRT and an adapted version of the QRI-5.

LARRC researchers noted, early on in their efforts, that comprehension instruction faces three consistent obstacles to success: (1) lack of teacher expertise for teaching the skills that support student comprehension, (2) the tendency to focus instruction on relatively easier-to-learn reading strategies and skills including decoding, and (3) insufficient time devoted to teaching the more challenging comprehension strategies (such as drawing inferences or using text structure to support making sense of text). The design of this RCT, because it included comprehensive professional development and detailed lesson scripts, substantial additions of instructional time, and instruction focused on comprehension components that are usually encountered in later grades, directly addressed these shortcomings.

The LARRC work is noteworthy for continuity over the years of the RfU funding. Thus, the results of the culminating RCT are tied closely to a long runway of LARRC studies that preceded it. Researchers began with creation of the LK curriculum, using the substantial literature on correlates and contributors to reading comprehension to inform development. They also examined shortcomings and obstacles to effective comprehension instruction. Researchers complemented this effort with design research that both informed curricular content and examined the needs of participating schools, teachers, and students. Iterations of LK were then tested—in relation to one another, and to BAU classrooms leading up to the culminating RCT.

Going forward, there are several possible paths for researchers to consider. The initial RCT study was conducted with highly experienced teachers teaching largely White, middle-to-higher income students. The follow-up study included a more diverse sample of students, but still underrepresented the full diversity of the American school-going population. Future inquiry should seek to gauge the effectiveness of LK with diverse student and teacher populations. In addition, the designation of control classrooms as BAU without accounting for the content of the reading curriculum or of time allotted to teaching and learning limits the value of comparisons—in part because the assessments used to gauge learning may have unstable or, at the very least, unknown instructional validity related to BAU content and learning goals. That young students demonstrated ability and growth in their comprehension-monitoring performance buttresses arguments for incorporating, early on, metacognitive instruction—a key correlate of reading comprehension. Finally, LARRC researchers approached their tasks from both theoretical and practical perspectives. The careful construction of the LK curriculum was informed by an iterative, detailed process of curriculum development that drew from relevant research. LARRC's use of a design study to build understanding of (and community with) teachers and students facilitated the customization of curriculum to best meet instructional needs within implementation settings.

Florida Center for Reading Research

Overview

The FCRR consortium developed a series of instructional approaches intended for pre-K through grade 4. The collection of interventions is called Comprehension Tools for Teachers (CTT). The interventions were developed by an interdisciplinary team of researchers, working closely with classroom teacher collaborators, who were united in

the goal of improving students' language and literacy outcomes. One of the guiding forces for FCRR's work was the Lattice Model, which draws from research on reading comprehension, children's language development, and literacy instruction to argue that development is facilitated through a series of "interacting, reciprocal and bootstrapping effects" involving a range of text-specific, linguistic, and social-cognitive processes (Connor et al., 2014, p. 380). Furthermore, the Lattice Model posits that unique child characteristics and instruction operate to yield interaction effects. Thus, it is not surprising that the FCRR portfolio of interventions reflects attention to the multitude of factors that can influence children's comprehension development, with a particular emphasis on oral language development. Given the important role of the Simple View of Reading (SVR; Reading comprehension = listening comprehension × decoding: RC = LC × DEC) in the conceptualization of this line of work, this emphasis on oral language is not surprising; along with decoding expertise, oral language is an important space in which to search for some of those malleable factors that might facilitate reading comprehension. In a sense, one might view the collection of language-focused interventions in the FCRR CTT as an attempt to build for the LC term in the SVR equation what the early literacy field had built for the DEC term over the three previous decades (Henbest & Apel, 2017; NELP, 2008; NICHD, 2000).

The sheer complexity of this extensive intervention portfolio defies easy summarization. However, looking broadly across the entire array of RCTs, a few patterns stand out. For each CTT intervention, the strongest significant effects were observed for proximal, researcher-designed measures that aligned most closely with the instructional emphases of each intervention. Even though effects on reading comprehension itself were null for all but grade 4 students in the Content Area Literacy Instruction (CALI) intervention, the results do suggest that the CTT interventions generally had the intended effects on specific measures aligned with the intervention, without any cost to reading comprehension. For an instructional approach like CALI, which integrates content learning and reading instruction, the presence of strong content learning effects (compared to BAU) with no detriment to reading comprehension is especially promising.

The full story for this suite of interventions is still to be told; however, pending the publication and release of currently embargoed data, indications from the FCRR team (C. Lonigan, personal communication, July 29, 2019) are that these yet-to-be-released results, which include some integrated pairing of the individual interventions, are even more encouraging than those currently in the archival literature (and hence summarized in this narrative).

Developing the CTT Portfolio

Most of the FCRR interventions reflected the Lattice Model preference for single components that, if enacted with students who need the very expertise emphasized by the component, should exhibit growth in reading and related skills. The set included Language in Motion (LIM—which emphasized understanding the role of the decontextualized features of the "printed" language of schooling), Morphological Awareness Training (MAT—which explicitly taught several common inflectional and derivational affixes), Teaching Expository Text Structures (TEXTS—a program that engaged students with common text structures, e.g., cause and effect, and the key words that

often signal them, e.g., because or so), and Enacted Reading Comprehension (ERC—a program that encouraged body movements as a way of anchoring abstract concepts such as tectonic plates). Each targeted explicit instruction and guided practice for its particular focus. As its name implies, Comprehension Monitoring and Providing Awareness of Story Structure (COMPASS) linked conceptually independent practices by monitoring one's ongoing "situation model" for sense making (operationalized as the ability to determine whether the sentences in a story are internally consistent with one another) and examined/exploited the prototypic infrastructure of the narrative genre so dominant in primary grade reading materials. Dialect Awareness (DAWS) was a targeted intervention designed to promote dialect awareness and versatility for speakers of nonmainstream American English. The Word Knowledge e-Book (WKeB) was a tablet-based intervention designed to improve students' vocabulary, their accuracy in estimating their vocabulary knowledge, and their use of metacognitive reading strategies. CALI, with its explicit attempt to deliver a multicomponent intervention (employing several reading and writing skills/practices in the service of acquiring science and social studies knowledge) was the exception to the componential emphasis among the FCRR interventions. It has the look and feel of the vast majority of the RfU interventions from other RfU teams, such as LK from LARRC and the range of multicomponent interventions from the secondary teams.

The interventions. All interventions included a structured format, professional development, semiscripted routines, and differentiation. While each intervention had consistent routines regardless of grade level, the content varied across grades to enable nonredundant use of the intervention over multiple years. Six of the seven interventions were intended for small-group, targeted interventions for students with specific weaknesses, while the seventh (CALI) was developed to be delivered to small, homogeneous groups within whole classes. The ultimate aim of CTT was to support students' development in the component processes and knowledge that constitute reading comprehension through small group instruction in short (20- to 30-minute) lessons provided by trained experts (not classroom teachers) 4 days per week for periods of several weeks. Only students who scored below the 45th percentile on the Expressive One Word Picture Vocabulary Test, Fourth Edition (EOWPVT; Martin & Brownell, 2010) participated in the intervention.

COMPASS. The Comprehension Monitoring and Providing Awareness of Story Structure intervention targeted comprehension monitoring and narrative text structure knowledge in pre-K through grade 3. The 8-week intervention consisted of two units, and lessons incorporated modeling, guided practice, and independent practice. The lessons were of increasing difficulty over time and across grade levels. Comprehension monitoring was taught in the context of very short narratives that children had to judge as making sense or not, while vocabulary and narrative text structures were taught in the context of longer narratives. Activities for the latter included read alouds, dialogic reading, retelling, teaching of target words, and visual and oral memory aids.

LIM. Language in Motion focused on knowledge and use of decontextualized language features, which are features that are different or more pronounced in written

language than in oral language. LIM targeted syntactic features like relative clauses, passive voice, anaphors, mental state verbs, and figurative language in pre-K through grade 3. LIM included 9- and 12-week versions with four units that were unique to each grade level but included common structural features. Units focused on scientific concepts involving motion and used stories, props, and visuals to maximize students' meaningful engagement with the target language.

MAT. Morphological Awareness Training focused on morphological awareness, an aspect of linguistic knowledge. Designed for use in kindergarten through grade 2, the 8-week intervention included 12 2-day lessons in inflectional and derivational affixes. Lessons included an orientation listening activity, followed by a story or word sort, a game or writing activity, and a summary activity. Review lessons occurred every 4 days. Kindergarten lessons focused solely on oral language, while grades 1 and 2 lessons covered both oral and written language.

TEXTS. Teaching Expository Text Structures targeted understanding and use of expository text structures and originally focused on kindergarten through grade 2, but was eventually expanded to grade 4. Developed for students with below-average listening or reading comprehension, TEXTS taught students that certain words can signal a specific expository text structure, including cause and effect, compare–contrast, problem–solution, and sequence. Activities included explicit instruction wherein students used graphic organizers and read texts with a target structure that included signal words. Guided practice included similar activities and added retellings calling for the use of signal words. Independent practice involved students completing and creating graphic organizers.

ERC. Enacted Reading Comprehension was developed for use in grades 3 and 4 based on the premise that comprehension involves mental simulations, at least in part. ERC built on prior research suggesting that acting out situations in text can support better comprehension (Glenberg, Gutierrez, Levin, Japunitich, & Kaschak, 2004). ERC extended this work by using enactment as a means of fostering comprehension of abstract situations and concepts for expository, persuasive, and narrative texts. In ERC, students use bodily movements to represent abstract ideas, like illustrating the movement of tectonic plates using one's hands.

DAWS. The Dialect Awareness intervention was one of the most targeted interventions, as it focused on metalinguistic awareness for children who use dialects other than mainstream American English in grades 2–4. The intervention used text editing as a means of promoting awareness of informal and formal language forms and code switching from one to the other. The 8-week intervention had weekly units where a new grammatical form was introduced the first day, the second day focused on receptive language, the third day on expressive language, and the fourth day on writing and editing. Explicit and implicit versions of DAWS were developed.

WKeB. The Word Knowledge e-Book intervention aimed to improve vocabulary and reading strategy use through a tablet-based, interactive book-reading program. The e-books in the program required students to select between two rare words at key

points in a narrative. Their choice determined how the plot evolves. Students had access to a digital dictionary throughout. The e-books also included occasional comprehension questions, and students received immediate feedback on the accuracy of their responses, as well as prompting to reread when they were incorrect. These interactive features were designed not only to improve vocabulary knowledge and strategy use, but also to improve students' metacognitive awareness of their own vocabulary knowledge and comprehension.

CALI. The Content Area Literacy Instruction was the most comprehensive of the interventions in that it was designed for use with all children in kindergarten through grade 4. CALI developed students' content-area knowledge in social studies and science, while building higher-order comprehension skills, use of comprehension strategies, and expository writing skills. CALI involved two 3-week units that included four lesson types. *Connect* lessons helped students connect the unit topic to their lives. *Clarify* lessons focused on learning to read to learn. *Research* lessons taught students how to read and use primary sources (for social science) or data (for science). *Apply* lessons wrapped up each unit through projects and writing.

The Design of the CTT Intervention Studies

The comparative efficacy studies. As suggested earlier, the design and implementation of this portfolio of interventions were quite complex, with many but not all of the interventions implemented in two large comparative efficacy studies (CE_1 and CE_2), and the rest, such as CALI, MAT, and DAWS, in free-standing RCTs.

CE_1. In the first comparative efficacy study (CE_1), which was carried out in the earlier years of FCRR, five of the interventions (LIM, COMPASS, MAT, TEXTS, and ERC) were compared against a common control, BAU, across pre-K through grade 4 to determine their effectiveness in promoting growth in both component processes (such as vocabulary, syntax, or comprehension monitoring, or in one case, decoding) or broader outcomes (such as listening comprehension, reading comprehension, or general knowledge). Only students who scored below the 45th percentile on the EOWPVT qualified for the study. To provide a robust counterfactual for the newly developed interventions at the preschool level (LIM and COMPASS), Dialogic Reading (DR), a well-studied intervention with well-established efficacy (Hargrave & Sénéchal, 2000), was added to the mix of interventions to which classrooms were assigned. CE_1 included large samples across many schools for pre-K through grade 4, with various interventions tested against BAU at different grade levels:

- Pre-K: LIM, COMPASS, DR
- Kindergarten through grade 2: LIM, COMPASS, MAT, TEXTS
- Grade 3: LIM, COMPASS, ERC
- Grade 4: TEXTS, ERC

Because of the complexity of examining the effects of the multiple interventions developed, FCRR researchers decided to disseminate CE_1 results by grade-level bands

rather than by component intervention, and dissemination of results is ongoing. As of when this synthesis was ready for publication, we had access to results only from grades 3 and 4 for CE_1 (Connor et al., 2018). In addition, some of the component interventions underwent earlier, smaller-scale RCTs prior to CE_1 and CE_2. Thus, where efficacy results were not yet published for a component intervention, we summarized the earlier, smaller RCTs (i.e., for LIM and MAT) whenever they were available. DAWS, WKeB, and CALI, as we said, were evaluated in free-standing RCTs.

CE_2. In the second comparative efficacy study (CE_2), FCRR researchers (C. Lonigan, personal communication, July 29, 2019) used the results of CE_1 to judiciously combine treatments into curricular approaches that bear a stronger resemblance to multicomponent interventions, hypothesizing that the combined approaches might overcome the somewhat sporadic pattern of mainly proximal effects observed for single-component interventions in CE_1 and yield, in their stead, more robust and consistent effects. The implementation design of CE_2 was limited to DR, LIM, and COMPASS in pre-K and kindergarten, with three combination interventions (DR/LIM, DR/COMPASS, and LIM/COMPASS) each compared to BAU. In grade 4, two versions of TEXTS (the CE_1 version and a newly created adaptive version [$TEXTS^A$], with provision for individualized journeys through the curriculum) were compared to BAU. As of the date this synthesis was ready for publication, no results from CE_2 were available for summarization. And as we intimated in the earlier overview, the trends for the paired interventions in CE_2 appear to be more promising than the results of individual components in CE_1.

CE_1 for grades 3 and 4. The main results available for CE_1 are for grades 3 and 4 (Connor et al., 2018). The sample consisted of 338 grade 3 students and 307 grade 4 students who qualified for reading comprehension intervention (meaning that they scored below the 45th percentile on the EOWPVT). Children came from 33 and 31 schools and 135 and 115 classrooms in grades 3 and 4, respectively. Students within schools were assigned to conditions (COMPASS, LIM, TEXTS, ERC, or BAU) using an incomplete-random-blocks design. The interventions were delivered not by classroom teachers but by members of the research team. BAU typically included reliance on core literacy curricula approved by the state of Florida: *Treasures*, *Wonders*, *Open Court Imagine*, or *Journeys*. Instruction in each of these curricula focused on reading comprehension, strategies, discussion, vocabulary, writing, decoding, and spelling, and researchers deemed it unlikely that this instruction included any intensive focus on the same components as the interventions under investigation.

Across all five interventions, vocabulary, syntactic and listening comprehension, narrative comprehension, comprehension monitoring, reading comprehension, and word reading were assessed with multiple standardized measures, except for narrative comprehension and comprehension monitoring. Narrative comprehension was assessed with a single standardized measure, while comprehension monitoring was assessed with a researcher-developed tool used in previous studies. A more detailed description of these common measures appears in Appendix 4-1.

Mixed models were used to analyze the data while accounting for the nesting of students within assigned block and school. Each intervention condition within each grade was compared to the BAU condition, but not to each other. All models controlled

for student age, raw scores at pretest on a vocabulary measure, and raw scores at pretest on the specific outcome analyzed. Each of the covariates in a model was also explored as a moderator of the intervention effects. Where moderation was significant, effect sizes were generated for the mean effect of the moderator and one standard deviation above and below the mean. All analyses for COMPASS, ERC, LIM, and TEXTS used a Benjamini-Hochberg (Benjamini & Hochberg, 1995) correction to reduce false discoveries (i.e., type II error). The design, method, and analyses for the additional RCTs are described under the relevant CTT interventions (namely, DAWS, LIM, MAT, and CALI).

Results

We summarize RCT results here for the most recent results available at the time of this writing for each of the CTT interventions. This approach is possible because FCRR analyzed all data by grade level and compared interventions only to BAU. There have been no direct comparisons among the interventions themselves in the work published thus far. Most of these results come from CE_1, but some from free-standing RCTs.

COMPASS. In grade 3, significantly better performance at posttest compared to BAU was found on one measure, comprehension monitoring, which demonstrated a marginal effect relative to BAU (see Table 4-3). However, once a statistical correction was applied to control for error due to multiple comparisons, this effect was no longer significant. Moderator analyses for COMPASS, based on student characteristics at pretest, were also conducted for all outcomes and revealed significant effects for three additional outcomes. Specifically, older (compared to average-aged and younger) students showed positive effects on narrative language skills relative to BAU. Also, students with poorer listening comprehension (on the Clinical Evaluation of Language Fundamentals,

TABLE 4-3 COMPASS Effect Size Summary by Assessed Construct and Measure in Grade 3

Target Constructs						
Listening Comprehension			Comprehension Monitoring	Vocabulary		Syntax
CELF	OWLS	TNLS	Inconsistency Detection	EOWPVT	CELF	CASL
0.14	0.04	0.13	0.31	0.14	0.08	0.12
Additional Constructs						
Reading Comprehension			Word Recognition			Knowledge
TOSREC	GMRT		WJ-III	TOWRE-SWE	TOWRE-PDE	WJ-III
0.04	–0.04		–0.10	0.03	–0.08	0.04

NOTES: Bold font indicates a significant effect at $p < .05$. The comprehension monitoring task, which was researcher designed, was marginally significant ($p < .10$), but only prior to correcting statistically for multiple comparisons. All effects represent Hedges's *g* contrasts with business as usual. CASL = Comprehensive Assessment of Spoken Language; CELF = Clinical Evaluation of Language Fundamentals; EOWPVT = Expressive One-Word Picture Vocabulary Test; GMRT = Gates-MacGinitie Reading Test; OWLS = Oral and Written Language Scales; TNLS = Test of Narrative Language Skills; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-PDE = Test of Word Reading Efficiency-2, Phonemic Decoding Efficiency; TOWRE-SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency; WJ-III = Woodcock Johnson.

Fourth Edition [CELF-4]) at pretest demonstrated better listening comprehension on the same measure relative to BAU, but students with better pretest listening comprehension on the same measure demonstrated a negative effect of COMPASS on their listening comprehension. Finally, students with lower expressive vocabulary at pretest benefited significantly from COMPASS on expressive vocabulary relative to BAU. However, once a statistical correction was applied to control for error due to multiple comparisons, only the effects on listening comprehension and narrative skills were significant. Aside from the narrative language result, the moderator analyses demonstrated a reversal of Matthew effects (Stanovich, 1986) such that the low performers benefited more than higher performers as a result of the COMPASS intervention in grade 3.

LIM. RCT results are available for LIM in pre-K and grade 3. The analytic sample, measures, design, and analyses for the grade 3 version of LIM were derived from CE_1 (Connor et al., 2018). An additional early RCT conducted in pre-K (Phillips et al., 2016) involved 82 children randomized to either pull-out LIM instruction or BAU. Children were drawn from Title I public schools with pre-K programs where 77 percent or more of students received free or reduced-price meals. The racially and ethnically diverse sample was drawn from 10 classrooms in 5 schools. To qualify for intervention, students had to perform below the 35th percentile on either a spoken language syntax measure or a listening comprehension measure, or both. In addition to the screening measure, five measures were administered at pretest and posttest. Students completed three standardized measures of expressive and receptive language and two researcher-developed, intervention-aligned measures of listening comprehension and of language targeted by LIM. Some of these measures were also administered mid-intervention. As with the grade 3 trial, moderation effects based on pretest scores were also examined.

Results for LIM (see Table 4-4) in grade 3 yielded no statistically significant main effects. The only moderation effect observed was a detrimental effect of LIM on listening comprehension (on the CELF-4) relative to BAU for students with stronger expressive vocabulary at pretest; it remained significant after correcting for multiple comparisons. LIM displayed a similar, marginal, detrimental effect on listening comprehension for students with stronger listening comprehension at pretest, but this effect was no longer significant after correcting for multiple comparisons. Although not an expected effect, LIM also exhibited a positive effect on sight word reading efficiency for students with poorer sight word skills at pretest, and this effect was significant after controlling for multiple comparisons. LIM results suggest it had effects that were inconsistent with the theory and intent behind the intervention.

Results for LIM in pre-K (Phillips et al., 2016) yielded several main effects at posttest. The largest effects were observed on the intervention-aligned measures of targeted language and listening comprehension, but these effects were moderated by pretest performance on the same measures. Students who performed above the mean on targeted language comprehension at pretest experienced larger gains at posttest, while those scoring below average at pretest experienced less benefit at posttest. Results were more mixed for listening comprehension. Those who performed better at pretest had no significant benefit at posttest. In contrast, those who performed below average at pretest demonstrated significant benefit at posttest. The results for standardized measures were not moderated, and listening comprehension showed marginally significant

TABLE 4-4 LIM Effect Size Summary by Assessed Construct, Measure, and Grade Level

	Target Constructs									
	Listening Comprehension				Vocabulary			Syntax		
Grade	CELF	OWLS	TNLS	Intervention-aligned	EOWPVT	CELF	WJ-III	CASL	CELF	Intervention-aligned
Pre-K	NA	0.31	NA	**0.79**	NA	NA	–0.14	NA	0.08	**0.88**
3	–0.14	0.08	0.04	NA	0.15	0.02	NA	0.09	NA	NA

Additional Constructs							
	Reading Comprehension		Comprehension Monitoring	Word Recognition			Knowledge
	TOSREC	GMRT	Inconsistency Detection	WJ-III	TOWRE-SWE	TOWRE-PDE	WJ-III
Pre-K	NA	NA	NA	NA	NA	NA	NA
3	0.04	–0.10	–0.18	–0.06	0.18	–0.02	0.16

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Hedges's g contrasts with business as usual. The comprehension monitoring and intervention-aligned tasks were researcher designed. CASL = Comprehensive Assessment of Spoken Language; CELF = Clinical Evaluation of Language Fundamentals; EOWPVT = Expressive One-Word Picture Vocabulary Test; GMRT = Gates-MacGinitie Reading Test; OWLS = Oral and Written Language Scales; TNLS = Test of Narrative Language Skills; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-PDE = Test of Word Reading Efficiency-2, Phonemic Decoding Efficiency; TOWRE-SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency; WJ-III = Woodcock Johnson.

improvement due to LIM, while vocabulary and syntax demonstrated no significant advantage over BAU.

MAT. Morphological Awareness Training was a part of CE_2 (for kindergarten through grade 2), hence we have no results available for this analysis. However, it was studied in a small-scale RCT in kindergarten through grade 2 (Apel & Diehm, 2013). Participating students came from several classrooms in a single school where 74 percent of students received free or reduced-price meals and were randomly assigned to MAT or BAU. MAT students received small-group pull-out instruction, whereas BAU students remained in class; content missed by MAT students varied depending on time of day but did not appear to include reading. MAT students took all of the same assessments as students in the other CE_1 study (see Appendix 4-1). Additionally, all MAT participants responded to two morphological awareness tasks, the Relatives and Rehit tasks. Relatives focused on students' awareness of the relation of base words to their inflected or derived forms, while Rehit focused on students' ability to explicitly combine two morphemes into a novel word, define that word, and then judge its semantic acceptability within the context of a spoken sentence. Two additional morphological awareness tasks were administered to only students in grades 1 and 2: the Affix Identification task, which measured students' conscious awareness of printed affixes and the orthographic changes that occur when those affixes are added to base words, and the Spelling Multimorphemic Words task, a spelling test of 26 multimorphemic words (e.g., washes, distaste, uneasy). Data were analyzed using analysis of covariance with pretest performance treated as a covariate (as opposed to a repeated measure); as a result, moderation effects could not be explored.

Results (see Table 4-5) indicated large significant effects of MAT on the researcher-designed measures of morphological awareness in all three grades, but no significant effects on word reading or reading comprehension. On a nonsense affix measure, students in kindergarten through grade 2 all demonstrated significant effects relative to BAU when controlling for pretest performance. On a derivational and inflectional morphology task, students in kindergarten and grade 2 demonstrated significant gains relative to BAU, but first graders did not. On a morphological spelling task,

TABLE 4-5 MAT Effect Size Summary by Assessed Construct, Measure, and Grade Level

	Reading Comprehension	Morphology				Word Recognition	
Grade	TOSREC	Nonsense Morphemic Blending	Derivational Awareness	Multimorphemic Spelling	Affix Identification	TOWRE-SWE	TOWRE-PDE
K	NA	**1.26**	**0.82**	NA	NA	0.00	0.00
1	0.26	**0.67**	0.41	**0.82**	**2.54**	0.11	–0.39
2	0.14	**0.86**	**1.07**	–0.03	**1.52**	0.12	0.28

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d contrasts with business as usual. The morphology tasks were researcher designed. TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-PDE = Test of Word Reading Efficiency-2, Phonemic Decoding Efficiency; TOWRE-SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency.

grade 1 students significantly outperformed BAU, but grade 2 students did not. Both grade 1 and 2 students in MAT significantly outperformed BAU students on an Affix Identification task. Post hoc exploratory analyses involving only MAT students suggested gains may have relied to some extent on pretest ability, but these results differed by grade and measure, making them difficult to interpret.

TEXTS. Teaching Expository Text Structures results (see Table 4-6) were available in grade 4 only (Connor et al., 2017) for CE_1. Although no longer significant after correcting for multiple comparisons, two positive main effects were observed prior to correction. The first was for listening comprehension (on the Oral and Written Language Scales [OWLS]) and the second, which was only marginally significant to begin with, was for academic knowledge. TEXTS also demonstrated three significant moderation effects based on incoming student characteristics, only one of which remained significant after correcting for multiple comparisons. The latter effect was for academic knowledge such that TEXTS students with poorer academic knowledge at pretest outperformed BAU students at posttest. While students with average academic knowledge at pretest only marginally outperformed BAU students in the initial analysis, this effect was significant after the multiple comparisons correction. By contrast, an effect on listening comprehension for students with lower listening comprehension at pretest was not maintained after correcting for multiple comparisons.

TABLE 4-6 TEXTS Effect Size Summary by Assessed Construct and Measure in Grade 4

Target Constructs						
Listening Comprehension				Syntax		Knowledge
CELF	OWLS	TNLS		CASL		WJ-III
0.09	0.25	0.04		0.11		0.20
Additional Constructs						
Reading Comprehension			Vocabulary		Comprehension Monitoring	
TOSREC	GMRT		EOWPVT	CELF	Inconsistency Detection	
–0.07	–0.08		0.16	0.07	–0.01	
Word Recognition						
WJ-III		TOWRE-SWE		TOWRE-PDE		
–0.07		0.02		–0.13		

NOTES: Bold font indicates a significant effect at $p < .05$. The effects for OWLS and Knowledge were significant or marginally so ($p < .05$ and $p < .10$, respectively), but only prior to correcting statistically for multiple comparisons. All effects represent Hedges's g contrasts with business as usual. The comprehension monitoring task was researcher designed. CASL = Comprehensive Assessment of Spoken Language; CELF = Clinical Evaluation of Language Fundamentals; EOWPVT = Expressive One-Word Picture Vocabulary Test; GMRT = Gates-MacGinitie Reading Test; OWLS = Oral and Written Language Scales; TNLS = Test of Narrative Language Skills; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-PDE = Test of Word Reading Efficiency-2, Phonemic Decoding Efficiency; TOWRE-SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency; WJ-III = Woodcock Johnson.

ERC. Enacted Reading Comprehension was studied in both grades 3 and 4 (see Table 4-7) under the umbrella of CE_1 (Connor et al., 2018). ERC did not show significantly better performance at posttest compared to BAU for any outcome except one measure of expressive vocabulary, which demonstrated a small positive effect for ERC relative to BAU in grade 3. However, once a statistical correction was applied to control for error due to multiple comparisons, this effect was no longer significant. Moderator analyses for ERC based on student characteristics at pretest revealed significant effects for two additional outcomes in grade 3 and one additional outcome in grade 4. In grade 3, students with poorer expressive vocabulary at pretest (on the EOWPVT) showed positive, significant differences compared to BAU students on two measures of expressive vocabulary, the CELF-4 and the EOWPVT. However, once a statistical correction was applied to control for error due to multiple comparisons, only the effect on EOWPVT was still significant. In addition, students with average pretest expressive vocabulary (on the EOWPVT) also showed a significant positive effect on the EOWPVT, but this effect was also no longer significant after controlling for multiple comparisons.

In grade 4, the only moderation effect observed was for the Woodcock Johnson measure of academic knowledge, where again students with poorer pretest scores showed strong positive effects relative to BAU, but students with better pretest scores showed negative effects relative to BAU. Both of these effects remained significant after the multiple comparison correction was applied. The moderator analyses for ERC

TABLE 4-7 ERC Effect Size Summary by Assessed Construct, Measure, and Grade Level

	Target Constructs				
	Comprehension Monitoring		Vocabulary		Knowledge
Grade	Inconsistency Detection		EOWPVT	CELF	WJ-III
3	0.07		0.33	0.14	0.14
4	–0.09		0.09	0.09	0.17
	Additional Constructs				
	Reading Comprehension		Listening Comprehension		
	TOSREC	GMRT	CELF	OWLS	TNLS
3	0.04	–0.09	0.08	0.09	0.10
4	0.04	–0.08	0.02	0.04	–0.09
	Syntax		Word Recognition		
	CASL		WJ-III	TOWRE-SWE	TOWRE-PDE
3	0.17		–0.05	–0.05	0.01
4	–0.15		–0.01	0.16	–0.05

NOTES: Bold font indicates a significant effect at $p < .05$. The effect for EOWPVT was significant ($p < .05$), but only prior to correcting statistically for multiple comparisons. All effects represent Hedges's g contrasts with business as usual. Only the comprehension monitoring task was researcher designed. CASL = Comprehensive Assessment of Spoken Language; CELF = Clinical Evaluation of Language Fundamentals; EOWPVT = Expressive One-Word Picture Vocabulary Test; GMRT = Gates-MacGinitie Reading Test; OWLS = Oral and Written Language Scales; TNLS = Test of Narrative Language Skills; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-PDE = Test of Word Reading Efficiency-2, Phonemic Decoding Efficiency; TOWRE-SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency; WJ-III = Woodcock Johnson.

demonstrated a reversal of traditional (rich get richer) Matthew effects (Stanovich, 1986) such that the students with lower pretest scores improved most relative to BAU as a result of the ERC intervention in grades 3 and 4; it is noteworthy that students with stronger pretest academic knowledge demonstrated a detrimental effect of ERC relative to BAU.

DAWS. Dialect Awareness was evaluated in grades 2–4 in two separate studies from a single publication (Johnson, Terry, Connor, & Thomas-Tate, 2017). The first sample for DAWS consisted of 116 students in grades 2–4; the sample for the follow-up study consisted of 374 students. Students were selected for DAWS participation based on pretest usage of nonmainstream English. Eligible students were randomly assigned to one of three conditions: BAU in both studies, DAWS in both studies, and only in the first study an editing program that could be construed as supporting implicit dialect awareness. Researchers used a measure of dialect variation (part I of the Diagnostic Evaluation of Language Variation-Screening test [DELV-S]; Seymour, Roeper, & deVilliers, 2003), first as a component of the screening protocol prior to instruction and then after the instructional program was completed. Students were asked to describe actions and respond to questions based on pictures, with the intent to elicit phonology and morphosyntactic features in students' spoken language. Researchers used students' written language samples to measure spontaneous dialect usage in writing. Students were shown a picture, given a prompt, and asked to write a story about what they thought happened in the picture. The written language samples were transcribed and analyzed using the Systematic Analysis of Language Transcripts software (Miller & Chapman, 2008). Then, researchers used a Dialect Density Measure in combination with the writing samples to determine the degree of students' nonmainstream American English. In addition, researchers used an editing task to measure students' ability to identify and then transform English-home language forms in sentences to school English. The editing program and DAWS used the same instructional materials and instructors and met for the same length of time. The second study also used a researcher-designed measure of morphosyntactic knowledge.

Hierarchical linear modeling was used to analyze data in both studies to control for the nesting of students in classrooms. Moderation effects were only examined in the second DAWS study. Structural equation modeling (SEM) was also used in the DAWS study to test the theory of change behind DAWS and examine whether DAWS effects generalized to more distal measures, such as reading comprehension.

In the first study (see Table 4-8), DAWS students demonstrated a significant difference from BAU students on the editing task. Students receiving the editing program

TABLE 4-8 DAWS Effect Size Summary by Assessed Construct and Measure in Grades 2–4

	Applications			
Study	Narrative Writing	Editing	Academic Language	Morphology
RCT_1	0.28	**0.69**	**0.44**	NA
RCT_2	**0.21**	**1.48**	NA	**0.33**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's *d* contrasts with business as usual. All measures were designed by FCRR researchers and were targeted constructs. NA = not applicable (i.e., not administered in a given year or study).

also outperformed BAU students on the editing task, and researchers also reported a significant difference with the editing program favoring DAWS. However, an effect size was not reported for this comparison. Results for the narrative writing task were not significantly different for either experimental condition. For the oral language use, only DAWS differed significantly from BAU, and the negative sign for this effect means that DAWS students used less nonmainstream dialect than did BAU students.

The second DAWS study, which excluded the editing comparison condition, also demonstrated significant effects. In the case of the editing task, the effect for DAWS was quite large. Positive effects were also observed for morphosyntactic knowledge and the narrative writing task. Moderation analyses revealed that students who performed more poorly at the editing pretest benefited more from DAWS relative to BAU students on both the editing and morphosyntactic knowledge posttests, but no effect sizes were reported for these analyses. Follow-up SEM analyses revealed that performance on the more proximal measures at posttest were predictive of better reading comprehension. Note that tests for grade-level differences in effects were conducted in the second study, and no significant grade-level differences were found.

WKeB. The RCT examining the Word Knowledge e-Book intervention (Connor et al., 2019) followed a review of 22 studies of e-books that demonstrated, when paper and e-books were directly compared, that results tended to favor e-books and that access to a digital dictionary was associated with better results. Based on two recent meta-analyses, the WKeB developers also noted that the affordances of e-books, especially in terms of interactive features that support but do not distract from comprehension, were associated with positive effects, whereas e-books that did not utilize the affordances of the digital format (i.e., used a linear organization akin to print books) actually resulted in negative effects. As a result, the WKeB developers determined to make their e-books interactive, but not excessively so, and to focus on two aspects of reading comprehension with strong research supporting their effectiveness for improving comprehension: vocabulary and metacognitive strategies.

The e-books were developed with the aid of a focus group of grades 3–5 students and their teachers. Based on pilot use of the e-books, the developers recruited teachers to collaborate in the development of a 15-minute weekly book club lesson plan that could support students' engagement with the e-books and utilization of their affordances.

Complete results from an RCT conducted after development was completed are still forthcoming. The RCT that utilized a delayed treatment design was conducted in grades 3–5 where nearly three-quarters of students were Hispanic and 70 percent received free or reduced-price meals. Classrooms were randomly assigned to implement WKeB immediately (i.e., the treatment condition) or after the first (treatment) cohort had completed the WKeB program (i.e., the BAU control/delayed treatment condition), which was 3 weeks long. Further randomized assignment protocols assigned children within the WKeB classrooms to participate in a weekly book club or not. The latter group still used WKeB but did not participate in the 15-minute weekly book club meetings. The book club sessions were implemented by trained research assistants rather than classroom teachers, but classroom teachers supported students during their reading of the e-books. All WKeB students engaged with the program 3 days per week. WKeB students in book clubs met as a group and were taught vocabulary learning strategies

1 day per week and spent the other 2 days reading the e-books. WKeB students not assigned to a book club read the e-books 3 days per week. Any student finishing the e-book before 3 weeks had elapsed were encouraged to reread the e-book and choose different paths (i.e., words) to see how the narrative differed. Initial results suggest that WKeB had positive effects that relied on weekly book club meetings.

CALI. The first RCT examining Content Area Literacy Instruction (Connor et al., 2017) followed a series of design-based implementation research activities focused on not only the development of CALI, but also on building understanding of student characteristics by intervention interactions with the hope of better targeting interventions for particular groups of students in subsequent RCTs. Researchers used CALI in an RCT focused on determining whether it is possible to improve students' science and social studies knowledge during literacy instruction without negatively affecting their reading development. Results indicated that CALI improved kindergarten through grade 4 students' social studies and science knowledge, and that CALI may also improve students' oral and reading comprehension.

The RCT was conducted with 418 kindergarten through grade 4 student participants from 40 classrooms in a large northern Florida school district. Student eligibility for free and reduced-price lunch averaged 57 percent across schools. Intervention teachers were employed by the research team, rather than by the participating schools. With the CALI focus, researchers used a combination of proximal content knowledge assessments and standardized measures. The proximal assessment consisted of 12 multiple-choice questions that focused on student knowledge of unit topics, as well as 3 open-ended, more application-oriented questions that sought to measure how well CALI supported students' ability to answer complex questions, or to talk or write about what they had learned. Standardized measures focused on vocabulary, letter-word identification, and passage comprehension, as assessed by the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001). These assessments were administered at the commencement of the design study in order to examine student characteristic by treatment interactions. Hierarchical linear modeling was used to analyze treatment effects on content-area knowledge, as students were nested within classrooms.

As documented in Table 4-9, researchers found significant treatment effects for both science and social studies knowledge, with students scoring significantly higher on proximal measures (which were administered in oral form for kindergarten and grade 1 students, and in written form for students in grades 2–4). Tests for child by instruction interactions garnered mixed results. In social studies, children with higher initial passage comprehension scores made greater gains in CALI social studies than did children who had lower scores. However, this interaction effect reversed for science: students with weaker pre-intervention passage comprehension scores made greater gains in science than did students with stronger scores. In addition, researchers found evidence that gains in the first unit (i.e., social science) predicted pretest scores in the second unit (i.e., science), suggesting some transfer of CALI effects across content areas.

Turning to distal measures, researchers found positive effects for treatment on students' picture vocabulary, oral comprehension, and passage comprehension for fourth graders, but no other treatment effects, positive or negative, in any other grades. A final series of analyses examined the researchers' theory of change and found that student

TABLE 4-9 CALI Effect Size Summary by Assessed Construct, Measure, and Grade Level

	Reading Comprehension		Listening Comprehension	Vocabulary	Knowledge	
Grade	WJ-III	Reading-2-Comprehension	WJ-III	WJ-III	Social Studies	Science
K	NR		NR	NR		
1	NR		NR	NR		
2	NR		NR	NR		
3	NR	NR	NR	NR		
4	**0.22**	**NR**	**0.47**	**1.20**		
K–4					**2.27**	**2.10**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent contrasts with business as usual, with all being estimated as Cohen's *d* except the researcher-designed measures of reading comprehension and content knowledge, where effects represent Hedges' *g*. NR = not reported; WJ-III = Woodcock Johnson III.

membership in CALI significantly predicted stronger performance on final unit posttest scores, which in turn predicted stronger performance on distal measures of vocabulary, oral comprehension, and passage comprehension.

Summary

The sheer volume of interventions developed and the complexity of some of the RCT designs make it difficult to do justice to FCRR work. Across the interventions, results were most positive for CALI, the one FCRR multicomponent intervention, and for the more targeted DAWS, LIM, and MAT interventions. Results were largely null for COMPASS, ERC, and TEXTS. However, in personal communications (C. Lonigan, personal communication, July 29, 2019), FCRR researchers shared the observation that CE_1 yielded positive results for LIM and COMPASS in pre-K through grade 2, as well as for a modified version of DR in pre-K. As a result, researchers explored combinations of pairs of three pre-K and kindergarten interventions: DR, LIM, and COMPASS in CE_2. In addition, FCRR tested two different version of TEXTS in grade 4. More details on the results of both comparative efficacy studies will be forthcoming from FCRR.

For each CTT intervention, the strongest significant effects were observed for proximal, researcher-designed measures that aligned most closely with the targets of each of the CTT interventions. Even though effects on reading comprehension itself were null for all but CALI grade 4 students, the results suggest that the CTT interventions generally had the intended effects without any cost to reading comprehension compared to BAU. For an instructional approach like CALI, which integrates content learning and reading instruction, the presence of strong content learning effects with no detriment to reading comprehension is especially promising.

It is particularly promising that the portfolio of instructional approaches that FCRR developed has the potential to provide teachers with an expanded comprehension instruction toolkit. The availability of interventions focused on dialect awareness, morphological awareness, and enactment of abstract concepts is a real asset to teachers

serving students with those specific needs. What remains to be seen is how the CTT interventions might best be integrated into everyday classroom practice. In other words, what will it take to support uptake of these and other RfU interventions outside the confines of an RCT and what will the effects look like in such cases?

Moreover, and to anticipate an issue for more extended discussion in our reflections in Chapter 5, even where effects were not statistically significant (sometimes only after controlling for family-wise error), many of the effect sizes observed in FCRR studies suggest the practical importance of the CTT interventions. For example, TEXTS demonstrated a Hedges's g of 0.25 on the Oral and Written Language Scales for grade 4 students. Also, ERC showed practically relevant effects on two different distal, standardized expressive vocabulary measures in grade 3 (g = 0.33 on the Expressive One-Word Picture Vocabulary Test and g = 0.14 on the Clinical Evaluation of Language Fundamentals). Existing guidelines for interpreting effect sizes (Hill, Bloom, Black, & Lipsey, 2008) suggest the average effect in RCTs for broad standardized tests is 0.07 with a standard deviation (SD) of 0.32 and for narrower standardized tests, which include the tests referenced here, is 0.23 with a SD of 0.35. The average effect in meta-analyses is 0.23 for grades 1–3 and 0.22 for grades 4–6 with a SD of 0.18 for both grade bands (Hill et al., 2008). Compared to these average effects, the FCRR results for standardized measures become more promising, statistical significance notwithstanding. Put more succinctly, examined in the context of an increasingly robust body of research on RCTs in education, the modest (and often nonsignificant) effects observed in FCRR studies are the norm for the class of standardized measures used. Nonetheless, we lack a common metric for interpreting effects on outcomes other than reading achievement because existing guidelines have been validated only for reading achievement tests.

Catalyzing Comprehension Through Discussion and Debate

Overview

CCDD engaged in a long-term curricular development effort to develop two multicomponent instructional programs: Word Generation for grades 4–8 and Strategic Adolescent Reading Intervention for grades 6–8. WG and STARI differ in three important ways. First, WG is a general education curriculum supplement intended for all middle grade students while STARI is an intervention intended for students struggling with reading comprehension. Second, WG in the RfU era continued a preexisting line of work by extending WG downward to the intermediate grades and outward to discipline-specific versions; the STARI effort expanded on a pilot previously developed by Hemphill in collaboration with Boston public school teachers. Third, WG, while nominally a vocabulary intervention, strives to engage students in deeper reading activity, including close reading, perspective taking, rich discussion and debate, and evidence-based argumentation; STARI, designed for students with weaker foundational skills, expands its similar focus on engaging questions and classroom discussion with specific procedures for attending to word attack, fluency, literal level comprehension, and facets of vocabulary.

In terms of overall results for the two CCDD interventions, effects for STARI (compared to BAU) were robust for indices of word recognition, comprehension efficiency, and morphological awareness. STARI researchers also found that both student

behavioral indices of engagement (how much of the curriculum they actually completed) and teacher judgments of their students' emotional and cognitive engagement moderated performance on all three of the outcomes. For WG, effects (compared to BAU) were more frequent and stronger in the second year of the study for reading comprehension, vocabulary, and perspective articulation and positioning. The most consistent effect was for taught vocabulary, which showed small but significant effects for both years across all grade bands. Nonetheless, it is notable that a vocabulary-centric curriculum, when enhanced with opportunities for classroom discussion, generated significant effects on the Global Integrated Scenario-Based Assessment (GISA), a deep and distal measure of comprehension that was not aligned to the curriculum.

Word Generation (WG)

The CCDD efforts were unique in that they represented an expansion of several years of earlier work on the WG intervention. CCDD, in partnership with public schools, developed WG several years prior to the RfU as a grades 6–8 schoolwide cross-curricular supplement. A major assumption of WG is that many students exhibit little mastery over the academic vocabulary and registers characteristic of "school talk" within and across the disciplines of ELA, sciences, social sciences, and math. The program focused on engaging students in rich weekly discussion and debate of short, provocative texts featuring five academic vocabulary words with high utility across the four disciplines, words like explanation, consistency, robust, and power, using curricular units called WordGen Weeklies. Teachers in each discipline led at least one lesson per week to emphasize the interdisciplinary merit of the target words. Words were introduced within the context of an article on an interesting topic that would easily spur debate. The week culminated with students writing a persuasive essay. The idea was that a rich set of engaging activities, including opportunities for students to use the target words in authentic ways, would deepen students' understanding of these words and their similarities and differences in everyday use across the disciplines.

Early (pre-RfU) quasi-experimental studies indicated WG resulted in better student learning of target words when compared to students in BAU schools, and that knowledge of the target words predicted performance on the state ELA accountability test (Snow, Lawrence, & White, 2009). Follow-up studies suggested that discussion played a mediating role in how much targeted vocabulary students learned (Lawrence, Crosson, Pare-Blagoev, & Snow, 2015), that reclassified English learners (ELs) benefited more than English-only learners (Hwang, Lawrence, Mo, & Snow, 2015), and that while better readers benefited more from WG, special education status did not moderate the benefit (Lawrence, Rolland, Branum-Martin, & Snow, 2014).

Revisions to WG undertaken as part of the RfU included expanding the grade levels served to include upper elementary grades and adapting the curriculum for use in the self-contained classrooms in these grades, amplifying the support for discussion in the curriculum, and adding six week-long middle school curricular units dedicated to science and social science for each of the middle grades (to be substituted for WordGen Weekly use when the topics match the larger curriculum).

The units for the upper elementary grades were extended to last for 10 days and focused on pertinent civic and social issues. Units were designed to be taught by the

classroom teacher and to last 45–50 minutes, a substantial extension beyond the original WG 20-minute lessons.

In contrast, the curriculum for the middle school grades was refined to add 12 content-area-specific units of a week's duration, to be used in some sequence with 12 of the WordGen Weekly units. The original units retained the 20-minute lessons executed across four content areas (i.e., ELA, math, science, and social science), but the new units were designed to be 45 minutes, six of them implemented in social studies classrooms and the other six in science classrooms (with aligned brief lessons for the other content area teachers provided, to sustain the distributed responsibility). The content-focused units included attention to discipline-specific argumentation and evidentiary criteria in the two subject areas as well as academic vocabulary, and were dubbed Science Generation (SciGen) and Social Studies Generation (SoGen).

Methods. CCDD conducted a single "grand" RCT (Jones et al., 2019) to evaluate the impacts of the two refined and extended versions of WG on grades 4–7 students' learning outcomes over 2 academic years. Outcomes included unit target vocabulary, which was assessed with the multiple-choice WG academic vocabulary test, and academic language, assessed with the Core Academic Language Skills-Instrument (CALS-I), a group-administered, multiple-choice assessment of core academic language structures and skills. Students' perspective taking was assessed with the Assessment of Social Perspective-Taking Performance (ASPP; Kim, LaRusso, et al., 2018), in which students were asked to construct written responses to questions about difficult social situations. Deep reading comprehension was assessed with GISA, in which students are placed in a simulated community of students and given a purpose, a suite of source materials to be read, and a reading-related application task.

A total of 7,752 grades 4–7 students in 25 schools in four districts in the Northeast participated in the study over 2 academic years. Two districts were located in major cities and served ethnically diverse, low-income students; one district in a small city served ethnically diverse and primarily low-income students; and one suburban district served a primarily White, low- to middle-income population. Researchers used a pairwise matching procedure prior to randomization to achieve demographic similarity between intervention and BAU schools within districts. Despite these efforts, students in BAU schools outperformed treatment students at pretest on several measures. Researchers developed instruction-aligned, proximal measures of taught vocabulary, academic language (CALS-I), perspective articulation, and perspective positioning (ASPP). GISA, developed by ETS as part of the RfU initiative (see Chapter 3), was used as a distal measure of reading comprehension, with a decided emphasis on applying the fruits of comprehension to address related but novel problems in a simulated collaborative setting (working with avatar students and a teacher). Students' workbook completion rates were used as a measure of student exposure to, and engagement with, the WG curriculum in treatment classrooms. Results were analyzed with grade levels collapsed and separate for the 2 years of the study.

Results. In year 1, significant effects of WG were limited, but in year 2 effects were more consistent and stronger across outcomes (see Table 4-10). In year 1, only taught vocabulary showed significant effects for both grade-level bands; the only other significant effect was

TABLE 4-10 WG Effect Size Summary by Assessed Construct and Grade Level

			Writing (SPTAM-R)							
	GISA		Perspective Articulation		Perspective Positioning		WG Vocabulary		Academic Language	
Grade	Y1	Y2	Y1	Y2	Y1	Y2	Y1	Y2	Y1	Y2
4–5	0.05	**0.15**	0.04	**0.12**	**0.14**	**0.19**	**0.22**	**0.28**	0.02	**0.06**
6–7	0.04	**0.10**	0.01	0.06	0.01	**0.19**	**0.13**	**0.16**	0.02	0.01

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and represent contrasts with business as usual. SPTAM-R = Social Perspective-Taking Acts Measure–Revised.

for perspective positioning in the elementary band. In year 2, vocabulary demonstrated slightly stronger significant effects than in year 1 for both grade bands. In addition, both grade levels had small significant effects on the distal index of reading comprehension on GISA. Modest significant effects were also observed for both grade levels on perspective positioning, but similarly modest significant effects were only observed in the upper elementary grades for perspective articulation and academic language.

Exposure to WG, the behavioral index of engagement, or more aptly exposure, was found to be a significant mediator of effects, such that students in the top tertile (one-third) for workbook usage showed the largest effects for taught vocabulary in the upper elementary and middle grade cohorts when compared to BAU students. Students in the middle tertile for WG exposure showed a significant difference from BAU students on vocabulary in the upper elementary grades. Students within the lowest tertile for exposure showed no significant effects on any outcome for either grade band.

These mediation effects were more pronounced in year 2 in that exposure mediated additional outcomes with stronger effects relative to BAU, although effects varied for the two grade levels. For example, vocabulary effects were significant for all three tertiles of exposure in the elementary band and showed a pattern of larger effects for more exposure. For reading comprehension, effects relative to BAU did not differ much in strength based on tertile in the elementary grades, even though all three levels were significantly different. In contrast, the top tertile of exposure in the middle grades again revealed significant differences from BAU, but middle and low levels did not. Finally, results for perspective positioning were inconsistent. In the elementary grades, the high and low but not middle tertiles showed significant effects, but in the middle grades, the opposite pattern was observed—only the middle level was significantly different from their BAU counterparts.

Discussion. The RCT findings for WG join a long line of research on WG as it has evolved over more than a decade. What began as a weekly, cross-disciplinary curriculum for use in grades 6–8 was extended down to grades 4 and 5 and expanded to cover disciplinary vocabulary and reasoning during the RfU. In addition, the RfU WG effort included an attempt to increase the gains found in earlier studies, specifically for vocabulary learning, and to accentuate the disciplinary aspects of the curriculum (Duhaylongsod, Snow, Selman, & Donovan, 2015).

In general, significant effects were more frequent and stronger in the second year of the study for reading comprehension, vocabulary, and perspective articulation and

positioning. The most consistent effect was for taught vocabulary, which showed small but significant effects for both years and grade bands. Nonetheless, it is notable that effects for such a specific vocabulary-centric curriculum generated effects on GISA, a deep and distal measure that is not aligned to the curriculum. Contrary to much existing research on vocabulary-focused interventions (Wright & Cervetti, 2017), WG evidenced effects on a distal measure of reading comprehension. Thus, despite the modest magnitude of these effects, they represent a promising departure from previous vocabulary-focused intervention research. What might explain WG's variance with the commonly found (Wright & Cervetti, 2017) null effect of vocabulary on comprehension? One plausible but speculative factor is the rich talk about text that was required as students were asked to develop and defend positions and perspectives on the thorny issues inscribed in the texts. In short, the texts were incidentally only vehicles to expose students to words; more likely, they provided occasions to engage in intellectual tussles about the ideas represented by the words, which contributed to better understanding of the words themselves.

Moreover, CCDD researchers gathered extensive data on the implementation of WG, including reports by curriculum coaches, teacher implementation challenge checklists, school administrator interviews, case summaries by literacy coaches, and teacher surveys and interviews (LaRusso, Donovan, & Snow, 2016). Interestingly, they also collected survey data from BAU teachers regarding general curriculum implementation challenges. WG teachers were significantly less likely than BAU teachers to report that class size, instructional materials, program "fit" with class, and unclear expectations were implementation challenges. Qualitative analyses indicated that among WG teachers, those in schools where administrators defined a specific period for WG implementation cited the challenge of managing time and balancing the WG with the school curricula far less than their colleagues in schools without this structure. Middle school teachers also spoke about the disruption that both the shorter, original WG lessons caused, as well as the newer, longer disciplinarily-focused lessons. As might be expected, teachers also voiced a great deal of innovation fatigue due to constantly having new curricula and initiatives foisted on them. The most common complaint was competition with time needed for testing and test preparation. In short, lack of alignment of WG with school, district, and state priorities caused considerable difficulty in its implementation at both elementary and middle school levels.

Analyses of the various versions of WG (those evaluated prior to the RfU as well as the CCDD version) have recurrently found increased growth for ELs and other language minority learners, as well as English-only students, in vocabulary (Lawrence, Capotosto, Branum-Martin, White, & Snow, 2012; Snow et al., 2009), academic language skills (Kim, Hsin, & Snow, 2018), and social perspective taking (Kim et al., 2018). In the first efficacy trial of WG, treatment-condition students who were from language-minority homes (i.e., who had parents who preferred to receive materials in a language other than English) demonstrated more growth in academic vocabulary than their English-only counterparts who received the treatment (Snow et al., 2009). The vocabulary items were those taught in the curriculum, which suggests that students from language-minority homes especially benefited from the instruction and its support for acquiring academically relevant vocabulary. Further exploration of a different subsample from that trial showed that the advantage in academic vocabulary

of English-proficient students from language-minority homes in the treatment condition, relative to their peers from English-speaking homes, persisted over 2 study years (Lawrence et al., 2012). But it also revealed that students with limited English proficiency did not experience the same differential gains as their initially English-proficient peers from language-minority homes (Lawrence et al., 2012).

More recently, however, the large-scale efficacy trial of WG found favorable differential effects for students *currently* classified as ELs (i.e., current limited English-proficient students) in academic language skills and in social perspective taking (Kim et al., 2018). In the second year of the trial, ELs in the treatment condition grew more than their English-proficient counterparts in their core academic language skills and in their social perspective articulation skills (Kim et al., 2018). A similar pattern was found among current ELs in the treatment condition on an argumentative writing assessment outcome, in which treatment ELs engaged in more social perspective articulation than did both control ELs and treatment English-only students (Hsin, Phillips Galloway, & Snow, n.d.). These findings offer good evidence that WG benefits proficient bilingual students (i.e., English-proficient students from language-minority homes) and emerging bilingual students in the process of learning English.

Strategic Adolescent Reading Intervention (STARI)

STARI (Kim et al., 2016), though not entirely new, had been much less fully developed at the start of the CCDD project. STARI was designed as a multicomponent, Tier 2, small-group intervention for students identified as specifically struggling with reading. As a supplemental program, STARI focused instruction on a wide swath of students' requisite skills—word reading, fluency, vocabulary, and comprehension—all situated within a peer discussion framework designed to promote comprehension and engagement with reading. STARI used thematic units (e.g., "How can we find a place where we really belong?") that combined disciplinary learning with reading instruction. Instructional materials included a range of texts, from poems to autobiographies to first-person accounts of events, and novels or full-length works of nonfiction. Reading materials were chosen using two primary criteria: (1) relevance to unit themes, and (2) accessibility and cognitive challenge for target students. Researchers hypothesized that "challenging text characteristics would promote classroom talk about text and help move struggling readers beyond very literal and limited responses to text" (Kim et al., 2016, p. 366), with discussion serving as a learning opportunity (and motivating factor) for students. Given the reader profiles of participating

> These kids are struggling readers. A lot of them don't want to read. It's an arduous task for a lot of kids … and I think the discussions help with that. It helps to get deep into the books and the characters and they can relate to a lot of them…. I had a student who … had some major behavioral issues. But … after the book Game, he like closed the book shut and said that was the first book he's ever read. I was also able to tap into [the book] and the life lesson of like "life is a game, you've gotta play it, there's obstacles you have to overcome" and he did.
>
> —*RfU Participating Teacher*

students, the STARI curriculum includes mini-lessons that focus on decoding, morphology, or comprehension, and students engaged in regular timed partner reading to build fluency (with brief texts that also provided requisite background knowledge related to the long texts). Students also regularly read silently in trade books, high-interest novels, and nonfiction and discussed their readings. Students also received alternating blocks of teacher-led guided reading, then partner reading and responding. At the middle and end of units, students engaged in classroom debates on issues related to unit themes.

Methods. Researchers used a randomized, treatment-versus-BAU, pretest-posttest design to address primary research questions. STARI students received three to five class periods of STARI instruction per week, across the entire school year. The research took place in eight middle schools located in four school districts in the northeastern United States and included two large urban districts and two rural/suburban districts. All participating school sites were Title I schools with moderate to high levels of family poverty, indicated in part by 49 percent to 90 percent of students eligible for free or reduced-price lunch. Participating students all scored "below proficient" (at or below the 30th percentile) on the state English language arts assessment. Excluded were students in the early stages of learning English and students whose specific special education designation required intensive phonics interventions. Student numbers, reported as treatment (BAU) groups, were 49 percent (51 percent) White, 19 percent (19 percent) Black, 26 percent (23 percent) Latino, 2 percent (3 percent) Asian, and 4 percent (4 percent) other designations. Students from low-income families comprised 69 percent (76 percent) of participants, and 30 percent (35 percent) of students were receiving special education services.

In addition to reading strategies and skills, which were often practiced by students in a STARI workbook, the research also focused on student engagement, indexed by the number of workbook pages that students completed during the school year, and the Reading Engagement Index-Revised (REIR; Wigfield et al., 2008), which asked teachers to rate the engagement of individual students. The researchers examined cognitive growth using the Reading Inventory and Scholastic Evaluation (RISE), originally developed by ETS in collaboration with CCDD (see Chapter 3), a multicomponent measure of the six domains (word recognition/decoding, vocabulary, morphological awareness, sentence processing, efficiency of reading for basic comprehension, and reading comprehension) that STARI was intended to improve. Thus, while RISE was a standardized measure, all of its subtests save for reading comprehension served more the role of a near-transfer and intervention-aligned measure than a far-transfer or distal measure, due to the exceptionally close alignment between the intervention and those assessments. CCDD researchers also examined whether levels of student behavioral engagement (both workbook completion and teacher ratings on the REIR) mediated the effects of STARI on reading outcomes.

Researchers used intention-to-treat estimates of the effects of STARI on different dimensions of reading skill, and compared "the posttest outcomes for STARI and BAU students regardless of individuals' amount of engagement with the STARI curriculum" (Kim et al., 2016, p. 370). The team also conducted instrumental variable analyses to examine how behavioral engagement in STARI predicted outcomes. In these analyses, students' proportion of completed workbook pages, which focused on essays, problems,

and responses to guiding questions, served as an index of behavioral engagement or exposure in STARI. Researchers also used hierarchical regression analysis to examine if teachers' ratings of STARI students' cognitive and emotional engagement in reading more generally explained significant and unique variance in posttest reading skill after pretest scores and school quality were controlled.

Results. STARI students (see Table 4-11) significantly outperformed BAU students on measures of word recognition, morphology, and efficiency of basic reading, which was a 3-minute maze task. Treatment students also performed at higher levels in sentence processing, vocabulary, and reading comprehension, although these differences from BAU students were not significant. Researchers also determined that BAU group students made little or no gain on reading skills, even though many of the students were enrolled in alternative literacy programs that stressed these skills.

Researchers examined how STARI workbook completion, an indicator of behavioral engagement in STARI instruction that could be conceptualized as dosage or opportunity to learn, predicted the same reading outcomes. Significant moderating effects were again found for word recognition, morphology, and efficiency of basic reading, and nonsignificant effects for sentence processing and reading comprehension. With the exception of reading comprehension, effect sizes were notably larger when behavioral engagement was used as an instrumental variable rather than the earlier intention-to-treat analysis. Note that because workbook completion was operationalized as the proportion of workbook pages completed for each student, these effect sizes can be interpreted as the projected effect for the hypothetical student who completed the entire STARI notebook; however, in the studied sample the highest proportion of completion was .89. Nonetheless, the significant findings suggest that completing more of the STARI intervention was significantly associated with stronger posttest scores in word recognition, morphology, and efficiency of basic reading. Finally, in an analysis limited to only STARI students, researchers examined whether student reading engagement, as measured by the REIR teacher engagement scale, predicted posttest scores after controlling for school quality and pretest scores. Reading engagement ratings significantly predicted word recognition, morphology, vocabulary, efficiency of basic reading, and reading comprehension, but not sentence processing; effect sizes were not reported for these outcomes.

Discussion. Students participating in the STARI program outperformed BAU students in word recognition, efficiency of basic reading comprehension, and morphological awareness. Follow-up analyses revealed that behavioral engagement predicted the

TABLE 4-11 STARI Effect Size Summary by Assessed Construct on RISE

	Reading Comprehension					
Grade	Multiple Choice	Maze	Vocabulary	Sentence Processing	Morphology	Word Recognition
6–8	0.08	**0.21**	0.16	0.15	**0.18**	**0.20**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and represent contrasts with business as usual.

same three outcomes, suggesting effects may have been stronger had students completed more of the STARI intervention.

As with WG, implementation proved challenging in the RCT. Although STARI teachers had high ratings of adherence to the curriculum and quality of delivery, a follow-up study digging deeper into these data and utilizing additional observations suggested that teachers adhered more faithfully to the fluency-building portions of the curriculum than to the comprehension portions (Troyer, 2017). Moreover, this same study revealed that adherence to fluency predicted student workbook completion, as well as total amount of reading during the year. In another follow-up study of the STARI RCT (LaRusso, Kim, et al., 2016), teachers reported student behavior and student absences as major barriers to implementation. As with WG, they pointed to test preparation and testing as additional barriers to implementation. This created a disequilibrium between forces of engagement and distractions related to mandated testing.

Consistent with the theme emerging from other RfU consortia, effects for STARI on the RISE assessment were more robust for the more intervention-aligned component measures (word recognition, efficiency of comprehension, and morphological awareness) than the more distal—and more general—indices (reading comprehension and vocabulary). Consistent with the findings of WG, engagement in the curriculum predicted performance, when measured by either behavioral (workbook pages completed) or judgment-based indicators.

Looking Across the Two CCDD Interventions

The work of CCDD focuses, in part, on student attainment of deep reading comprehension, a class of comprehension that is demanded by increasingly complex texts and tasks as students matriculate through the grades, and that is reflected in the Common Core State Standards. This work addresses what for some appears an intractable challenge—attending to two ends of a continuum of comprehension development. While all students are expected to build strategies and skills for advancement to deep comprehension, significant numbers of students must also work to shore up basic skills. Both the WG and STARI programs made progress in fostering student growth. WG is notable for tying together classroom discussions, vocabulary learning, and reading comprehension development as students experience deeper learning of academic words and participate in classroom talk. STARI helped struggling students shore up their reading strategies and skills, leading to increased comprehension efficiency. The behavioral engagement index of workbook pages completed, although a fairly rough measure (Is it engagement or compliance? Student interest or teacher rigor?), brings needed focus to the role of engagement and motivation in learning, especially for struggling readers. CCDD research also reminds us that improvement takes time—as evidenced by the superior student learning results for WG in year 2, compared with year 1.

While WG and STARI differ in significant features and goals, they share certain facets and outcomes. Both programs and related lines of research are informed by the results of design studies—in which the participants, actions, goals, and interventions are negotiated, examined, and determined. The studies also represent the joining of innovative comprehension curricular programs with assessments that describe more traditional (e.g., reading comprehension achievement) and more innovative (e.g.,

student perspective taking) foci of comprehension curriculum and instruction. In addition, the study of challenges to implementation addresses the fact that successful programs result not only from the quality of the reading comprehension instruction program but also from consideration of the school environments in which such instruction is delivered. In this case, attention to contextual variables (i.e., diverse adolescents and the different schools and classrooms they attend) allowed for tailoring the system to best meet student needs.

Promoting Adolescents' Comprehension of Text

During the course of their more than 5 years' tenure as an RfU research center, PACT researchers engaged in a wide range of research studies that directly addressed the need for teachers to "build students' content knowledge and reading comprehension skills" (Capin & Vaughn, 2017, p. 251). Their research portfolio, which was situated (mainly) in middle school, involved a family of interventions designed to promote both reading comprehension and knowledge acquisition:

- PACT (Promoting Adolescents' Comprehension of Text, and not to be confused with the name of the center) focused on acquiring both knowledge and disciplinary comprehension skills in grade 8 U.S. history.
- CCT (Comprehension Circuit Training), which as implemented in grades 6–8 ELA classes, was a broad-based approach to improving the set of comprehension and learning tools that students bring to any learning-from-text task.
- TBL (team-based learning) was a key component of both PACT and CCT, a context and support network to enhance the learning in both interventions.

Over the course of the 5-year RfU initiative, the work related to these interventions included design studies to devise, revise, and refine key instructional tools; pilot studies and smaller-scale efficacy studies to evaluate the contribution of particular facets of comprehension instruction; and, most important to our synthesis, RCTs to assess the magnitude of the effects of these multicomponent interventions on the comprehension and knowledge acquisition of key demographic groups (e.g., a general population of learners, students with learning disabilities, and, for some but not others, ELs). We examine the results of each and then discuss patterns and distinctions among the three.

> We were committed to identifying feasible comprehension practices that content area reading teachers could integrate into their teaching routines that would both promote content learning and comprehension. We think that the PACT intervention practices are on the right path to promote both content learning and reading comprehension in secondary settings.
>
> —*Sharon Vaughn, Steering Committee Representative from PACT*

Promoting Adolescents' Comprehension of Text (PACT, the Intervention)

This extensive line of work culminated in three key RCTs (Vaughn et al., 2013, 2015, 2017) in the area of grade 8 American history. Common to all three RCTs was a multicomponent intervention with five recurring features embedded in three experimenter-designed, multiweek, American history units—Colonial America, the Road to Revolution, and the Revolutionary War (see Vaughn et al. [2015] for a thorough discussion of the features, and Capin & Vaughn [2017] for exemplars). The five features of the intervention were as follows:

1. A **comprehension canopy** designed to build and/or invoke relevant background knowledge, motivation, and purpose. The canopy typically included a video overview of the unit, some guiding questions that might well support learning across the entire unit, and conversation about the issues prompted by the video and/or questions.
2. Initial and follow-up discussions/activities for a set of 6–10 **essential words** (defined as words/concepts central to the unit at hand and likely to reappear in future units).
3. Text-based **knowledge-acquisition** activities, delivered in a range of groupings from whole class to small group to pairs to independent work, including question-answering and note-taking activities that also linked back to the comprehension canopy and essential words.
4. **Team-based learning** activities focused on key understandings of the texts through a three-step cycle of responding to questions/tasks independently, reaching consensus on correct answers in small groups, and whole-class teacher-led reteaching of poorly understood ideas.
5. Culminating **team-based knowledge application,** "designed to clarify, apply, and extend understanding of text and content" within learning teams (Vaughn et al., 2015, p. 34).

Methods. In contrast to most RCTs, in which teachers are randomly assigned to treatment, a unique feature of the PACT studies is that treatment was operationalized as a within-teacher variable, with all teachers teaching both the PACT and the BAU curricula. Clearly, the PACT team was anticipating that the benefit of greater precision and power when treatment was nested within the teacher would outweigh the potential cost of between-condition contamination. Their careful fidelity observations (Vaughn et al., 2013, 2015, 2017) confirmed the fact that most teachers differentiated between PACT and BAU in implementing the curricula.

For the key RCTs, student performance on three primary outcome measures (two researcher-developed intervention-aligned assessments and one distal commercially available assessment) was used to assess the efficacy of this multifaceted intervention. The most intervention-aligned measure was the Assessment of Social Studies Knowledge (ASK)—a multiple-choice test of content knowledge covered in each unit, followed closely in alignment by the Modified Assessment of Social Studies Knowledge and Reading Comprehension (MASK)—a multiple-choice test measuring comprehension of passages topically related to the content of the unit but that had not been a part of the curriculum. Although the MASK assessment was aligned to the intervention, it

TABLE 4-12 PACT Effect Size Summary by Assessed Construct, Measure, and Study in Grade 8

	Reading Comprehension		Knowledge
Study	GMRT	MASK	ASK
RCT_1	**0.20**	**0.29**	**0.17**
RCT_2	0.01	0.02	**0.32**
RCT_3	0.12	**0.20**	**0.40**

NOTES: Bold font indicates a significant effect at $p < .05$. Study 1 and 2 effects represent latent model-based approach to Cohen's *d* except for the GMRT, which is reported as Hedges's *g*. Study 3 effects represent Hedges's *g*. Study 3 knowledge effects are for monolingual English students, followed by English learners. All effects represent contrasts with business as usual. ASK = Assessment of Social Studies Knowledge; GMRT = Gates-MacGinitie Reading Test; MASK = Modified Assessment of Social Studies Knowledge and Reading Comprehension.

covered novel content using released items from high-stakes history measures. The final measure was the Gates-MacGinitie Reading Test (GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2006), which measures reading comprehension in general and thus the possible generalization of skills learned in the intervention program to reading measured broadly. Additionally, a proximal, researcher-designed follow-up measure of the durability of unit content (patterned after the initial ASK measure of knowledge gained in each unit) was administered at 4 weeks, and sometimes 8 weeks, following the culmination of the intervention.

Results. Results are reported separately for the three main RCTs and for several follow-up studies that examined more nuanced facets of the data. Table 4-12 provides a summary of relevant effect sizes.

RCT_1. The first RCT (Vaughn et al., 2013) was the smallest scale ($N = 416$), involving five teachers teaching three units to 27 (16 PACT) sections of grade 8 American history. For the three immediate outcomes, effect sizes favoring PACT over BAU were found for the three major outcomes: knowledge (ASK), intervention-aligned reading comprehension (MASK), and distal reading comprehension (the latent GMRT distal measure). Results for the follow-up content ASK measure indicated that a continued advantage for PACT was still present 4 weeks later.[4]

RCT_2. In 2015, Vaughn and colleagues (2015) published the results of a much larger replication (19 teachers teaching 1,487 students in 85 sections, 47 of which implemented PACT) of the protocol used in the 2013 RCT_1. The intervention-aligned measure of knowledge (ASK and its follow-up versions) revealed a reliable advantage for PACT over BAU immediately after the treatment, after 4 weeks, and after 8 weeks. Moreover, the effect of PACT was found to be fully mediated by implementation fidelity. However, neither the intervention-aligned MASK nor the distal GMRT measure of comprehension revealed significant differences between PACT and BAU, nor did implementation fidelity mediate observed differences. One can view the lack of a significant effect on

[4] While Vaughn et al. (2013) did not report an effect size for this effect, we (the authors of this chapter) calculated it, using *M* and *SD* from the article, as representing an effect of d = 0.37.

reading comprehension either positively (suggesting that the gains in content learning came at no cost to students' reading comprehension) or negatively (suggesting that the students acquired the content but did not "learn how to learn" from text).

***RCT_3*.** A third RCT (Vaughn et al., 2017) involved 19 teachers teaching 94 sections, 49 of which were PACT. It focused intentionally on the performance of ELs (N = 1,629) by sampling from schools and classes with sizable EL populations (ranging from 42 to 52 percent ELs). What differed from the previous PACT RCTs is that RCT_3 was supplemented with tools (e.g., Baker et al., 2014; Francis, Rivera, Lesaux, Keiffer, & Rivera, 2006) designed "to enhance the features of instruction and promote best practice for teaching ELs" (Vaughn et al., 2017, p. 24). In a departure from the previous RCTs, the researchers fit hierarchical linear models to their data, including not only a main effect for the PACT treatment, but also main and interaction effects for EL status and the percentage of EL students in a class. As a result, where interactions with the PACT intervention were significant, effects must be interpreted in light of those interactions.

For the ASK measure, PACT effects depended on both a student's EL status and the percentage of EL students in their class. Specifically, for a class with 10 percent ELs, the effect for non-ELs was significant, and the effect for ELs calculated by the authors of the current chapter using additional data provided by the researchers was quite similar in magnitude. However, "the EL/non-EL difference in treatment classes widens as EL becomes more prevalent in a class," resulting in a lower average effect for ELs relative to non-ELs the higher the percentage of ELs in a class (Vaughn et al., 2017, p. 30). More specifically, the research team found that when the PACT classroom percentage of ELs was below 8.8 percent, ELs performed more similarly to non-ELs on ASK in comparison to BAU classes where ELs performed significantly more poorly than non-ELs, but only when the classroom percentage of ELs was below 8.8 percent. The difference in EL and non-EL scores was similar in PACT and BAU classrooms that had between 8.80 percent and 11.48 percent EL students. When classes had more than 11.48 percent ELs, the gap in performance between ELs and non-ELs was larger in PACT classes than in BAU classes. Thus, PACT reduced performance gaps between EL and non-EL students in low-percentage-EL classes (i.e., < 8.80 percent ELs), reproduced gaps in moderate-percentage EL classes (i.e., between 8.8 and 11.48 percent), and widened gaps in high-percentage-EL classes (i.e., > 11.48 percent). Nonetheless, regardless of the percentage of ELs in a class, PACT ELs outperformed BAU ELs. Finally, it should be noted that the reduction in benefit due to PACT was hypothesized by PACT researchers to be attributed to an "overreliance on discourse-based practices among peers whose language and vocabulary use in English were still developing would reduce the overall effects of the treatment" (p. 32).

In contrast to the ASK findings in the third RCT, ELs and non-ELs equally outperformed students in BAU classes on MASK regardless of the percentage of EL students in a class. As in the RCT_2, the PACT effects did not generalize to the distal GMRT measure of reading comprehension. Thus, effects for the modified PACT intervention were significant for both intervention-aligned measures, with the effect on reading comprehension extending to all students and classes; by contrast, the effect on intervention-aligned content learning (ASK) depended on student EL status and the percentage of ELs in a class.

Secondary analyses of the major PACT RCTs. The PACT team conducted secondary analyses of data from these RCTs to tease out more complex accounts of the impact of PACT on specific populations of learners, most commonly students with learning disabilities (e.g., Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015; Wanzek, Swanson, Vaughn, Roberts, & Fall, 2016). In general, the pattern of results for the overall population was replicated in that effects favoring PACT on knowledge were stronger and more consistent than those on content-based or general reading comprehension. In the Swanson et al. (2015) reanalysis of the Vaughn et al. (2013, 2015) RCT_1 and RCT_2 data sets, PACT students with learning disabilities (LDs) outperformed BAU students with LDs on the intervention-aligned ASK content measure and the intervention-aligned MASK comprehension measure but not on the distal GMRT measure of comprehension. In the Wanzek et al. (2016) reanalysis of the Vaughn et al. (2017) RCT_3 with ELs, PACT students with LDs scored higher than their BAU counterparts on the ASK but not on either of the comprehension measures—MASK or GMRT. An additional analysis corroborated the fact that the effect of PACT on the outcome measures was similar for both EL and non-EL students with learning disabilities, which led the PACT team to conclude that the curriculum was both accessible to and beneficial for all students, including those who had been diagnosed with a learning disability and were coping with a curriculum presented in a second language.

The PACT team has also conducted a follow-up analysis to examine the moderating effects of other individual difference variables. Wanzek, Roberts, Vaughn, Swanson, and Sargent (2019) reexamined the data from the Vaughn et al. (2015) RCT_2 replication to determine whether the typical PACT effect on content acquisition and content-related comprehension was moderated by the incoming class mean scores on prior knowledge of American history or incoming general reading achievement (the GMRT). They found no hint of any interaction effects. Students in classes with higher or lower levels of knowledge or achievement benefited equally from PACT instruction.

Team-Based Learning

Using the same design principles as PACT, Wanzek et al. (2014) randomly assigned the 463 students distributed across the 26 sections taught by the seven participating grade 11 American history teachers to 15 TBL and 11 BAU sections for three 15-week history units (Gilded Age, Imperialism and World War I, and The Twenties). Similar to the PACT studies, they compared outcomes on the intervention-aligned ASK content measure and the distal (GMRT) reading comprehension, but they did not employ the hybrid MASK comprehension measure. A significant main effect (see Table 4-13) was found for the ASK but not for GMRT, replicating a common finding in the multicomponent PACT work—it consistently improves content learning but only occasionally influences comprehension. They also found that the benefit of TBL for content knowledge growth was moderated by incoming content knowledge (pretest ASK scores), with TBL students possessing the greatest pretest knowledge, benefiting most in comparison to their BAU counterparts.

In a follow-up study (Kent, Wanzek, Swanson, & Vaughn, 2015) that focused on 24 students designated as LD from the Wanzek et al. (2014) grade 11 study, the team divided the 44-item ASK pool into 12 items focused more on vocabulary acquisition

versus the 32 items testing understanding of the content. Comparisons of the 16 LD students in the TBL treatment with the 8 in the BAU group indicated an effect favoring TBL for the vocabulary subset but not on the content subset (see Table 4-13). The effect size difference for the overall ASK measure was not statistically reliable.

Comprehension Circuit Training

Fogarty and colleagues (Fogarty et al., 2014, 2017; Simmons et al., 2014) focused on a parallel (to PACT) multicomponent intervention called Comprehension Circuit Training delivered for middle school students initially in conventional classroom plus printed text format (Fogarty et al., 2014) and then on a digital platform (Fogarty et al., 2017). They (Fogarty et al., 2014, 2017; Simmons et al., 2014) developed and tested CCT as a grades 6–8 intervention for English language arts classes over a several-year period, using the RfU practice of first developing and refining the curriculum with groups of stakeholders before subjecting it to efficacy studies and/or RCTs. Like PACT, CCT is a multicomponent reading comprehension intervention, based roughly on the direct and inferential mediation model (Cromley & Azevedo, 2007), with its emphasis on background knowledge, vocabulary, and inferential reasoning. CCT comprises both teacher- and student-directed practices. The set of teacher-directed practices included building/activating background knowledge, teaching key vocabulary through meaning-focused practices, and facilitating word identification of key words from the texts to be read in each unit. Student-directed practices, motivated by Kintsch's construction-integration theory (1998), focus on monitoring comprehension by previewing and setting personal comprehension checkpoints throughout the text. This student work is scaffolded by worksheets that aid in such stock taking. The student-directed activities were enacted in student pairs to facilitate talk about text and collaborative elaboration of ideas. Essentially, this mix of teacher- and student-directed activities was delivered in a sequence of learning stations through which the students cycled daily (hence the metaphor of "circuit" training) on a predictable schedule, usually working in pairs traveling together. As with PACT, each teacher taught both CCT and BAU sections. Both strong professional development (group teacher meetings during the summers and individual teacher coaching during the implementation) and the careful monitoring of treatment fidelity were employed to ensure fidelity of treatment. Results for all three studies are reported in Table 4-14.

TABLE 4-13 TBL Effect Size Summary by Assessed Construct, Measure, and Study in Grade 11

	Reading Comprehension	Knowledge		
Study	GMRT	ASK	ASK-Comprehension Items	ASK-Vocabulary Items
1	0.03	**0.19**	NA	NA
2	NA	0.50	0.38	**1.01**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Hedges's g and contrasts with business as usual. ASK = Assessment of Social Studies Knowledge; GMRT = Gates-MacGinitie Reading Test; NA = not applicable (i.e., not analyzed in a given study).

TABLE 4-14 CCT Effect Size Summary by Assessed Construct, Measure, and Study

		Reading Comprehension						Vocabu-		
Study	Grade	Narrative	Expository	GMRT	Latent	TOSREC	STAAR	lary	ORF	SWE
RCT_1	7–10	0.01	0.03	–0.01	NA	NA	NA	NA	NA	NA
RCT_2	6–8	0.06	NA	0.16	NA	NA	NA	NA	NA	NA
RCT_3	6–8	NA	NA	NA	**0.14**	**0.28**	0.10	**0.43**	–0.08	–0.04

NOTES: Bold font indicates a significant effect at $p < .05$. Study 1 effects represent Hedges's *g*, study 2 effects represent Structural Equation Model γs and study 3 effects represent gain-score adaptation of Cohen's *d*. All contrasts are with business as usual and control for pretest scores on the same measure. The vocabulary measure was for taught vocabulary in CCT. GMRT = Gates-MacGinitie Reading Test; Latent = a latent measure of reading comprehension based on GMRT, the Group Reading Assessment and Diagnostic Evaluation Comprehension composite, and the Gray Oral Reading Test-5 comprehension score; ORF = easy CBM oral reading fluency; STAAR = State of Texas Assessments of Academic Readiness; SWE = Test of Word Reading Efficiency-2, Sight Word Efficiency; TOSREC = Test of Silent Reading Efficiency and Comprehension.

RCT_1. The first study, RCT_1, conducted even before they had settled on the CCT moniker (Simmons et al., 2014), was more or less the proof of concept for the intervention, although no statistically significant main effects were found (see Table 4-14). A follow-up moderator analysis based on pretest performance yielded two small but provocative findings. First, when they compared students with GMRT scores below the 15th percentile with the rest of the sample, they found that the lower group made significantly more pre- to posttest progress on GMRT. On the Adolescent Literacy Inventory (ALI)-adapted passages, there were no differential effects attributable to pretest comprehension on the more narrative-like of the passages, but in a reversal of the GMRT findings, students who scored above the 15th percentile exhibited greater statistically reliable gains than those below the 15th percentile.

RCT_2. In year 2 of the RfU grant, Fogarty et al. (2014) conducted RCT_2 in 61 ELA classes involving 859 largely low-income (hovering at 67 percent) students taught by 14 middle school ELA teachers. The sections within each teacher's portfolio were randomly assigned to CCT or BAU. Two comprehension outcomes—the more distal GMRT and two adapted narratives from the ALI (Brozo & Afflerbach, 2011)—were used to measure the overall impact of CCT. Neither of the key comprehension measures yielded significant treatment effects. The team also examined the degree to which fidelity of treatment within the CCT condition mediated performance on the two comprehension outcomes; they found that, as fidelity improved, student outcomes improved within the CCT condition for both GMRT and the narrative measure.

RCT_3. By the time of the implementation of the second wave (in year 3 of the RfU grant), Fogarty et al. (2017) had converted CCT to a digital platform, with students cycling through digital stations with a plethora of teaching videos and computer-based practice activities rather than moving through physical stations and print material. Following the recommendation of Fletcher (2006), the RCT_3 used an array of reading comprehension measures to avoid "underrepresenting the complex reading comprehension construct" (Fogarty et al., 2017, p. 337). This array included commercial and researcher-developed assessments. The GMRT (MacGinitie et al., 2000) assessed

students' comprehension of short narrative and expository passages, and the Group Reading Assessment and Diagnostic Evaluation (Williams, 2001) examined student performance on Sentence Comprehension and Passage Comprehension subtests. Students were also administered the Gray Oral Reading Test, 5th edition (Wiederholt & Bryant, 2012) which focused on the amount of time needed to read the passage as well as reading errors, and open-ended response questions. In addition, researchers used extant student reading comprehension scores from the State of Texas Assessments of Academic Readiness (STAAR; Texas Education Agency, 2013). Component reading skills were also measured. Researchers used the Sight Word Efficiency (SWE) subtest from the Test of Word Reading Efficiency, 2nd edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012), and oral reading fluency (ORF) was measured using the EasyCBM system (Alonzo, Tindal, Ulmer, & Glasgow, 2006). Proximal Vocabulary Matching, a researcher-designed measure, was used to assess students' knowledge of CCT target words, while the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010) was used to assess students' silent reading fluency and sentence-level comprehension skills.

In short, the design was tightened and refined on both the treatment side and the outcome side of the RCT. Significant effects were found for the latent comprehension variable, but not on the state test. Interestingly, significant effects were found on some of the component skill measures, such as the proximal index of vocabulary and on one index of comprehension efficiency, the TOSREC, but not on another index of comprehension efficiency or the oral reading fluency index.

Summary

Across all PACT studies, the results for the portfolio of multicomponent interventions (PACT and CCT plus the common TBL component) were complicated. Regarding main effects, the most consistent finding, especially for PACT and TBL, is that the intervention often, and sometimes robustly, affected the acquisition of content knowledge for a range of secondary students. That gain in knowledge was sometimes accompanied by an increase in comprehension performance on texts that were related to the unit topics, but only occasionally by an increase on a distal measure of comprehension (GMRT). Importantly, the results indicate that incorporating reading comprehension instruction into content-area curriculum boosts content knowledge acquisition with no apparent cost to overall comprehension processes and practices.

For CCT, few effects materialized in its first print-based instantiation, but many effects were found for the smaller RCT_3 for year 3 with the digital delivery mechanism, namely on reading comprehension, its efficiency, and unit-related vocabulary. Regarding moderators and mediators, even more complexity arises.

With the first iteration of CCT, there was a trend for the lowest tier of students to benefit the most, in comparison to BAU students, on GMRT; however, these same students tended to exhibit lower relative growth on a comprehension measure for a topically related expository text. In the second iteration, post hoc analyses suggested marginally significant tendencies for students scoring the lowest on GMRT at pretest to benefit most from the intervention as evidenced by sizable effects accompanied by relatively high *p*-values.

Regarding language diversity, the results for PACT RCT_3 suggest that as language diversity in PACT classrooms increased, ELs' gains with PACT diminished, which the researchers hypothesize may be due to discourse patterns with a decreasing incidence of English academic language in these classrooms. As a result, PACT researchers suggest that for a discourse-based treatment like PACT to sustain positive impact on students' learning, additional supports are needed as the percentage of ELs increases. Results also indicate that PACT benefits students with LDs in similar ways to students without LDs, and neither prior class level of content knowledge nor reading achievement predicted responsiveness to the PACT intervention. With TBL work at the high school level, ASK pretest performance moderated ASK posttest scores, with the relative advantage over BAU students accruing to those who started with the most knowledge. For LD students, greater content learning growth attributable to TBL was evident for items focused on vocabulary compared to recognition of key content.

The cup-half-full story from this overall effort is that a family of multifaceted approaches tends to promote students' acquisition of new knowledge when compared to the usual diet of lecture and/or teacher-led presentation of ideas (BAU). These results hold for a wide range of students, including those who typically do not perform well on either external (e.g., standardized) or internal (class-related) measures of knowledge or comprehension. The common features in this family include (1) invoking students' prior knowledge, (2) key vocabulary, (3) (sometimes) enabling skills, (4) consistent collaboration among students, (5) robust talk about key text ideas, and (6) applying the fruits of comprehension to other tasks.

The cup-half-empty story is that the experimental effects are not consistent across a range of key student variables (e.g., existing language competency, background knowledge), curricular variables (e.g., text types, text topics, and disciplinary focus), and outcome variables (e.g., knowledge acquisition, intervention-aligned comprehension, distal comprehension, and vocabulary acquisition). The effects are interesting but not consistently robust. In short, there is still much more to learn. The range of student characteristics, texts, topics, and contextual factors addressed by PACT researchers should serve as able guides to future inquiries.

Reading, Evidence, and Argumentation in Disciplinary Instruction

Overview

Similar to LARRC (and in contrast to the multiple intervention approaches of PACT and FCRR), Project READI engaged in an articulated line of inquiry over the 5-plus-year life of the consortium, culminating in a single RCT study, which was carried out within a single discipline—grade 9 biology—in year 5. The program of research focused on fostering adolescents' literacy development and disciplinary expertise in grades 6–12 in three curricular domains—literary analysis, history, and science—through engagement in authentic but developmentally appropriate tasks in each discipline. Authentic tasks were defined as those consistent with the epistemic aims and goals of the discipline. For example, the work is science focused on explanatory modeling of science phenomena through text-based investigations. That is, the modules used authentic science texts to construct knowledge, draw on information and evidence, and develop explanations

and arguments that fit the data. This selection of texts contrasts with the typical textbook representation of science as a known body of facts. In the science community, information is presented in a wide range of representations, including verbal texts but also in static and dynamic visual displays. Data are tabulated, displayed, summarized, and reported in graphs, tables, and schematics, and there are conventional linguistic frames that constitute the rhetoric of argument in science (Lemke, 1998; Osborne, 2002; Park, Anderson, & Yoon, 2017; Pearson, Moje, & Greenleaf, 2010).

READI scholars worked in discipline-based collaborative design teams (comprised of teachers, learning scientists, and disciplinary experts) to develop the READI approach to achieving the learning goals in each discipline (see Goldman, Britt, et al., 2016). In addition, design-team teachers met with an expanded group of teachers in Teacher Inquiry Networks intended to promote within- and across-discipline exploration of key constructs in the READI definition of reading for understanding.

> Project READI's overarching aim was to engage students in reading, reasoning, and argumentation for purposes of accomplishing authentic disciplinary goals in literary reading, history, and science. Research and development staff collaborated with classroom teachers across iterative design cycles to create sequenced sets of materials, activities, participation structures, and implementation practices that supported students in achieving these goals. This accomplished a secondary goal of READI: to deepen and make teachers more self-aware of how they themselves read, reasoned, and argued in their disciplines. The enhanced awareness of their own ways of reading, thinking, and problem solving made it possible for them to make their processes visible to their students.
>
> —*Susan Goldman, Steering Committee Representative from Project READI*

The year 5 RCT, while carried out in the single domain of grade 9 biology, reflects the principles and practices developed by enacting the READI approach in all three disciplines. The results of the RCT suggested that both READI students and teachers distinguished themselves from BAU participants on important outcomes. READI students scored significantly higher than BAU students on GISA, a measure of deep comprehension that requires students to use knowledge gained from reading with (or in the context of) application tasks. READI students also significantly outperformed BAU students on a multiple-choice, near-transfer measure of within- and across-text integration and EBA. On other EBA tasks, READI students scored higher, but not significantly higher, than BAU students. READI teachers did not differ from BAU teachers at pretest on a science practices survey, but they scored reliably higher than BAU teachers at posttest. Classroom observation scales indicated that READI teachers also engaged in many more practices designed to promote deeper comprehension, thinking, and explanatory modeling than did BAU teachers.

The Development Process

Each READI discipline-based team began with careful study of the existing knowledge base within its discipline in concert with careful empirical study of exemplary practices in the classrooms of participating teachers. In this work, they relied heavily on decades of development and research by WestEd into the Reading Apprenticeship model of professional development (Greenleaf et al., 2011). Through analysis of these disciplinary practices, READI team members identified core constructs that, while shared across disciplines (e.g., the common claim-evidence-reasoning structure of arguments), are instantiated differently in each discipline (e.g., the nature of claims and evidence differ in science and literary analysis). READI's curricular and pedagogical interventions, like their descriptions of existing practice, reflect these twin axes of generic and discipline-specific features. READI researchers purposely did not study reading comprehension as a context- and discipline-free phenomenon, but rather focused on reading for understanding within specific disciplines. In other words, they studied and developed their intervention to address reading comprehension processes in the service of learning aims situated within disciplines. In this sense, READI work with comprehension reflected current ideas about the nature of reading espoused in the National Assessment of Educational Progress (NAEP) Reading Framework (NAEP, 2017) and the Common Core State Standards (NGA & CCSSO, 2010). Over the life of READI, each team of teachers, learning scientists, and disciplinary experts constructed, piloted, and revised instructional modules in small-scale field studies within the framework of design-based research. In year 5, READI scholars directed their focus to the ambitious RCT in grade 9 biology to assess the efficacy of the principles and practices that had guided the READI approach to the improvement of teaching and learning in *all* three disciplines.

Teacher learning focus. Teacher learning was an important feature of all five consortia, but in READI, it took on an even more central role in the research and development process. For READI, teacher learning was on par with student learning as an explicit and co-equal goal and outcome of the research based on a theory of action that teachers are the agents who provide the opportunities that students have to learn. In the READI RCT in grade 9 biology conducted in year 5, there was a pre- and postintervention survey that compared the READI intervention teachers with those in the control group on their attitudes and practices. In addition, at two time points during the implementation in both intervention classrooms, observations of classroom practices were conducted in intervention and control classrooms. Implicit in this approach is the assumption that, even if it does not cause student learning, teacher learning is on the pathway to improved student learning—an assumption examined, if not experimentally tested, in the culminating RCT.

The rationale for Project READI was two-fold: (1) citizens must engage with multiple information resources (e.g., traditional text, multimedia, graphics and other forms of visual representations) to accomplish academic, professional, and personal goals; and (2) national and international indicators show that current educational practices are not producing citizens with the skills to do so effectively. The READI team argued that there are multiple reasons for this, including increased demands of the information resources (hereafter referred to as texts) that convey disciplinary concepts and principles and the absence of explicit instructional attention to these conceptual and textual

demands, in conjunction with failure to recognize that different disciplines present different sources of conceptual and textual difficulty for adolescents (Goldman, 2012; Goldman & Snow, 2015; Goldman, Britt, et al., 2016; Lee & Spratley, 2010; Schoenbach & Greenleaf, 2009). Thus, the goal of the READI project was to develop and investigate approaches to improving learning in each discipline by focusing on the knowledge, heuristics, discourse, and reading practices relied upon in sense making and argumentation in literary analysis, history, and science.

Over the first 4 years of the project, there was a heavy emphasis on teacher learning through two primary activities of the project: collaborative design teams that involved researchers, subject-matter experts, and professional development facilitators and Teacher Inquiry Networks in two of the project locations (California and Chicago). The collaborative Teacher Inquiry Networks engaged in a range of activities intended to promote within- and across-discipline exploration of key constructs in the READI definition of reading for understanding. They read important conceptual and empirical papers within their discipline, examined best disciplinary and classroom discourse practices, developed prototype units and practices, tried them out in the crucible of the classroom, revised them, and began yet another cycle of this sort of design work. A key principle in their approach to teacher learning, consistent with the approach of the Strategic Literacy Initiative (Schoenbach, Greenleaf, & Murphy, 2016) and other efforts within the educative curriculum tradition (Davis & Krajcik, 2005), is that teachers must experience the planned curriculum and constituent practices in a way that gives them a vivid and personal sense of how their students experience the very curriculum that they (the teachers) are trying to teach. Thus, two goals for professional development in the biological sciences RCT (Goldman et al., 2019) were to

(a) "Raise teachers' awareness of their own practices for making sense of science" (p. 1169) when working with content that they find as challenging for them as adults as the grade 9 curriculum is for the students they teach, and
(b) "Immerse teachers as learners in the intervention they would subsequently implement with their students" (p. 1169).

These goals and the activities that were designed for the RCT intervention teachers' professional development were informed by the work with teachers over the first 4 years. Thus, although teachers who had participated in the design teams and inquiry networks were not allowed to participate in the year 5 RCT to avoid any bias in the assignment of teachers to treatments, they participated in the development of both of the modules that were taught by the freshly recruited RCT teachers and the professional development in which the READI intervention teachers participated.

Central to the READI instructional model is building students' awareness of *how* we know, rather than just *what* we know. Metacognitive conversations as well as teacher modeling protocols that emphasize making visible the what, how, and why are a linchpin of the READI instructional model (Lee, 2007; Schoenbach, Greenleaf, & Murphy, 2012). The modules were developed and tested by design teams consisting of classroom teachers, learning scientists, and experts in the relevant discipline. The modules were vetted and revised based on classroom experiences with the tasks, activities, and text sets through multiple cycles of design-based research. Each module began by engaging

students in an essential question authentic to the discipline and that motivated further text-based inquiry to address that question and those that emerged from it. Texts were selected and sequenced to enable students to develop the knowledge, reading, and reasoning practices needed to address the essential question of the module. The design of the text sequence along with scaffolds for disciplinary comprehension, reasoning, and oral and written discourse forms supported students in learning how to make sense of text by referring to their own prior knowledge, other text sources, and discussions with their peers.

Summarizing the research on the trajectory to the RCT. The legacy of the READI design teams and Teacher Inquiry Networks for all three disciplines (literary analysis, history, and science) is three-fold: (1) an extensive set of instructional modules that survived an intensive and extensive set of conceptual and empirical examinations, revisions, and refinements in the crucible of classroom implementation[5]; (2) a well-documented and, in the RCT, experimentally validated model of professional development that privileges long-term commitment to teacher learning by engaging teachers as active participants in the research and development process; and (3) an extensive research portfolio, consisting of *existence proofs* (classic short-term experiments to determine the relevance of key variables to inform the development of assessments, curriculum, and pedagogical routines) and *design experiments* to refine and revise and improve pedagogy—with both lines of work culminating in an RCT to test the efficacy of the modules and the professional development model.

The Randomized Controlled Trial

Based directly on the research and development activities and products (instructional modules, professional development routines, and assessments of key outcomes for both students and teachers) of the first 4 years of work, READI researchers (Goldman et al., 2019) tested the efficacy of its approach to student and teacher learning. Specifically, researchers conducted an RCT to determine the effects of a semester-long intervention on students' comprehension within an academic discipline—specifically, grade 9 students' creations of explanatory models of biological phenomena—using text-based investigations. Measures gauged comprehension and students' ability to transfer learning to apply information to biological modeling and EBA. Researchers also investigated the impact of the intervention, including professional development, on participating teachers' attitudes, beliefs, and practices.

Methods. Grade 9 science teachers and students who were recruited from six school districts from in and around a large Midwestern urban area participated in the research. READI researchers created a stratified sample, using family socioeconomic status and student achievement, ethnicity, and gender to equate READI and BAU control samples prior to intervention. The school student populations fit "three dominant demographic patterns": largely Black (defined as greater than 80 percent) with a mix of Latinx, White, Asian, or multiracial; largely Latinx (defined as greater than 80 percent), with a mix of

[5] Available through the Project READI case library at https://www.projectreadi.org/case-library.

Black, White, Asian, or multiracial; and mixed, defined as no single group constituting more than 60 percent of the student population. The EL population was 23 percent for the intervention group and 25 percent for the BAU control group. Regarding teachers, among the 24 treatment teachers, 33 percent were male and 66 percent were female; 79 percent were White, 12 percent were Black, and 8 percent were Asian. Among the BAU control teachers, 37 percent were male and 63 percent were female; 66 percent were White, 29 percent Black, and 4 percent Latinx.

READI researchers conducted a stratified RCT in which—after matching on a range of demographic variables—schools were randomly assigned to treatment. The intervention lasted 5 to 6 months (20 to 22 weeks of instruction), with professional development for teachers beginning 9 months prior.

Student intervention. The intervention consisted of a four-phase learning progression organized to enable students to build the science reading and reasoning practices needed to construct explanatory models of science phenomena through text-based investigations. Cutting across these four phases of the learning progression were six science related learning goals, all of which were enacted in each phase of the learning progressions.

The learning goals were (1) close reading and (2) analysis and synthesis of information within and across multiple information sources to (3) construct causal networks of phenomenon-relevant constructs and their relationships that they could (4) justify and (5) critique and evaluate explanatory models consistent with appropriate scientific principles and inquiry methods. A sixth goal was that students would be engaging in these practices in ways consistent with the epistemic commitments of science (e.g., Chinn & Sandoval, 2018).

Accordingly, the four-phase progression began with building classroom routines for close reading in science and then built toward the other practices:

1. Building classroom routines to support close reading of science information and class-wide knowledge-building discussions of the readings. Scaffolds included science reading and talking prompts, including metacognitive stems and evidence and interpretation note-takers. Content dealt with big ideas in biology including ecosystems and interdependence. The cycle of participation structures was established (independent reading, dyad and small group followed by whole-class discussion of reading, interpretations, and implications).
2. Building a repertoire of science literacy and discourse practices through repeated engagement in close reading of multiple texts and discussion of cell biology material, with attention to the kinds of evidence and the nature of interpretations and explanations that can be made from them. Students were introduced to and built understanding of conventions for models of science phenomena and criteria for evaluating them.
3. Deepening scientific literary and discourse practices for reasoned sense making through close reading and synthesis of multiple texts for purposes of building causal explanatory accounts of homeostatic processes and systems in the body. Students began to use models to clarify, refine, modify, and revise their scientific thinking.

4. Utilizing scientific literacy and discourse practices to deepen close reading and multiple-text synthesis for purposes of constructing, justifying, and critiquing causal explanatory accounts for scientific phenomena. Students studied MRSA as an example of evolution as a dynamic in living systems including natural selection, antibiotic resistance, and binary fission.

Teacher professional development. For the intervention teachers, professional development (10 days, approximately 60 hours) extended over a 9-month period prior to beginning the implementation of the intervention, with 2 days during the intervention. The professional development focused on building teachers' awareness of their own practices for making sense of science information, including their own reading and sense making of the various representational forms used in science (e.g., visual models, data tables, graphs, and simulations). READI curriculum modules (Reading Models, Homeostasis, MRSA) were used to immerse the teachers in the intervention they would implement, with attention focused on planning and anticipating what students would do and say and what that might mean with respect to further instructional moves.

Outcome measures. Student measures focused on a pretest measure of basic reading comprehension (RISE, described in Chapter 3), comprehension and application of information from multiple texts (GISA, described in Chapter 3), and an EBA assessment designed to align with the intervention in terms of the learning goals in science.

The EBA assessment was designed to closely align with the text-based inquiry intervention and involved constructing an explanatory model of a science phenomenon based on information distributed across a set of five texts, one of which was a graph and three of which included pictures as well as verbal information. Two phenomena were selected as topics—coral bleaching and sunburn—and were counterbalanced across pre- and posttests at the student and class level. Neither of these were topics that were covered in the intervention or the control classes, although the explanatory model for each drew on concepts and principles that were part of the biological sciences courses in both intervention and control classes. On day 1 of the EBA assessment, students were told that their task was to answer either the question "What leads to differences in the rates of coral bleaching?" or "What leads to differences in the risk of developing skin cancer?" based on information in the set of texts with which we provided them. They were also told that none of the texts contained all of the information they needed to answer the question. They read and annotated the texts on day 1, and on day 2 they responded to four types of assessment items. The essay task tapped their skill at using the information in the texts to write (or draw) an explanatory model; a multiple-choice test tapped inference making within and across texts; a peer-essay evaluation task assessed their awareness of criteria for critiquing and evaluating models (e.g., relevance, coherence); and a graphical model-evaluation task tapped their grasp of criteria for evaluating explanatory models. The EBA assessment was administered pre- and postintervention, with administration in control classrooms yoked to the timing of the assessments in the intervention schools. GISA, which was interestingly on the topic of mitochondrial DNA, was administered approximately 2 weeks after the EBA assessment.

In addition, a subset of students was administered a Science Epistemological Survey, which gauged students' epistemic knowledge and stances related to the use of multiple

sources in science inquiry, and a Science Self-efficacy Survey, which measured students' beliefs about confidence in learning and performing well in science class.

All teachers completed a self-report survey of attitudes toward science and science teaching practices. The preintervention survey was completed prior to the beginning of the professional development for intervention teachers and the postintervention survey was completed after all of the posttest student data had been collected from intervention and control classrooms. All READI intervention and control teachers were observed twice (3–4 weeks into the semester; 10–11 weeks after the first observation). From field notes of the observations, researchers rated the observed lesson on a six-construct rubric (Goldman et al., 2019).

Analyses. Preliminary data analysis employed exploratory factor analysis to examine the validity and reliability of student and teacher measures developed specifically for the RCT. READI scholars, after providing basic descriptive analyses, tested three multilevel models to examine treatment effects at the student level that reflected the nested character of the design; ultimately the team settled on the most parsimonious of the models (i.e., a three-level model with students nested within classrooms and classrooms within schools).

Student results. The major result of interest (see Table 4-15) is that READI students scored significantly higher than the BAU students on GISA, the main distal measure of multiple text comprehension, at posttest when controlling for a range of factors, including the pretest RISE assessment of basic comprehension, the preintervention scores on the two epistemology scales and the self-efficacy scale, and school-level demographic data. READI students scored higher, but not significantly higher, than BAU students on the various essay tasks related to explanations. In addition, there were no statistically significant differences between READI and BAU groups on topic prior knowledge, epistemology, or self-efficacy scales. READI researchers attribute the lack of transfer on the explanation tasks in the essay assessment to the complexity of learning required, coupled with insufficient instructional time for students to "master the rhetorical forms and language structures needed to express explanatory models" (Goldman et al., 2019, p. 1201) in writing.

Although the READI effect sizes qualify as small from a statistical point of view (Cohen, 1992), they are impressive in magnitude from a practical perspective. Specifically,

TABLE 4-15 READI Effect Size Summary by Assessed Construct for Grade 9 Students

	Reading Comprehension		Application: Evidence-Based Argumentation Essay	
	GISA	Multiple-Choice Evidence-Based Argumentation	Concepts	Connections
ES	**0.32**	**0.26**	0.11	0.08

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and represent contrasts with business as usual, and models controlled for pretest scores and school. READI application measures assessed evidence-based argumentation using multiple-choice items and an essay that was scored based on number of concepts represented and connections made. ES = effect size.

Hill and colleagues (2008) estimated the magnitude of change associated with 1 year of reading growth at the high school level to be 0.19. Although the effects are drawn from different measures, the magnitude of the READI effect sizes, which represent how much they gained over and above what BAU students gained, suggests that the READI students potentially demonstrated more than one year's improvement over that experienced by BAU students.

Teacher results. A unique facet of the READI RCT was the use of measures of teacher change over time, and results are summarized in Table 4-16. READI teachers changed their practices over the course of the intervention, shifting to practices more aligned with the Project READI approach, particularly the emphases on social support for reading and practices that promote reasoning and argument development from multiple information sources. On the Survey of Teacher Practices, READI teachers did not differ from BAU teachers at pretest. However, at posttest, the multilevel modeling approach revealed significant differences favoring the READI teachers on several of the scales grouped under science reading opportunities (i.e., learning structure, higher-order prompts, argumentation, multiple-source practices, content, metacognitive inquiry [for both teachers and students], and negotiating [with statistically significant effect sizes ranging from 1.34 to 2.24]). READI teachers scored higher than BAU teachers on observation-based indices of higher-order teaching practices (d = 1.28), as well as on all six of the subscales of higher-order teaching practices—opportunities, support, inquiry, strategies, argumentation, and collaboration (with a range of d from 0.65 to 1.49). Analyses of the observational data documented a tendency for READI teachers to employ a hybrid approach that balanced teacher-directed with student-collaborative activity, in contrast to the dominant BAU pattern of teacher lecture and PowerPoint presentations. Large effect sizes favoring the READI teachers were found on six instructional practices: opportunities, support, inquiry, strategies, argumentation, and collaboration.

TABLE 4-16 READI Effect Size Summary by Assessed Construct for Grade 9 Teachers

	Survey					
	CCSS	Attitude	Self-efficacy	Teaching Philosophy	Science Reading	Higher-Order Teaching
ES	0.45	0.53	0.41	0.46	**1.36**	**2.21**

	Argumentation Practices	Content Reading	Metacognitive Modeling	Metacognitive Practice	Negotiation Instruction
ES	**1.73**	**1.60**	**1.34**	**2.24**	**1.89**

	Practices						
	Higher-Order Teaching	Opportunities	Support	Inquiry	Strategies	Argumentation	Collaboration
ES	**1.28**	**1.49**	**1.09**	**1.37**	**1.07**	**0.65**	**0.83**

NOTES: Bold font indicates a significant effect at $p < .05$. All effects represent Cohen's d and represent contrasts with business as usual. Models for teaching practices controlled for pretest scores. All models controlled for school. CCSS = Common Core State Standards; ES = effect size.

The design did not permit an analysis of the impact of teacher practices on student performance.

Summary

The READI RCT represents the "tip of the iceberg" for the broader READI portfolio of research on the disciplinary literacies necessary for engaging in reading to gather and use evidence to construct arguments that satisfy the constraints of specific disciplines. The RCT did provide evidence of the efficacy of the overall approach—the instructional modules and the highly engaged approach to teacher professional learning—but our focus on the RCT in this chapter (a decision that was necessary in order to reign in the enormity of the scope of the five RfU consortia) obscured much of the texture of the READI research and development in the other two disciplines READI addressed, namely, literary analysis and history, and the extended collaborative design work with teachers, as well as the previous work on Reading Apprenticeship (Greenleaf et al., 2011) that preceded and inspired READI. In each of the three disciplines, READI researchers collaborated with participating teachers to conduct iterative, design-based research, including close observations of the implementations of designed modules followed by collaborative reflection to better understand the realities, merits, and gaps for purposes of improving the module designs and implementations (e.g., Cribb, Maglio, & Greenleaf, 2018; Shanahan et al., 2016; Sosa, Hall, Goldman, & Lee, 2016). This work was shared by the researchers and teachers during teacher inquiry network learning community meetings in which additional high school teachers in each of the three disciplines participated for purposes of transforming their classroom practices to support reading for understanding as manifest in interpretation, explanation, and argumentation in each discipline. Disciplinary similarities and differences emerged through exploration and discussion within disciplinary groups of the nature of argument, the demands of texts and tasks, and the various types of knowledge involved in evidence-based argumentation. At the same time, parallel studies explored cognitive processes of interpretation elicited by different types of tasks, task instructions, and response prompts (e.g., Blaum, Griffin, Wiley, & Britt, 2017; Burkett & Goldman, 2016; Goldman, McCarthy, & Burkett, 2015; Levine & Horton, 2015; Litman & Greenleaf, 2018; McCarthy & Goldman, 2015; Wiley, Jaeger, & Griffin, 2018).

One of the consistent challenges in the classroom implementations, as well as in the basic research, concerned the students' generation of written representations, including explanatory models for science phenomena, causal models for historical events, and interpretive essays in literature. The basic research, insights from the design-based research on curriculum modules, and the instructional model for implementation—in combination with lessons learned from the teacher inquiry networks—informed the culminating RCT summarized earlier. The point is that the instruction ultimately evaluated in the RCT in biological sciences was informed by a host of observational, design, and field implementation efforts not only in science but also in the context of history and literature instruction where much was learned about the nature of effective evidence-based argumentation and the careful, critical reading across sources that leads to it.

It should be noted that the READI RCT included 10 days of professional development beginning 9 months prior to implementation. Four years of design work laid the

foundation (and provided the warrants) for the design of the professional development in the RCT. It is also important to note as well that the design of the RCT professional development drew heavily on the model that the WestEd Strategic Literacy Initiative had developed through their Reading Apprenticeship work (Greenleaf et al., 2011). More than any other RfU team, READI had the explicit goals of changing teacher practices with respect to reading in the disciplines, and of focusing on reading for purposes of creating integrated models across multiple texts that would support evidence-based argumentation (see Goldman, Britt, et al., 2016). The practices of creating those integrated models were and are different in the three disciplines based on each discipline's epistemic aims, inquiry processes, underlying principles, frameworks, content, representational forms, and discourse practices. The READI work stands as a classic example of an intentional line of inquiry in which the development of the ultimate intervention was iteratively tested and refined in the crucible of classroom practice before it was tested in a large-scale RCT. It is the same long runway of research and development cited in our discussion of LARRC.

Looking Across the Array

So, what is one to make of this body of research as a whole? Having provided an account of the pedagogical work of each team that hopefully does justice to the importance and complexity of their work, we now turn to the central question of this synthesis in Chapter 5: regarding curriculum and instruction, what are the common findings, insights, trends, and implications for the various consumers of educational research? We hope that the report can speak to *all* the constituents of our educational system, starting with the general public, especially parents, and extending to those responsible for ensuring that our students learn to read well—the teachers and principals in our schools, the curriculum specialists in our districts, state departments, national educational agencies and organizations, curriculum developers and publishing houses, and the policy makers who set the goals and standards at every level in our educational system—from the national level right down to the classroom. That is the task of Chapter 5.

REFERENCES

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system*. Eugene, OR: Center for Educational Assessment Accountability.

Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education, National Institute of Education.

Apel, K., & Diehm, E. (2013). Morphological awareness intervention with kindergarteners and first and second grade students from low SES homes: A small efficacy study. *Journal of Learning Disabilities, 47*, 65–75.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289–300.

Blaum, D., Griffin, T. D., Wiley, J., & Britt, M. A. (2017). Thinking about global warming: The effect of policy-related documents and prompts on learning about causes of climate change. *Discourse Processes, 54*, 303–316.

Brozo, W., & Afflerbach, P. (2011). *Adolescent Literacy Inventory, Grades 6–12*. New York: Pearson.
Burkett, C., & Goldman, S. R. (2016). "Getting the point" of literature: Relationships between processing and interpretation. *Discourse Processes, 53*(5–6), 457–487.
Capin, P., & Vaughn, S. (2017). Improving reading and social studies learning for secondary students with reading disabilities. *TEACHING Exceptional Children, 49*(4), 249–261.
Chinn, C., & Sandoval, W. (2018). Epistemic cognition and epistemic development. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 24–33). New York: Routledge.
Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*(1), 35–70.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge E-Book. *Computers & Education, 128*, 284–311.
Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zaegar, E., Wood, T., & Petscher, Y. (2017). Acquiring science and social studies knowledge in kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of content-area literacy instruction (CALI). *Journal of Educational Psychology, 109*(3), 301–320.
Connor, C. M., Phillips, B. M., Kaschak, M., Apel, K., Kim, Y. S., Al Otaiba, S., Crowe, E. C., Thomas-Tate, S., Johnson, L. C., & Lonigan, C. J. (2014). Comprehension tools for teachers: Reading for understanding from prekindergarten through fourth grade. *Educational Psychology Review, 26*(3), 379–401. doi: 10.1007/s10648-014-9267-1.
Connor, C. M., Phillips, B. M., Kim, Y. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & Otaiba, S. A. (2018). Examining the efficacy of targeted component interventions on language and literacy for third and fourth graders who are at risk of comprehension difficulties. *Scientific Studies of Reading, 22*(6), 462–484.
Cribb, G., Maglio, C., & Greenleaf, C. (2018). Collaborative argumentation: 10th graders read modern Iranian history. *The History Teacher, 51*, 477–526.
Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325.
Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*, 3–14.
Duhaylongsod, L., Snow, C. E., Selman, R. L., & Donovan, M. S. (2015). Toward disciplinary literacy: Dilemmas and challenges in designing history curriculum to support middle school students. *Harvard Educational Review, 85*(4), 587–608.
Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323–330. doi: 10.1207/s1532799xssr1003_7.
Fogarty, M., Clemens, N., Simmons, D., Anderson, L., Davis, J., Smith, A., Wang, H., Kwok, O., Simmons, L. E., & Oslund, E. (2017). Impact of a technology-mediated reading intervention on adolescents' reading comprehension. *Journal of Research on Educational Effectiveness, 10*(2), 326–353.
Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N. H., & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review, 26*, 425–449.
Francis, D. J., Rivera, M., Lesaux, N., Keiffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions.* Portsmouth, NH: Center on Instruction. Retrieved from http://www.centeroninstruction.org/files/ELL1-Interventions.pdf.
Freebody, P., & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: An Australian Journal of TESOL, 5*(7), 7–16.
Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education, 25*, 99–125.
Glenberg, A. M., Gutierrez, T., Levin, J. R., Japunitich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Research, 96*(3), 424–436.

Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children, 22*(2), 89–116.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Shanahan, C., & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 51*(2), 219–246.

Goldman, S. R., Greenleaf, C., Yukhymenko-Lescroart, M., Brown, W., Ko, M. L. M., Emig, J. M., George, M., Wallace, P., Blaum, D., & Britt, M. A. (2019). Explanatory modeling in science through text-based investigation: Testing the efficacy of the Project READI intervention approach. *American Educational Research Journal, 56*(4), 1148–1216.

Goldman, S. R., McCarthy, K., & Burkett, C. (2015). Interpretative inferences in literature. In E. O'Brien, A. Cook, & R. Lorch (Eds.), *Inferences during reading* (pp. 386–415). Boston: Cambridge University Press.

Goldman, S., & Snow, C. E. (2015). Adolescent literacy: Development and instruction. In A. Pollatsek & R. Treiman (Eds.), *Handbook on reading* (pp. 463–478). New York: Oxford University Press.

Goldman, S. R., Snow, C., & Vaughn, S. (2016). Common themes in teaching reading for understanding: Lessons from three projects. *Journal of Adolescent & Adult Literacy, 60*(3), 255264. doi: 10.1002/jaal.586.

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

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75–90.

Henbest, V. S., & Apel, K. (2017). Effective word reading instruction: What does the evidence tell us? *Communication Disorders Quarterly, 39*(1), 303–311.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Hsin, L. B., Phillips Galloway, E., & Snow, C. E. (n.d.). *Social strengths amid linguistic limitations in the argumentative writing of English-language learners*. Unpublished manuscript.

Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism, 19*(3), 314–332.

IES (Institute of Education Sciences). (2009). *Request for applications: Reading for Understanding research initiative, CFDA Number: 84.305F*. Retrieved from https://ies.ed.gov/funding/pdf/2010_84305F.pdf.

Johanson, M., & Arthur, A. M. (2016). Improving the language skills of pre-kindergarten students: Preliminary impacts of the Let's Know! experimental curriculum. *Child & Youth Care Forum, 45*(3), 367–392.

Johnson, L., Terry, N. P., Connor, C. M., & Thomas-Tate, S. (2017). The effects of dialect awareness instruction on nonmainstream American English speakers. *Reading and Writing, 30*(9), 2009–2038.

Jones, S. M., LaRusso, M., Kim, J., Yeon Kim, H., Selman, R., Uccelli, P., Barnes, S. P., Donovan, S., & Snow, C. (2019). Experimental effects of Word Generation on vocabulary, academic language, perspective taking, and reading comprehension in high-poverty schools. *Journal of Research on Educational Effectiveness, 12*(3), 448–483.

Kent, S., Wanzek, J., Swanson, E. A., & Vaughn, S. (2015). Team-based learning for students with high-incidence disabilities in high school social studies classrooms. *Learning Disabilities Research & Practice, 30*(1), 3–14.

Kim, H. Y., Hsin, L. B., & Snow, C. E. (2018). Reducing academic inequalities for English language learners: Variation in experimental effects of Word Generation in high-poverty schools. *International Journal of Bilingual Education and Bilingualism,* 1–19. doi: 10.1080/13670050.2018.1535574.

Kim, H. Y., LaRusso, M. D., Hsin, L. B., Harbaugh, A. G., Selman, R. L., & Snow, C. E. (2018). Social perspective-taking performance: Construct, measurement, and relations with academic performance and engagement. *Journal of Applied Developmental Psychology, 57*, 24–41.

Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2016). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly, 52*(3), 357–382.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

LARRC (Language and Reading Research Consortium). (2016). Use of the Curriculum Research Framework (CRF) for developing a reading-comprehension curricular supplement for the primary grades. *The Elementary School Journal, 116*(3), 459–486.

LARRC, Arthur, A. M., & Davis, D. L. (2016). A pilot study of the impact of double-dose robust vocabulary instruction on children's vocabulary growth. *Journal of Research on Educational Effectiveness, 9*(2), 173–200.

LARRC, Jiang, H., & Davis, D. (2017). Let's Know! proximal impacts on prekindergarten through grade 3 students' comprehension-related skills. *The Elementary School Journal, 118*(2), 177–206.

LARRC, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research, 62*, 2812–2828.

LARCC, Johanson, M., & Arthur, A. M. (2016). Improving the language skills of pre-kindergarten students: Preliminary impacts of the Let's Know! experimental curriculum. *Child Youth Care Forum, 45*, 367–392. doi: 10.1007/s10566-015-9332-z.

LARRC, Pratt, A., & Logan, J. (2014). Improving language-focused comprehension instruction in primary-grade classrooms: Impacts of the "Let's Know!" experimental curriculum. *Educational Psychology Review, 26*(3), 357–377. doi: 10. 1 007/s 1 0648-0 1 4-9275- 1.

LaRusso, M. D., Donovan, S., & Snow, C. (2016). Implementation challenges for tier one and tier two school-based programs for early adolescents. *New Directions for Child and Adolescent Development, 154*, 11–30.

LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, C., Donovan, S., & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness, 9*(2), 201–222.

Lawrence, J., Capotosto, L., Branum-Martin, L., & Snow, C. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition, 15*, 437–451.

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal, 52*(4), 750–786.

Lawrence, J. F., Rolland, R. G., Branum-Martin, L., & Snow, C. E. (2014). Generating vocabulary knowledge for at-risk middle school readers: Contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk, 19*(2), 76–97.

Lee, C. D. (2007). *Culture, literacy and learning: Taking blooming in the midst of the whirlwind.* New York: Teachers College Press.

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York: Carnegie Corporation.

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87–113). London: Routledge.

Leslie, L., & Caldwell, J. (2011). *Qualitative Reading Inventory, 5th edition*. New York: Pearson.

Levine, S., & Horton, W. (2015). Helping high school students read like experts: Affective evaluation, salience, and literary interpretation, *Cognition and Instruction, 33*(2), 125–153.

Litman, C., & Greenleaf, C. (2018). Argumentation tasks in secondary English language arts, history, and science: Variations in instructional focus and inquiry space. *Reading Research Quarterly, 53*, 107–126.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests—directions for administration (levels 7/9 & 10/12)* (4th ed.). Rolling Meadows, IL: Riverside.

MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2006). *Gates-MacGinitie Reading Tests*. Itasca, IL: Riverside.

Martin, N., & Brownell, R. (2010). *EOWPVT-4: Expressive One-Word Picture Vocabulary Test–Fourth Edition*. Indianapolis, IN: Pro-Ed.

McCarthy, K. S., & Goldman, S. R. (2015). Comprehension of short stories: Effects of task instructions on literary interpretation. *Discourse Processes, 52*, 585–608.

Miller, J. F., & Chapman, R. S. (2008). *Systematic analysis of language transcripts (computer software).* Madison: University of Wisconsin–Madison, Waisman Center, Language Analysis Laboratory.

NAEP (National Assessment of Educational Progress). (2017). *Reading framework for the 2017 National Assessment of Educational Progress*. National Assessment Governing Board, U.S. Department of Education.

NAEP. (2019a). *NAEP reading stat facts.* Retrieved from https://nces.ed.gov/nationsreportcard/reading.

NAEP. (2019b). *NAEP report card: Reading*. Retrieved from https://www.nationsreportcard.gov/reading/?grade=4.

NELP (National Early Literacy Panel). (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Retrieved from http://www.nifl.gov/earlychildhood/NELP/NELPreport.html.

NGA & CCSSO (National Governors Association Center for Best Practices and Council of Chief State School Officers). (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Author.

NICHD (National Institute of Child Health and Human Development). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

NRC (National Research Council). (2014). *Developing assessments for the next generation science standards*. Washington, DC: The National Academies Press.

Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education, 32*(2), 203–218.

Park, M., Anderson, E., & Yoon, S. (2017). Learning biology coherently through complex systems: A model supported with scientific practices and agent-based simulations. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Proceedings of the 12th International Conference for Computer Supported Collaborative Learning* (pp. 199–206). Philadelphia: International Society of the Learning Sciences.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Science and literacy: Each in the service of the other. *Science, 328*, 459–463.

Petersen, D. B., & Spencer, T. D. (2012). The Narrative Language Measures: Tools for language screening, progress monitoring, and intervention planning. *Perspectives on Language Learning and Education, 19*(4), 119–129.

Phillips, B. M., Tabulda, G., Ingrole, S. A., Burris, P. W., Sedgwick, T. K., & Chen, S. (2016). Literate language intervention with high-need prekindergarten children: A randomized trial. *Journal of Speech, Language, and Hearing Research, 59*(6), 1409–1420.

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System [CLASS] manual*. Baltimore: Brookes.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. M. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431–451.

PISA (Programme for International Student Assessment). (2018). *PISA 2018 results.* Retrieved from https://www.oecd.org/pisa/publications/pisa-2018-results.htm.

Pratt, A., & Logan, J. (2014). Improving language-focused comprehension instruction in primary-grade classrooms: Impacts of the Let's Know! experimental curriculum. *Educational Psychology Review, 26*(3), 357–377.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum Associates.

Purcell-Gates, V., Perry, K., & Briseño, A. (2011). Analyzing literacy practice: Grounded theory to model. *Research in the Teaching of English, 45*, 439–458.

Schoenbach, R., & Greenleaf, C. (2009). Fostering adolescents' engaged academic literacy. In L. Christenburg, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 98–112). New York: Guilford Press.

Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). *Reading for understanding: How Reading Apprenticeship improves disciplinary learning in secondary and college classrooms* (2nd ed.). San Francisco: Jossey-Bass.

Schoenbach, R., Greenleaf, C., & Murphy, L. (2016). *Leading for literacy: A Reading Apprenticeship approach*. San Francisco: Wiley Press.

Seymour, H. N., Roeper, T. W., & deVilliers, J. (2003). *Diagnostic evaluation of language variation, screening test examiner's manual*. San Antonio, TX: The Psychological Corporation.

Shanahan, C., Heppler, J., Manderino, M., Bolz, M., Cribb, G., & Goldman, S. R. (2016). Deepening what it means to read (and write) like a historian: Progressions of instruction across a school year in an eleventh grade U.S. history class. *The History Teacher, 49*, 241–270.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Simmons, D., Fogarty, M., Oslund, E. L., Simmons, L., Hairrell, A., Davis, J., Anderson, L., Clemens, N., Vaughn, S., Roberts, G., Stillman, S., & Fall, A. M. (2014). Integrating content knowledge-building and student-regulated comprehension practices in secondary English language arts classes. *Journal of Research on Educational Effectiveness, 7*(4), 309–330.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325–344.

Sosa, T., Hall, A. H., Goldman, S. R., & Lee, C. D. (2016). Developing symbolic interpretation through literary argumentation. *Journal of the Learning Sciences, 25*(1), 93–132.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Swanson, E., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children, 81*(4), 426–442.

Texas Education Agency. (2013). *State of Texas assessments of academic readiness*. Austin, TX: Texas Education Agency.

Torgesen, J., Wagner, R., & Rashotte, C. (2010). *Test of word reading efficiency, 2nd ed*. New York: Pearson.

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *TOWRE-2 test of word reading efficiency, 2nd ed*. San Antonio, TX: Pearson.

Troyer, M. (2017). Teachers' adaptations to and orientations toward an adolescent literacy curriculum, *Journal of Curriculum Studies, 51*(2), 202–228. doi: 10.1080/00220272.2017.1407458.

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A.-M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology, 109*(1), 22–34.

Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A. M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review, 27*(1), 31–50.

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

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of silent reading efficiency and comprehension*. Austin, TX: Pro-Ed.

Wanzek, J., Roberts, G., Vaughn, S., Swanson, E., & Sargent, K. (2019). Examining the role of pre-instruction academic performance within a text-based approach to improving student content knowledge and understanding. *Exceptional Children, 85*(2), 212–228.

Wanzek, J., Swanson, E., Vaughn, S., Roberts, G., & Fall, A-M. (2016). English learner and non-English learner students with disabilities: Content acquisition and comprehension. *Exceptional Children, 82*, 428–442.

Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., Fall, A-M., Stillman-Spisak, S. J., & Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness, 7*(2), 183–204. doi: 10.1080/19345747.2013.836765.

Wiederholt, J., & Bryant, B. (2012). *Gray oral reading tests-Fifth edition (GORT-5)*. Austin, TX: Pro-Ed.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating the effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools, 45*, 432–445.

Wiley, J., Jaeger, A. J., & Griffin, T. D. (2018). Effects of task instructions on comprehension of multiple documents in history and science. In J. L. G. Braasch, I. Braten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 341–361). New York: Routledge.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. American Guidance Service. New York: Pearson.

Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia: Temple University Press.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of achievement*. Rolling Meadows, IL: Riverside.

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly, 52*(2), 203–226.

Appendix 4-1
Published Measures Used in the Reading for Understanding Portfolio of Efficacy Studies Represented in Chapters 4 and 5

TABLE APPENDIX 4-1 Published Measures Used in the Reading for Understanding Portfolio of Efficacy Studies Represented in Chapters 4 and 5

Construct	Assessment	Description	RfU Approach
Knowledge and learning	Woodcock-Johnson III (WJ-III) academic knowledge subtest	Norm-referenced, individually administered, ungraded, untimed test in which students answer aloud questions of increasing difficulty in science, social studies, and the humanities	COMPASS ERC LIM TEXTS
Reading comprehension	Gates-MacGinitie Reading Test (GMRT)	Norm-referenced, group-administered, grade-leveled, untimed test in which students read several passages and answer multiple-choice questions about each	CCT COMPASS ERC LIM PACT TBL TEXTS
	Global, Integrated, Scenario-Based Assessments (GISA)	See Chapter 3	READI WG
	Gray Oral Reading Test, 5th edition (GORT-5)	Norm-referenced, individually administered, ungraded, untimed test in which students read aloud and orally answer comprehension questions about a series of passages of increasing readability and complexity	CCT
	Group Reading Assessment and Diagnostic Evaluation (GRADE) sentence comprehension subtest	Norm-referenced, group-administered, grade-leveled, untimed test in which students choose a word among several choices that best completes a sentence	CCT
	GRADE passage comprehension subtest	Norm-referenced, group-administered, grade-leveled, untimed test in which students read several passages and answer multiple-choice questions about each	CCT
	Reading Inventory and Scholastic Evaluation (RISE) sentence processing subtest	See Chapter 3	STARI

TABLE APPENDIX 4-1 Continued

Construct	Assessment	Description	RfU Approach
Reading comprehension (continued)	RISE efficiency of basic reading comprehension subtest	See Chapter 3	STARI
	RISE reading comprehension subtest	See Chapter 3	STARI
	Test of Silent Reading Efficiency and Comprehension (TOSREC)	Norm-referenced, group-administered, grade-leveled, timed test in which students read sentences and judge them as true or false, completing as many as possible in 3 minutes	CCT COMPASS ERC LIM MAT TEXTS
	WJ-III passage comprehension subtest	Norm-referenced, individually administered, ungraded, untimed test in which students read texts of one to several sentences in length and verbally provide a word needed to make the passage complete	CALI
Listening comprehension	Clinical Evaluation of Language Fundamentals, 4th Edition (CELF-4) concepts and following directions subtest	Norm-referenced, individually administered, ungraded, untimed test in which students listen to, interpret, and follow directions of increasing difficulty	COMPASS ERC LIM TEXTS
	Oral and Written Language Scales (OWLS) listening comprehension scale	Norm-referenced, individually administered, ungraded, untimed test in which students point to a picture that correctly captures lexical/semantic, syntactic, pragmatic, and supralinguistic prompts of increasing difficulty	COMPASS ERC LIM TEXTS
	Test of Narrative Language (TNL) comprehension subtest	Norm-referenced, individually administered, ungraded, untimed test in which students answer literal and inferential open-ended questions about narrative texts	COMPASS ERC LIM TEXTS
	WJ-III oral comprehension subtest	Norm-referenced, individually administered, ungraded, untimed test in which students produce a missing word for an orally presented passage in increasing order of difficulty	CALI
Vocabulary	CELF-4 expressive vocabulary subtest	Norm-referenced, individually administered, ungraded, untimed test in which students name people, objects, and actions based on illustrations in increasing order of difficulty	COMPASS ERC LIM TEXTS

TABLE APPENDIX 4-1 Continued

Construct	Assessment	Description	RfU Approach
Vocabulary (continued)	Expressive One Word Picture Vocabulary Test, 4th Edition (EOWPVT)	Norm-referenced, individually administered, ungraded, untimed test in which students name objects, actions, and concepts based on illustrations in increasing order of difficulty	COMPASS ERC LIM TEXTS
	RISE vocabulary subtest	See Chapter 3	STARI
	WJ-III expressive vocabulary subtest	Norm-referenced, individually administered, ungraded, untimed test in which students name pictured objects	LIM
Syntax	Clinical Evaluation of Language Fundamentals Preschool, 2nd Edition (CELFP2) sentence structure subtest	Norm-referenced, individually administered, ungraded, untimed test in which students point to a picture that matches verbal prompts	LIM
	Comprehensive Assessment of Spoken Language (CASL) syntax construction subtest	Norm-referenced, individually administered, grade-leveled, untimed test in which students respond orally to a verbal prompt and picture with a grammatically and semantically appropriate word, phrase, or sentence	COMPASS ERC TEXTS
Morphology	RISE morphology subtest	See Chapter 3	STARI
Word recognition	easyCBM passage reading fluency subtest	Norm-referenced, individually administered, grade-leveled, timed test in which students read a passage aloud for 1 minute and are scored based on the number of words read aloud correctly	CCT
	RISE word recognition and decoding subtest	See Chapter 3	STARI
	Test of Word Reading Efficiency-2nd Edition (TOWRE2) phonemic decoding efficiency subtest	Norm-referenced, individually administered, ungraded, timed test in which students read nonsense words listed in order of increasing difficulty, reading as many as possible in 45 seconds	COMPASS ERC LIM TEXTS
	TOWRE2 sight word efficiency subtest	Norm-referenced, individually administered, ungraded, timed test in which students read real words listed in order of increasing difficulty, reading as many as possible in 45 seconds	CCT COMPASS ERC LIM TEXTS MAT
	WJ-III letter word identification subtest	Norm-referenced, individually administered, ungraded, untimed test in which students read letters and words in increasing order of difficulty	COMPASS ERC LIM MAT TEXTS

Appendix 4-2
Demographic Data for Reading for Understanding Teams' Randomized Controlled Trials

TABLE APPENDIX 4-2 Demographic Data for Reading for Understanding Teams' Randomized Controlled Trials

Team	Citation	Instructional Approach	Grades	N	Region	Race/Ethnicity	SES	Language Background	SPED
LARRC	LARRC et al., 2017	LK-B LK-D	pre-K–3	766	NR	86% White 12% Black 12% Hispanic/Latino 4% AsianAm 2% other	20% FARMS	NA	6%
	LARRC et al., 2019	LK	1–3	938	NR	29-43% ethnic minorities	NR	> 90% English at home	6–11%
FCRR	Connor et al., 2017	CALI	K–4	418	South	77% White 10% Black 13% other	50% FARMS	NR	NR
	Connor et al., 2018	COMPASS, ERC, LIM, TEXTS	3 and 4	645	South	53% White 39% Black 3% multiracial 5% other	≥ 40% FARMS	NR	NR
	Johnson et al., 2017	DAWS	2–4	Study 1: 126	South	95% Black 2% Latino/Hispanic 2% White 1% multiracial	93% FARMS	1% LEP	18%
				Study 2: 374	South	45% Black 33% White 7% multiracial 4% Latino/Hispanic 4% AsianAm	68–84% FARMS	5% LEP	8%
	Phillips et al., 2016	LIM	pre-K	120	NR	62% Black 28% White 8% Latino/Hispanic 2% AsianAm	77–100% FARMS	NR	NR
	Apel & Diehm, 2013	MAT	K–2	151	NR	76% Black 16% White 5% multiracial 3% Latino/Hispanic 1% AsianAm	74% FARMS	100% English only	8%

CCDD	Jones et al., 2019	WG	4–7	Year 1: 5,648	Northeast	42% Black 29% White 23% Latino/Hispanic 3% AsianAm 2% mixed/other	82% FARMS	10% ELL	18%
				Year 2: 5,317	Northeast	41% Black 29% White 24% Latino/Hispanic 3% AsianAm 2% mixed/other	83% FARMS	8% ELL	16%
	Kim et al., 2017	STARI	6–8	482	Northeast	50% White 24% Latino/Hispanic 20% Black 4% mixed/other 2% AsianAm	73% FARMS	16% ELL	33%
PACT	Fogarty et al., 2014	CCT	6–8	859	West	43% Latino/Hispanic 34% White 22% Black	67% FARMS	9% ELL	6%
	Fogarty et al., 2017	CCT	6–8	228	South	29% Black 27% White 26% Latino/Hispanic 14% multiracial 3% AsianAm 1% AmIndian/AKNative	63% FARMS	10% ESL	8%
	Simmons et al., 2014	CCT	7–10	921	West	35% Latino/Hispanic 32% White 31% Black 2% other/missing	69% FARMS	7% ELL	4%
	Kent et al., 2015	TBL	11	24	South	50% Black 46% White 4% NR	30% FARMS	NR	63%

continued

TABLE APPENDIX 4-2 Continued

Team	Citation	Instructional Approach	Grades	N	Region	Race/Ethnicity	SES	Language Background	SPED
PACT (continued)	Vaughn et al., 2013	PACT	8	419	South	53% White 30% Latino/Hispanic 9% Black 5% AsianAm 4% multiracial	23% FARMS	3% ELL	5%
	Vaughn et al., 2015	PACT	8	1,487	South, West	62% White 23% Latino/Hispanic 18% Black 9% AmIndian/ AKNative 4% AsianAm 4% multiracial	39% FARMS	5% ELL	8%
	Vaughn et al., 2017	PACT	8	1,629	South, West	61% Latino/Hispanic 60% White 14% Black 14% AmIndian/ AKNative 3% AsianAm 1% NativeHI/ PacIslander	49–83% FARMS	27% current or recent ELL	10%
	Wancek et al., 2014	TBL	11	463	South	47% White 45% Black 4% Latino/Hispanic 3% multiracial 2% AsianAm	38% FARMS	NR	6%
READI	Goldman et al., 2019	READI	9	979	Midwest	34% Latino/Hispanic 26% Black 19% White 16% multiracial/ other 5% AsianAm	47–94% FARMS	76% English first language	NR

NOTES: Race/ethnicity percentages may not total 100 due to rounding and whether ethnicity was treated separately from race. Where ranges are reported, the percentages represent school rather than student demographics. Report Language background uses the study's terms: LEP so far for FCRR, but others used different definitions. AmIndian/AKNative = American Indian/Alaska Native; AsianAm = Asian American; CALI = Content Area Literacy Instruction; CCT = Comprehension Circuit Training; CCDD = Catalyzing Comprehension through Discussion and Debate; COMPASS = Comprehension Monitoring and Providing Awareness of Story Structure; DAWS = Dialect Awareness; ELL = English language learner; ERC = Enacted Reading Comprehension; FARM = free and reduced-price meal; FCRR = Florida Center for Reading Research; LARCC = Language and Reading Research Consortium; LEP = limited English proficiency; LIM = Language in Motion; LK = Let's Know!; MAT = Morphological Awareness Training; N = sample size; NA = not applicable if students were excluded; NativeHI/PacIslander = Native Hawaiian/Pacific Islander; NR = not reported; PACT = Promoting Adolescents' Comprehension of Text; READI = Reading, Evidence, and Argumentation in Disciplinary Instruction; SES = socioeconomic status; SPED = special education; STARI = Strategic Adolescent Reading Intervention; TBL = team-based learning; TEXTS = Teaching Expository Text Structures; WG = Word Generation.