5

Teaching Reading for Understanding: Synthesis and Reflections on the Curriculum and Instruction Portfolio

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

CONTENTS

INTRODUCTION

Having summarized the pedagogical stories of each of the five Reading for Understanding (RfU) teams in Chapter 4, we now turn to the task of looking across those portfolios for trends, themes, insights, and implications for policy and practice. To accomplish this synthesis, we examine the evidence in two distinct but complementary ways.

First, building on the detailed site-by-site and intervention-by-intervention examination of experimental results from Chapter 4, we step back to take a more panoramic view of the experimental results for all five teams. We summarize the effect sizes across all of the randomized controlled trials (RCTs) and efficacy studies in two tables. The first table summarizes effect sizes for measures of comprehension, including listening comprehension and application tasks like writing, while the second table summarizes effect sizes for measures of component skills and knowledge that contribute to comprehension. Each table is organized with grade levels across the columns and measures constructs (e.g., reading comprehension, vocabulary, and word recognition) down the rows. For each measured construct in each table, we include two rows: one of effects for researcher-developed measures and one for published measures. This organization allows for a high-level examination of patterns of quantitative effects across the entire RfU portfolio of efficacy studies for the myriad approaches to curriculum and instruction.

Second, we traverse the same landscape of interventions, but to foreground the practices that cluster across teams in association with effective interventions. In the broadest terms, the first pass begins with the quantitative results and moves toward an account of the practices that were likely responsible for those results. The second pass, by contrast, begins with a careful description of consistently influential practices and moves toward the results that validate their efficacy.

PEDAGOGICAL EFFECTS ACROSS THE RFU PORTFOLIO

Judging the Magnitude of Effects

Finding a way to express the importance, or magnitude, of effects (as indexed by the difference between a treatment and an untreated control group) in everyday language rather than obscure technical terms has concerned researchers for at least three decades. Cohen (1992) suggested that standardized effect sizes in the mean difference, or *d* family, which includes Hedges's *g*, could be interpreted as indicators of the magnitude of quantitative results that furthermore could be expressed in everyday language such as weak to strong or small to large. He suggested that effects from 0.20 to 0.49 could be considered small in magnitude, 0.50 to 0.79 could be considered medium or moderate in magnitude, and effects of 0.80 or above could be considered large. However, in setting these standards, Cohen advised strongly that researchers consult typical effects in their particular field to more aptly define small, medium, and large effects more contextually.

Along these lines, Hill, Bloom, Black, and Lipsey (2008) provided guidance for interpreting effects in educational research with respect to reading and math achievement. They also provided guidance based on a number of criteria, including the population and type of assessment used to measure the effect. For example, when

examining average change over time in reading performance from kindergarten to grade 1, the mean effect they found was 1.52 with a margin of error of 0.21, whereas from grades 1–2 it was 0.97 with a 0.10 margin of error. Both of these outcomes are for growth on standardized, norm-referenced tests. In contrast, when examining effects for treatment versus control groups on randomized trials, which is the relevant frame for understanding the current body of RfU intervention work, much more modest average effects were found. The mean effect in the elementary grades was 0.33 with a standard deviation of 0.48, in the middle grades it was 0.51 with a standard deviation of 0.49, and in high school it was 0.27 with a standard deviation of 0.33. Because fewer randomized trials were available in the middle and high school grades when they did their analysis, they did not further break down those effects.

Hill et al. (2008) further disaggregated the mean effects observed in the elementary grades based on the grain size and type of test administered, that is, whether it was a "broad" standardized test, a "narrow" standardized test, or a highly specialized test (of the sort often constructed by a researcher to measure a construct of particular interest in a particular study). They found that the smallest mean effects were observed for the most general outcome measures ($M = 0.07$, $SD = 0.32$), larger for narrower standardized measures ($M = 0.23$, $SD = 0.35$), and largest for specialized tests ($M = 0.44$, $SD = 0.49$). However, even Hill et al. (2008) noted that these interpretive frames do not necessarily indicate what is *desirable* from a policy standpoint so much as they indicate what is *possible* to achieve based on prior research.

Our Decision

Given Hill et al.'s (2008) findings about the volatility of effect sizes depending on grade level and the grain size of the test, coupled with the fact that we have additional data from a full decade of research since they reported on these, we decided to adhere to Cohen's rule of thumb with the following amendments: Because effects on the broadest general outcome measures were typically so small in the Hill et al. (2008) work, we created another category for *weak effects*, defined as 0.07 to 0.19. We otherwise adopted Cohen's definitions of small (0.20 to 0.49), medium (0.50 to 0.79), and large (0.80 or above) effects. In interpreting these effects, however, we must emphasize that the average effects for randomized trials found by Hill et al. typically fall within the small category, making even medium effects impressive (or at least rare) in comparison.

In Tables 5-1 and 5-2, we present the effects found across the RfU consortia for constructs measured by at least two of the consortia. For more idiosyncratic effects, readers should refer to the site-by-site report of effect sizes for specific measures in Chapter 4. Note that we are missing effect sizes for interventions where effect sizes were not available to us, not reported by authors, or not derivable from the published report, and we also do not include mediated effects in the tables because of the diversity of approaches employed across the RfU teams and studies.

Effect Size Patterns

As noted in Chapter 4, the measured outcomes in the RfU ranged very widely within and across projects. With the exception of Reading, Evidence, and Argumentation in

Disciplinary Instruction (READI), all projects included measures of discrete component skills and knowledge, often representing near transfer of instructional targets. Some, though not all (e.g., the use of the Reading Inventory and Scholastic Evaluation [RISE] for the Strategic Adolescent Reading Intervention [STARI]), of these were developed by the researchers. All teams, including READI, also tackled measures tapping the orchestration of comprehension skills. In all cases, these measures included at least one assessment of desired far transfer of improvements on more discrete skills to reading or listening comprehension. In some cases, as with Dialect Awareness (DAWS), Content Area Literacy Instruction (CALI), Promoting Adolescents' Comprehension of Text (PACT), READI, team-based learning (TBL), and Word Generation (WG), they extended to applications (complex reading, writing, editing, and learning tasks) that required reading comprehension in their execution. While these tasks were often researcher developed, they also represented transfer measures in the sense that students were tasked with exercising their comprehension in the acquisition of knowledge and even applying that new knowledge in new ways (e.g., writing an essay). In essence, researchers on these teams developed transfer tasks that represented the orchestration of reading comprehension in pursuit of some other goal that was highly relevant to authentic reading tasks. It is nearly impossible to do justice to the wide range of outcomes and the measures used to assess them (see Appendix 4-1 for a summary table of measures used across the RfU efficacy studies) and the wide range in the populations served (see Appendix 4-2 for a summary table of demographics across the RfU efficacy studies).

As a result, in the tables described next, we decided to separate the measures and effect sizes based on whether measures tapped reading or listening comprehension directly, including the orchestration of comprehension for applied tasks (see Table 5-1), or measures tapped component skills and knowledge that undergird comprehension and its application (see Table 5-2). Within each, we also distinguish between effects on measures that were researcher designed (rows labeled "R" in Tables 5-1 and 5-2) and those that were more widely available and normed (rows labeled "P" in Tables 5-1 and 5-2). These decisions were informed primarily by Hill et al.'s (2008) findings regarding how the magnitude of effects typically depends on this distinction. Given the findings of Hill et al. (2008), the effects for researcher-developed measures ought to be larger than those for published measures; likewise, the effects in Table 5-1, which reports on broader measures, ought to be smaller than those in Table 5-2, which reports on more discrete measures.

Across both tables, the columns are defined by the grade levels targeted, running from pre-kindergarten (pre-K) through high school. Given Hill et al.'s (2008) findings that annual growth is larger in earlier grades and smaller in later grades, the effects running from left to right across columns ought to follow a similar pattern, with the largest effects observed for the youngest students.

To summarize, if the results of the RfU efficacy trials are consistent with what Hill et al. (2008) observed, then the reader should expect that effect sizes are greater in magnitude in Table 5-2 than in Table 5-1, greater in the left-hand columns than the right-hand columns in both tables, and greater in the top (i.e., R) rows than the bottom (i.e., P) rows for each construct in each table. That said, what these tables cannot capture well is how aligned the various discrete skills were with the various interventions. Thus, the pattern of larger effects in Table 5-2 than Table 5-1 should be less consistent than the differences observed between the two rows for each construct.

TABLE 5-1 Cross-Consortium Effect Size Summary of the RfU for Comprehension by Grade Level for Researcher-Developed (R) and Published (P) Measures

Construct		Pre-K	K	1	2	3	4	5	6	7	8	HS
Applied reading comprehension	R				**DAWS$^{\bullet\bullet\bullet\bullet}$**	**DAWS$^{\bullet\bullet\bullet\bullet}$**	**DAWS$^{\bullet\bullet\bullet\bullet}$** **WG$^{\bullet}$**	**WG$^{\bullet}$**	**WG$^{\bullet}$**	**WG$^{\bullet}$**		READI$^{\bullet}$
	P											
Knowledge and learning	R		**CALI$^{\bullet\bullet\bullet\bullet}$**	**CALI$^{\bullet\bullet\bullet\bullet}$**	**CALI$^{\bullet\bullet\bullet\bullet}$**	**CALI$^{\bullet\bullet\bullet\bullet}$**	**CALI$^{\bullet\bullet\bullet\bullet}$**				**PACT$^{\bullet\bullet}$**	**TBL$^{\bullet}$**
	P					COMPASS^{n} ERC$^{\bullet}$ LIM$^{\bullet}$	ERC$^{\bullet}$ TEXTS$^{\bullet\bullet}$					
Reading comprehension	R								CCT^{n}	CCT^{n}	CCT^{n} **PACT$^{\bullet\bullet}$**	CCT^{n} **READI$^{\bullet\bullet}$**
	P			MAT$^{\bullet\bullet}$	MAT$^{\bullet}$	COMPASS^{n} ERC^{n} LIM^{n}	**CALI$^{\bullet\bullet}$** ERC^{n} TEXTS^{n} **WG$^{\bullet}$**	**WG$^{\bullet}$**	**CCT$^{\bullet\bullet}$** **STARI$^{\bullet\bullet}$** **WG$^{\bullet}$**	**CCT$^{\bullet\bullet}$** **STARI$^{\bullet\bullet}$** **WG$^{\bullet}$**	**CCT$^{\bullet\bullet}$** **PACT$^{\bullet\bullet}$** **STARI$^{\bullet\bullet}$**	CCT^{n} **READI$^{\bullet\bullet}$** TBL^{n}
Listening comprehension	R	**LIM$^{\bullet\bullet\bullet}$** LK$^{\mathbf{D}\bullet\bullet\bullet}$ LK$^{D\bullet\bullet\bullet\bullet}$	**LK$^{D\bullet\bullet}$** LK$^{D\bullet\bullet}$	LK$^{B\bullet\bullet}$ LK^{D,n} LK^{n}	LK^{B,n} LK^{D,n} LK^{n}	LK$^{B\bullet\bullet}$ **LK$^{D\bullet\bullet\bullet\bullet}$** **LK$^{\bullet\bullet}$**						
	P	LIM$^{\bullet\bullet}$				COMPASS$^{\bullet}$ ERC$^{\bullet}$ LIM$^{\bullet}$	**CALI$^{\bullet\bullet}$** ERC^{n} TEXTS$^{\bullet\bullet}$					

NOTES: Only constructs assessed by more than one RfU consortium are included. **Bold font** indicates at least one significant ($p < .05$) effect found for the indicated curriculum. Bullets (•) indicate magnitude of effects on measure with largest effect size. All effects represent contrasts with business as usual. Applied reading comprehension measures included writing and other applications. Knowledge and learning measures included assessments of taught and untaught content knowledge. See Chapter 4 for specific effect sizes and more details on measures. HS = high school; ^{n} = nil or negative in magnitude (ES < 0.07); P = published measures; R = researcher-developed measures; • = weak (ES = 0.07–0.19); •• = small (ES = 0.20–0.49); ••• = medium (0.50–0.79); •••• = large (0.80+).

TABLE 5-2 Cross-Consortium Effect Size Summary of the RfU for Commonly Assessed "Component" Constructs by Grade Level for Researcher-Developed (R) and Published (P) Measures

Construct		Pre-K	K	1	2	3	4	5	6	7	8	HS
Comprehension monitoring	R	LK^{B}••• **LK^{D}••••**	**LK^{B}••••** **LK^{D}••••**	**LK^{B}••••** **LK^{D}••••** **LK••••**	**LK^{B}••••** **LK^{D}••••** **LK•••**	COMPASS•• ERC• LIM^{n} **LK^{B}••••** **LK^{D}••••** **LK•••**	ERC^{n} TEXTS^{n}					
	P											
Academic language	R				**DAWS••**	**DAWS••**	**DAWS••** **WG^{n}**	**WG^{n}**	WG^{n}	WG^{n}		
	P											
Vocabulary	R	**LK^{B}••••** **LK^{D}••••**	**LK^{B}••••** **LK^{D}••••**	**LK^{B}••••** **LK^{D}••••** **LK••••**	**LK^{B}••••** **LK^{D}••••** **LK••••**	**LK^{B}••••** **LK^{D}••••** **LK••••**	**WG••**	**WG••**	**CCT••** **WG•**	**CCT••** **WG•**	**CCT••**	
	P	LIM^{n}				COMPASS• ERC•• LIM•	**CALI••••** ERC• TEXTS•		STARI•	STARI•	STARI•	
Syntax	R	**LIM••••**										
	P	LIM•				COMPASS• ERC• LIM•	ERC^{n} TEXTS•					
Morphology	R		**MAT••••**	**MAT••••**	**DAWS••** **MAT••••**	**DAWS••**	**DAWS••**					
	P								**STARI•**	**STARI•**	**STARI•**	
Word recognition	R											
	P		MAT^{n}	MAT•	MAT••	COMPASS^{n} ERC^{n} LIM•	ERC• TEXTS^{n}		CCT^{n} **STARI••**	CCT^{n} **STARI••**	CCT^{n} **STARI••**	

NOTES: Only constructs assessed by more than one RfU consortium are included. **Bold font** indicates at least one significant ($p < .05$) effect found for the indicated curriculum. Bullets (•) indicate magnitude of effects on measure with largest effect size. All effects represent contrasts with business as usual. See Chapter 4 for specific effect sizes and more details on measures. HS = high school; ^{n} = nil or negative in magnitude (ES < 0.07); P = published measures; R = researcher-developed measures; • = weak (ES = 0.07–0.19); •• = small (ES = 0.20–0.49); ••• = medium (0.50–0.79); •••• = large (0.80+).

Patterns are best apprehended at a glance by attending to bold font (which indicates that the effect was associated with a statistically significant coefficient) and the number of bullets[1] (•••) following the acronym for the intervention. Thus, the notation of LK^{D}•••• in the R row for Listening Comprehension in pre-K in Table 5-1 tells us that the Language and Reading Research Consortium (LARRC) intervention Let's Know! (LK)-Deep produced a statistically significant, large effect in pre-K.

Measures of Comprehension and Beyond

Table 5-1 summarizes effects for four constructs that directly measured comprehension: listening comprehension, reading comprehension, knowledge and learning, and applications of reading comprehension. As is evident from Table 5-1, effect sizes were generally larger in earlier grades and for researcher-designed measures, which is consistent with typical findings in education (Hill et al., 2008; Lortie-Forgues & Inglis, 2019). Reversing the typical trend (Hill et al., 2008), for the more distal measures the results are better in grades 4–12 compared to the earlier grades.

Reading Comprehension

While the RfU consortia used a wide range of measures of reading comprehension, it is notable that nearly all consortia working in these grades saw at least one effect of 0.20 or above in reading comprehension. Keep in mind that effects on broad measures of constructs like reading comprehension are typically weak in magnitude, at least in the elementary grades, which is the only grade range for which we possess a distinction of effect in randomized research by grade level (Hill et al., 2008). Thus, despite being small in magnitude based on Cohen's rule of thumb, the nature of the measures used in Table 5-1 render the effects more impressive than they would otherwise be. Of particular promise are the findings of interventions that used RfU-developed measures of reading comprehension: namely, READI, STARI, and WG. These results could be attributed as much to the combination of improved intervention techniques as to the improved measurement approach of the Global Integrated Scenario-Based Assessment (GISA) and RISE. Indeed, the STARI results stand out for being both significant and practically meaningful across almost all targeted constructs, with the exception of vocabulary. It is important to consider, also, that these two distal measures (GISA and RISE) were developed as part of the RfU effort focused on new comprehension assessments, and during the same time frame; this may have worked to better align the RfU curricular content and goals with the RfU assessments.

Knowledge and Learning

Another fascinating finding comes from the results for knowledge and learning. The Florida Center for Reading Research (FCRR) was the only consortium to use a published measure of knowledge (the Woodcock-Johnson III [WJ-III]), and results here were unsurprisingly nil to weak in strength and universally nonsignificant. By contrast,

[1] We deliberately avoided asterisks (***) because of their long history of association with levels of statistical significance. Our rule is the more bullets, the larger the effect size.

CALI, PACT, and READI (see Chapter 4) all found small to large effects for researcher-developed measures of knowledge, and TBL found a similar but weak effect. Notably, all these effects were statistically significant. Most impressive about these findings is that knowledge gains were attained despite the merging of reading instruction and content instruction in these interventions—students were reading to learn, as well as continuing to learn to read. That is, despite what might be interpreted as a division of attention in instruction (between the demands of learning in a content area or discipline, versus the demands of continuing to learn higher-order reading strategies, for example), gains were observed in both reading and knowledge acquisition for CALI in grade 4 and for PACT and READI for grades 8 and 9, respectively. These results suggest the merger of reading and disciplinary instruction can yield benefits for both domains of learning, and address the perennial concern of teachers who have had to choose between teaching one or the other. Even in the case of TBL and CALI in grades 3 and below, where reading comprehension was not significantly affected, results suggest that integrating reading and content instruction can boost learning rather than hinder it.

Taken together, these results suggest not so much that every teacher can be a teacher of reading, but that teachers in the disciplines can attend to and teach comprehension processes and practices without sacrificing the primacy of the knowledge acquisition goals within their disciplines. If we cannot help students use knowledge gained from reading, we are stuck with approaches in which we either do the reading for them—a common practice in middle and even high school (Wanzek & Vaughn, 2016)—or we tell them (most likely with PowerPoint-propelled lecture) what they might have learned had they actually read the chapter.

But naturally this inference about the efficacy of integrated/orchestrated multicomponent interventions must be tempered by pointing out the extensive professional development and support that undergirded these efforts. Examination of teacher preparation within consortia and individual studies illustrates the effort and resources necessary; PACT, READI, and Catalyzing Comprehension through Discussion and Debate (CCDD) provided preexperiment training, as well as in situ and in-process training as experiments ran their course. That such professional development will be available in reading comprehension instruction projects that seek to emulate the RfU approaches, without the RfU's rich levels of funding and expertise, remains to be seen. At the same time, we need to recognize that none of the designs across the consortium examined the possible mediating effects of gains in teacher knowledge (as a function of professional development) on student performance.

Measures of Components of Comprehension

Table 5-2 presents results for several constructs that serve as components of comprehension. Immediately apparent are the larger effects for researcher-designed measures than for published ones and for the lower versus upper grades. Overall, the RfU reaped the most impressive effects from researcher-developed measures of vocabulary, both in terms of consistent statistically significant results and effects sizes. In general, effects on researcher-developed measures tended to be statistically significant and strong, though less so from grade 4 onward. Comprehension monitoring and morphology also demonstrated substantial effects, though more so in the earlier grades.

Vocabulary. Note that despite the strong results observed for vocabulary with researcher-developed measures, similar effects were not widely replicated for more distal measures of vocabulary or comprehension, with the exception of CALI's results on the WJ-III vocabulary assessment in grade 4. These findings are consistent with previous work suggesting that while students acquire taught vocabulary very well, gains in taught vocabulary infrequently translate into gains on more global measures of vocabulary (or comprehension or learning for that matter).

That several of the consortia framed their vocabulary work in relation to disciplinary literacy (READI) and academic language (CCDD) might suggest that it is not reasonable to expect transfer to more distal indices of vocabulary. The point of much of this instruction is to acquire broader and deeper knowledge of words related to a particular topic (e.g., earthquakes) or a particular genre of discourse (e.g., causal explanation). There are surely long-term benefits to advances in these specific phenomena, but they may lie not in the domain of vocabulary acquisition but rather in the domain of applying these words and the concepts they represent to novel tasks, projects, or other forms of learning, much like Bransford and Schwartz's (1999) construct of transfer as preparation for future work.

Rather than continue to seek effects on distal vocabulary measures, future research might focus instead on the degree to which acquisition of targeted vocabulary mediates effects on other, more distal and applied measures, such as reading comprehension or knowledge acquisition. For example, when LARRC researchers modeled the mediated effect of the combined Let's Know! on reading comprehension via vocabulary (not reported in the summary tables), the effects were significant and quite large (LARRC, Jiang, & Logan, 2019).

Morphology. Finally, the universally significant and small to large effects observed for interventions targeting morphology suggest a new avenue for reading comprehension intervention. Morphological Awareness Training (MAT), DAWS, and STARI all targeted and assessed effects of morphological awareness intervention to some extent. The MAT intervention produced consistently significant and large effects on proximal measures of morphology but failed to demonstrate effects on any standardized tests of word recognition or comprehension. STARI saw not only notable effects on the morphological structures that were taught, but also small to large effects on more distal measures, including reading comprehension and word recognition. The role of morphology in reading development and instruction has experienced a renaissance of late, and these results suggest that attention is not misplaced.

Moderating and Mediating Effects

As our knowledge about reading comprehension has expanded, so too has our knowledge of the different influences on students' comprehension development, particularly increased understanding of the nature and impact of individual differences (Afflerbach, 2016; Connor, 2016). However, developing detailed accounts of the relationship between the characteristics of individuals and the differential efficacy of interventions (what we used to call aptitude by treatment interactions but now talk about as the moderating effect of student variables on the effectiveness of the intervention—e.g.,

the treatment was superior to the control only for students with low pretest knowledge scores) is an ongoing challenge. It is important to note that even though we refer to these as individual differences, they are often, if not chiefly, characteristics that individuals possess due to their membership in different groupings (based on prior achievement or knowledge, cognitive capacities, language preferences or competencies, disability, socioeconomic status, gender, culture/ethnicity/race, and the like). The hope in this endeavor is usually to be able to make claims about the categories of students for whom an intervention is especially appropriate.

Across the RfU initiative, teams took a variety of approaches to understanding what works for whom and under what conditions. The LARRC team delivered its interventions to all students and used pretest skills as covariates, but did not examine any interactions of LARRC with pretest skills. LARRC results were consistent in terms of both significance and magnitude of effects regardless of the inclusion of statistical controls for pretest skills, but importantly the inclusion of controls in the absence of interaction terms leaves the question of whether effects were moderated unanswered. More promising were the results of the LARRC follow-up study, which collapsed the two versions of Let's Know! and 2 years of data to examine whether vocabulary mediated an indirect effect on reading comprehension in grades 1–3. Such was indeed the case; moreover, the effect sizes for this mediated effect were quite large. Thus, despite not elucidating which groups of students benefit differentially from LK, LARRC demonstrated a fairly unprecedented effect of vocabulary learning on distal measures of reading comprehension.

Within the RfU initiative, nowhere has the quest for understanding the impact of individual differences been more central than in the work of Connor and her FCRR colleagues (2018). Examining students who all scored below the 48th percentile on a vocabulary measure, they determined that, in many cases, those students with weaker pretest skills benefited more from intervention than did students with stronger pretest skills when compared to business-as-usual (BAU) groups. Connor et al. (2018) suggest that interventions should be informed by individual student profiles and related needs. The complex interactions between reading instruction and individual differences led these FCRR researchers to call for "a more complete model of reading comprehension" that incorporates "reciprocating effects among text specific, linguistic, social, and cognitive factors, that interact with instruction" and may impact reading comprehension. Such a resource in this work is the lattice model of reading comprehension development (Connor, 2016), which provides particular affordances for conceptualizing students' reading development. The assumption of the model is that interactivity of reading skill components varies in a highly individualistic manner; however, that interactivity can be predicted if one knows the key characteristics of particular individuals and groups.

The FCRR intervention portfolio, perhaps in part because it included so many interventions and so many measures administered at both pretest and posttest, yielded a host of moderation effects, some of which survived when the multiple comparison correction was applied in the analysis (see Chapter 4). Dealing only with those moderation effects that remained after the correction, several are noteworthy. For Comprehension Monitoring and Providing Awareness of Story Structure (COMPASS), older students made more relative growth than younger students on narrative language skills, and students with lower pretest scores on listening comprehension exhibited more relative

growth on that same measure at posttest. For Language in Motion (LIM), the key moderating effect was a differentially negative effect on the Clinical Evaluation of Language Fundamentals measure of listening comprehension for LIM students (compared to BAU) who scored high at pretest on expressive vocabulary. 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. For MAT, post hoc exploratory analyses involving only MAT students suggested gains may have been moderated to some extent on pretest ability, but these results differed by grade and measure, making them difficult to interpret. For the Teaching Expository Text Structures (TEXTS) intervention, it proved especially effective, compared to BAU, for students with poorer academic knowledge at pretest. For Enacted Reading Comprehension (ERC), the one moderating effect demonstrated that, among students with lower expressive vocabulary at pretest (on the Expressive One Word Picture Vocabulary Test), ERC students scored higher on that same measure at posttest than did the BAU students. In the second DAWS efficacy study, students who performed more poorly on the editing pretest benefited more from DAWS relative to BAU students on both the editing and morphosyntactic knowledge posttests. For CALI, diametrically opposed moderation effects were found for comprehension growth in social studies versus science. In social studies, children with higher initial passage comprehension scores made relatively greater gains in CALI social studies than did children who had lower scores. However, this interaction effect reversed for science: among students with weaker pre-intervention passage comprehension scores, CALI students made greater gains (relative to BAU students) in science than did students with stronger scores.

The quest to find moderating and mediating effects was a key part of the analysis for all of the teams, as detailed in Chapter 4. But there were not as many among the three adolescent teams. In fact, no moderating effects were reported for READI. Both of the CCDD interventions and PACT revealed moderating and mediating effects, as detailed in Chapter 4, and READI examined mediating effects.[2] Within the CCDD portfolio, variables that could be conceptualized as implementation or engagement served as significant mediators of the effects of WG and STARI (Jones et al., 2019; Kim et al., 2016). In addition, in follow-up analyses of the large-scale efficacy trial, WG demonstrated favorable differential effects for students *currently* classified as English learners (ELs; i.e., current limited English-proficient students) in academic language skills and in social perspective taking, and, in the second year of implementation, ELs in the treatment condition grew more than their English-proficient counterparts in core academic language skills and social perspective articulation skills (Kim, Hsin, & Snow, 2018). 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. Within the PACT portfolio, the PACT intervention was remarkable in that the results were so consistent across student variables, such as learning disability (LD) designation or language status (EL versus English only). If an outcome measure revealed a PACT advantage over the BAU for the population

[2] WG and STARI results were positively mediated by levels of student engagement, and PACT outcomes in RCT_3 were mediated by the proportion of ELs within classes (see Chapter 4) and by fidelity of treatment in all three RCTs.

as a whole, then the advantage held for LD students and ELs as well. By contrast, both Comprehension Circuit Training (CCT) and TBL researchers found significant differences in subgroups of students in the larger treatment groups. For example, Fogarty et al. (2017) determined that effect sizes for CCT were generally stronger for students with lower reading comprehension skills at pretest. Also, Wanzek et al. (2014) found that students with high or moderate pretest scores benefited more from TBL (relative to BAU) than students with low pretest scores. In addition, Simmons et al. (2014) determined that students' comprehension gains attributed to treatment varied according to individual's reading comprehension achievement levels prior to the experiment. With the first iteration of CCT, there was a trend for the lowest-performing tier of students to benefit the most, in comparison to BAU students, on the Gates-MacGinitie Reading Test (GMRT); however, these same students tended to exhibit lower relative growth on a comprehension measure for a topically related proximal expository text. In the second iteration, post hoc analyses suggested marginally significant tendencies for students scoring the lowest on the GMRT at pretest to benefit most from the intervention as evidenced by sizable effects accompanied by relatively high *p*-values.

While it is not in the tradition of examining student-based moderators, it should be acknowledged that several teams chose to address specific population interests by going out of their way to situate their intervention in sites that would draw heavily upon samples whose interests are not always well served in American schools. Recall that for FCRR only students who scored below the 45th percentile (CE_1 or CE_2) on a relevant language or literacy measure qualified for participation; similarly, STARI limited participation to the students scoring below the 30th percentile on the state English language arts (ELA) examination. PACT's RCT_3 was placed intentionally in sites with high proportions of ELs. READI's sampling process for grade 9 science RCT guaranteed that they would be working in schools with many linguistic, ethnic, and racial minority students. Thus, while teams could not examine differential effects for different groups in these situations, they were optimally situated to determine whether the intervention proved efficacious for these often-underserved populations of learners.

Moderation Across the Entire RfU Portfolio

In the very broadest sense, the interaction between existing student characteristics and particular interventions was a complex story for the RfU. The dominant pattern for moderating effects is one of idiosyncrasy. First, many interventions revealed no stable moderating effects, implying that if an intervention worked, it worked equally well for a range of categories of student variables. Second, when examining the array of moderating effects that did surface, it was found that they vary dramatically by grade level, intervention, and outcome measure. For some groups (say, initially low-achieving students) in some grade-level groupings (say, pre-K and grade 2), a treatment was much more effective than BAU, but the interaction patterns did not hold for students in grade 1. This makes it hard to establish differential policy recommendations for particular populations of students, such as students with learning disabilities, ELs, primary grade students, or low (or high) achievers. One is left with one of two options—broad recommendations for all students or bringing the recommendations down to the level of

the individual student, as is the goal of initiatives like Connor's lattice model. It is not clear that the overall findings from the RfU, in and of themselves, can help us decide which of these models (or perhaps a hybrid model) should prevail.

The search for moderation has been a story of complexity and inconsistency. The promise, we think, resides in three pockets of possibility. First, there were, for several of the interventions, indications that initially low-performing students reaped more benefit from certain interventions—CCT for comprehension, CALI for science learning, ERC for expressive vocabulary, and TEXTS for academic knowledge. Second, interventions that were situated in sites with high proportions of potentially vulnerable learners (FCRR, PACT, READI, and STARI) demonstrated consistent advantages (relative to BAU) for the interventions. The third possibility lies in the future, when the field applies what was learned from both the successes and shortfalls of the various projects to new, revised, and refined pedagogical practices.

Mediation Across the Entire RfU Portfolio

Although mediation was examined far less often, the findings here are striking and overtly promising. First, teams like LARRC and FCRR provided evidence that gains in vocabulary and other components of comprehension can act as significant mediators of effects on comprehension—in the case of LARRC, rather large effects. Second, teams like CCDD, FCRR, and PACT provided evidence that indicators of implementation, dosage, learning, and student engagement with interventions can also act as significant mediators of effects on comprehension.

Together these findings suggest that reading comprehension may be most malleable when approached indirectly. In fact, the mediation findings to date suggest ripe avenues for continued analysis of the RfU data. The current state of results for the RfU instructional portfolio invites further scrutiny. As but one example, READI researchers showed significant effects not only for students, but also for teachers. An analysis begging to be conducted is whether teacher learning mediated effects of READI for students. More importantly, beyond the RfU, future investigations of reading comprehension instruction ought to plan and statistically power for analyses that can elucidate these indirect, but important, pathways by which comprehension can be improved.

Moving the Needle on Reading Comprehension

It is easy to look across the results presented in this and the previous chapter with a glass-half-empty perspective. The effects could have been stronger and significant results more plentiful and consistent across subgroups and outcome measures. But, we believe that aggregate RfU results can contribute to cautious optimism and guidance for future reading comprehension instruction. To abuse a hoary idiom, we would argue that a half-empty perspective misses the forest for the trees. Although many results were uneven and varied across multiple RCTs, some promising patterns emerge when we take a broader view of the collective work accomplished under the RfU. The RfU results suggest that carefully developed and orchestrated multicomponent (and intersectional if you will) instruction, when implemented with fidelity by teachers who are supported by robust professional development, can yield effects that are strong enough to move

the dial on reading comprehension and a host of related measures, such as vocabulary, knowledge acquisition, application, and many enabling skills. The hands on the dial may not move radically, but they most certainly have moved in a positive direction. With continued investment in coordinated, collaborative, and extended efforts like the RfU, the field of education is much more likely to witness significant progress in instruction and resultant reading comprehension.

EXAMINING THE PEDAGOGICAL FEATURES OF THE RFU PORTFOLIO

Having examined the empirical patterns of performance across sites, interventions, and measures, we turn now to a more conceptual analysis of the pedagogical practices themselves, trying to ferret out shared curricular and instructional features across this highly varied landscape of interventions. In a sense, this analysis is the logical complement of the previous account of statistically reliable effects; it answers the question, "What did we learn about the consistency of features of effective reading comprehension pedagogy?"

We have organized our analysis as a set of assertions about the legacy of the RfU portfolio of efforts to improve curriculum and instruction. Mostly they are claims about what we know now that we did not know before the RfU effort began. However, sometimes they are restatements of claims we could have made a decade or two ago, but can now make with greater confidence, nuance, or both.

We also note that, as we move into this new epistemological frame, we shift the standards of evidence and argument used to warrant our claims. In the previous section, when we traversed the landscape of effect sizes, the evidence to support our generalizations was the consistency of the direction of effects (treatment versus BAU) across interventions. In this section, as we traverse the landscape of common practices, our standard of evidence is not effect sizes, but more of a class inclusion standard: How frequently was a given feature or component associated with an effective intervention, one that outperformed the BAU control? It is not a standard that permits causal inferences, but it does suggest that if the preponderance of evidence points to a particular variable or feature, it is probably worth our attention and maybe our support. Given that important constraint, what follows is a set of claims that deserve our consideration—perhaps our support.

The Relationships Among Enabling Skills, Knowledge, Language, and Reading Comprehension Are Dynamic and Synergistic

When we consider antecedent strategies, skills, and dispositions for reading comprehension, we might ask, "What kind of comprehension?" The RfU research reminds us that listening comprehension generally precedes reading comprehension (LARRC, Arthur, & Davis, 2016), and that reading comprehension can be categorized, variously, as literal and low-level inferential (Connor et al., 2018), higher order (Kim et al., 2016), or discipline based (Goldman et al., 2019).

If students lack any prerequisite skills, strategies, or knowledge demanded by a particular text-task combination, reading comprehension instruction can and should help students develop and incorporate these into their reading. For example, a key premise

in LARRC research is that listening comprehension is a key pathway toward reading comprehension. Thus, a major thrust of LARRC research was investigating the role of language skills, strategies, and knowledge in the development of children's comprehension. LARRC's LK curriculum fostered pre-K students' vocabulary, comprehension monitoring, and language comprehension skills (Johanson & Arthur, 2016), contributing to young children's reading comprehension.

The RfU research that focused on middle and high school students (i.e., CCDD, PACT, and READI) provided a detailed catalog of strategies and skills, as well as different types of knowledge, that students must bring to acts of reading to comprehend increasingly challenging texts. Strategies helped students learn new content, decode academic language, and achieve higher-order comprehension, while content-area knowledge informed students' disciplinary and epistemic reading and related tasks. In school, texts are regularly used to introduce the new topics and concepts that comprise content-area and disciplinary knowledge (Vaughn, Roberts, et al., 2019). Goldman et al. (2016) noted that student success at comprehension in the upper grades is contingent on understanding unfamiliar content that is often embedded in complex language forms.

> While I've always valued the knowledge and experience my students bring to the classroom, I hadn't begun to think about how to leverage their everyday experiences with language, symbols, argument, and reading for the benefit of disciplinary learning in my classroom; the use of cultural data sets made clear how important it was to provide invitations for students to surface and build upon this knowledge. In supporting students to make explicit their understanding about symbols through [one text] and then providing opportunities for them to use this knowledge in an analysis of symbols in [two other texts], I was able to understand the critical role that cultural data sets played in helping students to bring their everyday interpretative understandings to bear on literature.
>
> *—RfU Participating Teacher*

Across the history of comprehension instruction and across content areas, there has been the common assumption that students can use their relevant prior knowledge to assist in the construction of meaning. But when content is new, students' strategy of using their prior knowledge to make inferences and connections, which may have served them well for texts about more familiar topics and situations, may fail (Fodor, 1975). One implication is that curriculum and instruction must attend to specifying, invoking, and, when needed, providing the most relevant declarative knowledge to allow students to bridge from what they know to what is new in the text (Pearson & Johnson, 1978).

A further need relates to academic language and the relation to complexity and challenge in comprehension (Kim et al., 2016). Students must understand how to read disciplinary texts—replete with diverse syntax and unfamiliar words—to be able to fully comprehend them. Students must also develop reading comprehension strategies that support and reflect higher-order thinking. For example, READI examined comprehension in disciplinary reading and determined that there are numerous, complex strategies—including analysis, integration, and critique—necessary for secondary students to succeed (Goldman et al., 2019). This work was based, in part, on the assumption that the more "basic" reading comprehension strategies, such as simple inferencing, are

operating and providing a foundation for more complex strategy use and comprehension at some later, more sophisticated, stage of the comprehension process.

In summary, the effectiveness of the RfU comprehension instruction is based in part on the determination of what students bring to the classroom—their antecedent knowledge, their incoming strategies and skills, and their commitment to doing well on the tasks set before them. This grounding offers opportunities to engage students just in time for curricular activities if particular knowledge and strategies are missing—or to bootstrap (use them as a stepping stone to more sophisticated instantiations) them when they are present but ineffective.

Many Kinds of Knowledge Play a Role in Reading Comprehension

Knowledge resides at the core of reading comprehension processes and products. The RfU research focused on the different types of knowledge that can be prerequisites for successful reading, results of successful reading, or both. We have long known that students understand what is new in a text by connecting to and building on what they already know, that is, by using relevant prior knowledge (Anderson & Pearson, 1984; Bartlett, 1932; Moll, Amanti, Neff, & González, 1992). In classrooms, this process involves activating (or when students do not possess it, providing) relevant prior knowledge to build those connections. To do so, teachers and curriculum have largely been focused on declarative knowledge, which, along with strategies and skills (Duke & Pearson, 2002; Pressley, 2001), enhance reading comprehension. Students must have the means for relating new information to existing information, and for making the many inferences that are central to the construction of meaning. The RfU research focused on this critical role of declarative knowledge. But the RfU went well beyond declarative knowledge (Goldman et al., 2019) to catalog additional types of knowledge involved in acts of student reading and learning in history, science, and literature: declarative, procedural, conditional, disciplinary, and epistemic.

Declarative Knowledge

It is commonplace to think of declarative knowledge as the preexisting foundation of comprehension; we understand what is new in terms of what we know (Anderson & Pearson, 1984), but more recent perspectives have also documented knowledge or, more accurately, increases in knowledge, as the consequence of comprehension. As indicated by an impressive array of effect sizes, gains in declarative knowledge were a resounding outcome in many RfU interventions, ranging from pre-K through high school. For example, researchers from the LARRC determined that the newly developed curriculum and instruction (LK), while ostensibly about language, also entailed gains in knowledge and had significant impact on young children's (pre-K and kindergarten) vocabulary learning (Johanson & Arthur, 2016; LARRC, Arthur, & Davis, 2016). Researchers from CCDD determined that the WG curriculum contributed to significant vocabulary growth, which may be little more than an alias for knowledge, for students in grades 4–7 (Jones et al., 2019). All three of the PACT interventions—PACT, TBL, and CCT, instruction of 11th graders that included team-based learning—led to increased social studies learning (Wanzek et al., 2014).

Procedural Knowledge

Students' ability to comprehend increasingly challenging text and to apply what is learned in increasingly challenging tasks is fostered by the teaching and learning of *procedural knowledge*—the how of reading comprehension. Procedural knowledge includes strategies for constructing meaning, monitoring the ongoing construction process (is what I just understood consistent with what I just read or what I know to be true about the world?), as well as strategies for using meaning constructed through reading to perform another task. STARI researchers found that middle school students receiving instruction that targeted procedural knowledge about how to engage in a range of strategies increased their achievements on several outcomes, including word recognition and decoding, vocabulary, morphological awareness, sentence processing, and basic reading comprehension (Kim et al., 2016).

PACT researchers determined that struggling middle school readers benefited from the CCT curriculum, which featured reading strategies as one of its key components, as they exhibited significant gains on reading comprehension tests (Fogarty et al., 2017). Related, Greenleaf and Valencia (2017) warned that student development of procedural and declarative knowledge is impeded by the simple fact that texts may be missing in content-area classrooms. Teachers' need to cover content, combined with the fact that some students' levels of reading development are not up to the task of comprehending disciplinary texts, results in classrooms in which teachers, via lecture and PowerPoint-guided discussions rather than text, are the main sources of information. A result is that students have restricted opportunity to develop declarative knowledge by applying the procedural knowledge they might be gaining through some form of strategy instruction or teacher-scaffolded encounters with text.

Conditional Knowledge

A third type of knowledge—*conditional knowledge*—is also featured in the RfU research. Much of conditional knowledge in reading relates to managing acts of reading: goal setting, monitoring meaning making, noting challenges, fixing problems, and comparing ongoing construction of meaning with the goals readers set for reading. The centrality of conditional knowledge to complex cognitive undertakings such as reading is widely recognized. However, the onset of children's metacognition and the related optimal initiation of metacognition instruction are debated. Research from across the RfU consortia reveals a clear focus on the development of conditional knowledge in support of reading comprehension. At the earliest levels of formal schooling, researchers from LARRC developed instruction that fostered comprehension monitoring in pre-K and kindergarten students (Johanson & Arthur, 2016). FCRR researchers developed the World Knowledge e-Book (WKeB) technology platform and curriculum that focused, in part, on promoting metacognition (Connor et al., 2019). The WKeB intervention led to students' enhanced word calibration—a key index of metacognitive monitoring—and improved students' reading comprehension performances. In addition, PACT researchers had middle schoolers ponder and repeatedly revisit framing questions, which prompted reflection and metacognition (Vaughn et al., 2013) as students worked through texts and related tasks. Finally, as conditional knowledge involves knowing when to use particular reading strategies, READI researchers (Goldman et al., 2019)

focused on helping students use disciplinary and epistemic lenses to determine when it is suitable (or advantageous, or acceptable) to adopt particular stances toward texts and tasks, and to use related strategies.

Disciplinary Knowledge

All of the RfU research revolves around students' acquisition and use of declarative and procedural knowledge, and several studies focused on conditional knowledge. However, the unprecedented contribution of the RfU research is to alert us to additional types of knowledge that contribute to students' reading comprehension success, the most prominent being *disciplinary knowledge*. READI and CCDD researchers engaged in deep dives into the disciplinary knowledge needed to understand, vet, critique, and use texts within the disciplines of science, history, social studies, and literature (Goldman et al., 2016, 2019; Kim et al., 2016). Disciplinary knowledge was also featured in PACT (Capin & Vaughn, 2017), though in a more embedded manner, in the tasks that students were asked to complete for recurring unit features, such as text-based knowledge acquisition, team-based learning, and team-based application.

Within each of the content areas that comprise disciplinary school learning are agreed-upon means of representing knowledge using specialized reading comprehension strategies, employing discourse practices and ways of explaining and arguing, and pursuing goals representative of the discipline. This disciplinary knowledge complements the declarative and procedural knowledge that is necessary for literal and inferential interpretation of text. Furthermore, it allows student readers to move from such literal levels to analytic and evaluative forms of reading comprehension (Shanahan, Fisher, & Frey, 2016).

READI colleagues (Goldman et al., 2016) proposed that reading comprehension requires both general reading strategies and strategies particular to specific disciplines (e.g., history, science, and literature). These specialized strategies focus on investigation of the nature of evidence that is used in arguments, the reasoning principles that undergird argumentation, the foci of claims, and the nature of disciplinary knowledge. CCDD also focused on disciplinary knowledge by engaging grades 4 and 5 students in the WG curriculum, an intervention program intended to build students' academic language (including both vocabulary and discourse), perspective taking, and ultimately their deep reading comprehension. Students made gains in perspective articulation and positioning skills in the second year of implementation of the WG curriculum (Jones et al., 2019), along with academic language and deep reading comprehension, although researchers cautioned that generalization of results was not warranted because of variability in implementation and duration of the WG curriculum.

> When I first started doing [historical inquiry], I noticed that students began with the idea that everything that's printed is true. Especially like textbooks are true. You know, if I asked that question on day one, [students] will say, "Yeah, everything in a textbook is true." Pretty much 100 percent of them will say that. And so, then I understood that part of my role was to move them from that to something that was a little bit more deep historical thinking than that.
>
> *—RfU Participating Teacher*

Epistemic Knowledge

Epistemic knowledge, a more recent focus for our understanding of reading comprehension (Alexander, 2012), involves students developing theories of knowledge about what we can know and how we come to know it. Its major functions in reading are to help readers frame tasks, decide on particular stances they will assume for different types of texts, and guide their use of declarative and procedural knowledge in carrying out tasks (Lee, Goldman, Levine, & Magliano, 2016). READI colleagues (Goldman, 2018) noted the centrality and power of epistemology in disciplinary inquiry, as it provides students with both purpose and motivation for reading. The READI curriculum included an overall focus on epistemic knowledge, how it develops, and how it evolves to reflect readers' growth. Working in the history discipline, Shanahan, Fisher, and Frey (2016) noted that students find their encounters with epistemic knowledge challenging because it forces changes in what are often well-established student schemas. An example is changing students' conceptualization of history from a "basket of facts" to be memorized for a test to one that requires inquiry, interpretation, judgments about relevance and trustworthiness, and, ultimately, the production of an argument (Shanahan et al., 2016). Accordingly, disrupting students' notions of "what history is" was accomplished by presenting them with accounts of the past that were incompatible with one another and requiring them to reconcile both the texts' content and the students' underlying assumptions about how knowledge is constructed. In CCDD's WG, students were required to engage in "perspective taking"—learning and using skills relevant to "reading" the world—as is required in comprehending social discourse or interpreting characters' or authors' intentions. In summary, and in aggregate, research results from the RfU consortia serve to expand our view of the knowledge that students must possess to comprehend successfully, as well as the knowledge that results from successful comprehension.

Learning to Read and Reading to Learn Are Better Regarded as Complementary Processes Than Separate Stages of Development

The aggregate research from the RfU teams allowed for examination of the proposition that students generally progress from learning to read to reading to learn (Chall, 1996). Recognizing that there are no clearly drawn boundaries between learning to read and reading to learn, it is commonplace in characterizing stages of reading development to assert that first you learn to read and then you read to learn. In fact, it is well-nigh canonized in Jeanne Chall's (1983) classic stage theory. The RfU initiative challenged that assumption by showing us that even our youngest readers can successfully read to learn while they are still learning to read, and middle school and high school readers are still learning about reading when most of the reading they do is in the service of reading to learn.

The two "early" sites, LARRC and FCRR, provided us with compelling evidence that young readers acquire considerable vocabulary (LARRC's LK^D and LK^B) and declarative knowledge on various topics (FCRR's CALI) even as they are still in the business of learning foundational skills of phonemic awareness, decoding, and fluency. We would also point out that even though the LK curriculum is organized around improving language skills (vocabulary, text structure, and story grammar elements)

and comprehension monitoring, the instructional activities are bonded to topical units that also provide an opportunity for students to acquire topical knowledge about the world around them, especially in the expository text units.

An ongoing challenge in many classrooms—especially content-area classrooms—is achieving appropriate balance between what are most often competing goals: students' acquisition of content-area knowledge versus their continued learning of reading comprehension strategies, skills, and stances. This challenge increases as students move from the upper elementary grades to middle school and then high school, as opportunities for dedicated reading comprehension instruction often diminish. Also, students who are not reading at expected levels for a particular grade level will face difficulties constructing meaning, regardless of content area, because they lack the strategic infrastructure to persevere in the face of weak knowledge of the topic at hand.

The three secondary RfU teams (CCDD, PACT, and READI) addressed the challenge of reading to learn while learning to read with multiple projects. CCDD researchers developed STARI with the intent to involve middle school students who scored below proficient on a statewide ELA assessment with challenging texts and tasks at the very same time as they further developed more foundational decoding and fluency strategies and skills. As the STARI results in Chapter 4 document, it worked well, with students in the STARI condition outperforming BAU students on growth in word recognition, morphological awareness, and efficiency of basic reading comprehension (Kim et al., 2016). The CCDD WG program was designed so that discipline-based curricular materials in language arts, science, social studies, and math were geared to students' level of development, with the goal of rendering disciplinary reading and thinking tangible, and to engage all students in related discussion and debate on controversial but accessible topics. If STARI focused on bringing the foundational skills along with basic level comprehension (of the ilk measured by RISE), WG focused on more advanced "learning to read" processes, such as those involved in academic discourse (including vocabulary), critiquing and constructing arguments, and taking multiple perspectives on text interpretation (Kim et al., 2018).

The PACT intervention in American history demonstrated that students gained considerable knowledge about the content in their modules while improving performance on proximal measures of comprehension and sometimes but not consistently on distal comprehension measures (Vaughn et al., 2013, 2015, 2017). Fogarty and colleagues (2014) found that Comprehension Tools for Teachers could provide instruction targeted at two related, often incompatible goals: building foundational reading skills (e.g., word identification, vocabulary knowledge, and reading fluency) and boosting reading comprehension achievement. The major thrust of READI was to understand and improve the advanced reading skills of disciplinary literacy. These are the strategies and skills needed when the role of reading shifts from getting the author's message to evaluating the relevance and trustworthiness of authorial claims on the pathway to distilling nuggets of information and perspective to use in the service of evidence-based argumentation or other more application-oriented tasks.

Comprehension Can Be Conceptualized as a Waypoint, or the End State, of a Journey

Recent conceptualizations of reading (e.g., National Assessment of Education Progress [NAEP], RAND) emphasize not only the construction of meaning (the waypoint), but also readers' subsequent use of that meaning (the end state). For example, NAEP defines reading as "an active and complex process" that includes "using meaning as appropriate to type of text, purpose, and situation." Similarly, recent influential initiatives such as the Common Core State Standards for ELA cast basic comprehension as a benchmark on the pathway to higher-order thinking that can include evaluating, analyzing, comparing, and synthesizing. Successful reading requires comprehension, and successful comprehension facilitates student engagement with real-world tasks. In this view, reading is not complete until the information and insight resulting from the act(s) of comprehension are redeployed to engage in one or more of these applications.

Several of the RfU research projects were well aligned with this conceptualization of reading—that students' reading development is indexed by both what is comprehended and what students do with the fruits of that comprehension. Among the RfU teams, READI research focused, in part, on text comprehension as a prerequisite for—and complement to—learning in the disciplines (Goldman et al., 2019). READI researchers developed Disciplinary Core Constructs consisting of five categories, or types of knowledge, that members of particular disciplines use during inquiry and argument. While these core constructs extend across all disciplines, they are customized within particular disciplines. For example, a major thrust in the READI work (Goldman et al., 2016) was designing instructional units intended, in part, to engage students in evidence-based argumentation. The primary "stuff" of this disciplinary argumentation is information and insight gained in acts of comprehension within and across individual texts, but almost always integrated with information gathered through other media as well as knowledge that students bring to their initial encounters with text.

Likewise, CCDD developed the WG curriculum that required students to both comprehend and evaluate—and ultimately construct—arguments (Kim et al., 2016). Students also learned to debate ideas they had initially comprehended via text. A related finding was that while comprehension was a prerequisite for engaging in discussion and informed debate, students' fundamental comprehension of the ideas initially encountered in text almost inevitably evolves as a result of engaging in these subsequent interactions and applications. This is a dynamic view of comprehension, one in which it becomes interwoven with and nearly inseparable from learning.

PACT researchers also focused on students' application of knowledge gained through comprehending text. In fact, the final activity in the PACT intervention cycle for each of its modules is an explicit application activity implemented in small project groups in which the ideas originally encountered in texts are transformed in the service of completing the project. For example, in the colonialism module, students prepared a written tract to entice immigrants to settle in a particular colony. In CCT (Fogarty et al., 2017), the knowledge flex "station" required students to work in teams to synthesize information from recently read texts. In the TBL intervention (Wanzek et al, 2014), grade 11 students used routines that included engaging in dialogue about course content, application of content to solve problems, and use of evidence to support responses to

comprehension and explanation prompts. Even in one of the primary interventions, FCRR's CALI, primary grade students used what they had understood from the texts they read in what they called research lessons, which involved writing activities in social studies and science.

Implicit if not explicit in this family of interventions is an emerging expectation for the field: The job of comprehension may not be complete until the insights and information gleaned from it are put to work in the service of some other process, goal, or product. It is almost as though comprehension has assumed a new, more enabling, role in the learning process. Paris (2005), in his description of constrained and unconstrained skills, conceptualized just such a role for foundational skills like phonemic awareness, decoding accuracy and automaticity, and fluency; they are enabling skills on the pathway to comprehension. Their value was fostering the more worthy goal of text understanding. In this new vision for comprehension (Anderson, 2018), comprehension may have assumed just such an enabling role. The job of comprehension is not complete until some significant action occurs—a story is told, a phenomenon is explained, an argument is constructed, a bias is unearthed and laid bare, a text is composed, or a product is created.

Metacognitive Processes Play a Role in the Comprehension Instruction Repertoire

Students who successfully employ reading strategies and skills (routines that help you build a text base and a situation model) also depend on metacognitive resources in order to initiate, work through, and complete acts of reading (Vaughn, Martinez, et al., 2019; Veenman, van Hout-Wolters, & Afflerbach, 2006), mainly in order to assure themselves that the models they have built are valid (or that they stand in need of revision). Despite an extensive portfolio of research and theory documenting its importance, metacognition (a salient form of conditional knowledge) has not been a consistent focus of comprehension instruction. The RfU research is notable for its attention to metacognition as both an important learning *outcome* of comprehension instruction, and as an *influence* on comprehension performance—a mediator of comprehension that operated across the developmental continuum from novice early readers to sophisticated adolescent readers.

At the early end of the continuum, LARRC scholars incorporated comprehension monitoring as a key component in the LK curriculum (LARRC, Farquharson, & Murphy, 2016). In LK, comprehension monitoring is co-equal with two other key components—young children's vocabulary and language comprehension skills—and LK instruction produced gains in both students' vocabulary and comprehension monitoring. Furthermore, Johanson and Arthur (2016) determined that comprehension monitoring instruction contributed to both vocabulary and language comprehension development. Also working with early readers, FCRR researchers examined the role that metacognition and comprehension monitoring played in students' overall comprehension development. For example, FCRR researchers (Connor et al., 2018) developed COMPASS, which was used with students in pre-K through grade 3. Connor et al. (2019) also marshalled the benefits of using technology (WKeB platform and curriculum) to promote metacognition. Technology allowed for consistent metacognitive prompting of students while they read, and the use of game rules that prompted and fostered student attention to the reading task.

Turning to adolescents, PACT scholars included metacognition in their fundamental definition of comprehension, operationalizing reading for understanding as acquiring knowledge (vocabulary and concepts) from text, and *monitoring understanding using text structure as a standard for evaluating progress*. Simmons et al. (2014), in the initial studies of CCT, focused on curriculum that supported student-regulated metacognitive strategies to monitor and repair misunderstanding. Researchers were also interested in building, through metacognition, students' independent ability to *activate* prior knowledge, *adjust* cognitive processes, *make* inferences, and *integrate* information in text (Simmons et al., 2014)—all key facets of metacognitive activity. Researchers determined that metacognition instruction had beneficial influence, especially for students who had already demonstrated competence in reading (i.e., higher standardized reading test scores). In addition, CCT documented the benefits of learning about and practicing comprehension monitoring and fix-up strategies (Fogarty et al., 2017). Wanzek et al. (2014) used team-based learning to encourage students to establish habits of accountability across team members. In this intervention, the metacognitive monitoring component focused less on text understanding and more on students' ability to self-monitor and self-evaluate key principles and practices of the team discussions.

Across pre-K through high school, the RfU research demonstrated that curriculum and instruction that includes a metacognitive component—including comprehension monitoring, self-regulation, and word calibration—boosted student performance on metacognitive tasks and, more importantly, elicited transfer effects on measures of language comprehension, reading comprehension, and vocabulary development. Finally, that instructional inroads were made for metacognition in the early grades represents a fairly new frontier for the metacognitive reading curriculum. Theory and related instruction are unsettled as to predictable onsets of metacognitive ability in young children, diminishing early curricular attention to this vital aspect of reading development. That the RfU teams implemented metacognitive instruction early on and continued investigation of different aspects of metacognition throughout the course of pre-K through grade 12 school reading is notable, as are the related student development and noted contributions of metacognition to student growth in reading achievement.

Collaboration Is Often a Key Element of Effective Interventions

Historically, both basic and applied research on comprehension development, and reading development more generally, has assumed that most comprehension action takes place "behind the eyes and between the ears" (McDermott & Varenne, 1995). Learning to read, and continuing to read throughout the school years with the attendant strategies, skills, and stances, has been typically conceptualized as a solitary undertaking (see Pearson & Cervetti, 2015; RRSG, 2002); in this individualistic paradigm, students learn and apply reading knowledge to become better readers. By contrast—and especially after the rediscovery of Vygotsky's (1978) more socially grounded views of mind, language, and learning and the beginning of the social turn in reading (Pearson & Cervetti, 2015)—there is increased interest in the social and collaborative contexts of schooling in which reading development is nurtured. The question is, to what degree do these social supports provide benefits for students' comprehension development and academic learning?

The RfU research portfolio includes many instances of collaborative learning, each of which was included as a part of a larger, multicomponent program to enhance comprehension and learning. Connor et al. (2019) provided a telling example of the way the social face of instruction is entwined with other facets of instruction designed to promote cognitive or linguistic growth. They examined the impact of a program that wedded students' word calibration (a metacognitive task and ability), WkeB technology, and book club participation (the social face) on their both proximal measures (word calibration, strategy use, and word knowledge) and distal outcomes (standardized test scores). Significant effects for the curriculum package surfaced on the proximal measures of word knowledge, word knowledge calibration, and strategy use; these, in turn, predicted student performance on the more distal standardized reading comprehension and vocabulary measures. Most relevant to this discussion, the positive effects were greater for students in weekly book clubs; social interaction benefited performance on the distal outcomes.

The CCDD programs, Word Generation and STARI, don't actually teach reading comprehension—they introduce topics and issues sufficiently motivating and complex that students engaging with them think, argue, read, and write at high levels. The programs provide the curricular resources and supports that enable students to learn to comprehend while thinking, arguing, reading, and writing. The key support is teacher facilitation of peer discussion, during which critical thinking is modeled and promoted in socially and cognitive scaffolded ways. Learning to read with Word Generation and STARI takes reading comprehension off the list of skills to be mastered and puts it back where it belongs—at the center of learning, analyzing, and engaging in civil discourse.

—Catherine Snow, Steering Committee Representative from CCDD

The impact of social aspects of learning is also present, to varying degree, in the work of the three adolescent RfU teams—CCDD, PACT, and READI. STARI (Kim et al., 2016) was built with social interaction as a core design principle explicitly to promote social interactions that foster student engagement, which contribute to cognitive growth. STARI used four types of peer collaboration: partner-assisted fluency practice, reciprocal teaching of comprehension strategies, partner reading and responding to novels and nonfiction texts, and peer debate, in which teams gathered text evidence and built arguments. The theory of action in STARI was that these collaborations, in which partners work together on meaning construction, would help move readers—especially struggling readers—beyond literal and limited responses to text. A hierarchical regression analysis indicated that engagement, including engagement in collaborative groups, was a malleable factor that contributed to gains in multiple dimensions of reading skill for STARI students.

The PACT team found that TBL routines—including dialogue about course content, application of content to solve problems, and the use of evidence to support responses (Wanzek et al., 2014)—produced reliable effects on measures of content area knowledge acquisition, especially for students who began the intervention with medium and high scores on a distal reading comprehension measure. In addition, Vaughn et al. (2013) included TBL as a feature of the PACT intervention in grade 8 American history;

students conducted collaborative comprehension checks for reading comprehension and social studies knowledge, which also influenced individual and team accountability. This collaboration was intended not only to inform learners' construction of meaning, but also to help student teams apply the knowledge gained from reading. That said, the research and statistical design did not permit direct inferences about the specific TBL activities.

Many of the RfU projects were intended to disrupt traditional practice—where the teacher does the interpretive knowledge construction work and hands it to the students in the form of a lecture—by replacing it with a more active and responsible role for students. Accordingly, the READI curriculum was designed with the expectation that students would be active—constructing knowledge through thinking, reasoning, and questioning. These activities were supported by specific student participation structures and instructional routines. A related challenge is providing students with the resources and tools they need to make it possible for them to meet these higher expectations, and this was achieved through teachers' support of students. Such support focused on different aspects of students' comprehension and learning.

Brown and Shanahan (2017) examined teacher support in science classrooms, in relation to disciplinary literacy practices. Support was intended to boost students' opportunity to learn and teachers did so through strategies of orchestrating, demonstrating, and assessing. Teacher mediation was examined using field notes and video recordings. Analysis led to detailed descriptions of how teachers supported student engagement in science reading practices. Furthermore, teachers provided flexible supports for students who were facing the challenges related to learning to read science texts and learning to justify and critique science models. Additional support focused on students' epistemic development and the fact that many students do not have appropriate schema for innovative instruction and curriculum within the disciplines. As students read, debated, interrogated, and sourced texts in history class, teachers reminded students that texts might be oppositional, that understandings of history might be unresolved, and that constructing meaning might be challenging. These verbalizations helped students better understand the specific culture of each discipline and the novel nature of learning within the disciplines.

> I have been teaching for 25 years. Only after using Word Generation in my classroom did I realize how badly I had been underestimating my students all those years.
>
> *—RfU Participating Teacher*

Complementing the empirical evidence accumulated through the RfU studies, and in relation to READI research initiatives, Greenleaf and Valencia (2017) posited that promoting engaged academic literacy involves supporting collaborative meaning making through text-based discussions. This requires that teachers orient students away from teacher-dominated question-and-answer sessions and toward fruitful discussions with fellow students. It also demands that students have discussion tasks that are grounded in the material learned from texts. In addition to students' collaborative efforts, another facet of "social" interventions centers on the teacher's role as a part of learning in groups. Teacher scaffolding and built-in curricular support were apparent

across the RfU research projects. But teacher scaffolding, as Greenleaf and Valencia (2017) point out, can be a double-edged sword, when teachers supplant any need for students to actually read the text by either reading it for the students or telling them in a lecture or PowerPoint what they might have learned had they actually read the text. The boundaries between knowledge that is enabling of students' comprehension and is provided by teachers or ancillary materials, and knowledge that students might take responsibility for acquiring independently should be delineated.

Going forward, specificity regarding who is learning, what is learned, and how it is learned within collaborative environments will help researchers tease out how, when, and with whom these collaborative activities promote individual student participation and performance.

Engagement with Texts and Tasks Supports Comprehension

Students' motivation and engagement influence reading comprehension (Guthrie & Klauda, 2014). Attention to the role of motivation and engagement in reading development and reading comprehension is relatively recent, and the RfU consortia made strides in examining specific effects and interactions involving motivation and engagement, and reading comprehension. Moreover, several RfU research projects positioned motivation and engagement as potentially potent and malleable variables in acts of reading comprehension. The three adolescent teams (CCDD, PACT, and READI) created curricula that used student engagement as a touchstone, from the start to finish of individual lessons and for series of lessons. These curricula positioned engagement prompts throughout their modules and instructional routines—at the beginning, in the midst of, and toward completion of units of instruction. Essential questions or problem statements provided a clear purpose for reading a text or texts. Furthermore, as students encountered new content in text, engagement was promoted through emphasis on the relationship of new knowledge to students' lives—their existing, experiential knowledge. Finally, engagement was maintained as students worked in personally meaningful reading-related tasks and activities.

> It can work for anyone. The naysayers who say kids can't discuss or have discourse at this level should see my class. I had groups that worked better than EVER!!!! They argued, debated, proved their points.
>
> *—RfU Participating Teachers*

From the STARI results in Chapter 4, we know student performance on more proximal-like RISE outcomes (morphological awareness, word recognition, and reading comprehension efficiency) were mediated by both behavioral (percentage of workbook pages completed) and perceptual (teacher judgments about student engagement in the curriculum, using the Reading Engagement Index–Revised; Wigfield et al., 2008) indicators of engagement. STARI also featured a system that sought to match content-area texts with students' current reading achievement levels, with the intention of building student self-efficacy.

Attending to students' engagement and motivation also featured within the Social Studies Generation (SoGen) program offshoot of WG. Duhaylongsod, Snow, Selman, and Donovan (2015) describe design principles for SoGen that focused on curriculum

comprised of engaging topics and materials, and instruction geared to a student's specific level of reading development, in order to render disciplinary reading and thinking accessible. The researchers concluded that the SoGen curriculum facilitated student engagement with high-interest topics that had a degree of relevance to students' lives, especially when combined with classroom discussions and debates—activities that further accelerated that engagement with texts.

Vaughn et al. (2013) developed the PACT program wherein each unit included a motivational springboard and opportunities for students to access relevant background knowledge so that student engagement might be optimized. Researchers also built in group discussions and collaborative work, as these have been shown to positively influence motivation and engagement. While the study reported significant experimental treatment effects for students' content-area learning and reading comprehension development, the design did not permit an assessment of the independent influence of motivation on student performance.

Motivation was often enmeshed with other factors in the READI work. Goldman et al. (2016), for example, noted that the READI approach views "epistemology as central, providing purpose and motivation to the ways in which inquiry is conducted" (p. 6). However, there were no direct measures of motivation and engagement, nor were there analyses of the influence of motivation and engagement on student performance. In a more qualitative vein, Brown and Greenleaf (2017) used field notes and video recordings to determine how teachers supported student engagement in science reading practices. Texts in this study were sequenced with the intention of building and maintaining engagement, while inquiry questions were designed to encourage student engagement with scientific inquiry.

LESSONS LEARNED

We are eschewing the "usual suspects" framework for a discussion section of a research report (summary, limitations, and future directions) in favor of a two-part approach—a section labeled *lessons learned* followed by a very brief summary that serves as a coda for the RfU's portfolio of curriculum and instruction research. The *lessons learned* section combines limitations and future directions by looking back and forward in a single scan of the landscape. We hope the points we stress are a forward-looking set of reflections about what might have been done "if we knew then what we know now"—a sort of Monday morning quarterbacking. And the instant one utters something that sounds like a limitation, it also gains entry to our collective wish list for where we hope the field looks in the future for the next big boost in the phenomenon under study—to wit, reading comprehension pedagogy. This account is offered in the spirit of how the good might be rendered even better, and with the assumption (which we believe is real) that the best legacy for any research initiative—big or small—lies in the grist for creative and critical thinking it leaves for others to build on. So in that spirit, we offer a small set of observations. Other suggestions (incorporating a greater emphasis on digital text and reading or multiliteracies, for example) appear in Chapter 6 because they pertain not just to pedagogy but to the entire reading comprehension enterprise. But here are the most salient that have captured our attention in reading across the pedagogy research discussed in Chapters 4 and 5.

Our research needs to be laser focused on diversity, especially on the welfare of emergent bilingual learners. While there are gaps in the research base regarding the instruction of all underserved, often minority, and almost always low-income students, that gap is especially apparent for ELs (what we now more accurately refer to as emergent bilingual learners [EBs]). To the credit of the RfU initiative, it did reinforce and extend our understanding of the complex and dynamic nature of language competencies (LARRC, FCRR, and CCDD) as well as our understanding of the relationships between oral language and comprehension (FCRR, CCDD, and READI). Some interventions explicitly targeted EBs, most notably PACT's RCT_3; PACT researchers even added a unit on pedagogical tools uniquely suited to the needs of EBs and to the professional development curriculum for the teachers in RCT_3. Others, as we suggested earlier, often included high proportions of EBs by virtue of the sites in which they placed their studies.

Thinking ahead to the next generation of research on comprehension instruction, we would be remiss not to pay more attention to EBs. The increasing numbers of students, across the world, who are learning through a language other than their "mother tongue" has spurred interest in issues related to language in *all* classrooms (e.g., Beacco et al., 2015; Lucas, Villegas, & Freedson-Gonzalez, 2008). In the United States, for example, there has been dramatic growth in the numbers of students who come to school speaking a language other than English. Between 1990 and 2010, the population of ELs in the United States increased by 80 percent and ELs now represent 10 percent of student enrollment (Valdés & Catellón, 2011). This trend is characteristic across the United States and not just of coastal or border states, with states such as Indiana, North and South Carolina, and Tennessee each realizing a 300 percent increase in the population of ELs between 1995 and 2005.

Data regarding current academic achievement levels of EBs are troubling. For example, NAEP results from 2009 indicate that in California and New York only a small proportion of ELs were able to achieve at or above the basic level in reading in grade 4 (25 and 29 percent, respectively; Samson & Collins, 2012).

In the United States, the vast majority (80 percent) of ELs speak Spanish as a home language. Confounding any consideration of the appropriate education of EBs is the fact that newly arrived immigrants from Spanish-speaking countries are typically coming from lower economic and educational backgrounds. For example, nearly 24 percent of immigrants from Central America and Mexico have family incomes below the poverty line, compared with 9 to 14 percent of immigrants from other areas of the world, and 11.5 percent of the native-born population.

EBs are triply at risk. First, their comparatively low scores on traditional achievement measures are painfully apparent. Second, these poor educational outcomes are accompanied by two significant challenges, language and socioeconomic status; compared to middle class and affluent English speakers, they have a lot more work to do to achieve even at a basic level. Third, knowing what we know about the maldistribution of resources and expertise (Darling-Hammond, 2019; Wilburn, Cramer, & Walton, 2019), EBs are even more at risk because they are often denied access to the "good stuff" in curriculum, which is more likely to be reserved for more affluent mainstream learners. Ironically, this disparity is exacerbated by a "first things first" disposition among policy makers and educators—a well-meaning attempt to make sure that EBs are well grounded in the basics of reading and writing before they get to the more interpretive,

critical, and creative facets of the ELA curriculum. Many EBs spend their entire school careers "catching up" with these foundational skills and never get to the "good stuff." What does this mean for those of us who are trying to improve access to literacy and learning for this population in particular? First, it means that we need to ensure that the samples of schools and students with whom we do **all** of our work (whether it focuses on pedagogy, assessment, development, or even theories of basic processes) on reading comprehension include proportions of EBs that reflect their distribution in the broader population. We cannot afford theories, tests, or instructional tools that are based on evidence gathered from any narrow demographic category, especially mainstream language-majority learners. Second, it means that we should make sure that the pedagogical interventions we develop are as much informed by what we already know, as a field, about approaches that are responsive to the needs and assets that EBs bring to the classroom. Third, in addition to statistical analyses that use demographic variables as covariate control variables, we should, wherever feasible, conduct secondary analyses that can tell us whether an intervention, or even key features of an intervention, are particularly helpful for EBs. Granted, we have substantial evidence that approaches that work well for one group also work for other groups; even so, we should, as a matter of course, be on the lookout for interactions between interventions and student characteristics.

We need to describe and measure BAU instruction as diligently as we describe and measure instruction in our interventions. Reading for Understanding was intended to produce positive change in teaching and learning reading comprehension. Necessarily, this required a change of the status quo. In many of the reviewed RfU studies, this status quo is referred to as "business as usual," or BAU. We interpret this phrase as meaning "reading comprehension instruction as it has been," or "as it is" in control classrooms. While BAU is a handy and widely used referent, it implies a sameness of curriculum and instruction across BAU classrooms that is probably inaccurate—and this assumption of "sameness" in BAU classrooms can lead to difficulties in interpreting research findings. First, lack of detail about control classrooms can diminish researchers' ability to accurately interpret results—the significant and insignificant findings, the interactions, and the site-specific features and anomalies that, if known, could add greater precision to the research narratives we employ to interpret findings and implications. In effect, if we do not move beyond the BAU label to more detailed knowledge of control classrooms, we may inhibit the ability to interpret results. Second, using BAU to label control classrooms and groups prevents the determination of the suitability of measures used by researchers in treatment and control classrooms. With no sense of the constructs guiding reading comprehension instruction, nor the curricular focus in BAU classrooms, assessments cannot be gauged for their construct validity vis-à-vis control (i.e., "business-as-usual") classrooms, or for their instructional sensitivity. This is especially so when treatment-control comparisons revolve around proximal measures that are especially shaped to be sensitive to the very features present in the treatment. As a result, we are not in a position to evaluate the opportunity cost of an intervention.

To the credit of the RfU community, many projects did describe the instruction in the BAU as carefully as the intervention. And many projects also evaluated plausible

opportunity costs; we recall several conclusions of the ilk, "students gained greater knowledge of the topic under study with no appreciable loss in their reading comprehension acumen." This advice is especially important when we encourage educators to do something out of the ordinary, such as offer a more challenging curriculum to more vulnerable or lower-achieving students. In those situations, it is incumbent on us to demonstrate that any increase in higher-order reasoning they accrue from the treatment does not come at a cost to more foundation skills, strategies, or dispositions. And the converse is also true: when the treatment emphasis is on foundational skills, we need to demonstrate that there is no opportunity cost for higher-order skill development.

These concerns are even more important when the intervention involves component practices that may already be operative, sometimes even prevalent, in ordinary classrooms. Collaboration offers a good case in point, precisely because it was a common feature of successful interventions. What we do not know, unless we measure it, is how common it was in BAU classrooms. We have made a lot of progress in measuring teaching practices, via surveys, observations (Pianta, La Paro, & Hamre, 2008), and teacher activity logs (Rowan & Correnti, 2009), so it seems wise, in pedagogical studies, even when we would rather adopt the causal inference affordances of an intent-to-treat approach, to know what was really going on in the BAU classrooms.

Going forward, we recommend that researchers make efforts to describe the instruction that students receive in control groups and classrooms, beyond "business as usual." This helps both researcher and research audience best interpret findings, accept or challenge these findings and interpretations, and compare innovative reading comprehension instruction in relation to more traditional or habitual instruction.

We need to find ways of better embedding engagement and motivation, as *inputs* (malleable factors), *outcomes* (measuring the constructs), and *mediators* (catalysts for accelerating comprehension and learning outcomes). The RfU research described in detail the workings of reading comprehension and successful reading comprehension instruction. Going forward, we need to pay more attention to conative and affective factors that are, variously, precursors of, influences on, and outcomes of improved reading comprehension. This requires identifying the affective and conative "surrounds" that operate during students' reading comprehension development and designing studies that focus, in part, on conation and affect as both supporting and resulting from reading comprehension. Consider motivation in relation to reading. Prior to reading, motivation can lead a student toward, or away from, engaged reading. This motivation is the result of students' prior experiences (and successes and failures) with acts of reading. During reading, student motivation may increase, decrease, or remain in steady state. This ebb and flow of motivation is influenced, in part, by the student's ongoing performance, along with feedback from the teacher and self-monitoring of the cognitive and affective facets of the reading act. Following reading, a reader will include an account of the just-completed reading in something like a mental diary of reading experiences. Research that continues to chart and explicate the relationships of reading comprehension development and achievement in relation to student conation and affect, consistent with the READI investigation of literature learning (Lee et al., 2016), will help the field better understand this sort of situated cognition. This could well lead to interventions that keep their eye on the prize of cognitive gain for students as they

enlist student motivation or self-efficacy in the effort. We are not likely to learn more about the role of this cluster of factors if we do not systematically attempt to examine, change, and measure them.

Successful reading experiences help students maintain long-term motivation and positive affect. Negative experiences reinforce lack of motivation and poor self-efficacy. These are fairly predictable outcomes for many student readers. Most important, particular acts of reading—where a student demonstrates learning by accomplishing what could not be done earlier—can be transformative. The student who lacks self-esteem, viewing himself as a poor reader, and who then actually learns and excels in a particular episode of reading comprehension has gained not only in relation to reading achievement, but also in relation to regarding the self as a reader. Instructional features and classroom contexts that support this development should be a feature of future studies. As indicated in many studies in the RFU repertoire, success in reading and establishing comprehension is not a solely cognitive story.

We need to expand the role of critique in our comprehension interventions. A telling finding from PACT, unearthed by Wancek and Vaughn (2016) in an analysis of treatment fidelity, was that teachers were much more likely to implement the more basic elements of the PACT intervention (building background knowledge within the comprehension canopy and teaching essential words) than the higher-order and critical elements involved in text discussions and knowledge application. For WG, LaRusso, Donovan, and Snow (2016) found that the biggest challenge for teachers was finding time for the critical reading, debate, and argument generation activities of WG in a system with so much competition coming from pressures to "cover" the required school curriculum and to prepare students to take the state test. That said, it is clear that engaging students in one form or another of critical thinking was an essential part of the work of the three adolescent teams (CCDD, PACT, and READI), and there are traces of it in LK (the comprehension monitoring activities require students to determine what is puzzling about a text and how to fix it). More specifically, there are examples of both the internal (to the text) stance of critical reading in the liberal humanist tradition (How good an argument did the author make for the impact of greenhouse gases?) or the more external (to the text) critique coming from critical literacy approaches (What ideologies and assumptions about government are inscribed into the text? Or whose interests are served by this text?) (Vasquez, 2017). However, for critique to find firm footing in reading programs there needs to be a rebalancing of instructional or cognitive targets. Using the NAEP trichotomy (NAGB, 2017)—locate and recall (literal comprehension tasks), integrate and interpret (interpretive comprehension tasks), and critique and evaluate (critical comprehension tasks)—as a benchmark for the types of tasks students are asked to complete in reading assignments, what is needed is a shift from more literal and even interpretive to more critical tasks.

In the next era of comprehension research, it would be useful to extend this work in four ways: (1) simply increasing the frequency of tasks that invite either internal or external critique, (2) building composite tasks that require students to understand a text on the way to critiquing it (or starting with an invitation to critique and dragging along the comprehension required to carry out the critique), (3) bringing critical tasks down to the primary level to learn more about what even 5- and 6-year-olds are capable of,

and (4) moving into a multiliteracies (Cope & Kalantzis, 2015; NLG, 1996) framework for what counts as text and what it means to engage with text. Regarding these suggestions, two facts must be acknowledged. First, the Request for Application for the RfU initiative never asked applicants to directly address "critique" aspects of reading. That we had as much emphasis on critical reading and thinking as we did within the RfU portfolio is noteworthy. Second, all the while that the RfU work was playing itself out, a parallel movement within literacy research and practice was unfolding and expanding in our journals and classrooms. It is time to wed these parallel movements. The multiliteracies perspective would surely benefit from the rigorous application of the research tools developed and implemented by the RfU research teams.

We need more robust and more nuanced analyses of the role that text plays in interventions. Text was involved in most of the interventions in the overall RfU portfolio. But it played a highly variable role, especially as a function of the age level of the students receiving the intervention. For the secondary interventions (READI, PACT, CCT, STARI, and WG), students were expected to read and be accountable for demonstrating their personal understanding of the texts they read as a part of their instructional modules or units. Moreover, in WG, READI, PACT, and CCT, they were expected to use the knowledge gained while reading texts to accomplish other goals, most often a writing-from-sources task. At the other end of the developmental continuum, with primary students in FCRR and LARRC, when texts were involved, they often served as opportunities for listening, not reading, comprehension; in only one early intervention, CALI, were students expected to apply what they had learned from text in a new task. However, inside the interventions we reviewed, text was a fixed factor, not a variable, even when the intervention focused on text structure (e.g, TEXTS or Let's Know!). So we did not learn much about how variations in text content, structure, or purpose affected comprehension or learning. This observation parallels a similar conclusion about the lack of emphasis on text from Cervetti's review of the developmental work in Chapter 2. Text was always there, but it was seldom examined.

Going forward, text deserves a more central role in our pedagogical research—as a malleable factor, a curricular tool—rather than simply an artifact in the instructional ecology or a medium for hosting other malleable factors, such as close reading routines or variations in discussion practices. This inclusion is especially important if we suspect, as we do, that pedagogical routines may interact with text elements, such as genre, challenge, or structure.

We need an ambitious program of research focused directly on the tension between assembled (one-component-at-a-time) and orchestrated (multicomponent) approaches to improving comprehension. A persistent tension across the RfU teams centered on fundamental assumptions about the optimal grain size of an intervention. Anchoring the atomistic components end of the continuum was FCRR, with its theoretical grounding in the lattice model (and its implicit search for the ideal set of components for a given student), and its quest, along with LARRC, to populate the listening comprehension (LC) factor in the Simple View of Reading formula (RC = DEC × LC; where RC is reading comprehension and DEC is decoding) with a curated collection of language structures and routines that might ultimately drive reading comprehension. Anchoring

the orchestrated activity end of the continuum was READI, with its commitment to situating comprehension practices within the context of discipline-based learning modules that employed collaborative learning, close reading of texts to acquire knowledge to use in constructing evidence-based arguments, and engagement in the discourse practices of the discipline. The other three teams fit somewhere in between FCRR and READI, with, in our reading of the work, CCDD and LARRC leaning toward the READI end of the continuum and PACT somewhere in the middle. If one values transfer effects to learning or distal measures of comprehension, then the nod for effectiveness goes to the orchestrated end of the continuum. But, given the sporadic distribution of main and interaction effects favoring treatments over the BAU, it is wise, we think, to devote more resources and conceptual energy to understanding and managing, if not resolving, these tensions. We have all too many convictions on this tension and way too little empirical evidence. We need more.

A CODA FOR THE PEDAGOGICAL PORTFOLIO OF THE RFU

The RfU work on curriculum and instruction was designed with the overall goal of moving the needle on students' reading comprehension achievement. Not all treatments led to statistically significant student gains of remarkable magnitude. Even so, innovative multicomponent approaches to comprehension instruction, when supported by teacher professional development and evaluated with relevant measures, led to a range of significant effects of respectable magnitude on comprehension and related outcomes—especially for older students. It would have been ideal, from the point of view of making precise, specific, and highly generalizable recommendations, if the contributions of specific components—the emphasis on different types of knowledge, the rich talk about text prompted by collaborative settings, the salutary contribution of motivation, metacognition, specific skills or strategies, and more—could be isolated. That would tell us how much emphasis to place on each element. Perhaps, however, it is more important that we know that when these components are integrated into engaging and consequential curriculum activities, good outcomes are possible for knowledge development, either at no cost to comprehension (the more common finding) or in concert with advances in comprehension. And, as a bonus, in many cases, other kinds of development (vocabulary, morphology, metacognition, perspective taking, or constructing/evaluating arguments, for example) are enhanced as well. In terms of a legacy, the RfU work on curriculum and instruction taught us much about what works and, equally as important, left us a catalog of insights, hunches, and unfinished business that will keep many of us occupied as school-based researchers, particularly in those schools working with currently underserved students, for the foreseeable future.

REFERENCES

Afflerbach, P. (2016). *Handbook of individual differences in reading: Reader, text, and context*. New York: Routledge.

Alexander, P. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist, 47*(4), 259–280. doi: 10.1080/00461520.2012.722511.

Anderson, R. C. (2018, June). *Reassessing constructivist perspectives on reading comprehension: Comprehension as an act of reflective action*. Paper presented at the Webfestschrift of P. David Pearson, Berkeley, CA.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 255–291). New York: Longman.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Beacco, J., Fleming, M., Goullier. F., Thurmann, E., Vollmer, H., & Shiels, J. (2015). *The language dimension in all subjects: A handbook for curriculum development and teacher training*. Strasbourg, France: Council of Europe.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research Education, 24*, 61–100.

Brown, W., & Greenleaf, C. (2017, November). *Disciplinary literacy with text based investigations: Reading to model science phenomena.* Paper presented at the annual conference of the Literacy Research Association, Tampa, FL.

Capin, P., & Vaughn, S. (2017). Improving reading and social studies learning for secondary students with reading disabilities. *Teaching Exceptional Children, 49*(4), 249–261. doi: 10.1177/0040059917691043.

Chall, J. (1983). *Stages of reading development.* New York: McGraw-Hill.

Chall, J. (1996). *Learning to read: The great debate.* New York: McGraw-Hill.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Connor, C. 2016. A lattice model of the development of reading comprehension. *Child Development Perspectives, 10*(4), 269–274.

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., Phillips, B. M., Kim, Y. S. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & Al Otaiba, S. (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.

Cope, B. & Kalantzis, M. (2015). *A pedagogy of multiliteracies: Learning by design*. London: Palgrave.

Darling-Hammond, L. (2019). *Investing for student success: Lessons from state school finance reforms*. Palo Alto, CA: Learning Policy Institute. Retrieved from https://learningpolicyinstitute.org/product/investing-student-success-school-finance-reforms-brief.

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.

Duke, N., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. Farstrup & J. Samuels (Eds.), *What research has to say about reading instructio*n (3rd ed., pp. 205–242). Newark, DE: International Reading Association.

Fodor, J. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

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., & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review, 26*(3), 425–449. doi: 10.1007/s10648-014-9270-6.

Goldman, S. R. (2018). Discourse of learning and the learning of discourse. *Discourse Processes, 55*(5–6), 434–453.

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

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.

Greenleaf, C., & Valencia, S. (2017). Missing in action: Learning from texts in subject-matter classrooms. In K. A. Hinchman & D. A. Appelman (Eds.), *Adolescent literacies: A handbook of practice-based research* (pp. 135–155). New York: Guilford Press.

Guthrie, J., & Klauda, S. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly, 49*, 387–416.

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.

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.

Jones, S. M., LaRusso, M., Kim, J., Kim, H. Y., Selman, R., Uccelli, P., Barnes, S., 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*, 448–483.

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. doi: 10.1002/rrq.171.

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.

LARRC (Language and Reading Research Consortium), 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, Farquharson, K. & Murphy, K. (2016). Ten steps to a large, multi-site, longitudinal investigation of language and reading in young children. *Frontiers in Developmental Psychology, 7*, 1–16.

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.

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.

Lee, C. D., Goldman, S. R., Levine, S., & Magliano, J. P. (2016). Epistemic cognition in literary reasoning. In J. Green, W. Sandoval, & J. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 165–183). New York: Routledge.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher, 48*(3), 158–166. doi: 10.3102/0013189X19832850.

Lucas, T., Villegas, A., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English language learners. *Journal of Teacher Education, 59*, 361–373.

McDermott, R., & Varenne, H. (1995). Culture as disability. *Anthropology & Education Quarterly, 26*, 324–348.

Moll, L. C., Amanti, C., Neff, D., & González, N. (1992). Funds of knowledge for teaching: A qualitative approach to connect households and classrooms. *Theory into Practice, 31*, 132–141.

NAGB (National Assessment Governing Board). (2017). *Reading framework for the 2017 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board and U.S. Department of Education. https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/reading/2017-reading-framework.pdf.

NLG (New London Group). (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*, 60–93.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202.

Pearson, P. D., & Cervetti, G. N. (2015). Fifty years of reading comprehension theory and practice. In P. D. Pearson & E. H. Hiebert (Eds.), *Research-based practices for teaching Common Core literacy* (pp. 1–39). New York: Teachers College Press.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Towson, MD: Paul H Brookes.

Pressley, M. (2001, September). Comprehension instruction: What makes sense now, what might make sense soon. *Reading Online, 5*(2). Retrieved from www.readingonline.org/articles/art_index.asp?HREF=/articles/handbook/pressley/ index.html.

Rowan, B., & Correnti, R. (2009). Measuring reading instruction with teacher logs. *Educational Researcher, 38*, 449–551.

RRSG (RAND Reading Study Group). (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica, CA: RAND Corporation.

Samson, J. F., & Collins, B. A. (2012). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness*. Washington, DC: Center for American Progress.

Shanahan, T., Fisher, D., & Frey, N. (2016). The challenge of challenging text. In M. Sherer (Ed.), *On developing readers* (pp. 100–109). Alexandria, VA: ASCD.

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.

Valdés, G., & Castellón, M. (2011). English language learners in American schools: Characteristics and challenges. In T. Lucas (Ed.), *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators* (pp. 18–34). New York: Routledge.

Vasquez, V. M. (2017). Critical literacy. *Oxford Research Encyclopedia, Education*. Oxford, UK: Oxford University Press.

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., Martinez, L. R., Williams, K. J., Miciak, J., Fall, A-M. & Roberts, G. (2019). Efficacy of a high school extensive reading intervention for English learners with reading difficulties. *Journal of Educational Psychology, 111*(3), 373–386. doi: 10.1037/edu0000289.

Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children, 85*(2), 180–196. doi: 10.1177/0014402918782618.

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*, 77–93.

Veenman, M., van Hout-Wolters, B., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological issues. *Metacognition and Learning, 1*, 3–14.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wanzek, J., & Vaughn, S. (2016). Implementation of a text-based content intervention in secondary social studies classes. *New Directions for Child and Adolescent Development, 154*, 31–48.

Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., Fall, A-M., Stillman-Spisak, S., & 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.

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.

Wilburn, G., Cramer, B., & Walton, E. (2019). The great divergence: Growing disparities between the nation's highest and lowest achievers in NAEP mathematics and reading between 2009 and 2019 [Blog post]. NAEPPLUS+ The Official Blog of the Nation's Report Card. Retrieved from https://nces.ed.gov/nationsreportcard/blog/mathematics_reading_2019.aspx.