

The image is a vertical composition of three distinct educational scenes. The top scene shows two children, a girl and a boy, looking intently at a red tablet held by the boy. The middle scene features a woman with long dark hair, smiling warmly at a young man in a classroom setting. The bottom scene is a close-up of a library bookshelf filled with books of various colors. The entire image is overlaid with a large, semi-transparent red circular graphic on the right side, which contains the text.

# REAPING THE REWARDS

*of the*

# READING FOR UNDERSTANDING INITIATIVE

NATIONAL ACADEMY *of* EDUCATION

REAPING THE  
REWARDS  
*of the*  
READING FOR  
UNDERSTANDING  
INITIATIVE

P. David Pearson, University of California, Berkeley

Annemarie Sullivan Palincsar, University of Michigan

Gina Biancarosa, University of Oregon

Amy I. Berman, National Academy of Education

National Academy of Education  
Washington, DC

NATIONAL ACADEMY OF EDUCATION 500 Fifth Street, NW

Washington, DC 20001

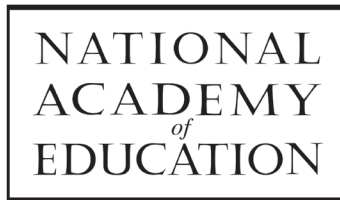
NOTICE: This project and research reported here were supported by grant R305U160002 from the U.S. Department of Education's Institute of Education Sciences (IES). The opinions expressed are those of the editors and authors and do not represent the views of IES.

Digital Object Identifier: 10.31094/2020/2

Additional copies of this publication are available from the National Academy of Education, 500 Fifth Street, NW, Washington, DC 20001; <http://www.naeducation.org>.

Copyright 2020 by the National Academy of Education. All rights reserved.

Suggested citation: Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (Eds.). (2020). *Reaping the Rewards of the Reading for Understanding Initiative*. Washington, DC: National Academy of Education.



The **National Academy of Education** (NAEd) advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars.

# Dedication

To Carol McDonald Connor, who sowed many of the seeds of the Reading for Understanding (RfU) initiative. We wish that she were with us to witness the harvest, to reap the rewards with the rest of the RfU community. Thank you, Carol, for a life well lived and a career committed to improving children's reading.

**REAPING THE REWARDS OF THE  
READING FOR UNDERSTANDING INITIATIVE**

*Steering Committee*

**Annemarie Sullivan Palincsar** (*Co-Chair*), University of Michigan  
**P. David Pearson** (*Co-Chair*), University of California, Berkeley  
**Donald Compton**, Florida State University  
**Susan Goldman**, University of Illinois at Chicago  
**Kenji Hakuta**, Stanford University  
**Glynda Hull**, University of California, Berkeley  
**Laura Justice**, The Ohio State University  
**Christopher Lonigan**, Florida State University  
**John Sabatini**, The University of Memphis  
**Catherine Snow**, Harvard University and Strategic Education Research Partnership  
**Sharon Vaughn**, The University of Texas at Austin

*Staff*

**Amy I. Berman**, Deputy Director



# Contents

<b>ACKNOWLEDGMENTS</b>	<b>ix</b>
<b>EXECUTIVE SUMMARY</b>	<b>1</b>
<b>1 INTRODUCTION TO THE READING FOR UNDERSTANDING INITIATIVE</b>	<b>11</b>
<i>Annemarie Sullivan Palincsar, P. David Pearson, Amy I. Berman, and Peter Afflerbach</i>	
APPENDIX 1-1 DESCRIPTION OF PROCEDURAL ELEMENTS FOR DEVELOPING THIS REPORT	31
<b>2 THE NATURE AND DEVELOPMENT OF READING FOR UNDERSTANDING</b>	<b>41</b>
<i>Gina N. Cervetti</i>	
<b>3 THE ASSESSMENT OF READING FOR UNDERSTANDING</b>	<b>67</b>
<i>Panayiota Kendeou</i>	
APPENDIX 3-1 BRIEF REVIEWS OF READING FOR UNDERSTANDING ASSESSMENTS	99
<b>4 TEACHING READING FOR UNDERSTANDING: SUMMARIZING THE CURRICULUM AND INSTRUCTION WORK OF THE FIVE CORE READING FOR UNDERSTANDING TEAMS</b>	<b>143</b>
<i>Gina Biancarosa, Peter Afflerbach, and P. David Pearson</i>	
APPENDIX 4-1 PUBLISHED MEASURES USED IN THE READING FOR UNDERSTANDING PORTFOLIO OF EFFICACY STUDIES REPRESENTED IN CHAPTERS 4 AND 5	205
APPENDIX 4-2 DEMOGRAPHIC DATA FOR READING FOR UNDERSTANDING TEAMS' RANDOMIZED CONTROLLED TRIALS	209



<b>5</b>	<b>TEACHING READING FOR UNDERSTANDING: SYNTHESIS AND REFLECTIONS ON THE CURRICULUM AND INSTRUCTION PORTFOLIO</b>	<b>215</b>
	<i>Peter Afflerbach, Gina Biancarosa, Matthew Hurt, and P. David Pearson</i>	
<b>6</b>	<b>TAKING STOCK OF THE READING FOR UNDERSTANDING INITIATIVE</b>	<b>251</b>
	<i>P. David Pearson, Annemarie Sullivan Palincsar, Peter Afflerbach, Gina N. Cervetti, Panayiota Kendeou, Gina Biancarosa, Jennifer Higgs, Miranda Fitzgerald, and Amy I. Berman</i>	
	<b>BIOGRAPHICAL SKETCHES OF STEERING COMMITTEE MEMBERS AND AUTHORS</b>	<b>293</b>

# Acknowledgments

Reading for understanding is surely one of our most important educational goals, especially given the importance of reading for so many activities in other fields that require students to build knowledge from text and critically evaluate the trustworthiness, relevance, and utility of the information they encounter in those texts. That shared common purpose no doubt accounts for the enthusiasm we encountered when we asked colleagues, both inside the National Academy of Education (NAEd) and in the broader educational research community, to join us in this effort to synthesize the amazing portfolio of work constructed in the Reading for Understanding (RfU) initiative, under the moniker *Reaping the Rewards* (of the RfU). Given the privilege we were accorded in conducting this synthesis, it is important to acknowledge the contributions of these colleagues as we make this synthesis public.

First and foremost, we thank our funder, the Institute of Education Sciences (IES), for its foresight in realizing that the RfU initiative might be appreciated by a larger readership if the work could be broadly shared in a synthesis that highlighted key findings and significant themes across the six teams. Key individuals within the IES community who championed the RfU cause were Elizabeth Albro, Karen Douglas, Rebecca Kang McGill-Wilkinson, and Vinita Chhabra. Elizabeth Albro, who currently serves as the Commissioner of the National Center for Education Research, had the original vision for the RfU initiative and shepherded it into being in the first decade of this century. Karen Douglas was the program officer for the RfU during its 5-year funding cycle—the person who convened the RfU leaders on a regular basis to promote cross-team sharing of insights, conundrums, and solutions. Elizabeth Albro and Karen Douglas worked together to champion a special funding allocation to bring this synthesis into reality. Rebecca Kang McGill-Wilkinson and Vinita Chhabra served as our program officers, making sure that we received the support and oversight needed to complete our task. We also want to thank Catherine Snow, an Academy member, for her suggestion and encouragement to the Academy to pursue a grant from IES for this synthesis project.

Second, we thank the cadre of folks who gave direction to the synthesis, the steering committee. Joining the co-chairs and co-principal investigators (co-PIs) (David and Annemarie), the steering committee included the site team directors—Susan Goldman, University of Illinois at Chicago (Reading, Evidence, and Argumentation in Disciplinary Instruction [READI]); Laura Justice, The Ohio State University (Language and Reading Research Consortium [LARRC]); Christopher Lonigan, Florida State University (Florida Center for Reading Research [FCRR]); John Sabatini, The University of Memphis (Educational Testing Service [ETS]); Catherine Snow, Harvard University and Strategic Education Research Partnership (Catalyzing Comprehension through Discussion and Debate [CCDD]); and Sharon Vaughn, The University of Texas at Austin (Promoting Adolescents' Comprehension of Text [PACT]). Other steering committee members not directly involved in the RfU included Donald Compton (Florida State University), Kenji Hakuta (Stanford University), and Glynda Hull (University of California, Berkeley). The steering committee met for 2-day meetings on three separate occasions—at the outset to provide direction for the work, at a midpoint to assess the progress of the three core draft chapters, and very close to the end of the process, with a full dossier of chapters available for review. They all went the extra mile to ensure direction, access to work, and rigorous oversight of the synthesis.

Third, we gratefully acknowledge the team of writers and editors who brought the work to life. The key writers for the core chapters on comprehension development, assessment, and curriculum and instruction were Gina Cervetti (University of Michigan), Panayiota Kendeou, (University of Minnesota), and Peter Afflerbach (University of Maryland). The three of them produced the core of our work in the first stage of the effort and got us off to a flying start. Advising the authors of the core chapters were a pair of consulting editors assigned to each: Walter Kintsch (University of Colorado) and Laura Justice for Gina Cervetti's development chapter; John Sabatini (then ETS, but now The University of Memphis) and Joan Herman, (University of California, Los Angeles) for Panayiota Kendeou's assessment chapter; and Richard Anderson (University of Illinois at Urbana-Champaign) and Carol Lee (Northwestern University) for Peter Afflerbach's curriculum and instruction effort. Additionally, we want to thank three Academy members who were involved in an internal review of the synthesis on behalf of the Academy: Judith Warren Little, who chairs the Standing Review Committee, recruited Kenji Hakuta and Glynda Hull to review the entire manuscript on behalf of the Academy.

We (David and Annemarie) were joined by Gina Biancarosa (University of Oregon) and Amy Berman (NAEd Deputy Director) to form the editorial team. The four editors also served as co-authors of chapters when it was appropriate. Finally, joining us as co-authors of particular chapters were Matthew Hurt (University of Maryland; Chapter 5), Jennifer Higgs (University of California, Davis; Chapter 6), and Miranda Fitzgerald (University of North Carolina at Charlotte; Chapter 6). The two of us owe an even larger thank-you to Amy Berman and Gina Biancarosa for the heavy editorial burden they bore in moving us through many drafts along the pathway from the initial to the final draft.

Fourth, we thank the Academy staff who worked wonders for this volume. Without their effort, this project would not have been remotely possible. Abigail Bell, Tess Bonnette, Daniel Chavez, Naomi Chudowsky, and Dian Dong provided support throughout the project, and a special thank-you goes to Linda Loughran for our cover design. Gregory White, the Executive Director of the Academy, assumed most of the responsibility for project and grant management and communications with our funder (thank you, Greg!). And the single person to whom we, as co-PIs and steering committee co-chairs, owe our heartiest and most heartfelt thanks is Deputy

Director Amy Berman. Amy was “communication central” (with so many constituents), nudge extraordinaire (keeping us on track with the gentlest of touches), and editor supreme (making sure we did not go off too many deep ends and were always APA compliant). She was—and is—the soul of the project.

Additionally we thank each other. We have known one another since roughly 1980 (imagine that—four decades!), but this represents our first real opportunity to work closely together on a project of this magnitude.

David, in thanking Annemarie, had this to say: Even though I went into this collaboration with the greatest admiration for all that Annemarie has done for the literacy research profession and for the Academy, I left it even more astounded by her intellect, knowledge, perseverance, and generosity.

Annemarie, in thanking David, had this to say: I am so grateful that David invited me to partner with him in this effort. I am most fortunate to have witnessed, first-hand, David’s commitment to the field of literacy scholarship; his capacity to bring ideas to life and to make the numerous conceptual connections this work demanded. Finally, I am awed by his indefatigable passion to support the work of educators.

We hope, when you read this synthesis, you will appreciate just how much we owe one another—and how much we owe the entire *Reaping the Rewards* team—in bringing this synthesis to completion.

Finally, you will, we are certain, appreciate the incredible scope and depth of work of the more than 200 scholars who did the work that made this synthesis possible in the first place. Bravo!

It appears to take a village to raise a synthesis!

Annemarie Sullivan Palincsar and P. David Pearson, *Co-Chairs*  
Reaping the Rewards of the Reading for Understanding Initiative



# Executive Summary

In 2009, the U.S. Institute of Education Sciences (IES) allocated \$120 million to establish the Reading for Understanding (RfU) initiative. This initiative responded to concern that children’s improvement in reading comprehension had leveled off over the previous few decades, coupled with the observation that research on reading comprehension had sufficiently matured to warrant a major investment in improving student performance. The RfU initiative involved a research and development network of six interconnected teams focused on improving reading comprehension for students in pre-kindergarten (pre-K) through grade 12. The rationale for such a major investment, based on a direct analogy to the United States’ highly successful 1960s networked approach to accelerating the goal of a moon landing, was that the severity of the problem, and the likelihood of finding a solution, rendered reading comprehension a wise investment.

Thus, in 2010, six teams of researchers (one focused on assessment and five charged with understanding and improving the development and pedagogy of reading comprehension) were funded to carry out the initiative. Two teams (the Florida Center for Reading Research [FCRR] and the Language and Reading Research Consortium [LARRC]) focused on early reading levels (pre-K through grade 4); three teams focused on older readers from grades 5–12 (the Catalyzing Comprehension through Discussion and Debate [CCDD], Promoting Adolescents’ Comprehension of Text [PACT], and the Reading, Evidence, and Argumentation in Disciplinary Instruction [READI]); and one team (the Educational Testing Service [ETS]) focused on assessment. Collectively, the teams studied the development, instruction, and assessment of reading comprehension from pre-K through grade 12. The funding mandate called for a network, a unique feature of this effort that brought site directors and scholars from the six teams together on a recurring basis to share collegial critique and common experiences, and to promote synergies across teams.

In 2016, following the 5-year award period, as the RfU teams continued to analyze data and add to the portfolio of more than 200 publications already generated, IES

funded an invited proposal from the National Academy of Education (NAEd) to synthesize findings, themes, principles, and barriers related to this ambitious attempt to understand and improve U.S. reading comprehension performance. Through this *Reaping the Rewards of the Reading for Understanding Initiative*, the NAEd was charged with answering the question: What has been the yield from this investment? More specifically, the Academy's charge was to synthesize, from this substantial and unprecedented effort, what had been learned about understanding and improving reading comprehension.

To guide the NAEd in answering this question, a steering committee was established; its membership included NAEd members knowledgeable about literacy and reading, the leaders of the six funded teams, and two NAEd members (Annemarie Sullivan Palincsar and P. David Pearson) whom the Academy had recruited as co-chairs of the project. With the steering committee's guidance about the scope and methods of the review, the NAEd staff, with the advice of the co-directors, recruited scholars to assist with the synthesis in three large "buckets" of research—the nature and development of reading comprehension, reading comprehension assessment, and curriculum and instruction to promote reading comprehension. That collective—the steering committee, the scholars serving as authors of the report, the NAEd staff, and the co-chairs—worked on this effort from 2017 through 2019.

## THE YIELD

The synthesis revealed that the RfU initiative was successful in advancing knowledge for all three strands—development, assessment, and curriculum and instruction. Highlights from the synthesis include *key findings* and many *lessons learned* about (1) how we think differently about reading comprehension now than we did in the pre-RfU period, (2) how to implement ambitious efforts such as research networks, and (3) the direction of future research inspired by the RfU.

In this Executive Summary, we offer highlights from this effort that are documented in the chapters that follow. We begin with the three most important contributions of the RfU initiative, the "headlines." Then we move to a more elaborate and specific set of key findings across the work of the six teams, which is followed by a set of lessons learned and, finally, an agenda for future work.

## HEADLINES

**Knowledge is cause, consequence, and covariate of reading comprehension.** How we think about the role of learners' knowledge in explaining, assessing, and facilitating reading comprehension is broader and deeper than it was before the RfU initiative. Our understanding of the types of knowledge necessary for particular acts of reading have expanded beyond the familiar triad of declarative, procedural, and conditional knowledge to also include disciplinary and epistemic knowledge. In particular, disciplinary knowledge about topics—such as how explanation and argumentation operate, what count as claims and evidence, how oral and written discourse conventions shape those processes, and how we come to know what we know—are central to students' acquisition of knowledge and inquiry practices within disciplines. Additionally, the RfU research provides a deeper understanding of the role that conventional knowledge

sources play in fundamental processes such as inferencing (filling in gaps, such as the motive of a character, left unsaid by the author) and comprehension monitoring (evaluating how well you *really* understood the last paragraph). Finally, the RfU highlighted the “other side” of the all-important relationship of knowledge to comprehension. For decades, we have emphasized how knowledge shapes comprehension, but only more recently have we focused more on how comprehension shapes knowledge—knowledge that is then available to use in other learning and application tasks. Much of the RfU work focused on using the fruits of comprehension to apply to other tasks, such as writing an argument, telling a story, or solving a problem. Nowhere is this progress better reflected than on the assessment front, where the RfU work successfully validated a comprehension assessment, the Global Integrated Scenario-Based Assessment (GISA). GISA measures both “close reading” of texts plus the ability to use knowledge gained from reading to carry out application tasks within a contextualized scenario that privileges purpose-driven activity within a simulated social setting.

**Language drives every facet of reading comprehension.** As with knowledge, the RfU has helped us both to broaden and deepen the ways we think about the role of language in explaining, assessing, and facilitating reading comprehension. We have known for almost a quarter century that different facets of language provide strong explanations for the nature and quality of reading performance at different levels of development. Early on, in kindergarten through grade 2, subword processes like letter-sound knowledge and phonemic awareness tend to explain the majority of the variance in reading achievement, while more meaning-based language variables, including receptive and expressive vocabulary, explain increasing proportions of the variance as students move into grades 2 and 3. What we did not know before the RfU was how important the more sophisticated facets of academic and disciplinary language would become in explaining and improving advanced levels of reading comprehension, such as those we encounter in middle and high school. But even for more traditional facets of language, such as more basic lexical and grammatical elements, the RfU teams were able to unpack and evaluate their contributions to comprehension performance in greater detail than ever before. As with the knowledge agenda, the RfU teams also made progress in the assessment of some of these more sophisticated facets of language.

**Reading is an inherently cultural activity.** On the face of it, this headline is old news, but the RfU portfolio breathes new life into the claim that all facets of reading are contextualized. Development always occurs in a particular situation—in a classroom, at a community center, or around a kitchen table. Decontextualized assessment may not be the best way to monitor development over time or to ascertain pedagogical effects. Assessments like GISA represent a step in the right direction. Most importantly, successful classroom-level comprehension interventions require fundamental changes to classroom cultures, not just changes to routine instructional practices. These changes in classroom cultures, which are inherently situated (they look a little different in every classroom), include alternative expectations for the tasks, social supports, talk, and purposes that surround reading. The most successful interventions in the RfU portfolio, particularly for older students, involved *collaborative* work groups that undertook close reading and *dialogically-based discussion* of challenging, often controversial, texts



with the immediate goal of mining the texts for information that students could use to meet the longer-term goal of *applying what they learned* to new problems or situations. Conceptualizing the implementation of interventions as needing to affect classroom cultures, rather than only improving technical proficiencies, suggests a different stance toward promoting classroom and school change. This sort of change demands teacher learning as well as student learning, and many RfU teams required teachers to learn new approaches to pedagogy as prologue to effective teaching. Teacher learning involved viewing professional development and one's own learning as a long-term, continuous journey within professional learning communities. In pursuing an even more ambitious goal, teachers were involved in the design, delivery, and critique and revision of curricular materials, pedagogical routines, and professional development activities in a design-based laboratory where teachers worked alongside researchers and curriculum designers in a continuous improvement enterprise.

### KEY FINDINGS

A high-level summary of key findings adds detail to the headlines, offering new understandings across the three major strands of development, assessment, and curriculum and instruction.

With respect to the *nature and development* of comprehension, the RfU portfolio of work:

- Described the heightened importance of both word and world knowledge in explaining comprehension development, especially for inferential reasoning and comprehension monitoring.
- Rendered the Simple View of Reading more complex by proposing different models of how the broad components of listening comprehension and decoding interact at various stages of development and adding additional variables (facets of knowledge, language, and other internal processes) to account for the complexity of comprehension during the adolescent years.
- Demonstrated that language is most productively regarded as a single construct, or perhaps as a cluster of closely related skills.

Regarding *assessment*, the RfU portfolio of work:

- Demonstrated that standards of authenticity, complexity, and psychometric adequacy can be achieved in a single assessment system that assesses text comprehension, learning, and application.
- Instantiated knowledge as an integral component of reading comprehension that should be integrated into the assessment of comprehension, not simply controlled.
- Developed specialized tests of subcomponents of reading that can, and in some cases do, contribute to larger batteries that address a range of comprehension-related variables—prior knowledge, academic language, perspective taking, inference making, evidence-based argument, and reading and self-regulatory strategies.

For *curriculum and instruction*, the RfU portfolio of work:

- Produced a range of positive, but often inconsistent, results on a wide range of measures across the K–12 continuum.
- Revealed that effects were greater and more consistent for curriculum-aligned than for curriculum-independent measures of key outcomes.
- Demonstrated that the strongest effects were observed for measures of vocabulary, morphology, comprehension monitoring, and knowledge acquisition.
- Revealed that the interventions that “moved the needle” on reading comprehension and a host of related measures (such as vocabulary, knowledge acquisition, application, and enabling skills) were characterized by well-orchestrated, multi-component instruction.
- Established that reading comprehension interventions were often (if not always) coordinated with content-area learning goals, usually with comprehension activity enacted in the service of content acquisition.
- Provided evidence that when positive outcomes did not emerge on both comprehension and content learning, advances in one did not come at a cost to the other.

### LESSONS LEARNED

We learned a great deal from the RfU initiative about the nature of the research process as well as specific issues related to the general question of “what works.” More specifically, several lessons stand out as unique and significant.

**Being able to design research with a long runway for implementing projects enables more robust and credible research.** The research model enacted in the RfU initiative provides a demonstration of what is possible in the design, implementation, and analysis of lines of inquiry with the affordances of adequate funding, more generous time frames, and a diverse array of expertise to carry out the work. When there is a sufficiently long runway, scholars have the opportunity to exploit the complementarity of research methods, scholarly traditions, and academic disciplines. Add to that mix the opportunity of the RfU network to serve as a crucible for sharing collegial critique and insight, and the affordances multiply.

**Teacher professional learning can serve as either a bridge or a barrier to successful implementation.** Within the RfU, we learned much about facets of pedagogy that are easier and harder to learn, the barriers to teacher learning and uptake, and the contextual supports that account for positive changes in teacher knowledge and practice. Three observations are warranted from the study of teacher learning and uptake: (1) the more complex the pedagogy, the lower the likelihood of implementation; (2) the more teachers are embedded in all aspects of the intervention, the greater their uptake of important aspects of the intervention; and (3) a major roadblock to teacher uptake of new practices is the accountability infrastructure of reform movements. The more test scores matter, the less the likelihood that teachers will adopt novel teaching practices.

**The RfU research portfolio increased our understanding of the barriers to “moving the needle” on comprehension achievement.** Because the randomized controlled trials and efficacy studies in the RfU were well designed and well implemented, the typical explanations for failing to move the needle (shortcomings related to design, duration, and measurement issues) could be ruled out. What remain as more plausible explanations are the inherent difficulty of this sort of work (researchers, professional developers, teachers, and students are being asked to undertake more challenging agendas) and unrealistic expectations (i.e., we might believe that moderate [0.50] if not large [0.80] effect sizes are achievable when the more realistic expected value for work of this sort is nearer the small [0.20] standard).

**Learning to read and reading to learn surfaced in the RfU portfolio as complementary goals, rather than separate stages of development.** The conventional wisdom in reading is that first students learn to read and then they read to learn. Within the RfU work, to the contrary, researchers found that these two complex processes were more likely to be interwoven across students’ school careers. In the primary grades, even as early as kindergarten, students can read to learn as they learn to read. The case for complementarity between reading to learn and learning to read is stronger than the case for separate, encapsulated stages. Conversely, there is evidence that, even in middle school, when reading to learn is prominent in the disciplines of history, science, and literature, there is still much to learn about how to read effectively, such as language and vocabulary, the special nature of academic discourse, and strategies for unpacking dense grammatical structures. Also, while both learning to read and reading to learn have much in common across history, literature, and science, they also differ within each discipline.

**The RfU research advanced understanding of both general and specific aspects of reading comprehension.** In summarizing contributions to development, we noted how the RfU complicated the Simple View of Reading. Regarding the RAND heuristic model,<sup>1</sup> with its emphasis on the independent and joint influence of the reader, task or activity, and text within a sociocultural context on comprehension, the RfU made progress on all four of these key constructs. That said, in our view, the RfU work taught us more about reader and activity (task) variables than it did about text and context variables. Regarding adolescent/disciplinary literacy, the RfU initiative shifted the emphasis of comprehension instruction to an emphasis on students actively and collaboratively constructing and extracting meaning from texts, using language in the form of rich conversations about text to sharpen and deepen their understanding, and using the knowledge gained from reading, thinking, and talking to solve problems and explain how and why things in the world work the way they do.

---

<sup>1</sup> RRSF (RAND Reading Study Group). (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Corporation.

## FUTURE WORK

Perhaps one of the most important contributions of the RfU is the legacy of defining the work that remains to be completed in the name of understanding and improving reading comprehension. In terms of a specific agenda, several initiatives and recommendations are particularly important.

**As a field, literacy research must incorporate scholarship about new literacies, digital literacies, and multiliteracies into its analysis of reading comprehension processes and practices.** Because the Request for Application for the RfU was grounded in the strong cognitive tradition of the RAND Reading Study Group (2002), it did not require sites to address these new literacies. Thus, it is not surprising that these new perspectives are not well represented in the RfU portfolio, although both READI and the assessment work by ETS did embrace elements of the digital/new literacies agenda. Perhaps in 2010, when these teams were funded, such an omission was understandable, but from where we sit in 2020, and as we move into the future, comprehension research must incorporate the texts, tasks, and affordances and constraints of these additional, alternative representations of meaning.

**More of our work on comprehension needs to be directed toward populations currently underserved in U.S. schools.** This suggestion is first and foremost a call for equity in the conduct of research. The list of currently marginalized populations is long because it includes cultural and linguistic minority groups and children of poverty irrespective of race, ethnicity, or home language. At the top of the list should be emergent bilingual learners, a growing but still underserved population. The particular irony of this population is that, even though they bring rich language experiences to the classroom, we seem unable to exploit their first language or interlingual (first to second language connections) linguistic resources to craft effective programs for deep reading experiences in English as a second language. Developing curriculum, and for that matter assessments, that exploit their linguistic resources, brought into relief by increasingly prominent and deeper understanding of the role of translanguaging and interlingual expertise (the special knowledge that accrues to students who work in more than one language), represents a real opportunity for scholars of comprehension to embrace in order to better exploit the special resources of bi- and multilingual students.

**We need to develop more precise tools for evaluating the implementation of interventions by incorporating tools from the relatively new field of improvement science.** We have learned much about how to implement and sustain change in the past decade. Important in the field of improvement science is moving toward metrics that assess not only what individuals are learning (e.g., measures of *student* learning or *teacher* fidelity) but also indicators of *system* learning, where entities like schools, districts, and collaboratives are also assessed for the enhancements or barriers they create in reform efforts. We think literacy research efforts, even tightly controlled randomized controlled trials, would benefit from a more ecologically sensitive approach to examining the constraints and affordances of implementation, especially when we have compelling evidence of the consequential influence of the ecological context on research outcomes.

**We need to add both breadth and depth to our study of the knowledge-comprehension relationship.** We need to move beyond the aphorism that we learn what is new in terms of what we already know in favor of more complex, even reciprocal views of the knowledge-comprehension relationship.

**Writing in response to reading and learning from text is a likely candidate for improving reading comprehension.** Writing and reading bear an inherently complementary relationship. We know that reading informs writing, but we do not know as much about how writing, as the natural complement to and outcome of reading comprehension, improves reading. This relationship was implicit in all of the middle and high school interventions—CCDD, PACT, and READI. Much work remains to be completed about the role that writing can play in promoting integration and analysis of key textual ideas. It is time to address this important pedagogical agenda.

**Given the tension within the RfU between the assembly and orchestration models of skill acquisition, the field (perhaps with the leadership of IES) should undertake a major national initiative, including meta-analyses and new research studies, to evaluate the relative merits of competing theories of the process and pedagogical models of delivery.** Albeit with different terminology, the issue of which metaphor—assembly or orchestration—better captures the character of reading (and reading comprehension) development is one that arose in each strand of this review. It is time for the field, and IES, to allocate more conceptual and financial energy to this important but underanalyzed question. It makes a difference in how we design interventions to improve both comprehension and foundational word-level skills.

**Affect and conation deserve more emphasis in our research on comprehension development, assessment, and pedagogy.** The facets of learning that entail engagement, motivation, self-efficacy, and social well-being deserve more attention in our study of comprehension and learning. We need to know more than what we learned from the RfU about how these affective, dispositional, and social factors moderate and/or mediate learning from text in the short term, and shape students' reading in the long term.

#### OVERARCHING CONTRIBUTION OF THE RFU INITIATIVE

On a final note, as we think about the legacy of the RfU initiative, there are, by our collective reading, two complementary lessons. First is a lesson about making clear the theory of reading comprehension at play in our work. What the RfU demonstrates is that whether we are studying the nature and development of reading comprehension, creating assessments of reading comprehension, or working actively to improve reading comprehension, how we conceptualize reading comprehension will necessarily shape what we examine and, ultimately, what we achieve. The RfU made fundamental strides in elaborating what it means to comprehend what we read and, thus, in how we understand its development during schooling, how we can better assess the nuances and sources of comprehension, and what it means to improve comprehension and learning from text.

Second is a lesson focused more specifically on improving reading comprehension in school-based settings. The RfU initiative taught us about how much it takes to achieve even small effects for increases in student reading comprehension performance. It is a matter of commitment and sustenance. We witness the most impressive effects when we see strong and supportive professional learning communities that hold high standards and provide continuous support, in the form of coaching and careful monitoring, to help teachers acquire practices that promote the widest student engagement in higher-order talk within intentionally collaborative discussions about interesting and thought-provoking texts—all moving toward a target of applying what students learn in such a process to some issue, problem, or project worth addressing.



# Introduction to the Reading for Understanding Initiative

Annemarie Sullivan Palincsar, *University of Michigan*

P. David Pearson, *University of California, Berkeley*

Amy I. Berman, *National Academy of Education*

Peter Afflerbach, *University of Maryland*

## CONTENTS

INTRODUCTION .....	12
THE SIX TEAMS .....	13
The Language Bases of Reading Comprehension, The Ohio State University, LARRC.....	13
Examining Effective Intervention Targets, Longitudinal Intensity, and Scaling Factors in Pre-K to Grade 5 Student Comprehension, Florida State University, FCRR.....	14
Catalyzing Comprehension Through Discussion and Debate, Strategic Education Research Partnership Institute, CCDD.....	14
Understanding Malleable Cognitive Processes and Integrated Comprehension Interventions for Grades 7–12, The University of Texas at Austin, PACT .....	15
Reading for Understanding Across Grades 6–12: Evidence-Based Argumentation for Disciplinary Learning, University of Illinois at Chicago, Project READI.....	16
Assessing Reading for Understanding: A Theory-Based, Developmental Approach, ETS .....	17
SETTING THE CONTEXT FOR THE RFU EFFORT .....	17
The National Research Council Report <i>Preventing Reading Difficulties in Young Children</i> .....	18
The National Reading Panel .....	19
Educational Policy Specific to Reading Instruction in the United States.....	20
The New London Group’s <i>A Pedagogy of Multiliteracies</i> .....	22
The RAND Reading Study Group .....	23
Initiatives Regarding Adolescent Literacy .....	24



Evolving Ideas About the Nature of Reading Comprehension Reflected in Assessments .....	25
Static Reading Scores in the Face of Rising Demands for Comprehension . .	26
Theoretical Perspectives of Reading Comprehension .....	27
CONCLUSION .....	27
REFERENCES .....	27

## INTRODUCTION

In 2009, the U.S. Institute of Education Sciences (IES) announced the Reading for Understanding (RfU) research initiative. The RfU was a remarkably ambitious project. By educational research standards, it represented a huge investment (approximately \$120 million) in a fairly well-specified scope of work, which was identified as “(a) examining underlying processes of reading comprehension and identifying malleable processes that may be targets of interventions for enhancing reading comprehension, and (b) developing and testing interventions intended to improve reading comprehension” (IES, 2009, p. 5). The ultimate goal defined in the call was to redress the disappointing performance of students in the United States on national assessments of reading.

Grant applicants were to identify whether they were applying to become a core team or an assessment team; core teams were to propose reading comprehension research that covered a range of at least five grades, while assessment teams were to design reading comprehension measures for pre-kindergarten (pre-K) through grade 12. Core team applicants were required to propose an iterative design process that would culminate in a reading comprehension intervention that would be the subject of an efficacy study; furthermore, core teams were expected to use the measures designed by the assessment team in their research.

Invoking the model used by the National Aeronautics and Space Administration (NASA) in its mission to the moon, the RfU research was to be conducted by multidisciplinary, networked groups of researchers in partnership with practitioners. In the call, IES signaled that it would foster ongoing collaboration across the research groups for the duration of the 5-year awards in an effort to accelerate the pace of the research.

Ultimately, awards were made to five core teams and one assessment team. Collectively, the teams studied the development, instruction, and assessment of reading comprehension from pre-K through grade 12. In 2016, following the 5-year award period, and as the RfU teams were continuing to analyze data and add to the more than 200 publications already generated, IES funded a National Academy of Education (NAEd)-invited proposal, *Reaping the Rewards of the Reading for Understanding Initiative*, to lead an effort to:

- Articulate findings and common themes across the RfU projects to contribute to a full-range view of reading development;
- Identify obstacles to on-time reading achievement, as well as factors supporting success;
- Examine cross-project findings to identify areas of agreement and productive tension; and
- Find common principles underlying instructional programs across projects.

In that spirit, NAEd engaged in a collaborative effort, bringing together representatives of each of the six teams, joined by others, to produce a summary report informed by the publications prepared by the RfU teams, as well as proceedings of meetings. This volume reports on the results of the NAEd effort.

We began our work by convening a 3-year working committee with leaders of the six RfU teams and scholars who are in the field but not directly involved in the RfU initiative. The 11-member steering committee was co-chaired by Annemarie Sullivan Palincsar (University of Michigan) and P. David Pearson (University of California, Berkeley). The six team directors were members of the steering committee: Susan Goldman, University of Illinois at Chicago (Reading, Evidence, and Argumentation in Disciplinary Instruction [READI]); Laura Justice, The Ohio State University (Language and Reading Research Consortium [LARRC]); Christopher Lonigan, Florida State University (Florida Center for Reading Research [FCRR]); John Sabatini, The University of Memphis (Educational Testing Service [ETS]); Catherine Snow, Harvard University and Strategic Education Research Partnership (SERP) (Catalyzing Comprehension through Discussion and Debate [CCDD]); and Sharon Vaughn, The University of Texas at Austin (Promoting Adolescents' Comprehension of Text [PACT]). Other steering committee members not directly involved in the RfU included Donald Compton (Florida State University), Kenji Hakuta (Stanford University), and Glynda Hull (University of California, Berkeley).

The steering committee guided the work of this report, including organizing the synthesis around the three main topics of development, assessment, and instruction; producing relevant research articles to ground each topic; and providing necessary feedback to identify common themes and findings. Specific details regarding the processes we used to conduct the RfU synthesis are presented in Appendix 1-1. At this point, we introduce the reader to the stars of this volume: the six Reading for Understanding teams.

## THE SIX TEAMS

This section provides a brief description of each team that was awarded IES funding through the RfU research initiative. Each team is identified by the original title provided to IES, the awardee, and, where appropriate, the team or project name commonly used in the research literature to identify the team's work.<sup>1</sup> We present the teams in the order that reflects the age/grade span that was central to their work, beginning with pre-school and concluding with the assessment team that addressed all age/grade spans.

### **The Language Bases of Reading Comprehension, The Ohio State University, LARRC**

This project investigated the role of lower- and higher-level language skills in the development of listening and reading comprehension for pre-K through grade 3 general education students, as well as English learner students. The team explored which language skills had the greatest leverage in promoting reading comprehension

---

<sup>1</sup> These descriptions are based on IES grant summaries, team websites, and research publications.

in grade 3. The Language and Reading Research Consortium (LARRC) team, based on the results of their longitudinal cognitive studies, developed a set of classroom-based interventions designed to increase comprehension skills; the intervention was called Let's Know! (LK). The LK curriculum used core content as a base for developing foundational reading comprehension with a systematic scope and sequence of instruction designed to build students' language skills with units spanning one academic year for each grade, pre-K through grade 3.

LARRC researchers conducted a culminating randomized controlled trial (RCT) to investigate the influence of the LK curriculum on students' comprehension and comprehension-related skills (comprehension monitoring, understanding narrative and expository text through inferencing and text structure knowledge) and vocabulary.

Research partners in this team include researchers at The Ohio State University, University of Nebraska–Lincoln, University of Kansas, Arizona State University, Florida State University, Lancaster University in the United Kingdom, and Massachusetts General Hospital Institute of Health Professions.

#### **Examining Effective Intervention Targets, Longitudinal Intensity, and Scaling Factors in Pre-K to Grade 5 Student Comprehension, Florida State University, FCRR**

The Florida Center for Reading Research (FCRR), using a cross-sectional longitudinal study, identified critical linguistic, cognitive, and basic word-level components of reading for understanding in pre-K through grade 6. They investigated differences in oral and text comprehension explained by these components and examined how earlier developing skills are related to later skills. Their goal was to look at early developing correlates of reading comprehension—the key points at which educators are able to make a difference—with a commitment to designing and evaluating the impact of interventions aimed at those key points.

Based on the data from the longitudinal study, the FCRR team created and evaluated several integrated, component (or multicomponent) instructional interventions, most focusing on one or more linguistic or cognitive skills to support students' proficient oral and text comprehension and reading for understanding in pre-K through grade 4. The collection of interventions is called Comprehension Tools for Teachers.

FCRR also worked with the ETS team to develop assessment tools (as described below) as well as developed the Florida Center for Reading Research Reading Assessment (FRA). With FRA K–2 and FRA Grades 3–12 assessments, FRA assesses the alphabetic principle, knowledge of word meanings or lexical quality, syntactic awareness, and reading comprehension.

#### **Catalyzing Comprehension Through Discussion and Debate, Strategic Education Research Partnership Institute, CCDD**

The Catalyzing Comprehension through Discussion and Debate (CCDD) team developed and evaluated multiple programs that rely on discussion and debate to catalyze the growth of academic language skills, perspective-taking ability, and complex reasoning for students in grades 4–8. The researchers argued that, compared to

elementary school, at the middle school level, students read topics that may be less compelling to them, and sentences and words are used to present more complex ideas, with more unfamiliar and polysemous words, and more metaphors.

The team developed two cross-content programs to use discussion and debate to support and develop reading comprehension skills. The first suite of programs, Word Generation (WG), is a set of tier 1, cross-content-area programs for students in grades 4–8. The WG suite is comprised of WordGen Weekly, Science Generation, Social Studies Generation, and WordGen Elementary. WordGen Weekly is a middle school program that exposes students to academic vocabulary, builds perspective-taking skills by providing multiple viewpoints on high-interest, controversial topics, and motivates complex reasoning through the demands of discussion, debate, and writing. To extend WG into more in-depth treatment of content-area topics in middle school, the team developed units in social studies and science. The team also developed a tier 2 program—the Strategic Adolescent Reading Intervention (STARI)—that targets middle school students reading several grade levels below expectation. It is intended to build their deep comprehension skills at the same time as more basic reading skills are addressed. STARI also relies on high-interest topics and uses discussion and debate to actively engage students in perspective taking, complex reasoning, and the use of academic language. The team also developed professional development materials needed to implement WG and STARI. CCDD work culminated with an RCT to evaluate the impacts of two refined and extended versions of WG on grades 4–7 students’ learning outcomes over two academic years and an RCT to examine the impact of STARI versus business as usual in middle schools.

Partners in this team include researchers at SERP, Harvard, and Lectica.

**Understanding Malleable Cognitive Processes and  
Integrated Comprehension Interventions for Grades 7–12,  
The University of Texas at Austin, PACT**

The Promoting Adolescents’ Comprehension of Text (PACT) project was designed to study the roles of cognitive processes, motivation, and engagement in reading comprehension and develop interventions, based on the understandings of these roles, to improve reading comprehension specifically for students with reading comprehension difficulties in grades 7–12. In order to design the interventions, the team had to gain a better understanding of “malleable” factors that distinguish good from poor comprehenders and aim interventions at those factors. Thus, prior to developing interventions, PACT spent significant time identifying these distinguishing factors and determined, among other findings, that motivation is important (students need to believe that they can become better readers) and that support for inference making from texts is critical.

The team focused interventions on English language arts classrooms as well as in the content area of social studies for grade 8 students. PACT researchers developed two major multicomponent interventions: PACT, with a focus on reading comprehension and knowledge acquisition within history classes, and Comprehension Circuit Training (CCT), which incorporated word identification, vocabulary enhancement, and comprehension and metacognition strategy development, within English language arts

classrooms. A major component of both PACT and CCT was team-based learning, a collaborative structure for promoting student-to-student support of learning. The team's work culminated with three RCTs in the area of grade 8 American history.

Research partners in this team include researchers from The University of Texas at Austin, Texas A&M University, University of Texas Health Science Center, University of Houston, and Florida State University.

**Reading for Understanding Across Grades 6–12:  
Evidence-Based Argumentation for Disciplinary Learning,  
University of Illinois at Chicago, Project READI**

Project READI (Reading, Evidence, and Argumentation in Disciplinary Instruction) took up the challenge of designing and researching learning environments that would support adolescent students in building the requisite knowledge, strategies, and dispositions that comprise 21st-century competencies as learners engage in evidence-based argumentation across multiple information resources. They proposed to develop instructional interventions in three disciplines (history, science, and literary analysis) for adolescent learners in grades 6–12. This team was concerned with how students select, analyze, synthesize, and evaluate information from text for purposes of accomplishing tasks that are authentic to the epistemic aims of each discipline. The rationale for Project READI was two-fold: (1) citizens must engage with multiple information resources (e.g., traditional text, multimedia, and graphics and other forms of visual representations) to accomplish academic, professional, and personal goals; and (2) national and international indicators show that current educational practices are not producing citizens with the skills to do so effectively. The READI team argued that there are multiple reasons for this, including increased demands of the information resources (hereafter referred to as texts) that convey disciplinary concepts and principles, and the absence of explicit instructional attention to these conceptual and textual demands, in conjunction with failure to recognize that different disciplines present different sources of conceptual and textual difficulty for adolescents (Goldman, 2012; Goldman et al., 2016; Schoenbach & Greenleaf, 2009). The consequences of lack of attention to differences among disciplines in the literacy demands of the texts, tasks, and purposes of reading more often than not result in content-area teachers assuming that “reading is reading” and reading instruction is the job of the English teacher. Consequently, the READI team asserts that adolescents are never taught *how* to read in the various disciplines in which they are being asked to read, nor how the goals of reading are different in different disciplines. The goal of Project READI was to develop and investigate approaches to improving learning in each discipline by focusing on the knowledge, heuristics, discourse, and reading practices relied upon in sense making and argumentation in literary analysis, history, and science.

Primary research partners in this team include researchers at the University of Illinois at Chicago, Northern Illinois University, Northwestern University, WestEd, and Inquirium LLC. Additional partnering researchers were at DePaul University, University of Chicago, and University of Pennsylvania.

### **Assessing Reading for Understanding: A Theory-Based, Developmental Approach, ETS**

When funding the RfU initiative, IES determined that it would fund one team to develop a new summative assessment of reading comprehension in pre-K through grade 12. This team, led by ETS, developed and evaluated a new system of assessments for pre-K through grade 12 students. ETS strove to design assessments that are aligned with current theoretical constructs and empirical findings pertaining to both reading comprehension and performance moderators, are sensitive to changes in development in reading comprehension, emphasize strategic reading processes empirically supported in the literature, provide greater information for guiding instruction (especially for students struggling to reach proficiency), and are comprised of texts and tasks that represent a range of purposeful literacy activities in which 21st-century students are expected to read texts for understanding. The assessments culminating from their research and development are scenario based and technology rich, focus on collaboration and communication, include meaningful structure and sequence, and include component measurements.

The construct ETS chose to evaluate was identified broadly as *reading literacy* and was measured by two assessment types: (1) components of reading, focusing on foundational skills, assessed with the Reading Inventory and Scholastic Evaluation, and (2) global reading literacy, focusing on higher-level and goal-directed reading comprehension, assessed with the Global Integrated Scenario-Based Assessment (GISA).

Research partners include researchers at Florida State University/FCRR, Arizona State University, and Northern Illinois University.

### **SETTING THE CONTEXT FOR THE RFU EFFORT**

The RfU call did not spring from fallow ground; in fact, there were a number of initiatives that provided context for and, indeed, motivated the call for the RfU project. As Dr. Karen Douglas, the project officer for the RfU grants, noted:

We knew researchers had made progress on the more fundamental skills of reading (e.g., decoding) but that did not lead to better reading comprehension across the age levels. IES had the opportunity to bring major resources to a problem and they chose reading because it is so important for learning and life, but also because they felt that the reading field was deep and broad enough, and sufficiently advanced as a research area. (personal communication, June 7, 2017)

We review a sample of these projects for the purpose of providing backdrop and characterizing the Zeitgeist at the time of the RfU call. However, not everything that was “in the air” at the time of the RfU Request for Application (RfA) influenced the RfU call or the RfU research. Furthermore, there was not a natural “progression” to the unfolding of ideas across these efforts. We present the work as a chronology, selecting those deliberations and findings that are germane to setting the stage for the RfU grants.

For a comprehensive treatment of the history of reading comprehension, the reader is referred to Pearson and Cervetti (2017). They propose four periods:

[the first of which] tracks the evolution of reading comprehension instruction before the beginning of the revolution in cognitive psychology that led to a paradigm shift in how we think about comprehension and its instruction—roughly the first 75 years of the 20th century. The second period is a short 15 years, from 1975 to the early 1990s; it examines the theoretical and research bases of the instructional activities and routines spawned by the cognitive revolution. The third period is even shorter, from the early 1990s, but with strong roots in the 1980s and even the 1970s, to the end the Bush administration and the dominance of No Child Left Behind. And the fourth and final period, while it has roots in the 1970s, 1990s, and 2000s, comes into relief in 2010 with the publication of the CCSS [Common Core State Standards]. (p. 13)

In this chapter we focus principally on the third period, which was most contemporaneous with the RfU initiative.

### **The National Research Council Report *Preventing Reading Difficulties in Young Children***

The committee that generated the 1998 *Preventing Reading Difficulties in Young Children* report (NRC, 1998) was convened by the National Academy of Sciences at the request of the U.S. Department of Education and the U.S. Department of Health and Human Services. The charge to this interdisciplinary group was to identify, through a consensus-building process, the effectiveness of interventions designed for young children at risk of having difficulties learning to read and to translate those research findings for parents, educators, publishers, and others.

The study group concluded that effective reading instruction is built on a foundation that assumed that reading ability is determined by multiple factors. Furthermore, they asserted that adequate initial reading instruction required that children use reading to obtain meaning from print; have frequent and intensive opportunities to read; are exposed to frequent, regular spelling-sound relationships; learn about the nature of the alphabetic writing system; and understand the structure of spoken words (NRC, 1998, p. 3). They further suggested that adequate progress to learn to read beyond the initial level depended on having a working understanding of how sounds are represented alphabetically, sufficient practice to achieve reading fluency with a range of texts, sufficient background knowledge and vocabulary to render written texts meaningful and interesting, control over procedures for monitoring comprehension and repairing misunderstandings, and interest and motivation to read for a variety of purposes (NRC, 1998, p. 4).

Given its charge, the committee further suggested that children most likely to experience reading difficulties were those who entered school with less prior knowledge and fewer skills in the areas of general verbal ability, attending to the sounds of language, familiarity with the basic purposes and mechanisms of reading, and letter knowledge. At the heart of the committee's recommendations was the critical importance of providing excellent reading instruction to all children—instruction that would only be enabled by well-prepared, knowledgeable, and well-supported teachers. Furthermore,

adopting a systems perspective, the committee acknowledged that schools needed to be organized to optimally support the instruction advocated (through curriculum and support services) and that students' home languages needed to be taken into consideration when planning instruction.

Specific to comprehension, the committee referenced van Dijk and Kintsch (1983), calling out the distinction between the reader's understanding of the text base (i.e., what the text says) and the situation model (i.e., what the text is about). They acknowledged that concept development and knowledge of word meanings are important parts of comprehension. Prefiguring a prominent focus in the RfU research, the *Preventing Reading Difficulties* report acknowledged that many basic cognitive processes were shared during reading and listening, including syntactic and inferential processes, as well as background knowledge and word knowledge. They noted that the correlation between reading and listening rose from grades 1–6. However, the committee introduced three cautions when interpreting data regarding the relationship between listening and reading comprehension. First, there are fundamental differences between written and oral language in terms of their social processes. Second, high correlations between reading and listening comprehension occur after the child has learned to decode. And the final caution is that correlations are useful to understanding variations across a population and not within specific individuals; hence, the gap between specific children's listening and reading comprehension could, in fact, be quite large even while the correlation between the two, generally speaking, is quite high.

### **The National Reading Panel**

The National Reading Panel (NRP) began its work in 1998, which was the year that *Preventing Reading Difficulties in Young Children* was published. In an unprecedented move, the National Institute of Child Health and Human Development (NICHD) was charged by President Clinton and the U.S. Congress to gather a diverse group of scientists and practitioners to identify research findings specific to the best ways to teach reading. In contrast to being charged to come to consensus (as was true of the *Preventing Reading Difficulties* committee), the NRP was charged with conducting a systematic review of the empirical literature germane to reading instruction.

While 30 topics were initially considered for inclusion, the panelists determined that there was an adequate research base to address findings in six areas: phonemic awareness, phonics, oral reading fluency, encouraging children to read, vocabulary, and comprehension strategies. They focused on research in grades K–12.

In this report, reading comprehension was defined as the act of understanding and interpreting the information in text. The panelists concluded that there were many avenues to enhancing reading comprehension, including through the teaching of phonemic awareness, phonics, oral reading fluency, and vocabulary. The panel reviewed 205 studies of reading comprehension instruction; typically, these were studies of strategy instruction taught singly or in a combination. The panel reported finding evidence for positive effects of teaching: question asking, monitoring, summarizing, story mapping, the use of graphic organizers, and cooperative grouping. Furthermore, they reported that the most powerful effects were obtained when multiple strategies were taught together. In contrast to the *Preventing Reading Difficulties* report, which treated reading



in a wholistic manner, discussing the interplay of the component skills of reading, the NRP report was organized by the six areas on which it focused, muting the synergistic nature of these areas. The 2000 NRP report became the cornerstone of the Reading First program, which we describe next.

### **Educational Policy Specific to Reading Instruction in the United States**

By the early 1990s, recognition of policy incoherence as an obstacle to educational reform in the United States led to a systemic reform movement: a logic of improvement focused on using a small set of policy instruments (e.g., content standards, performance standards, and accountability assessments) to coordinate system-wide reform activity (Fuhrman, 1991; Smith & O'Day, 1991). The logic of systemic reform was instrumental in shaping a series of federal policies that sought to effect coordinated, integrated improvements, including the Reading Excellence Act of 1998, and the reauthorization of the Improving America's Schools Act as the No Child Left Behind Act of 2001 (NCLB). One of the cornerstone programs of NCLB was the Reading First (RF) program. NCLB has been widely recognized as the most ambitious federal intervention into K–12 schooling in the history of U.S. public education, with operational implications for states, districts, and schools.

The Reading First program sought to promote instructional practices that had been validated by scientific research, which was explicitly defined in the legislation (NCLB). The Act legislated that RF funding was to be used for (1) reading curricula and materials that focus on the five essential components of reading instruction, as identified by the NRP (NICHD, 2000): (a) phonemic awareness, (b) phonics, (c) vocabulary, (d) fluency, and (e) comprehension; (2) professional development and coaching for teachers regarding how to use scientifically based reading practices, and how to work with struggling readers; and (3) diagnosis and prevention of early reading difficulties through student screening, interventions for struggling readers, and monitoring of student progress. States were permitted flexibility with regard to allocating resources across these three categories, and local decisions could be made regarding specific choices within the categories (i.e., which curricula, assessments, and models of professional development would be used). The RF grants were made available to states between July 2002 and September 2003. By April 2007, states had awarded subgrants to 1,809 school districts, which had provided funds to 5,880 schools. By design, districts and schools demonstrating the greatest need, as measured by student reading proficiency and poverty status, were to receive the highest funding priority.

What did we learn? The Reading First Impact Study (Gamse, Jacob, Horst, Boulay, & Unlu, 2008) used a regression discontinuity design to control statistically for all systematic preexisting differences between the two groups of schools being compared in the study: those that received RF funds, and those that were eligible for funding but did not receive funds. In this manner, non-RF schools played the same role as control schools would play in a randomized experiment. There were 18 study sites: 17 school districts and 1 statewide program.

Direct observations and surveys to assess instruction and program implementation revealed that RF produced a positive and significant impact on the amount of instructional time spent on the five essential components in grades 1 and 2. The impact was

equivalent to an effect size of 0.33 standard deviation in grade 1 and 0.46 in grade 2. In addition, RF produced a positive and significant impact on multiple practices promoted by the program, including professional development, support from coaches, amount of reading instruction, and supports for struggling readers. RF produced a positive and significant impact on decoding (using the Test of Silent Word Reading Fluency) among first graders tested in one school year (spring 2007), with an effect size of 0.17 standard deviation. However, RF did not produce a significant impact on student reading comprehension test scores (measured using the Stanford Achievement Test) in grades 1, 2, or 3. Furthermore, the average grades 1, 2, and 3 student in RF schools was reading at the 44th, 39th, and 39th percentile, respectively, on the end-of-year assessment.

The failure to find any effect of RF on reading comprehension was, of course, a disappointing outcome considering the \$6 billion investment that RF represented. On the other hand, there were lessons to be learned from this initiative that had important implications for future large-scale efforts to improve instruction for struggling readers. While RF did change the amount of time dedicated to reading instruction, as well as the nature of teacher practices, RF had no statistically significant impact on student engagement with print; it has long been recognized that opportunities for students to read self-selected text, and to read widely, has a significant effect on reading achievement (e.g., Nagy, Herman, & Anderson, 1985). Furthermore, the RF impact study found no evidence of differentiated instruction for struggling readers. As we will see when reviewing the RfU research, differentiation of reading instruction is necessary to optimize student achievement.

In addition, the National Reading Panel report, which shaped the architecture of RF interventions, represented comprehension instruction in terms of the teaching of individual strategies, noting seven strategies in particular. A number of reading researchers have expressed concern regarding the appropriate place of strategy instruction in the teaching of reading comprehension (e.g., McKeown, Beck, & Blake, 2009). When first conceived, strategy instruction was designed to engage readers in monitoring how well they were understanding text and to support them in regulating their reading of text for the purpose of building meaning. Strategies were designed to be a means to an end—comprehension—and not an end in and of themselves. The RF Impact Study was not designed to assess the quality of comprehension instruction; hence, a hypothesis to be explored in the RfU research was whether teachers engaged in forms of comprehension instruction that did not promote understanding and learning from text, and how to optimize the teaching of strategic reading. Related to this point is the fact that the focus on language arts instruction, legislated by RF, reduced the amount of time that primary grade students spent learning science and social studies content. Given the role that knowledge plays in supporting comprehension, this was an unfortunate by-product of RF.

An additional conceivable explanation for the disappointing finding regarding comprehension is that the primary measure of reading deployed across districts implementing RF was the Dynamic Indicators of Basic Early Literacy Skills, a measure that places a premium on reading speed rather than comprehension. Finally, while the *Preventing Reading Difficulties* report signaled the importance of attending to students' home language in planning reading instruction, the National Reading Panel report, as well as RF, was virtually silent on the instruction of English language learners.

In closing, RF changed the struggling reader landscape; despite its limitations, it focused attention on the need for teachers to receive professional development specific to early reading instruction and struggling readers. In addition, it acknowledged that classroom-level change in curriculum and instructional practice is key to improving the performance of struggling readers. It certainly provided grist for the RfU effort in terms of the questions it raised about what constitutes quality comprehension instruction and how comprehension should be assessed.

### **The New London Group's *A Pedagogy of Multiliteracies***

In 1996, an international group of 10 scholars met and spent more than 1 year developing new ways of thinking and talking about the rapidly changing social contexts of literacy learning and teaching. Building on concepts advanced by the critical literacy and New Literacy Studies movements, the New London Group (NLG) proposed the term *multiliteracies* to summarize the confluence of increased cultural and linguistic diversity, globalization, and rapid advancements in technologies for multimodal communication.

The NLG focused on the multiplicity and value of diverse linguistic, cultural, communicative, and technological resources and argued that one literacy—that is, academic literacy—cannot support all of the communicative needs of varied social groups. A multiliteracies approach viewed literacy as dependent on contexts, purposes, tools, and skills available for meaning making, and being literate as the ability to create and comprehend meanings made available through multimodal forms of communication.

One outcome of the work of this group was the articulation of a pedagogic model that addressed how these significant sociocultural changes were affecting and, indeed, challenging classroom teaching and learning. Referred to as *A Pedagogy of Multiliteracies* (NLG, 1996), it promoted a critical, socially just orientation to teaching that explicitly acknowledges possible convergences of diversity and multimodal communication channels that can empower young people and pave the way for new social futures.

There were numerous implications from the work of this group that relate to comprehension, including expanding notions of text; attending to the relationships among school-based learning, work life, citizenship, and private life; and calls for literacy pedagogies that turn interpretive authority over to students, support students to connect goal-driven meaning making to action, and foster students' critical understanding of text.

A central concern of the NLG was "the plurality of texts that circulate" in "increasingly globalized societies" (NLG, 1996, p. 61). Countering traditional notions of texts (e.g., alphabetic texts), the NLG advanced a more expansive notion of text that encompassed "the burgeoning variety of text forms associated with information and multimedia technologies," as well as "representational forms that are becoming increasingly significant in the overall communications environment, such as visual images" (p. 61). Drawing on a social semiotic lens (Kress & Van Leeuwen, 2001), the NLG underscored the salience of multiple communicative modes for meaning making, including image, gesture, sound, written language, speech, gaze, and music. An aim of the multiliteracies pedagogic model was to attend to this plurality of meaning-making resources and to expand conceptions of text to include the multiple and multimodal texts that young

people construct and comprehend through a wide range of everyday literacy practices. As the NLG explained, multiliteracies “create a different kind of pedagogy, one in which language and other modes of meaning are dynamic representational resources, constantly being remade by their users as they work to achieve their various cultural purposes” (p. 64).

What does the NLG have to do with the RfU initiative? The NLG makes salient the fact that the RfU call was focused on traditional academic texts, rather than the expansive view of texts proposed by the NLG; there was, however, some RfU research and development that did hew to the NLG call, namely, the work conducted by READI. We revisit the NLG in Chapter 6, in which we consider future directions for reading comprehension research.

### The RAND Reading Study Group

The RAND Reading Study Group (RRSG) was charged by the U.S. Department of Education’s Office of Educational Research and Improvement to propose strategic guidelines for a long-term research and development program specifically for the purpose of supporting the improvement of reading comprehension. The 14 experts who served on this study group represented a range of disciplinary and methodological perspectives. The group convened in 1999 and published its report, *Reading for Understanding: Toward an R&D Program in Reading Comprehension*, in 2002.

As we describe the contributions of the RfU teams, we make reference to this document; in this introduction, we identify several contributions of the RRSG. One contribution was its definition of reading comprehension as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (RRSG, 2002, p. 11). Furthermore, the study group proposed a heuristic suggesting that comprehension is influenced by the interaction of the reader who is doing the comprehending, the text that is to be comprehended, and the activity in which comprehension is a part, all of which occur within larger sociocultural contexts that influence and are influenced by the reader and that interact with each of the three elements. The Study Group wondered what a focused program of research might reveal about the influence of each of these factors on comprehension. In fact, in a 36-page appendix, the Study Group identified a host of dimensions associated with variability in readers, text, and activity that might be productively examined in depth.

The RRSG called for a significant focus on classroom instruction grounded in the belief that good instruction is the most powerful means of developing proficient comprehenders and preventing the development of reading comprehension problems. Indeed, the RfU RfA required that each grantee’s program of research culminate in an efficacy study of comprehension instruction, informed by the findings of each team’s research on the malleable factors that contribute to comprehension. Dr. Karen Douglas, the IES project officer for the RfU grant, noted that “The RAND report wasn’t explicitly referred to in the RfA, but . . . it was an important resource for thinking about the aspects of reading instruction that should be considered in improving reading comprehension” (personal communication, June 17, 2017).

### Initiatives Regarding Adolescent Literacy

Snow and Moje (2010) published an essay in *Phi Delta Kappan* titled “Why Is Everyone Talking About Adolescent Literacy?” Such a title might sound a bit hyperbolic, but, indeed, as the RfU research was launching, there was a good deal of activity in the area of adolescent literacy, much of it driven by the recognition that adolescent literacy had received short shrift in preceding initiatives. In addition, there was increased disenchantment with generic content-area reading approaches. Snow and Moje (2010) argued that there were three components essential to the design of literacy instruction for adolescents: (1) continued development of general language and literacy skills, (2) incorporating literacy into content-area instruction, and (3) supporting struggling readers. In addition, the field was beginning to attend to the developmental needs of adolescents in terms of what would be both motivating and engaging to youth, as well as likely to prepare them for postsecondary education and career readiness. For example, Moje and colleagues documented the ways adolescents used reading and writing to participate in social networks and to enhance their social capital (e.g., Moje, Overby, Tysvaer, & Morris, 2008).

Disciplinary literacy, as contrasted with content-area literacy, arose, in part, in response to an emerging consensus that decades of research and best practices regarding content-area literacy had not done enough to influence instruction or outcomes in content-area learning at the secondary level (Carnegie Council on Advancing Adolescent Literacy, 2010; Heller & Greenleaf, 2007; Schoenbach & Greenleaf, 2009; Wise, 2009). These reports pointed to the importance of teaching discipline-specific reading and writing, as well as continuing to support students’ development of literacy skills beyond the early elementary years. These calls were bolstered by research, such as the study reported by Shanahan and Shanahan (2008), documenting how reading approaches differ among the disciplines and demonstrating the need for reading strategies that would engage readers in internalizing the core principles of the disciplines. Moje (2008), for example, argued that a disciplinary perspective involved a “turn toward literacy as an essential aspect of disciplinary learning” such that literacy “becomes an essential aspect of disciplinary practice.”

There was another landmark report and policy initiative germane to adolescent literacy that also deserves attention. *Reading Next*, commissioned by the Carnegie Corporation of New York (Biancarosa & Snow, 2004), reported on the results of a consensus study of adolescent literacy research. The report begins by making a persuasive case for the importance of focusing on adolescent literacy, including the facts that (around 2004) more than 8 million students in grades 4–12 were classified as “struggling readers,” and that the majority of the more than 3,000 students who dropped out of high school were most often identified as lacking the literacy skills to meet increasingly complex demands associated with literacy in contemporary society. The authors concluded that the problem of adolescent literacy attainment was both complex and multifaceted and called for the careful blending of instructional and educational infrastructure solutions. Instructional improvements meriting attention included direct, explicit comprehension instruction; effective instructional principles embedded in content; attention to motivation and self-directed learning; opportunities for text-based collaborative learning; the use of strategic tutoring; the inclusion of diverse texts; intensive writing;

the use of technological tools as scaffolds and technology as content; and ongoing formative assessment of students. The suggested infrastructure improvements included extended time for literacy engagement, professional development for middle school and secondary educators, ongoing summative assessment of students and programs, and leadership committed to a comprehensive and coordinated literacy program.

In 2006, shortly before the launch of the RfU initiative, Striving Readers Comprehensive Literacy discretionary grants were awarded by the U.S. Department of Education, on a competitive basis, to states, which in turn awarded funding to local educational agencies. The awards were initially to be used to mount comprehensive school-wide literacy programs to advance literacy skills—including preliteracy skills, reading, and writing—for students from birth through grade 12, including limited-English-proficient students and students with disabilities. By 2009, the awards were dedicated to supplementary literacy interventions targeted at children and youth reading significantly below grade level with a particular focus on middle and high school students' literacy levels in Title I-eligible schools. Furthermore, the goal for the duration of the grant was to build a strong, scientific research base for identifying and replicating strategies that improve adolescent literacy skills. There was a total of 16 Striving Readers grantees and there were 10 different reading interventions that were studied in grades 6–10. Independent evaluators, using the What Works Clearinghouse criteria, determined that 12 studies met criteria without reservations, 3 met criteria with reservations, and 2 did not meet criteria. The results were complex and spoke clearly to the contextual issues that stand to influence the outcomes of any intervention; of the 10 interventions studied, 6 had no discernible effects, and the remaining 4 had mixed effects. These findings prefigure outcomes of the RfU studies that inform our understanding of contextual features that influence reading comprehension.

### **Evolving Ideas About the Nature of Reading Comprehension Reflected in Assessments**

In addition to the landmark reports and policy initiatives described thus far, another source of evidence regarding evolving conceptualizations of reading comprehension can be derived from large-scale assessments. The Reading Framework of the National Assessment of Educational Progress (NAEP) (NAGB, 2017) is regularly updated using expert consensus regarding relevant research findings. Examination of the NAEP definition of reading demonstrates the field's evolving conceptualization of comprehension. The 1992–2000 NAEP Reading Framework proposed that reading comprehension was comprised of the following "Reading Stances": (a) *initial understanding*, the preliminary consideration of the text as a whole; (b) *developing an interpretation*, discerning connections and relationships among ideas within the text; (c) *personal reflection and response*, relating personal knowledge to text ideas; and (d) *critical stance*, standing apart from the text to consider it objectively (NAGB, 1992). This characterization reflects the influence of research and theories from the fields of information processing, cognition, and literary criticism and reflects a static product of reading; reading "ends" when comprehension of text is attained—a sort of "comprehension for comprehension's sake."

In contrast, the current NAEP Reading Framework, which is the same as the 2009 NAEP Reading Framework, maintains a focus on reading comprehension as the construction of meaning with text, but includes the following:

“Reading is an active and complex process that involves

- Understanding written text,
- Developing and interpreting meaning, and
- *Using meaning as appropriate to type of text, purpose, and situation*”<sup>2</sup> (NAGB, 2017).

The current definition represents a significant change in perspectives on comprehension, and reading for understanding; reading involves not only the *construction of meaning*, but also the *use of the meaning that is constructed*. Readers’ strategies, skills, reading experiences, and domain prior knowledge fuel the meaning making that results in comprehension. Furthermore, the current NAEP Framework acknowledges what readers do with comprehension; for example, they analyze text content (Prior & Bazerman, 2004), identify claims and supporting evidence (NRC, 2005), apply what they learn from text to solve problems and ask questions (Hinchman & Appleman, 2017), establish epistemic stances toward the processes and contents of reading (Bråten, Braasch, Strømsø, & Ferguson, 2015), synthesize information from within and across texts (Coté, Goldman, & Saul, 1998), interrogate author motive and craft (Beck, McKeown, Sandora, Kucan, & Worthy, 1996), and critique and evaluate text contents and structures (Vasquez, Harste, & Albers, 2010). This expanded notion of comprehension has important implications for instruction as well as the assessment of comprehension, which we will see reflected in a number of the RfU studies.

### Static Reading Scores in the Face of Rising Demands for Comprehension

To further appreciate the impetus for the 2009 RfU call, it is instructive to consider trends in NAEP reading scores in grades 4, 8, and 12. While the 2009 average grade 4 reading score (221 on a 0 to 500-point scale) was statistically significantly higher than the 2002 score (219), the average grade 8 score was the same (at 264) in 2002 and 2009, and the average grade 12 score was statistically the same in 2002 and 2009, according to data from the NAEP. Moreover, on the NAEP in 2009, only 33 percent of grade 4 students, 33 percent of grade 8 students, and 38 percent of grade 12 students were determined to be proficient or higher in reading skills.<sup>3</sup> In both reading and content domains that demand reading, NAEP student scores are flat and suggest that substantial numbers of students struggle to achieve basic levels of reading comprehension. For example, on the 2010 NAEP in the area of United States history, only 20 percent of fourth graders, 17 percent of eighth graders, and 12 percent of 12th graders demonstrated proficiency (NCES, 2011). The intractability of performance on comprehension measures, in hand with increasing expectations both within school and in the postsecondary worlds of work and tertiary education (NGA & CCSSO, 2010), regarding (a) the types of complex

<sup>2</sup> Italic in quote added for emphasis.

<sup>3</sup> Retrieved from [https://www.nationsreportcard.gov/reading\\_2017/nation/scores?grade=4](https://www.nationsreportcard.gov/reading_2017/nation/scores?grade=4); [https://www.nationsreportcard.gov/reading\\_2017/nation/scores?grade=8](https://www.nationsreportcard.gov/reading_2017/nation/scores?grade=8); [https://www.nationsreportcard.gov/reading\\_math\\_g12\\_2015/#reading](https://www.nationsreportcard.gov/reading_math_g12_2015/#reading); and [https://nces.ed.gov/programs/coe/pdf/coe\\_cnb.pdf](https://nces.ed.gov/programs/coe/pdf/coe_cnb.pdf).

texts students will encounter, and (b) the complex purposes for using text, brought into sharper relief the importance of a rigorous and comprehensive program of research to study the development, instruction, and assessment of comprehension.

### Theoretical Perspectives of Reading Instruction

The RfU RfA noted that the prevailing theory informing reading instruction at the time of the call was grounded in the “Simple View of Reading” (Gough & Tunmer, 1986). This model suggests that reading comprehension emerges from two distinct—but requisite—strands of knowledge: (1) word recognition and language comprehension skills, and (2) the skills necessary to integrate oral language knowledge with word recognition skills. In conceptualizing the RfU RfA, IES extended the Simple View to include text processing skills. Citing Perfetti’s (1999) cognitive model of reading comprehension, the architects of the call noted that reading comprehension depends on word knowledge to support both word recognition and comprehension processes, with word recognition including a mapping from the visual presentation of the word to the phonological representation of the same word. Furthermore, Perfetti’s model signaled the role of mapping the visual representation of the word to the word’s meaning. In this manner, the process of word recognition informs (and is informed by) comprehension processes. As the RfU RfA concluded: “comprehension processes, in turn, depend upon the reader’s ability to use word level information to build a representation of the text being made, to draw inferences from the text, and to represent the meaning of the text . . . comprehension depends upon the reader’s linguistic and general knowledge” (IES, 2009, p. 7). As the reader will soon discern, RfU researchers continued to invoke—and problematize—the Simple View of Reading.

### CONCLUSION

In summary, as we entered the RfU era, the field was informed by substantial research-based knowledge of reading comprehension. From the 1970s to the 1990s, we had gained increased understanding of how comprehension was orchestrated by readers as a process with many constituent parts (Anderson et al., 1985; Pressley & Afflerbach, 1995). We were, with the help of sociocultural perspectives (Gee, 2001; Luke, 1991; Purcell-Gates, Melzi, Najafi, & Orellana, 2011), gaining knowledge of the contexts in which comprehension may be best taught—or learned, and used. Yet, this research and theory had not mattered much in relation to improving many students’ comprehension performance. That was the context in which the RfU initiative was initiated. The RfU teams were asked to change the pattern of performance that fell short of expectations; it is to their work that we now turn our attention.

### REFERENCES

AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.



- Anderson, R. C., Hiebert, E. H., Scott, J. A., Wilkinson, I. A. G., Becker, W., & Becker, W. C. (1985). Becoming a nation of readers: The report of the Commission on Reading. *Education and Treatment of Children, 11*(4), 389–396.
- Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal, 96*(4), 385–414.
- Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.
- Bråten, I., Braasch, J. L., Strømsø, H. I., & Ferguson, L. E. (2015). Establishing trustworthiness when students read multiple documents containing conflicting scientific evidence. *Reading Psychology, 36*(4), 315–349.
- Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York: Carnegie Corporation of New York.
- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*(1), 1–53.
- Fuhrman, S. H. (Ed.). (1991). *Designing coherent education policy: Improving the system*. San Francisco: Jossey-Bass.
- Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B. & Unlu, F. (2008). *Reading First impact study final report (NCEE 2009-4039)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Gee, J. P. (2001). Reading as situated action: A sociocognitive perspective. *Journal of Adolescent & Adult Literacy, 44*(8), 714–725.
- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children, 22*(2), 89–116.
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., et al. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 51*(2), 219–246.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.
- Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Washington, DC: Alliance for Excellent Education.
- Hinchman, K. A., & Appleman, D. A. (Eds.). (2017). *Adolescent literacies: Handbook of practice-based research*. New York: Guilford Press.
- IES (Institute of Education Sciences). (2009). *Request for applications: Reading for Understanding research initiative, CFDA Number: 84.305F*. Retrieved from [https://ies.ed.gov/funding/pdf/2010\\_84305F.pdf](https://ies.ed.gov/funding/pdf/2010_84305F.pdf).
- Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Edward Arnold.
- Luke, A. (1991). Literacies as social practices. *English Education, 23*(3), 131–147.
- McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44*(3), 218–253.
- Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*(2), 96–107.
- Moje, E., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review, 78*(1), 107–154.
- NAGB (National Assessment Governing Board). (1992). *Reading framework for the 1992 National Assessment of Educational Progress*. Washington, DC: Author.
- NAGB. (2017). *Reading framework for the 2017 National Assessment of Educational Progress*. Washington, DC: Author. Retrieved from <https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/reading/2017-reading-framework.pdf>.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253.
- NCES (National Center for Education Statistics). (2011). *The nation's report card: U.S. history 2010 (NCES 2011-468)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

- NGA & CCSSO (National Governors Association Center for Best Practices and Council of Chief State School Officers). (2010). *Common Core State Standards*. Washington, DC: National Governors Association Center for Better Practices and Council of Chief State School Officers.
- NICHHD (National Institute of Child Health and Human Development). (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Department of Health and Human Services.
- NLG (New London Group). (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66, 60–92.
- NRC (National Research Council). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- NRC. (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: The National Academies Press.
- Pearson, P. D., & Cervetti, G. N. (2017). The roots of reading comprehension instruction. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (pp. 12–56). New York: Guilford Press.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). New York: Oxford University Press.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Prior, P. A., & Bazerman, C. (2004). *What writing does and how it does it: An introduction to analyzing texts and textual practices*. Hillsdale, NJ: Erlbaum.
- Purcell-Gates, V., Melzi, G., Najafi, B., & Orellana, M. F. (2011). Building literacy instruction from children's sociocultural worlds. *Child Development Perspectives*, 5(1), 22–27.
- RRSG (RAND Reading Study Group). (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Corporation.
- Schoenbach, R., & Greenleaf, C. (2009). Fostering adolescents' engaged academic literacy. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 98–112). New York: Guilford Press.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40–59.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman and B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 233–267). New York: Falmer Press.
- Snow, C., & Moje, E. (2010). Why is everyone talking about adolescent literacy? *Phi Delta Kappan*, 91(6), 66–69.
- van Dijk, T., & Kintsch, W. (1983). *Strategies for discourse comprehension*. San Diego, CA: Academic Press.
- Vasquez, V., Harste, J., & Albers, P. (2010). From the personal to the world wide web: Moving teachers into positions of critical interrogation. In E. A. Baker (Ed.), *The new literacies: Multiple perspectives on research and practice* (pp. 265–284). New York: Guilford Press.
- Wise, B. (2009). Adolescent literacy: The cornerstone of student success. *Journal of Adolescent & Adult Literacy*, 52(5), 369–375.



# Appendix 1-1

## Description of Procedural Elements for Developing This Report

This report summarizes hundreds of journal articles and technical reports produced over the course of several years and supported by approximately \$120 million in research funding from the U.S. Institute of Education Sciences (IES). Given the importance, broad scope, and mandate of this project, we, as the editors and authors of the report, believe it is imperative to create an archive in which we describe in detail the methods we employed in compiling and synthesizing this extensive body of work. We begin by restating our mandate (as expressed in Chapter 1), but we expand on the work of the steering committee and the editorial and writing teams. We then briefly describe specific methodological decisions made for each chapter. We hope, through this appendix, both to provide a transparent roadmap of what we did to maximize the integrity of the effort and to provide guidance for future projects.

### OVERVIEW

In 2009, IES announced the Reading for Understanding (RfU) research initiative. The RfU was a remarkably ambitious project. By educational research standards, it represented a huge investment (approximately \$120 million) in a fairly well-specified scope of work, which was identified as “(a) examining underlying processes of reading comprehension and identifying malleable processes that may be targets of interventions for enhancing reading comprehension, and (b) developing and testing interventions intended to improve reading comprehension” (RfA, 2009, p. 5). The ultimate goal defined in the call was to redress the disappointing performance of students in the United States on national assessments of reading.

Grant applicants were to identify whether they were applying to become a core team or an assessment team; core teams were to propose reading comprehension research that covered a range of at least five grades, while assessment teams were to design reading comprehension measures for pre-kindergarten (pre-K) through grade 12. Core team applicants were required to propose an iterative design process that would culminate in a reading comprehension intervention that would be the subject of an efficacy study or a randomized controlled trial (RCT); furthermore, core teams were expected to use the measures designed by the assessment team in their research.

Invoking the model used by the National Aeronautics and Space Administration (NASA) in its mission to the moon, the RfU research was to be conducted by multidisciplinary networked groups of researchers in partnership with practitioners. In the call, IES signaled that it would foster ongoing collaboration across the research groups for the duration of the 5-year awards in an effort to accelerate the pace of the research. Ultimately, awards were made to five core teams and one assessment team. Collectively,

the teams studied the development, instruction, and assessment of reading comprehension from pre-K through grade 12.

In 2016, following the 5-year award period, and as the RfU teams were continuing to analyze data and add to the more than 200 publications already generated, IES funded a National Academy of Education (NAEd) invited proposal, *Reaping the Rewards of the Reading for Understanding Initiative*, to lead an effort to:

- Articulate findings and common themes across the RfU projects to contribute to a full-range view of reading development;
- Identify obstacles to on-time reading achievement, as well as factors supporting success;
- Examine cross-project findings to identify areas of agreement and productive tension; and
- Find common principles underlying instructional programs across projects.

In that spirit, the NAEd engaged in a collaborative effort, bringing together representatives of each of the six teams, joined by others, including members of the Academy, to produce a summary report that was informed by the publications prepared by the RfU teams, as well as proceedings of meetings. This volume reports on the results of the NAEd effort.

We began our work by convening a 3-year steering committee with leaders of the six RfU projects and scholars in the field who were not directly involved in the RfU initiative. The 11-member steering committee was co-chaired by Annemarie Sullivan Palincsar (University of Michigan) and P. David Pearson (University of California, Berkeley). The six team directors were members of the steering committee: Susan Goldman, University of Illinois at Chicago (Reading, Evidence, and Argumentation in Disciplinary Instruction [READI]); Laura Justice, The Ohio State University (Language and Reading Research Consortium [LARRC]); Christopher Lonigan, Florida State University (Florida Center for Reading Research [FCRR]); John Sabatini, The University of Memphis (Educational Testing Service [ETS]); Catherine Snow, Harvard University and Strategic Education Research Partnership (SERP) (Catalyzing Comprehension through Discussion and Debate [CCDD]); and Sharon Vaughn, The University of Texas at Austin (Promoting Adolescents' Comprehension of Text [PACT]). Other steering committee members not directly involved in the RfU included Donald Compton (Florida State University), Kenji Hakuta (Stanford University), and Glynda Hull (University of California, Berkeley).

The steering committee guided the work of this report, including organizing the synthesis around the three main topics of development, assessment, and curriculum and instruction (C&I); securing relevant research articles to ground each topic; providing necessary feedback to identify common themes and findings; and reviewing work products, including the chapters in this volume.

The steering committee met on three occasions, each a 2-day, in-person meeting. The agenda of the first steering committee meeting (February 22–24, 2017) included several goals: (a) provide an update of the work of each RfU team; (b) fine-tune the goals of the project, including syntheses and dissemination activities; and (c) form three subcommittees and identify potential experts for commissioned papers. The steering committee identified three topics for the commissioned papers (development, assessment,

and C&I). The steering committee also expanded the scope of the final synthesis report, which, in addition to covering the work of the RfU projects through commissioned papers, would also address what we now know about reading comprehension that we did not know when the seminal RAND report was written in 2002 as well as reviewing literature on multimodal and digital literacies, which did not figure prominently in the RfU Request for Applications (RfA). In preparation for this additional coverage, Palincsar and Pearson agreed to oversee a literature review conducted by Jennifer Higgs (University of California, Berkeley) and Miranda Fitzgerald (University of North Carolina at Charlotte).

The steering committee decided to form subcommittees to guide the work of the commissioned papers, which were originally written to guide the work of the final synthesis, but, as described below, were in fact turned into the chapters of this volume. The paper topics and authors are (a) development, by Gina N. Cervetti, University of Michigan; (b) assessment, by Panayiota Kendeou, University of Minnesota; and (c) C&I, by Peter Afflerbach, University of Maryland. The subcommittees guiding the development of the commissioned papers consisted of (a) development: Annemarie Sullivan Palincsar (co-chair), Catherine Snow, Laura Justice, and Chris Lonigan; (b) assessment: P. David Pearson (co-chair), John Sabatini, and Don Compton; and (c) C&I: Annemarie Sullivan Palincsar (co-chair), P. David Pearson (co-chair), Susan Goldman, and Sharon Vaughn. Additionally, each paper author was provided with two consulting editors, one RfU team member and one NAEd member with expertise in the topic. The consulting editors were (a) development: Laura Justice (RfU member) and Walter Kintsch, University of Colorado Boulder (NAEd member); (b) assessment: John Sabatini (RfU member) and Joan Herman, National Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles (NAEd member); and (c) C&I: Carol Lee, Northwestern University (RfU member who is also a NAEd member) and Richard Anderson, University of Illinois (NAEd member).

In preparation for the work of drafting the commissioned papers, representatives from all of the RfU teams provided seminal papers in each of the three paper topic areas. These articles served as the foundation for the commissioned papers.

The second steering committee meeting was held March 1–2, 2018, to discuss the commissioned papers as well as the objectives for the final report. Prior to the second committee meeting, the synthesis paper authors shared drafts with their consulting editors, obtained feedback, and incorporated the feedback, as appropriate, into second drafts of their papers. These papers were then shared with the steering committee members prior to the second meeting in order to prepare for discussion and critique. We also engaged the assistance of Gina Biancarosa, University of Oregon, as a consultant to support the research, analysis, and written work for this initiative.

During the second steering committee meeting, the paper authors presented their papers, providing the following information to the committee: (a) what they examined, (b) how they synthesized/addressed the materials, (c) what they planned to continue to address to complete their papers, and (d) how the committee could assist them in completing the synthesis effort. Pre-identified steering committee members served as discussants for particular papers and provided initial feedback at the meeting. Discussion by the entire steering committee ensured that insights and implications for policy and practice across the RfUs were identified in each paper topic area.

After the second steering committee meeting, the editors of the final report (P. David Pearson, Annemarie Sullivan Palincsar, Gina Biancarosa, and Amy I. Berman) continued to provide feedback to the commissioned paper authors, as did the consulting editors. Additionally, the model for the final summary report was modified in response to issues that surfaced in the process of drafting the various chapters. The original model was to have three free-standing commissioned papers that would inform a single synthesis document. Instead, the co-chairs, after reaching out to the entire steering committee for feedback as well as the commissioned paper authors and Rebecca Kang McGill-Wilkinson (IES program officer), determined that a better course would be to have an edited volume with an introduction and conclusion written by the editors and the three papers repurposed as core chapters within the volume. This decision was largely informed by the realization that summarizing the commissioned paper summaries was not a worthwhile endeavor. The editors worked closely with the chapter authors to ensure consistent, accurate, and informed chapters. Moreover, it allowed the editorial team to focus on summarizing across chapters (see Chapter 6).

The editorial team met regularly by videoconference to oversee the development of the report and also convened for an intensive drafting and editing session in Berkeley, California, December 14–15, 2018. In particular, during this meeting, the editors focused on the content for the introduction and conclusion chapters of the report, as well as overseeing the successful review and integration of the three commissioned chapters.

The full steering committee met for the third time October 3–4, 2019. Prior to the meeting, a draft of the full volume (except for the Executive Summary) was shared with the steering committee. The purpose of this meeting was to closely review the completed and in-progress draft chapters and to develop an outline for the Executive Summary. The steering committee was charged with ensuring (a) the accuracy of the factual statements, (b) the validity and trustworthiness of the interpretive and evaluative claims and recommendations, and (c) the accessibility and usefulness of the report. Over 2 days, the steering committee discussed the report, chapter by chapter. The committee also discussed potential additions to the report, such as quotes from RfU team leaders and teachers. During the meeting and after it, the committee members continued to share relevant articles and information.

In addition to the extensive review delineated above and significant additional editing and review across chapters by authors and editors, the volume underwent NAEd internal review. Judith Warren Little, the chair of the NAEd Standing Review Committee, requested that NAEd members Kenji Hakuta and Glynda Hull review the entire volume paying special attention to the Executive Summary, and introductory and concluding chapters. Hakuta and Hull reviewed the volume and provided feedback that was incorporated into the report.

Table Appendix 1-1 provides a summary of the individuals and groups who were involved in writing and reviewing various parts of the report.

**TABLE APPENDIX 1-1** Authors and Reviewers for the Chapters of the Report

Chapter	Chapter Title	Author(s)	Review/Editorial Team	
			Steering Committee	Consulting Editors
	Executive Summary		Collaborated on by the entire committee as well as part of the final review of the volume by outside peer reviewers selected by the NAEd standing review committee (Kenji Hakuta and Glynda Hull)	
1	Introduction to the Reading for Understanding Initiative	Annemarie Sullivan Palincsar, P. David Pearson, Amy I. Berman, and Peter Afflerbach	Reviewed by the entire committee as well as part of the final review of the volume by outside peer reviewers selected by the NAEd standing review committee (Kenji Hakuta and Glynda Hull)	
2	The Nature and Development of Reading for Understanding	Gina N. Cervetti	Annemarie Sullivan Palincsar, Catherine Snow, Laura Justice, and Chris Lonigan	Laura Justice and Walter Kintsch
3	The Assessment of Reading for Understanding	Panayiota Kendeou	P. David Pearson, John Sabatini, and Don Compton	John Sabatini and Joan Herman
4	Teaching Reading for Understanding: Summarizing the Curriculum and Instruction Work of the Five Core Reading for Understanding Teams	Gina Biancarosa, Peter Afflerbach, and P. David Pearson	Annemarie Sullivan Palincsar, P. David Pearson, Susan Goldman, and Sharon Vaughn	Carol Lee and Richard Anderson
5	Teaching Reading for Understanding: Synthesis and Reflections on the Curriculum and Instruction Portfolio	Peter Afflerbach, Gina Biancarosa, Matthew Hurt, and P. David Pearson	Annemarie Sullivan Palincsar, P. David Pearson, Susan Goldman, and Sharon Vaughn	Carol Lee and Richard Anderson
6	Taking Stock of the Reading for Understanding Initiative	P. David Pearson, Annemarie Sullivan Palincsar, Peter Afflerbach, Gina N. Cervetti, Panayiota Kendeou, Gina Biancarosa, Jennifer Higgs, Miranda Fitzgerald, and Amy I. Berman	Reviewed by the entire committee as well as part of the final review of the volume by outside peer reviewers selected by the NAEd standing review committee (Kenji Hakuta and Glynda Hull)	

NOTE: All chapters were reviewed by the entire steering committee.



Now we turn to a more detailed description of the process followed by the core chapter authors.

## **Chapter 2: The Nature and Development of Reading for Understanding**

This review of the RfU research is based on a set of studies nominated by the RfU teams as those that addressed issues of the development of comprehension. From among the nominated papers, a small number were ultimately excluded because they focused on outcomes other than reading comprehension, such as word reading skills, reading fluency, or oral language. Following the initial nomination process, additional requests for newer studies were extended to the RfU teams, and searches of academic journal databases were used to identify relevant studies.

We summarized each of the identified studies. In addition, key information was recorded in a table, using descriptive statements and, in some cases, descriptive codes. The table included information regarding the focus (e.g., language, cognitive skills), participants, design, and findings.

A series of clusters was created by sorting the studies according to broad focus. We wrote descriptively about the findings within these clusters and also used a question-driven, qualitative examination to derive a set of overall findings. The analytic questions included the following:

- What skills and knowledge are concurrent correlates of reading comprehension at different stages of reading development?
- What characteristics distinguish more and less successful comprehenders?
- What are early predictors of later comprehension?
- What is the dimensionality of enabling skills that underlie successful comprehension?

Guided by these questions, we used a variation of the constant comparative method, in which we developed themes, read further into the data, and either added evidence or refined the themes accordingly.

## **Chapter 3: The Assessment of Reading for Understanding**

Writing of the assessment chapter followed a systematic, iterative, and integrative approach that focused on the minimum assessment criteria put forth by the RAND Research Study Group (2002), current trends in reading comprehension research, and an in-depth review of each assessment. The review of each assessment (what is now included as an appendix to Chapter 3) focused on the conceptual framework guiding development, content and sample items, administration and scoring guidelines, and evidence for technical quality focusing specifically on validity, reliability and precision, fairness in testing, and intended use of scores. A distinction was made between the assessments that emerged from the core assessment mission versus those developed to allow researchers to measure key facets of their interventions. Nonetheless, we applied the same standards to all of the assessments introduced in this chapter. Through this iterative, integrative evaluative process, nine themes emerged that summarize the contributions of the RfU assessment research. The chapter includes a discussion of each of those theme contributions as well as directions for future research.

**Chapter 4: Teaching Reading for Understanding: Summarizing the Curriculum and Instruction Work of the Five Core Reading for Understanding Teams and Chapter 5: Teaching Reading for Understanding: Synthesis and Reflections on the Curriculum and Instruction Portfolio**

When we began this effort we anticipated one chapter on C&I. However, once the scholar recruited to draft the C&I chapter, as well as the editors, began to grasp the C&I portfolio it became clear that one chapter could not capture the volume and complexity of the literature.

Additionally, in trying to follow the model of synthesizing across the work of the five teams, we felt we were losing the “identity” of the C&I portfolio of each team. When we were operating at a very high level of synthesis, we found ourselves moving toward high-level generalizations that, in our view, did not convey a sense of the vividness and vitality of the approaches that had proved successful within each site. While we wanted to document high-level generalizations that held across the work of the five teams, we felt that such claims would be more credible if readers knew more about the work of each team.

So we revised our plan for the C&I portfolio of work. We settled on a two-stage synthesis of the C&I work of the five teams. In stage one, we would first synthesize the work of each team. Our reasoning was that if we could tell the story and reveal the essence and core of each team’s effort, we would set the stage for a more meaningful cross-team synthesis. Our narrative would be better grounded—more firmly situated in the details of the RfU work. Optimistic about the utility of such an approach, we went back to reread the pool of studies we had gathered in 2017 and 2018; we scoured the archival literature for additional work that we would need to take into account, and repeatedly checked in with each team to acquire new work.

When we started to reread old (and read new) entries, we were faced with another realization: within each team, we found another layer of heft and complexity, making even site-specific summaries challenging. The work of two of the teams, Language and Reading Research Consortium (LARRC) and Reading, Evidence, and Argumentation in Disciplinary Instruction (READI), was reasonably focused and integrated across the 5-year cycle of work; they had what we came to call a “long runway.” The work of another team, Florida Center for Reading Research (FCRR), anchored the “diverse portfolio” end of the continuum, with at least eight “variations” on its C&I theme. Catalyzing Comprehension through Discussion and Debate (CCDD) and Promoting Adolescents’ Comprehension of Text (PACT), each with at least two major strands of parallel research, landed somewhere in the middle. So we searched for a way to provide greater focus for these site-specific syntheses.

In the end, we settled on a very specific review strategy based on a contractual requirement of the grant. Each team was required, by the final (fifth) year of its life, to conduct an RCT on a significant pedagogical intervention; more specifically, it was supposed to be an intervention that represented the insights that had been gained from other sorts of preliminary efforts (their developmental work examining relations among key pedagogical and outcome variables) and their experimental lead-ins to the RCT(s) (usually some combination of design-based research to fine-tune the interventions, pilot studies to test specific components, and short-term efficacy studies of early versions of the intervention(s)). Given this requirement, along with the knowledge that each team

had, in fact, conducted one or more major RCTs, we decided to use the RCT effort of each team as a focus for conveying the core of each site's efforts in C&I. We therefore began with the RCT(s) and worked our way back into the C&I efforts that led up to the RCT(s). This approach subsequently led to Chapter 4. Chapter 5 follows with a cross-site synthesis, informed by our work in Chapter 4, of what the RfU accomplished, both in developing and evaluating new ways of improving reading comprehension from pre-K through grade 12.

In summarizing the work of each team's C&I work in Chapter 4, we did not impose a common organizational framework on the five narratives, mainly because we felt that each site followed a unique trajectory. The variability was reflected in the sheer number of experimental and quasi-experimental analyses completed; two sites completed a single RCT, while other sites conducted more than a dozen. Some sites emphasized the development of the interventions over their evaluation, while others employed the reverse emphasis. Accordingly, we let the structure of each narrative follow from an attempt to capture the site's pedagogical identity and research process. That said, we did impose two common "text features" on the story of each site: (1) each narrative begins with an overview of the key goals, findings, and conclusions from that site's pedagogical work—a kind of mini Executive Summary for the site, and (2) to ease the readerly burden of all of the "numbers" that are needed to convey the magnitude and significance of the findings (all of the *p*-values and effect sizes), we have organized them into tables that report the most relevant effect sizes and statistical significance of each pedagogical project that underwent an efficacy study for each RfU team. We would note that the number of tables for any particular team does not reflect the amount of work a team conducted, but rather it reflects the team's approach to the research process (i.e., multiple pedagogical products and/or multiple efficacy studies versus one focal product and/or study).

Chapter 5 served as a synthesis of the work across the sites. It begins with a statistical analysis of intervention effects from the RCTs across all five sites before moving on to a synthesis of themes—what we learned when we read across the work in the pedagogical portfolios of the five teams. They take the form of important findings, themes, and insights about how to improve comprehension, focusing on the common threads that inform the design, delivery, and effectiveness of practices and programs. Chapter 5 concludes by addressing a set of dilemmas and limitations in conducting this sort of pedagogical research. The chapter is essentially an account of what was learned, along with a discussion of ongoing issues, concerns, and directions to consider in light of the progress achieved by this unprecedented effort to improve reading comprehension pedagogy and achievement.

Methodologically, the authorship team for Chapter 4 engaged in these steps, more or less in this order, but with a lot of traversing up and down the steps as needed to summarize the work of each site:

1. Ask the five sites for a list of their most important curriculum and instruction studies—those that absolutely had to be included in our summary.
2. Scan the key journals in the reading pedagogy archival literature at regular intervals, looking for additional publications from the various teams.

3. Divide the sites across reviewers and do a close reading, resulting in a short summary of key methods, findings, and conclusions for each entry in the list.
4. Assign the entries to one of two broad categories: intervention development (studies that lay the groundwork for an eventual intervention) and intervention evaluation (mainly RCTs, or large efficacy studies), in which the efficacy of an intervention was compared to one or another counterfactual (usually a business-as-usual control group but sometimes an alternative treatment group).
5. In cases in which there were multiple interventions, group the entries by program. For example, READI completed a single RCT, while LARRC completed a few RCTs on a single program. CCDD studied two interventions, PACT three, and FCRR at least eight.
6. Record (or compute if necessary) effect sizes for key outcome variables for each intervention for each site.
7. Summarize the effect sizes in a table for each intervention.
8. As an authorship team, discuss and interpret the results for each intervention and then for each site.

For Chapter 5, we took a distinct approach, both for our statistical synthesis and our thematic synthesis. The two syntheses employed completely different methods.

For the statistical synthesis, we created two grand synthesis tables of effect sizes, one each for direct effects on comprehension (listening or reading; see Table 5-1), including the orchestration of comprehension for applied tasks, and components of comprehension (see Table 5-2), such as vocabulary, morphology, or metacognition. Within each table, we 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), and we note the magnitude and statistical significance of effects. For the latter, we adhered to Cohen’s rule of thumb about small, moderate, or large effect sizes, with the following amendments: Because effects on the broadest general outcome measures were typically so small in the empirical studies, 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 typically fall within the small category, making even medium effects impressive (or at least rare) in comparison.

For the thematic synthesis, we adopted a completely different and, we think, complementary set of methods. It was a classical discovery of themes driven by conceptual analysis of the pedagogical practices themselves. We read and reread the very same manuscripts that formed the basis of our earlier summary of the five teams. But in this reading, we read across teams, trying to ferret out shared curricular and instructional features across this highly varied landscape of interventions. In a sense, this decidedly qualitative analysis (it was akin to an ethnography of the research articles themselves) was designed to answer the following question: What did we learn about the consistency of features of effective reading comprehension pedagogy across the RfU initiative?



# The Nature and Development of Reading for Understanding

Gina N. Cervetti, *University of Michigan*

## CONTENTS

EXECUTIVE SUMMARY .....	41
INTRODUCTION .....	44
MODELS OF READING COMPREHENSION AND THE ADEQUACY OF THE SIMPLE VIEW OF READING AS AN EXPLANATION FOR READING COMPREHENSION .....	45
THE STRUCTURE OF LANGUAGE AND ITS RELATIONSHIP TO READING COMPREHENSION .....	48
COGNITIVE SKILLS AND DISPOSITIONS IN RELATION TO READING .....	51
Attention, Memory, and Self-Regulation.....	51
Inferential Processes .....	53
Motivations and Reading Comprehension .....	55
TEXT CHARACTERISTICS, GENRE KNOWLEDGE, AND READING COMPREHENSION .....	56
DISCUSSION .....	57
Overall Implications .....	57
Key Limitation of This Work.....	62
REFERENCES.....	63

## EXECUTIVE SUMMARY

In 2002, the RAND Reading Study Group (RRSG) described the state of our knowledge about reading comprehension and outlined a research program designed to support the improvement of reading comprehension in children. In the context of the report, *Reading for Understanding*, the RRSG called for “high-quality,” “long-term and cumulative” research to inform policies and programs to address the underperformance of U.S. students in reading and the persistent gaps in reading performance among

students in different demographic groups. The RAND report informed the development of the Reading for Understanding (RfU) research initiative from the U.S. Department of Education's Institute of Education Sciences. A competitive process was used to select six teams, described in Chapter 1, to participate in the initiative and collectively address three dimensions of reading comprehension: underlying, malleable processes, interventions, and assessments. Over the past 8 years since the grants were awarded, the RfU researchers have produced hundreds of publications describing the results of their efforts. This chapter is part of the effort to synthesize this vast body of research, articulating key findings and implications for research and practice moving forward.

Several of the RfU teams made a better understanding of the nature and development of reading comprehension a central goal of their work. This chapter summarizes the work of the RfU teams as it relates to these issues. Collectively, the research teams examined language skills, cognitive skills, social skills, and forms of knowledge that may relate to reading comprehension, guided by the Simple View of Reading (SVR; Gough & Tunmer, 1986; Hoover & Gough, 1990) and cognitive models of reading comprehension, including the Construction-Integration Model (Kintsch, 1988; Perfetti, 1999). These examinations extend our collective understanding about comprehension through improvements in research designs. They extend previous research, also, by attending to an array of linguistic, cognitive, and dispositional characteristics that have been shown to influence comprehension in past research, but are not yet entirely understood.

This research offers a set of insights that both support our collective understandings about the processes that underlie the development of reading comprehension and can serve as the basis for design research on instruction and assessment. Key findings from this work include the following:

- **The array of skills and knowledge that support reading comprehension and their relative importance shift as students move through school.** The RfU research has documented the contributions of a broad set of skills and knowledge that influence concurrent reading comprehension and comprehension development from the earliest years of schooling. In doing so, this research both confirms the significance of letter- and word-level skills in the development of comprehension and represents successful reading comprehension as dependent on the coordination of an array of different kinds of skills and knowledge. Among preschool and elementary students, these “other” phenomena include linguistic, cognitive, and behavioral skills. Among adolescent readers, these include background knowledge, vocabulary knowledge, strategy use, and inference making—and may include discipline-specific reading and reasoning skills and epistemological perspectives.
- **Language skills underlie successful listening and reading comprehension.** The RfU research offers evidence about the significance of early language skills, including those related to orthography, phonology, morphosyntax, and vocabulary, as concurrent and longitudinal predictors of listening and reading comprehension. This research suggests that children who struggle with listening comprehension and reading comprehension in the elementary grades may have had underlying difficulties with components of language early in school. However, the findings of this research also offer the caution that language may be best conceptualized as a single skill or closely interrelated cluster of skills in young children, calling into

question assessments and instructional approaches that target discrete language skills. The RfU studies also document the role of academic language skills and other complex skills, such as reasoning and inferencing, in sophisticated forms of reading comprehension in adolescence.

- **Many cognitive skills play a role in listening and reading comprehension, but some are more pivotal than others.** The RfU studies suggest that listening and reading comprehension are associated with and may depend on an array of cognitive skills, such as the ability to activate information relevant to the situation described in a text, to suppress irrelevant information (*inhibitory control*), to evaluate one's own ongoing understanding during reading (*comprehension monitoring*), to form connections within a text and between textual information and prior knowledge (*inferencing*), and to remember and follow sets of directions (an aspect of *self-regulation*). The RfU research offers insight into both the significance and the nature of the relationship between cognitive skills and comprehension, documenting, for example, that some cognitive skills seem to make more substantial contributions to comprehension than others. Skills related to attentional control including inhibitory control and self-regulation make small but significant contributions to comprehension. Other skills, such as comprehension monitoring and inferencing, seem to make more substantial contributions to comprehension and to distinguish stronger and weaker comprehension. The latter may prove more fruitful as the basis of intervention and assessment research.
- **Word and world knowledge enable successful reading comprehension.** The RfU studies bolster substantial prior research demonstrating the significance of word and world knowledge in listening and reading comprehension, particularly as students move into adolescence. The studies also extend our understanding by shedding new light on the nature of the relationship between knowledge and comprehension, suggesting, for example, that word and world knowledge support comprehension, at least in part, by aiding readers' inferencing and monitoring. In spite of these advances, the connections among word and world knowledge, cognitive processes, and comprehension are not yet entirely understood and merit further research.

By looking at a broad and complex array of skills and knowledge, the RfU research has contributed to understanding about the nature of reading comprehension and consequential weaknesses in reading development with sufficient clarity to inform the design of interventions and diagnostic assessments. However, there are key limitations to this research. First, the RfU developmental research largely focuses on reader skills (both cognitive and linguistic) and knowledge as explanations for reading comprehension. Few of the studies consider other contributors to comprehension, including textual and contextual factors. Second, although a number of the RfU studies include large numbers of students, some of the samples do not reflect the racial, linguistic, or economic diversity that characterizes the U.S. school population. Third, the studies have collectively identified statistically reliable correlates and predictors of comprehension. However, like most developmental studies, they do not consistently provide clarity about the relative importance of each element or consider each element in light of the potential for malleability (responsiveness to instruction).



## INTRODUCTION

The development of proficiency in reading comprehension is an important milestone for children in part because it underlies success in many other academic, personal, and, later, professional endeavors. The significance of proficient reading comprehension and the fact that it still eludes many students makes it a worthy subject for careful examination. Several of the RfU teams focused on the nature and development of reading comprehension. This chapter is an attempt to answer the question, “What did we learn from the RfU research about the nature and development of comprehension?” We address this question by reviewing a set of papers nominated by the research teams as those that offer insights into the development of comprehension.

Collectively, these teams examined the “usual suspects” involved in reading development, especially the word recognition and listening comprehension that comprise the SVR’s depiction of reading comprehension (Gough & Tunmer, 1986), in an effort to better understand the structure and impact of those dimensions. In addition, the teams examined language skills, cognitive skills, social skills, and forms of knowledge that may relate to reading comprehension, guided by cognitive models of reading comprehension, including the Construction-Integration Model (Kintsch, 1988).

This work sheds light on questions about the concurrent correlates of reading comprehension at different stages of reading development, early predictors of later reading comprehension, and the characteristics that distinguish more and less successful comprehenders.

It is well understood that children who master foundational skills and knowledge of reading, including letter-sound knowledge, phonological awareness, and word reading, are more likely to become successful comprehenders. However, skill at the letter and word levels does not ensure students will become strong comprehenders of text (Connor et al., 2015). Successful reading comprehension ultimately requires the coordination of an array of different kinds of skills and knowledge. The RfU teams largely focused on potential explanations for good and poor comprehension beyond letters and words, including the linguistic, cognitive, and dispositional characteristics, which have been shown to influence comprehension in past research but are not yet entirely understood. By looking at a more replete and complex array of skills and knowledge, the teams have helped us to better understand consequential weaknesses in reading achievement at different developmental stages with sufficient clarity to inform the design of interventions and diagnostic assessments.

The teams brought a high level of methodological precision and power to the study of comprehension development with studies involving large samples of students across grades and geographic regions, sometimes employing longitudinal designs, using comprehensive sets of measures for target constructs, and leveraging advanced statistical techniques in the analysis (e.g., Francis, Kulesz, & Benoit, 2018; LARRC, 2015b; Murphy, LARRC, & Farquharson, 2016). In several cases, these advances provided a means to resolve past controversies and to explain more variance in comprehension than previous studies.

The RfU studies discussed in this report can best be characterized as “basic” research designed to refine existing models of comprehension and inform the development of instructional interventions and assessments. The development of diagnostic assessments and effective interventions for students with comprehension difficulties

depends on such knowledge. Nevertheless, where possible, implications for practice—assessment and instruction—as well as future research, are discussed.

This chapter is organized by major insights related to the following themes:

- Models of reading comprehension and the adequacy of the SVR as an explanation for reading comprehension;
- The structure of language and its relationship to reading comprehension;
- Cognitive skills and dispositions in relation to reading; and
- Text characteristics, genre knowledge, and reading comprehension.

### MODELS OF READING COMPREHENSION AND THE ADEQUACY OF THE SIMPLE VIEW OF READING AS AN EXPLANATION FOR READING COMPREHENSION

Several RfU studies examined the validity of theoretical models of comprehension, especially the SVR (Gough & Tunmer, 1986; Hoover & Gough, 1990), which posits that reading comprehension is the product of decoding, or word recognition, and listening comprehension. Together, these studies largely validate the SVR's conceptualization of early reading, but they also offer insights into challenges of using the SVR to guide instruction and assessment. Furthermore, the RfU research documents important limitations of the SVR model for understanding reading beyond the early grades, offering theoretical and empirical support for alternative models of comprehension in adolescence.

The Language and Reading Research Consortium (LARRC) (2015a) examined the SVR model in grades 1–3, bringing improved approaches of measurement and data analysis to the question of the validity of the SVR. In particular, the researchers used multiple measures for each construct (word recognition, listening comprehension, and reading comprehension) and used structural equation modeling to examine the overall adequacy of the model and shifts in the relative contributions of word recognition and listening comprehension to reading comprehension across the grades. LARRC found, in line with previous research, that the SVR provides a good estimate of reading comprehension in these grades. LARRC also extends understanding of the model by documenting a shift as early as grade 2 from word recognition to listening comprehension as the leading predictor of reading comprehension. With respect to the word recognition component of the SVR, LARRC found that the role of word-reading *accuracy* (i.e., ability to accurately read words) declines after grade 1, but the role of word-reading *fluency* (i.e., speed of accurate word reading) becomes significant at grade 3. This suggests that fluency may become a better indicator of word recognition as students develop greater accuracy in their word recognition.

LARRC and Chiu (2018) similarly examined the utility of the SVR for understanding grade 3 reading comprehension and pre-kindergarten (pre-K) predictors of comprehension in grade 3. As in LARRC (2015a), the SVR constructs were found to account for a large proportion of the variance in reading comprehension at grade 3 (about 94 percent). In addition, the longitudinal analysis confirmed that oral language and code-related skills in pre-K explained substantial variance in grade 3 reading comprehension. In the longitudinal analysis, a developmental pathway emerged in

which preschool oral language strongly predicted later reading comprehension through listening comprehension.

Lonigan, Burgess, and Schatschneider (2018), of the Florida Center for Reading Research (FCRR) team, examined the validity of the SVR in grades 3–5. As in the studies of younger students, decoding and language factors (vocabulary and syntax) accounted for approximately 90 percent of the variance in reading comprehension. Across the grade levels, decoding and language skills shared substantial variance, and language accounted for a larger proportion of unique variance (24 to 33 percent) than did decoding (10 percent). In addition, decoding explained less variance with older students than younger students, echoing a pattern similar to that found by LARRC (2015a). In addition, vocabulary was a stronger predictor of comprehension for students with higher levels of comprehension skill.

While the overall influence of decoding on comprehension attenuates over time, decoding remains an important factor for students whose word-level skills are underdeveloped. In a study of students in grades 5–12, the Educational Testing Service (ETS) RfU team (Wang, Sabatini, O'Reilly, & Weeks, 2019) identified a threshold of decoding skill below which the decoding and comprehension were only weakly correlated, and they found that grades 5 and 8 students who fell below this threshold made little progress in reading comprehension over 3 years. Among students who performed above the decoding threshold, comprehension accelerated across the grades. For students who struggle with decoding, progress in reading comprehension may depend on interventions that help them reach the decoding threshold.

Taken together, these studies both validate the SVR and suggest a pattern in which language skills become more influential for comprehension once decoding skills are more developed. This suggests the importance of helping students consolidate their word-level skills early in school and providing simultaneous support for oral language as they do.

Although studies conducted by the RfU teams validate the SVR, they also illuminate challenges in using the SVR as a framework for understanding reading comprehension and diagnosing comprehension difficulties in young children. Most notably, the FCRR team (Lonigan & Burgess, 2017) tested the separability of decoding and reading comprehension in kindergarten through grade 5, finding that reading comprehension is not measurable separately from decoding until grade 3. That is, existing measures are unable to distinguish challenges with decoding and challenges with other aspects of reading comprehension in the earliest grades. This finding highlights the need for studies like those discussed later in this report that identify additional potential underlying factors in reading comprehension—including cognitive, linguistic, and dispositional factors that may emerge as obstacles later in school. Currently, children who will experience challenges with reading comprehension related to factors other than decoding may not be identified early in school, because our understanding and measurement of these factors has been limited. This research also points to the need to identify or develop measures of these comprehension-related processes outside of reading tasks so that students who will later struggle with reading comprehension, in spite of adequate word-reading abilities, can be identified early in school.

The RfU studies also add to recent research raising questions about some components of the SVR, particularly the role of vocabulary knowledge in the model. In the

SVR, vocabulary knowledge has been traditionally conceptualized as part of listening comprehension. LARRC (2017) offers confirmatory evidence for this conceptualization in their finding that oral language (vocabulary and grammar) and listening comprehension are only measurable as a single construct, oral language, in pre-K through grade 3. However, two RfU studies (LARRC, 2015a; Wagner, Herrera, Spencer, & Quinn, 2015 [FCRR]) call the SVR's conceptualization into question by offering evidence that vocabulary knowledge may contribute to word recognition, as well as listening comprehension, in grades 1–3. The inconsistent findings suggest the need for additional research and further call into question the adequacy of the SVR as a singular guide for instruction and assessment.

Additional RfU studies sought to move beyond the SVR in their efforts to understand adolescent comprehension. For example, Ahmed et al. (2016) of Promoting Adolescents' Comprehension of Text (PACT) sought to understand sources of variance in reading comprehension for middle and high school students who exhibit a range of reading comprehension skill. Ahmed et al. examined the dimensions of reading comprehension in older students through a validation study of Cromley and Azevedo's (2007) Direct and Inferential Mediation (DIME) theory of reading comprehension. The DIME model describes background knowledge, vocabulary knowledge, and word reading as having a direct influence on comprehension, and it describes background knowledge and vocabulary as also influencing comprehension through inference making and reading strategies. Background knowledge and vocabulary knowledge have the strongest influence on comprehension in the DIME model. The analysis of Ahmed et al. (2016) supports the original DIME model and a second model in which background knowledge, vocabulary knowledge, reading comprehension, word-reading skill, inference making, and the use of reading strategies all make significant direct contributions to comprehension. Moreover, Ahmed et al. (2016) documented a shift after grade 6 in which the role of vocabulary knowledge attenuates, but inferencing skill and background knowledge exhibit an increase in their contributions to reading comprehension.

O'Reilly, Wang, and Sabatini (2019 [ETS]) further examined the role of background knowledge in high school students' reading comprehension. The researchers assessed knowledge of ecology by asking students to evaluate the relatedness of a series of words to the topic. Comprehension was then measured using a multitext, scenario-based assessment on the topic of ecosystems. The researchers identified a knowledge threshold at which the relationship between background knowledge and reading comprehension shifted. For students whose knowledge fell below the threshold, there was a flat relationship between knowledge and comprehension. For students who fell above the threshold, increases in knowledge were associated with increases in comprehension, suggesting a facilitative role for knowledge. This suggests that students may need a minimum amount of topic knowledge to comprehend texts on that topic.

Francis, Kulesz, and Benoit (2018 [PACT]) sought to address limitations of the SVR by proposing a new model that accounts for variation in readers and texts, the Complete View of Reading (CVR*i*). The researchers modeled reading fluency as a proxy for reading comprehension in grades 6–8, using measures of reader characteristics (word-reading efficiency, decoding, verbal knowledge, and listening comprehension) and text characteristics (average word frequency, average sentence length, narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion).

Francis et al. (2018) found evidence that the development of fluency is heterogeneous across readers, with varying rates of growth over time, and that text characteristics affect readers differently. For example, expository texts and more difficult texts have a negative impact on fluency (i.e., cause students to read more slowly), particularly for better readers who may adjust their reading rate as they encounter more challenging texts. According to Francis et al., these findings suggest that models like the SVR that attribute comprehension entirely to component skills may overlook important variation in how individuals approach comprehension across situations and texts, and they may thus overlook potential pathways for intervention.

Goldman et al. (2016), of the Reading, Evidence, and Argumentation in Disciplinary Instruction (Project READI) team, further augment theoretical conceptualization of adolescent reading by examining underlying processes through a disciplinary lens. Goldman et al. developed a conceptual framework that describes the reading, reasoning, and argumentation practices of disciplinary learning in literature, history, and science. They used an examination of empirical and theoretical literatures to articulate a set of core constructs within each discipline (e.g., epistemological considerations and types of text structures) and a set of related goals that describe reading and reasoning in each discipline. The goals are designed to articulate processes that may be challenging for adolescent readers but are necessary for authentic forms of disciplinary engagement, such as forming intertext generalizations about theme and characterization in literature or evaluating historical interpretations for their completeness and quality of evidence in history. Relatedly, in a study of students in grades 4–7, LaRusso et al. (2016), of the Catalyzing Comprehension through Discussion and Debate (CCDD) team, found that academic language was the strongest predictor of deep comprehension, but that the disciplinary skill social perspective taking (expressing thoughts and feelings of individuals in a scenario and positioning based on contextual and other considerations) accounted for significant variance beyond academic language.

While largely validating the SVR in the early grades and the significance of word recognition in early reading, this RfU research also adds to evidence about the early importance of oral language and the later importance of inferencing skill, vocabulary knowledge, background knowledge, and disciplinary knowledge for successful comprehension. In doing so, it suggests directions for future research on assessment and instruction with a focus on the skills underlying successful reading comprehension. Several of these constructs were examined in subsequent RfU studies. This work is described in the sections that follow.

### **THE STRUCTURE OF LANGUAGE AND ITS RELATIONSHIP TO READING COMPREHENSION**

The RfU studies add to existing evidence regarding the significance of language skills for reading comprehension, suggesting that early language skills likely serve as a foundation for proficient reading comprehension in the elementary grades and that sophisticated forms of linguistic knowledge and skill are associated with reading comprehension in early adolescence.

Previous research had established that language skills are significant concurrent and longitudinal predictors of reading comprehension (e.g., de Jong & van der Leij, 2002;

Ouellette, 2006). In addition, contemporary models of reading comprehension—from the SVR to the Construction-Integration Model (Kintsch, 1988)—have long posited an important role for language. However, oral language and academic language have often been operationalized narrowly as vocabulary knowledge. The research of the RfU teams extends our conceptualization of language by documenting the predictive relationships of a broader array of language skills, including grammatical skill and morphological knowledge, to comprehension in the early elementary grades (Apel, Diehm, & Apel, 2013 [FCRR]; LARRC & Logan, 2017). This work produced several major findings.

First, in addition to documenting the significance of knowledge for comprehension, longitudinal examinations of reading comprehension conducted by the RfU teams have identified early language-related skills and profiles of skills that predict later listening and reading comprehension. Quinn, Wagner, Petscher, and Lopez (2015 [FCRR]) found that students with higher levels of vocabulary knowledge in grade 1 made greater growth in their reading comprehension across grades 1–4, supporting an instrumental view of vocabulary knowledge in which knowledge of word meanings leads to better comprehension over time (Anderson & Freebody, 1981). Murphy et al. (2016 [LARRC]) examined profiles of lexical quality in pre-K as predictors of grade 1 reading comprehension, listening comprehension, and word recognition. They found that students' orthographic, phonological, morphosyntactic, and vocabulary skill accounted for substantial variance in grade 1 reading comprehension. They also found that students within a particular band of grade 1 reading comprehension performance (low-average) had somewhat different underlying skill profiles in pre-K compared with other groups. Students who had low letter knowledge in pre-K had similar grade 1 word recognition as students who had been low in language, but the students who had lower language skills in pre-K were lower on listening comprehension at grade 1. This suggests that low language skills are a better predictor of later reading comprehension difficulties than low letter knowledge. Alonzo, Yeomans-Maldonado, Murphy, Bevens, and LARRC (2016) examined pre-K predictors of grade 2 listening comprehension. They used a variety of language-related predictors, including listening comprehension, and found that the broad set of language measures used in the study accounted for substantial variance (55 percent) in grade 2 reading comprehension. However, only a pre-K measure of listening comprehension and a measure of working memory and language skills predicted grade 2 listening comprehension. It is possible that some additional aspects of language, such as vocabulary knowledge, were captured by the listening comprehension measures. Taken together, these findings point to the significance of early oral language for later reading comprehension and suggest that language development early in school may set the stage for later success with comprehension.

Second, the RfU studies found that, among students in the upper elementary through middle school grades, additional academic language and reasoning skills predict

Our team went “all-in” to better understand the role of language in skilled reading comprehension, doing work to examine the potentially causal relations between language and reading via a range of methodologies, including experimental design.

—*Laura Justice, Steering Committee Representative from LARRC*

sophisticated forms of reading comprehension. Uccelli, Galloway, Barr, Meneses, and Dobbs (2015 [CCDD]) validated a measure of academic language skills that includes understandings about register and argument, as well as higher-level grammar and morphology. The measure, the Core Academic Language Skills Instrument (CALSI-I), predicted students' reading comprehension beyond grade level, English-proficiency designation, socioeconomic status, word reading, and vocabulary knowledge in grades 4–6, accounting for 12 percent of the variance in reading comprehension. LaRusso et al. (2016 [CCDD]) found that the academic skills measured by the CALSI-I predicted students' deep comprehension in grades 4–7, as well as students' ability to position actors (or characters) in a text based on their roles and contexts. Deep comprehension was measured using the Global Integrated Scenario-Based Assessment (GISA) (O'Reilly & Sabatini, 2013 [ETS]), a multitext, problem solving–focused comprehension assessment (see Chapter 3 in this volume). Phillips Galloway and Uccelli (2019 [CCDD]) examined growth on the CALSI-I and its association with reading comprehension among emergent bilingual students and English-proficient students across grades 6 and 7. They found that emergent bilingual students had significantly lower initial scores on both measures but exhibited similar rates of growth compared with their English-proficient peers. They also found that students who had higher initial scores on the CALSI-I also had higher levels of reading comprehension and higher growth in reading comprehension over time. These studies offer a promising measure of academic language that specifies a range of skills and knowledge needed for engagement with content-area texts. In addition, these studies highlight the need to consider complex acts of comprehension, such as deep (intertextual, problem-oriented) comprehension of challenging texts, in constructing models of comprehension, and they point to the sophisticated knowledge and reasoning skills that may support success with these tasks.

Third, while the significance of language skills for reading comprehension is evident as early as grade 2, different aspects of language are challenging to distinguish in the youngest students. Five studies in this review examined the relationships among dimensions of oral language in the primary and elementary grades with some differing results. LARRC (2017) found that oral language (grammar and vocabulary) and listening comprehension are best characterized as a single oral language construct in pre-K through grade 3. LARRC, Jiang, Logan, and Jia (2018) also found that grammar and vocabulary scores are closely associated in preschool through grade 3. LARRC (2015b) supported a single-factor model (i.e., grammar, vocabulary, and discourse were not distinguishable) at pre-K and kindergarten, a two-factor model (i.e., vocabulary and grammar comprising one dimension and discourse skills comprising a second) at grades 1 and 2, and a three-factor model (grammar, vocabulary, and discourse) at grade 3. By contrast, Lonigan and Milburn (2017 [FCRR]) found dimensionality in oral language with separate factors for vocabulary and syntax/listening comprehension for students in pre-K through grade 5. LARRC (2015c) found that dimensionality of oral language was evident in pre-K students who were Spanish-English dual language learners. The best model for these students included a dominant general language factor and two highly correlated factors representing word knowledge and grammatical knowledge. In addition, Spencer, Muse, et al. (2015 [FCRR]) found that vocabulary knowledge and morphological knowledge are best understood as a single construct in grade 4.

Some of the differences in the results of these studies are likely attributable to the use of different measures to represent core constructs. In particular, comprehension monitoring and inferencing are treated differently across studies. For example, LARRC (2015b) used an inferencing task as part of the discourse construct along with measures of comprehension monitoring and text structure knowledge, whereas LARRC (2017) used an inferencing task as part of listening comprehension. What emerges, however, is a conceptualization of oral language as dominated by an overall, or general, language factor in the earliest grades, becoming increasingly separable into word-level, grammatical, and higher-level (discourse- and inferencing-related) constructs as students move from primary into elementary grades. These findings have implications for the assessment of oral language—suggesting that an omnibus oral language assessment may be unable to distinguish particular language-related obstacles but may be useful in identifying students with less developed oral language skills that may undermine the development of reading comprehension.

## COGNITIVE SKILLS AND DISPOSITIONS IN RELATION TO READING

### Attention, Memory, and Self-Regulation

Several RfU studies examined the contributions of cognitive skills, especially attention, memory, and self-regulation, to listening and reading comprehension. The results of these studies suggest that a range of cognitive skills is associated with listening and reading comprehension as early as kindergarten and that some of these skills may have reciprocal relationships with reading comprehension development.

Among the attention-related skills studied in the RfU research is the ability to suppress irrelevant information or meanings during reading or listening, a skill that is sometimes labeled *inhibitory control* or *cognitive inhibition*. In the RfU research, this skill was found to contribute to listening comprehension in kindergarten and grade 1 (Kim & Phillips, 2014 [FCRR]) and reading comprehension in grades 6–12 (Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014 [PACT]; Barnes, Stuebing, Fletcher, Barth, & Francis, 2016 [PACT]). The ability to maintain sustained attention and focus on task-relevant goals also predicted reading comprehension in the Arrington et al. study, as did working memory. LARRC, Jiang, and Farquharson (2018) studied behavioral attention and working memory in relation to listening comprehension and reading comprehension. Both attention and memory were concurrent predictors of listening comprehension at each grade and, with the exception of attention at grade 3, both predicted reading comprehension at each grade. The effects on listening comprehension were direct, but the effects on reading comprehension were mediated by listening comprehension and/or word reading both concurrently and longitudinally.

In a study of struggling readers in grades 7–12, Swanson, Barnes, Fall, and Roberts (2018 [PACT]) found that vocabulary and inference-making ability, but not decoding ability, predicted reading comprehension among students with different levels of inattention and hyperactivity. In addition, working memory predicted comprehension for two of the three groups of students: those with low inattention and low hyperactivity and those with high inattention and low hyperactivity. However, for students with the highest levels of inattention (those in the high-inattention and high-hyperactivity



group), working memory did not predict comprehension, suggesting the possibility that high levels of inattention mediate the relationship between working memory and comprehension.

A second, related cognitive skill, comprehension monitoring, operationalized as the ability to evaluate one's own understanding of a story, predicted students' listening comprehension of narrative text in kindergarten and grade 1 (Kim & Phillips, 2014 [FCRR]) and, when considered as part of higher-level language skills, predicted reading comprehension in grade 3 (LARRC & Logan, 2017; see above). Denton, Enos, et al. (2015 [PACT]) also found that comprehension monitoring skill distinguished adequate comprehenders from poor comprehenders in middle and high school. LARRC and Yeomans-Maldonado (2017) found that vocabulary predicted comprehension monitoring in grade 1, and grade 1 comprehension monitoring contributed to grade 3 reading comprehension, even after controlling for vocabulary, decoding, and working memory. Connor et al. (2015 [FCRR]) used eye-tracking technology to examine students' growth in comprehension monitoring in grade 5. They found evidence that students with higher academic language skills made more attempts to repair their understandings of text than did students with lower academic language skills. These attempts were considered an indicator of comprehension monitoring. Specifically, students with higher academic language skills had a greater gap in rereading times between sentences with implausible words as opposed to plausible words (e.g., "Last week Kyle flew to visit his family in another city. The large **plane/truck** was spacious and quickly transported them" [p. 117]). Rereading time is the total time spent focusing on the target word (bolded in the example) after the initial fixation. Higher rereading times (more time spent fixating on the implausible word) is assumed to indicate attempts to repair understanding. As such, the higher rereading times for students with higher academic language skills suggest that these students were better at monitoring their comprehension, or noticing a breakdown in comprehension and attempting to repair it. These studies suggest that comprehension monitoring may play an important role in the development of comprehension and may be associated with language skills.

A third attention-related skill, self-regulation, was examined in four RfU studies included in this review. Day and Connor (2017 [FCRR]) developed a new measure of self-regulation, the Remembering Rules and Regulations Picture (RRRP) task, which required students to remember and follow directions (e.g., students are given blocks and pictures and are asked to "Put a blue block on the squirrel by the rock.") (p. 100). Day and Connor found concurrent and predictive relationships between scores on the measure and some reading skills in grade 3. For example, one part of the two-part RRRP predicted fall-to-spring gains in reading comprehension and vocabulary. Connor (2016 [FCRR]) tested a "lattice" model of reading in which cognitive, linguistic, and text-specific processes have reciprocal and interacting effects on reading development. Connor found that, in grades 1 and 2, self-regulation and word and world knowledge have reciprocal relationships with reading comprehension. Importantly, Connor found that classrooms with teachers receiving reading-related professional development had higher and less stable growth in reading comprehension, meaning that students' reading skills in grade 1 were less likely to predict skills in grade 2 than in the comparison group. This suggests that instruction influenced students' trajectories to a greater degree in the classrooms of teachers receiving the professional development. Day,

Connor, and McClelland (2015 [FCRR]) examined behavioral self-regulation skills in relation to *noninstructional* classroom time and reading skill in grade 1 students. Day et al. were particularly interested in whether children's behavioral self-regulation skills in fall were related to growth in reading skill (word identification and passage comprehension) and time spent in productive or unproductive noninstructional activity over the course of the year. Noninstructional activity that related to learning was considered *productive* (e.g., explaining the importance of a lesson) while activity not related to learning was considered *unproductive* (e.g., waiting time or time spent dealing with disruptions). Day et al. found that students with high self-regulation and reading skills in the fall had stronger reading skills in spring. They also found, among other things, that students with weak regulation skills were more likely to be in classrooms with higher levels of teacher-managed unproductive time. When teachers decreased the amount of productive noninstruction time over the course of the year, students had stronger spring reading scores; this was particularly true for students with weaker behavioral regulation skills. Students in these classrooms may have benefited when the teachers were able to establish clear routines and devote more time to instructional activities across the year. Denton, Wolters, et al. (2015 [PACT]) found that adequate adolescent comprehenders also had stronger self-regulation skills than did struggling comprehenders (see below).

It is important to note that while these studies collectively document the contribution of specific cognitive skills to reading comprehension, some of these contributions are small. For example, Barnes et al. (2016 [PACT]) suggest that the contribution of inhibitory control (suppression of irrelevant information, accounting for about 1 percent of the variance in reading comprehension) is too small to warrant the development of interventions targeting this skill in particular. Moreover, the nature of these relationships is not entirely understood; for example, we do not know precisely how self-regulation (as the ability to follow directions) contributes to comprehension and how it relates to other dimensions of executive functioning. Given the large number of cognitive skills that make meaningful, but small, contributions to reading comprehension and preliminary evidence that they have reciprocal relationships with reading comprehension, there is reason to consider the possibility that multicomponent interventions that focus on supporting reading comprehension may better support students than those that target weaknesses in specific skills (see Chapter 5). Nevertheless, the understandings about the role of cognitive processes in decoding and comprehension offered by this work may lead to a better understanding of reading proficiency and may lead to more effective and well-rounded reading interventions.

### Inferential Processes

Several studies conducted by the RfU teams focus on inferential processing, offering insights about the associations between inferencing skills and reading comprehension. For example, the RfU studies found that the ability to make inferences about the states or perspectives of actors (i.e., characters) in text predicts listening comprehension in kindergarten and grade 1 (Kim & Phillips, 2014 [FCRR]) and reading comprehension in grades 4–7 (LaRusso et al., 2016 [CCDD]; see above). In addition, the ability to use visual-spatial information acquired from studying a three-dimensional model of a space

before reading a text set in that space is related to comprehension in elementary-age and adolescent readers (Barnes, Raghobar, Faulkner, & Denton, 2014 [PACT]).

Overall, inferencing skill distinguished more and less proficient adolescent comprehenders across several studies. In particular, adequate comprehenders are more skilled at forming the text-based and knowledge-dependent inferences that establish and maintain local coherence and global coherence during reading. Local coherence refers to understanding across information that appears close together in a text, for example in neighboring sentences, such as resolving anaphora (the “he” refers to Henry) or connecting what is reported in one proposition as a plausible cause or explanation of an event or outcome reported in a preceding or following proposition. Global coherence refers to constructing meaning across longer distances in a text (connecting them all to a theme or topic, for example). Denton, Enos, et al. (2015 [PACT]) found that adequate comprehenders in middle and high school have higher levels of acceptable inferences, acceptable paraphrasing, and comprehension monitoring than poor comprehenders when reading informational text. Barth, Barnes, Francis, Vaughn, and York (2015 [PACT]) found that, compared with struggling adolescent comprehenders, adequate comprehenders were able to evaluate the consistency of causal bridging (intratextual) inferences more quickly and were better able to rely on working memory to form connections across adjoining propositions in order to maintain local coherence. Barnes, Ahmed, Barth, and Francis (2015 [PACT]) similarly found that weaker comprehenders were less able than more skilled comprehenders to make inferences that maintain local coherence. Weaker comprehenders struggled to maintain local coherence even when they had relevant knowledge to bring to bear on knowledge-dependent inferences. Denton, Wolters, et al. (2015 [PACT]) found that, compared with struggling comprehenders, adequate comprehenders reported more frequent use of strategies associated with text- and knowledge-based inferences and text evaluation, as well as regulation strategies associated with adjusting reading to enhance comprehension, including rereading and changing reading rate. Taken together, these studies support prior research documenting the contributions of inferencing skill to successful reading, including reading comprehension, and they suggest that weak inferencing skill is a potential source of reading comprehension difficulty in struggling comprehenders.

The RfU research also points to text and reader characteristics that are associated with more and less successful inferencing. Three themes emerge from this research.

First, students may have greater difficulty with inferencing when reading informational text compared with narrative text. Denton, Enos, et al. (2015 [PACT]) found that, regardless of reading comprehension proficiency level, students produced fewer acceptable inferences and paraphrases from informational text versus narrative text.

Second, some kinds of relationships among propositions in text are more challenging to process than others. Barth et al. (2015 [PACT]) found that middle school students at all levels of comprehension proficiency had more difficulty making successful bridging (intratextual) inferences across longer spans of text compared with inferences that depend on information in adjoining sentences. In addition, Vorstius, Radach, Mayer, and Lonigan (2013 [FCRR]) found that, for grade 5 students, negative causal relationships (e.g., “although”) were more challenging than positive causal relationships (e.g., “because”). Barnes et al. (2015 [PACT]) found that, in forming bridging inferences that

rely on connections to prior knowledge, texts that included stronger causal connections across pairs of sentences were easier for students to process than those that relied on weaker causal (temporal) connections.

Third, the RfU studies provide suggestive evidence that, in addition to comprehension skill, background knowledge and vocabulary knowledge are associated with ease or speed of processing inferential relationships in text and with comprehension monitoring. Barnes et al. (2015 [PACT]) found that high school students with higher levels of background knowledge read pairs of causal sentences more quickly. Connor, Radach, et al. (2015 [FCRR]) found that fifth graders' academic language skills predicted their comprehension monitoring behavior during reading. LaRusso et al. (2016 [CCDD]; also discussed above) found that academic language skills predicted "deep comprehension," or comprehension that demands sophisticated inferential processing within and across texts in the interest of problem solving.

A challenge identified by the RfU research regards the measurement of inferencing. Comprehension theory and research have long relied on the understanding that there are distinguishable types of inferences (e.g., Diakidoy, Mouskounti, and Ioannides, 2011; Graesser, Singer, & Trabasso, 1994). LARRC and Muijselaar (2018) found that, while different types of inferences have been distinguishable in some research, local and global inferences cannot yet be measured reliably as two distinct types. LARRC and Muijselaar examined whether local and global inferences could be distinguished statistically. While the two types of inferences were distinguishable, a general inference-making factor explained most of the variance across items.

### **Motivations and Reading Comprehension**

One study included in this review looked at the associations between motivation and reading comprehension. Wolters, Denton, York, and Francis (2013 [PACT]) examined a range of motivational factors, including those related to competence and expectancies, valuing of reading, achievement goals, and socially mediated aspects of motivation for reading. Wolters et al. (2013) found that most of the factors were positively and significantly related to each other. In addition, the researchers found that, among aspects of motivation, feelings of competence and expectancies for success were most closely associated with reading comprehension. Compared with their highly skilled peers, adolescent students who were less skilled comprehenders tended to express lower levels of belief in their ability to do well at reading if they chose to do so. Importantly, stronger and weaker comprehenders were similar in their assessments of importance and utility value and in their goal orientations. Although substantial previous research has addressed the relationship between reading comprehension and motivations in young children, few studies had looked at adolescent students. This study suggests that adolescent students who have struggled with comprehension have less adaptive reading motivations, particularly perceived control or ability to succeed, but similar value-related judgments. In pointing to specific, and potentially malleable, aspects of comprehension, this study provides possible directions for future intervention research.

### TEXT CHARACTERISTICS, GENRE KNOWLEDGE, AND READING COMPREHENSION

Although text characteristics were not a central focus of the RfU research, three studies nominated for this report add evidence to previous research regarding the role of text characteristics and genre knowledge in comprehension. In a study of middle and high school students' performance on a standardized comprehension measure, Kulesz, Francis, Barnes, and Fletcher (2016 [PACT]) examined how reader characteristics (decoding accuracy, fluency, working memory, background knowledge, and vocabulary), test properties (passage genre, passage cohesion, word difficulty, sentence length, and test memory/recall versus inferential questions), and their interactions influenced students' performance in grades 7–9 and grades 10–12. Kulesz et al. (2016) found that expository passages were more difficult than narrative passages and that passages that were high in deep cohesion were easier than those that were low in cohesion. In addition, vocabulary knowledge and background knowledge were significant reader characteristics above word reading, reading fluency, and working memory for both grade bands. For students in the upper grade band, background knowledge was particularly influential when text cohesion was low.

Weak comprehenders' challenges with informational text may be related to low levels of knowledge or low motivation, but they may also be an indication that these students have underdeveloped understandings about what it means to comprehend informational text. Denton, Enos, et al. (2015 [PACT]) used a think-aloud methodology to examine middle and high school students' text processing with texts of different genres (narrative and informational) and different difficulties (on level and above challenging). Poor comprehenders made fewer acceptable inferences than did stronger comprehenders when reading informational text. Denton, Enos, et al. (2015) suggest that poor comprehenders may believe that the development of a text base, or remembering information for the sake of a test, is sufficient when reading informational text.

The RfU research offers some evidence that genre knowledge influences comprehension even among the youngest children. Barnes, Kim, and Phillips (2014 [FCRR]) examined pre-K through grade 1 students' use of literate language features (adverbs, conjunctions, mental and linguistic verbs, and elaborated noun phrases) in narrative retellings and narrative production. They found that use of the literature language features was not related to listening comprehension or narrative production. However, awareness of the features of narrative text structure (as indicated by the proper introduction of characters in a narrative) was related to the quality of grade 1 students' narrative comprehension (ability to recall narrative features after hearing a story), their production of narrative text (ability to produce a story based on a set of illustrations including key narrative elements), and their listening comprehension.

Overall, this work confirms and extends prior research suggesting that informational text and less cohesive text may be more challenging for readers than narrative text and more cohesive text (e.g., Best, Floyd, & McNamara, 2008; O'Reilly & McNamara, 2007). Moreover, these text characteristics may be particularly salient among students who struggle with comprehension. What is not clear from this work is the source of the difference between these two broad "genres." It is not entirely clear why informational text is often more challenging than narrative. It may be that students have more experience with the features and structures of narrative genres and their underlying

structures, or it may be that they have less familiarity with the topics and language often featured in informational text.

## DISCUSSION

### Overall Implications

The research reported in this chapter is best characterized as “basic” research. The goal of such research is to build our collective understanding about basic processes—in this case, the underlying processes that influence the development of comprehension across the age span—but basic research does not alone offer a sufficient basis for the development of recommendations regarding the practice of reading instruction or assessment. Recommendations regarding practice are best formulated through intervention research conducted with students, such as the large body of work synthesized in Chapters 4 and 5 (this volume). However, all things being equal, interventions based on solid developmental research are more likely to be effective, because they are grounded in sound understandings about the nature of the reading process and how it evolves as children learn to read. While basic developmental studies do not speak directly to the practice of reading instruction or assessment, they do provide foundational insights that can support the design of intervention research that would yield such recommendations. As such, while caution is warranted when discussing implications for practice, this section outlines key contributions of the RfU studies with an eye to the formulation of design research on comprehension instruction and assessment.

**The RfU research sheds light on the wide array of skills and knowledge that underlie successful reading comprehension, concurrently and longitudinally.** In current educational practice, many early reading assessments and interventions focus on word reading as the primary enabling skill for proficient reading comprehension, but they often lack attention to other skills and knowledge that may enable successful reading and listening comprehension as students advance into the upper elementary grades and beyond. The RfU researchers used sophisticated research designs—often large-scale, longitudinal, and analytically sophisticated approaches—in order to better understand a wider array of concurrent and longitudinal contributors to comprehension among students at different stages of development.

The RfU research has documented the contributions of a broad set of skills and knowledge that influence reading development from the earliest years of schooling. In doing so, this research suggests that, while the SVR provides a good concurrent and longitudinal explanation of reading comprehension in the early elementary years, there are some important limitations to the model. Although the model accounts for most variance in reading comprehension in the primary grades, it may not provide sufficient guidance for the development and application of interventions. In focusing on two broad predictors of comprehension that are difficult to distinguish in the earliest grades (Lonigan & Burgess, 2017 [FCRR]), the model may overlook underlying factors that will affect some students’ reading comprehension later in school. Addressing the underlying skills for successful comprehension in later elementary school and adolescence requires a more expansive and forward-looking gaze than that provided by the

SVR. Among elementary students, this may involve consideration of linguistic, cognitive, and behavioral skills, including attention and self-regulation.

Preliminary evidence within the RfU portfolio suggests that word-level skills contribute less to comprehension over time as decoding is consolidated and students encounter texts and tasks that require sophisticated forms of academic language and knowledge. As such, explaining comprehension for older students may involve unpacking the infrastructure of the Simple View (e.g., What is entailed in the listening comprehension component of the Simple View?) or augmenting it with additional facets, such as those investigated by Ahmed et al. (2016 [PACT]) and Francis et al. (2018 [PACT]). Characterizing the reading comprehension of middle and high school students requires a more complex model that includes background knowledge, vocabulary knowledge, strategy use, inference making, and disciplinary reading and reasoning skills. More research is needed to understand the nature of these skills and knowledge and their interactions with other reader, text, and task variables. While the explanations for reading comprehension become more complex in adolescence for most students, the RfU research also suggests there are students whose underdeveloped decoding skills markedly suppress progress in reading comprehension in grade 5 and beyond (Wang et al., 2019 [ETS]).

Overall, the results of the RfU studies suggest that a focus on word reading in early interventions is an important foundation for the development of reading comprehension and should be accompanied by attention to other important skills and knowledge that are necessary for reading comprehension once word-level skills are well developed (e.g., LARRC, Jiang, & Farquharson, 2018). In addition, the RfU research suggests that the focus on word reading in early assessment may be driving the underidentification of students who will later experience difficulties with listening comprehension and reading comprehension due to challenges with other enabling skills related to language and cognition (Alonzo et al., 2016 [LARRC]).

**The RfU research documents the nature and significance of language skills for comprehension development.** The RfU research suggests that language skills are an important foundation for skilled comprehension, and the studies provide potential insights about the nature, measurement, and instruction of language. Several RfU studies attest to the significance of early language skills as concurrent and longitudinal predictors of listening and reading comprehension. These studies suggest that children who struggle with listening comprehension and reading comprehension in the elementary grades may have had identifiable underlying difficulties with components of language years earlier, including lower-level knowledge and skills related to orthography, phonology, morphosyntax, and vocabulary (e.g., Murphy et al., 2016 [LARRC]). In addition, academic language skills and other complex skills, such as reasoning and inferencing, predict sophisticated forms of reading comprehension in adolescence and distinguish stronger and weaker adolescent comprehenders (e.g., Uccelli et al., 2015 [CCDD]).

These findings and prior research showing that older children with adequate word-reading skills, but poor reading comprehension, may have had low language skills earlier in school suggest the need for assessment and instruction early in school. The RfU studies shed light on how weaknesses in a broad skill domain, such as language and comprehension, relate to profiles of specific skills and knowledge, potentially positioning educators

to do stronger diagnostic work and to develop richer conceptualizations of the skills and knowledge that enable successful comprehension in childhood and adolescence.

However, the RfU research also offers cautions about the assessment and instruction of language skills. In particular, although the language skills associated with reading comprehension ultimately include several different dimensions, these dimensions (e.g., grammar, vocabulary, and higher-level discourse skills) are difficult to distinguish at the preschool and early elementary levels, becoming increasingly separable as students move further into elementary grades. It is unclear whether the lack of differentiation is more attributable to children's development (i.e., skills in fact become more distinct over time through exposure to more sophisticated language structures in text or through a consolidation of lower-level language skills that leads to the development of higher-level skills), or the limitations of current assessments.

Either way, the RfU research suggests that, although many commercially available language and literacy assessments include numerous subtests, the subtests may measure just a few highly related language skills. As such, approaching variation in scores on different subtests as evidence of specific strengths and weaknesses may not produce valid interpretations. That is, because language skills appear to be largely undifferentiated early on, it is not clear that multiple subtests provide information about distinct aspects of students' language development. Additional studies in this report suggest possible candidates for language assessment, including an omnibus assessment of oral language (see LARRC, 2017) or listening comprehension (Alonzo et al., 2016 [LARRC]), which may be sufficient to identify students at risk for later reading difficulties due to underdeveloped language skills.

The lack of differentiation among language skills early in reading development may also offer implications for instruction. For example, the results documenting that various dimensions of language, such as grammar and vocabulary skills, tend to be closely associated in children—both in their initial scores and in their growth trajectories across the early years of schooling—suggest that language may be best conceptualized as a single skill or closely interrelated cluster of skills in young children (e.g., LARRC, Jiang, Logan, & Jia, 2018). There are several possible implications for instruction that might be explored in future research. For example, it may be that rich language experiences can be used to develop multiple aspects of language concurrently, providing better support for successful reading comprehension than instruction targeting discrete language skills. It may also be that targeted interventions, which are focused on the more malleable dimensions of language (e.g., grammar rather than vocabulary), affect other associated dimensions of language. Given the significance of language for reading comprehension, future intervention research should explore these possibilities.

**The RfU research offers insights regarding the role of cognitive skills in comprehension.** The RfU studies suggest that listening and reading comprehension are associated with and may depend on cognitive skills, such as the ability to activate information relevant to the situation described in a text, to suppress irrelevant information (inhibitory control), to monitor comprehension, to engage in successful inferencing, and to remember and follow sets of directions (self-regulation).

This research offers insight into both the significance and the nature of the relationship between cognitive skills and comprehension. For example, challenges with



inhibitory control may undermine comprehension because irrelevant information that remains active interferes with meaning construction for poor comprehenders (Arrington et al., 2014 [PACT]; Kim & Phillips, 2014 [FCRR]). In addition, for some students, poor reading comprehension may involve an inability to maintain sustained attention to the text and thus an inability to develop a coherent understanding of the text (Arrington et al., 2014 [PACT]).

Although the RfU studies have identified a broad array of cognitive skills that make significant contributions to comprehension, it is important to continue to characterize these skills in terms of the magnitude of their contribution to comprehension, their significance in distinguishing students at different levels of comprehension skill, their relationships with comprehension, and their malleability. Based on the evidence in the RfU studies, several cognitive skills, including those related to attentional control and self-regulation, make small but significant contributions to comprehension while others, such as monitoring and inferencing, seem to have substantial and intrinsic relationships with comprehension and distinguish stronger and weaker comprehension. In adolescence, for example, less skilled comprehenders have difficulty maintaining

Our team's efforts were two-fold. First, we focused on identifying skills that contributed in meaningful ways to reading comprehension in the elementary school grades. Our approach to these questions employed a broader focus and allowed more rigorous methods to identify unique and important contributors to reading comprehension than most prior studies. Second, we were strongly committed to developing and evaluating instructional interventions informed by our ongoing investigations of skill areas with sizable contributions to reading comprehension. We developed multiple interventions that were revised and refined over several year-long cycles of evaluation. We found that our interventions, which focused on building language, knowledge, metacognitive skills, and text structure in preschool, kindergarten, and early elementary school grades, directly affected the intervention targets and were more effective when multiple component interventions were combined.

—*Christopher Lonigan, Steering Committee Representative from FCRR*

both local and global coherence through text- and knowledge-based inferences (Barnes et al., 2015 [FCRR and PACT]; Barth et al., 2015 [PACT]; Denton, Enos, et al., 2015 [PACT]), particularly when reading informational text.

Taken together, this research suggests that efforts to improve comprehension should include attention to the development of cognitive skills. However, given the large number of cognitive skills that make meaningful, but small, contributions to reading comprehension and preliminary evidence that they have reciprocal relationships with reading comprehension, there is reason to consider the possibility that multicomponent interventions that focus on supporting reading comprehension may better support students than those that target weaknesses in specific skills. For example, attentional issues may be best addressed as part of holistic interventions designed to support students' reading comprehension, while comprehension monitoring and inferencing may merit more focused instructional attention. It is also possible that some cognitive and attentional skills have a critical role in comprehension for particular students, suggesting different

approaches to intervention. More research on variation among readers is needed. In spite of remaining uncertainties, the understandings about the role of cognitive processes in decoding and comprehension offered by this work may lead to a better understanding of reading proficiency and may lead to more effective and well-rounded reading interventions. And, of course, this is precisely the role that good basic research on development should play—generating plausible interventions that can be tested in the crucible of classroom curriculum and instruction.

**The RfU research adds evidence regarding the significance of word and world knowledge for reading comprehension.** The RfU studies bolster substantial prior research demonstrating the significance of word and world knowledge for comprehension, particularly as students move into adolescence. The studies also extend our understanding by, for example, suggesting that vocabulary may contribute to both word reading and listening comprehension early in school (LARRC, 2015a; Wagner et al., 2015 [LARRC]); that knowledge and vocabulary support comprehension, at least in part, by supporting readers' inferencing and monitoring (Ahmed et al., 2016 [PACT]; Connor et al., 2015 [FCRR]); and that the relationship between world knowledge and comprehension is reciprocal (Connor, 2016 [FCRR]). They also find that, at least in adolescence, comprehension may be dependent on a minimum amount of knowledge about the topic of the text (O'Reilly et al., 2019 [ETS]). These studies suggest the need to redouble efforts to build students' vocabulary knowledge and to develop approaches to building students' general and text-specific knowledge. The studies also point to specific features of design work related to word and world knowledge. For example, Spencer, Muse, et al. (2015 [FCRR]) provide evidence that different aspects of word knowledge are acquired simultaneously and that a comprehensive understanding of students' knowledge requires the use of multidimensional approaches to vocabulary assessment. In addition, this suggests that students may benefit from vocabulary instruction that extends beyond the instruction of definitions to include many kinds of information about the words and should attend to students' morphological knowledge and skill.

These findings regarding the significance of knowledge, along with those related to inferencing, support contemporary cognitive models of comprehension (e.g., Kintsch, 1988; van den Broek, Risen, Fletcher, & Thurlow, 1996). These models describe development of a coherent mental representation of a text as dependent on forming connections between the propositions in a text and the knowledge stored in long-term (and short-term) memory. However, the connections among word and world knowledge, cognitive processes, and comprehension are not yet entirely understood. As Barnes et al. (2015 [FCRR and PACT]) discuss, it is possible that the accessibility of readers' knowledge in long-term memory affects the knowledge integration process. That is, knowledge that is well elaborated and connected to other concepts in memory may result in more efficient retrieval and thus ease inferencing. Denton, Enos, et al. (2015 [PACT]) and Connor et al. (2015 [FCRR]) raise the possibility that comprehension monitoring may depend on word and world knowledge, and lower levels of these knowledge sources may be partially responsible for struggling comprehenders' difficulties with comprehension monitoring. As Connor et al. discuss, the ability to monitor comprehension and resolve breakdowns in meaning (such as the appearance of implausible words in sentences) may depend in part on students' academic language skills, including their vocabulary

knowledge and background knowledge. Understanding more about these interactions may support the development of interventions.

### **Key Limitations of This Work**

In their report outlining a research agenda on reading comprehension, the RRSG described reading comprehension as consisting of three key elements: the reader (skills and dispositions), the text, and the activity or purpose for reading, all interacting in a larger sociocultural context that relates separately with each element and in combination across elements. In addition, the RRSG's vision of the reader describes individuals as influenced by a broad constellation of cognitive capabilities, motivations, types of knowledge, and experiences. Relative to the RRSG's expansive vision of comprehension and its influences, the RfU teams addressed a narrower set of individual student characteristics as explanations for the development of comprehension. The preponderance of the RfU teams' research offers insights about reader skills and knowledge. Only a small number of the RfU studies look outside the reader to text and task in characterizing the development of comprehension over time. Even within the studies looking at readers, attention was directed toward students' component skills and knowledge with less focus on motivations or life experiences or even consolidated assemblages of skills that may influence comprehension development. It is notable that, in line with previous research, the few RfU studies (by the FCRR team) that did include environmental variables, such as characteristics of classroom environments, found significant associations with students' skills and literacy development (e.g., Connor, 2016 [FCRR]; Connor et al., 2015 [FCRR]; Day et al., 2015 [FCRR]).

In their description of a research agenda for reading comprehension, the RRSG discussed the significance of sociocultural and other contextual factors for understanding comprehension. They point out, for example, that readers' skills and dispositions are "shaped by cultural and subcultural influences, socioeconomic status, home and family background, peer influences, classroom culture, and instructional history" (p. 20). As a result, they call for understanding factors that influence "both the inter- and intraindividual" dimensions of reading (p. 20). In particular, the RRSG notes that one motivating factor for the development of a research agenda is to address persistent and unacceptable gaps in reading performance between students in different demographic groups. The RfU teams only partially realized this vision in their research. While the instructional research reported later in this volume involved diverse samples, some of the research that was most pointedly about development was conducted with fairly homogeneous student samples that do not reflect the racial, linguistic, or economic diversity that characterizes the U.S. school population (Alonzo et al., 2016 [LARRC]; McIlraith & LARRC, 2018; Murphy, LARRC, & Farquharson, 2016). In addition, diversity was often treated as a covariate, rather than investigated to identify potential differences in the development of comprehension as a function of such factors as first language and socioeconomic level. Including diverse samples of students is critical when characterizing developmental patterns in reading.

One risk in examining underlying processes in a complex task such as reading with comprehension is that the complex task will be disaggregated into a multiplicity of small components, leading to an assumption that each of these components plays an

equally important role in the comprehension process and is therefore equally important in the design of high-quality instructional routines and assessments. Although the RfU studies have collectively identified statistically reliable correlates and predictors of comprehension, they do not provide a consistent portrait about the relative importance (i.e., asking, Is this element absolutely pivotal for comprehension development? Is it uniquely important?) and malleability (i.e., Is this element amenable to instruction?) of each element. As Ahmed et al. (2016 [PACT]) point out, it is important to understand which of the many factors that are related to comprehension are actually most integral to comprehension and which are most malleable. Future research is needed to determine the significance and malleability of different skills at different points in students' development and in relation to particular text genres and characteristics.

One important advantage of the RfU research reported here is that multiple measures were often used to capture underlying constructs for concurrent and longitudinal prediction of comprehension. In addition, several of the studies of middle and high school students used measures of deep comprehension, requiring sophisticated inferential understandings within and across texts. However, the studies of young students modeled comprehension using well-established standardized measures of comprehension that largely capture literal comprehension of short passages. We should not assume that the contributors to comprehension would be identical had the measures required more complex forms of textual and intertextual comprehension and application.

## REFERENCES

- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M. A., & Kulesz, P. (2016). Validation of the direct and mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44*, 68–82.
- Alonzo, C. A., Yeomans-Maldonado, G., Murphy, K., Bevans, B., & LARRC (Language and Reading Research Consortium). (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Child Language Disorders, 36*, 312–333.
- Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.
- Apel, K., Diehm, E., & Apel, L. (2013). Using multiple measures of morphological awareness to assess its relation to reading. *Topics in Language Disorders, 33*, 42–56.
- Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*, 325–346.
- Barnes, A., Kim, Y.-S., & Phillips, B. (2014). The relations of proper character introduction to narrative quality and listening comprehension for young children from high poverty schools. *Reading and Writing: An Interdisciplinary Journal, 27*, 1189–1205.
- Barnes, M. A., Ahmed, Y., Barth, A., & Francis, D. J. (2015). The relation of knowledge-text integration processes and reading comprehension in seven to twelfth grade students. *Scientific Studies of Reading, 19*, 253–272.
- Barnes, M. A., Raghubar, K. P., Faulkner, H., & Denton, C. A. (2014). The construction of visual-spatial situation models in children's reading and their relation to reading comprehension. *Journal of Experimental Child Psychology, 119*, 101–111.
- Barnes, M. A., Stuebing, K. K., Fletcher, J. M., Barth, A. E., & Francis, D. J. (2016). Cognitive difficulties in struggling comprehenders and their relation to reading comprehension: A comparison of group selection and regression-based models. *Journal of Research on Educational Effectiveness, 9*, 153–172.

- Barth, A. E., Barnes, M. A., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing, 28*, 587–609.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137–164.
- Connor, C. M. (2016). A lattice model of the development of reading comprehension. *Child Development Perspectives, 10*, 269–274.
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading, 19*, 114–134.
- Cromley, J., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325.
- Day, S. L., & Connor, C. M. (2017). Examining the relations between self-regulation and achievement in third-grade students. *Assessment for Effective Intervention, 42*, 97–109.
- Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade classroom environment. *Journal of School Psychology, 52*, 409–428.
- de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6*, 51–77.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., Fletcher, J. M., & Carter, S. (2015). Text processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and informational text. *Reading Research Quarterly, 50*, 393–416.
- Denton, C. A., Wolters, C. A., York, M., Swanson, E., Kulesz, P., & Francis, D. J. (2015). Adolescents' use of reading comprehension strategies: Differences related to reading proficiency, grade level, and gender. *Journal of Learning and Individual Differences, 37*, 81–95.
- Diakidoy, I. A. N., Mouskounti, T., & Ioannides, C. (2011). Comprehension and learning from refutation and expository texts. *Reading Research Quarterly, 46*, 22–38.
- Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the Simple View of Reading to account for variance within and across readers: The Complete View of Reading (CVRi). *Remedial and Special Education, 39*, 274–288.
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., et al. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework of core processes and constructs. *Educational Psychologist, 51*, 219–246.
- Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395.
- Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.
- Kim, Y.-S., & Phillips, B. M. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49*, 269–281.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A Construction-Integration Model. *Psychological Review, 95*, 163–182.
- Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology, 108*, 1078–1097.
- LARRC (Language and Reading Research Consortium). (2015a). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*, 151–169.
- LARRC. (2015b). The dimensionality of language ability in young children. *Child Development, 86*, 1948–1965.
- LARRC. (2015c). The dimensionality of Spanish in young Spanish-English dual-language learners. *Journal of Speech, Language, and Hearing Research, 58*, 754–766.
- LARRC. (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*, 1273–1284.
- LARRC & Chiu, Y. D. (2018). The Simple View of Reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education, 39*(5), 289–303.

- LARRC, Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Reading and Writing: An Interdisciplinary Journal*, 31, 1449–1477.
- LARRC, Jiang, H., Logan, J. A., & Jia, R. (2018). Modeling the nature of grammar and vocabulary trajectories from prekindergarten to third grade. *Journal of Speech, Language, and Hearing Research*, 61, 910–923.
- LARRC & Logan, J. (2017). Pressure points in reading comprehension: A quantile multiple regression analysis. *Journal of Educational Psychology*, 109, 451–464.
- LARRC & Muijselaar, M. L. M. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading*, 22, 117–136.
- LARRC & Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Reading and Writing: An Interdisciplinary Journal*, 30, 2039–2067.
- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., Donovan, S., & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, 9, 201–222.
- Lonigan, C. J., & Burgess, S. R. (2017). Dimensionality of reading skills with elementary-school-age children. *Scientific Studies of Reading*, 21, 239–253.
- Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the Simple View of Reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, 39, 260–273.
- Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of typically developing preschool through fifth-grade children. *Journal of Speech, Language, and Hearing Research*, 60, 2185–2198.
- McIlraith, A. L., and LARRC. (2018). Predicting word reading ability: A quantile regression study. *Journal of Research in Reading*, 41, 79–96.
- Murphy, K. A., LARRC, & Farquharson, K. (2016). Investigating profiles of lexical quality in preschool and their contribution to first grade reading. *Reading and Writing*, 29, 1745–1770.
- O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes: A Multidisciplinary Journal*, 43, 121–152.
- O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (Research Report No. RR-13-31). Princeton, NJ: Educational Testing Service.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, 30(9), 1344–1351.
- Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554–566.
- Perfetti, C. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). New York: Oxford University Press.
- Phillips Galloway, E., & Uccelli, P. (2019). Examining developmental relations between core academic language skills and reading comprehension for English learners and their peers. *Journal of Educational Psychology*, 111(1), 15–31.
- Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, 86, 159–175.
- RRSG (RAND Reading Study Group). (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA: RAND Corporation.
- Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, 28, 959–988.
- Swanson, E., Barnes, M., Fall, A., & Roberts, G. (2018). Predictors of reading comprehension among struggling readers who exhibit differing levels of inattention and hyperactivity. *Reading & Writing Quarterly*, 34(2), 132–146.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50, 337–356.

- van den Broek, P., Risdien, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Hillsdale, NJ: Erlbaum.
- Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review, 42*, 191–206.
- Wagner, R. K., Herrera, S. K., Spencer, M., & Quinn, J. M. (2015). Reconsidering the Simple View of Reading in an intriguing case of equivalent models. *Journal of Learning Disabilities, 48*, 115–119.
- Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology, 11*(3), 387–401.
- Wolters, C. A., Denton, C. A., York, M. J., & Francis, D. J. (2014). Adolescents' motivation for reading: Group differences and relation to standardized achievement. *Reading and Writing, 27*, 503–533.

# The Assessment of Reading for Understanding<sup>1</sup>

Panayiota Kendeou, *University of Minnesota*

## CONTENTS

EXECUTIVE SUMMARY .....	68
RECENT HISTORY OF LITERACY INITIATIVES .....	69
The RAND Reading Study Group: Needs in Reading Comprehension Assessment .....	69
The Reading for Understanding Initiative .....	70
CONTRIBUTIONS OF THE RFU RESEARCH INITIATIVE AND A FUTURE RESEARCH AGENDA .....	72
Authenticity: Complicating the Reading Comprehension Construct .....	72
Theoretically Based: Component and Process Theories of Reading Comprehension .....	73
Developmental Sensitivity: A Dynamic Construct .....	75
Instructional Sensitivity: Reflect the Effects of Intervention .....	77
Instructional Value: Identify Student Strengths and Weaknesses to Inform Instruction .....	81
Increased Complexity: Texts and Tasks .....	83
Prior Knowledge: An Integral Component .....	83
Technical Adequacy .....	84
Standardization and Efficiency .....	85
CONCLUSION .....	86
REFERENCES .....	87

---

<sup>1</sup> The writing of this report was funded in part by the Guy Bond Chair in Reading from the College of Education and Human Development at the University of Minnesota to P. Kendeou and by grant R305U160002 from the U.S. Department of Education to the National Academy of Education. P. Kendeou would like to thank Elly Orcutt, Reese Butterfuss, and Jasmine Kim at the University of Minnesota for their assistance. The opinions are those of the author and do not represent the policies of the U.S. Department of Education.



## EXECUTIVE SUMMARY

The U.S. Department of Education's Institute of Education Sciences (IES) created the Reading for Understanding (RfU) research initiative with the ultimate goal being to improve reading comprehension across pre-kindergarten (pre-K) through grade 12 in U.S. schools. The initiative funded a set of six connected projects (teams) that designed, developed, and tested new interventions and assessments in pre-K through grade 12. This chapter focuses primarily on the three main assessments developed by the assessment consortium, which consisted of the Educational Testing Service (ETS) in collaboration with the Florida Center for Reading Research (FCRR) at Florida State University (FSU). This consortium was tasked specifically with the development of a new summative assessment of reading comprehension across all grades. The five teams that designed and tested intervention programs in different age groups (elementary, middle, and high school) also developed assessments of various reading-related constructs. When relevant, this chapter also includes a discussion of a selected set of these measures because they showed evidence for innovation, technical adequacy, and promise for further development.

To address the RfU core assessment mission, the assessment consortium defined the construct of reading comprehension as *reading literacy*, which was measured by two assessment types: components of reading and global reading literacy. Two assessment systems were developed to assess *components of reading* in K–12: the Reading Inventory and Scholastic Evaluation (RISE) and the FCRR Research Reading Assessment (FRA). One assessment system was developed to assess *global reading literacy* in grades 3–12: the Global Integrated Scenario-Based Assessment (GISA). In addition to these three main assessments, a variety of measures were developed by the other teams (Language and Reading Research Consortium [LARRC], Catalyzing Comprehension through Discussion and Debate [CCDD], Promoting Adolescents' Comprehension of Text [PACT], and Reading, Evidence, and Argumentation in Disciplinary Instruction [READI]) to assess reading-related constructs, such as inference making, social perspective taking, knowledge acquisition, evidence-based argumentation, epistemic beliefs, and academic language, as well as classroom survey tools to assess teaching strategies and student strategies.

Our review and evaluation of these assessments and tools led to the conclusion that the RfU research initiative had a profound impact in the area of reading comprehension assessment. The initiative enabled innovative, large-scale work in diverse populations and contexts. Collectively, the set of assessments developed by ETS and FCRR can be characterized as a *new generation* of reading assessments. These assessments reflect a broader and more authentic conceptualization of reading comprehension, are developmentally sensitive, emphasize instructional sensitivity and value, and reflect the consequences of reading with comprehension. All assessments, those developed by the assessment consortium and the other teams, have a strong theoretical basis and defensible psychometric properties. The overall result is a set of *forward-thinking assessments* that promise to advance both research and practice in reading comprehension for years to come.

An important goal in the future research agenda would be to use these assessments in place of more traditional standardized reading comprehension measures. The use of these assessments in various populations and contexts will, in turn, inform further

development and refinement of reading comprehension theories and models, help evaluate with better precision additional aspects of reading comprehension in younger and older readers, and help understand more deeply the implications of integrating important moderators (such as prior knowledge) into the assessment design. Finally, because these new assessments reflect some of the inherent complexities of the comprehension process that only now have been realized in assessment, they open new possibilities for a future research agenda that can significantly advance theories of reading comprehension.

### RECENT HISTORY OF LITERACY INITIATIVES

In 1999, the U.S. Department of Education's Office of Educational Research and Improvement (the predecessor office to the Institute of Education Sciences) charged the RAND Reading Study Group (RRSG) with developing a research agenda to address pressing issues in literacy over the next 10 years. This initiative materialized in a 2002 publication (RRSG, 2002), in which the RRSG made recommendations for a future research agenda that focused on three areas: comprehension instruction, teacher education, and assessment. Pertinent to this report were the recommendations with respect to the assessment of reading comprehension. The RRSG proposed a new approach to assessment, advocating for a strong theoretical basis that is at the same time flexible to adapt and change in the presence of new empirical evidence. The group also advocated for using assessment to directly inform and improve instruction. Specifically, the call was for the design of technically adequate measures of reading comprehension that are sensitive to instructional interventions as well as to specific forms of reading instruction for all readers. The research agenda put forth by the RRSG informed the research focus and priorities set by the RfU research initiative 10 years later.

#### **The RAND Reading Study Group: Needs in Reading Comprehension Assessment**

The findings of the RRSG report were consistent with persistent criticisms of widely used reading comprehension assessments. These assessments have long been criticized for inadequately representing the complexity of reading comprehension and its development, lacking instructional utility (Klingner, 2004; Pearson & Hamm, 2005; Snyder, Caccamise, & Wise, 2005), and not meeting technical adequacy criteria (Mislevy, 2006, 2008). These assessments depend primarily on immediate recall and basic literal and inferential multiple-choice questions. Most important, none of these assessments are based on a current theory of reading comprehension (RRSG, 2002).

According to the RRSG, new assessments of reading comprehension needed to (a) reflect the dynamic, developmental nature of comprehension; (b) represent adequately the interactions among the dimensions of reader, activity, text, and context; and (c) satisfy criteria set forth by psychometric theory. Furthermore, these new assessments needed to also reflect the consequences of reading with comprehension, such as acquiring and applying knowledge. Most important, developing new assessments was of the highest priority as good assessments are a prerequisite to making progress with all other aspects of the research agenda on reading comprehension.

The minimum criteria for the development of new assessments put forth by the RRSg were the following:

1. Capacity to reflect authentic outcomes;
2. Consistency with actual comprehension processes;
3. Developmental sensitivity;
4. Capacity to identify poor comprehenders;
5. Capacity to identify subtypes of poor comprehenders;
6. Instructional sensitivity;
7. Openness to intraindividual differences;
8. Usefulness for instructional decision making;
9. Adaptability to individual, social, linguistic, and cultural variations; and
10. A basis in measurement theory and psychometrics.

It is important to note that the RRSg acknowledged that no single assessment could meet all of these criteria. Rather, the research agenda called for an *assessment system or systems* that would address different purposes, audiences, and populations.

### The Reading for Understanding Research Initiative

In 2010, IES funded the RfU research initiative (IES, 2010) to provide rigorous research to guide the development of better interventions and assessments across pre-K through grade 12. The Institute funded a set of connected projects that would design and test new interventions and assessments to improve *reading for understanding* across all readers in U.S. schools (Douglas & Albro, 2014). The RfU not only renewed professional interest in reading comprehension across the entire pre-K through grade 12 range, but also presented a unique opportunity to develop a community of researchers who undertook innovative work in the area of reading comprehension, with the potential to advance both research and practice.

#### *Core Assessment Mission*

To address the need for the development of a new reading comprehension assessment system, the RfU funded one assessment consortium, consisting of the ETS in collaboration with the FCRR at FSU. This consortium was tasked specifically with the development of a new summative assessment of reading comprehension in pre-K through grade 12. In this context, the assessment consortium expanded the definition of the construct of reading comprehension. The construct was identified as that of *reading literacy* and was measured by two assessment types: components of reading and global reading literacy (O'Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012; Sabatini & Bruce, 2009; Sabatini, Bruce, & Steinberg, 2013; Sabatini, O'Reilly, & Deane, 2013). The *components of reading* were assessed with RISE (Sabatini, Bruce, Steinberg, & Weeks, 2015; Sabatini, Weeks, et al., 2019) and with the FRA (Foorman, Petscher, & Schatschneider, 2015a, 2015b). *Global reading literacy* was assessed with the GISA (Sabatini, O'Reilly, Weeks, & Steinberg, 2016; Sabatini, O'Reilly, Weeks, & Wang, 2019).

*Additional Assessment Development*

It is important to note that the RfU also resulted in a set of additional measures and survey tools that were developed by the other teams in the context of their intervention work; that is, the teams needed to develop additional, often more specific, measures of reading comprehension or related language, knowledge, or cognitive processes in order to fully evaluate the impact of their interventions. For the purposes of this report, a selected set of these assessments were reviewed because they showed evidence for promise and technical adequacy for further development. Specifically, the LARRC developed an Inference Task (LARRC & Muijselaar, 2018) to assess local and global *inference processes*. The CCDD team developed two measures, the Assessment of Social Perspective-taking Performance (ASPP; Kim, LaRusso, Hsin, Selma, & Snow, 2018) to assess *social perspective taking* and the Core Academic Language Skills Instrument (CALs-I; Phillips Galloway & Uccelli, 2019; Uccelli et al., 2015a, 2015b) to assess *academic language*. The PACT team developed a Causal Inference Task to assess *inference making* (BRIDGE-IT; Barth, Barnes, Francis, Vaughn, & York, 2015) and a Background Knowledge measure (ASK; Vaughn et al., 2013) to assess *knowledge acquisition*. The READI team developed the Evidence-Based Argument (EBA) assessment (Goldman et al., 2016, 2019) to evaluate *evidence-based argumentation* and the literature epistemic cognition measure (Yukhymenko-Lescroart et al., 2016) to evaluate domain-specific *epistemic beliefs* in content areas. With respect to survey tools, the PACT team developed the Contextualized Reading Strategy Survey (CRSS; Denton, Wolters, et al., 2015) to evaluate *students' strategy use*, and the READI team developed a teacher survey scale to evaluate *attitude, self-efficacy, and argument/multiple source practices* as well as a classroom observation scale to evaluate *teaching practices and student activities* (Goldman et al., 2019). All assessments and surveys reviewed are listed in Figure 3-1.

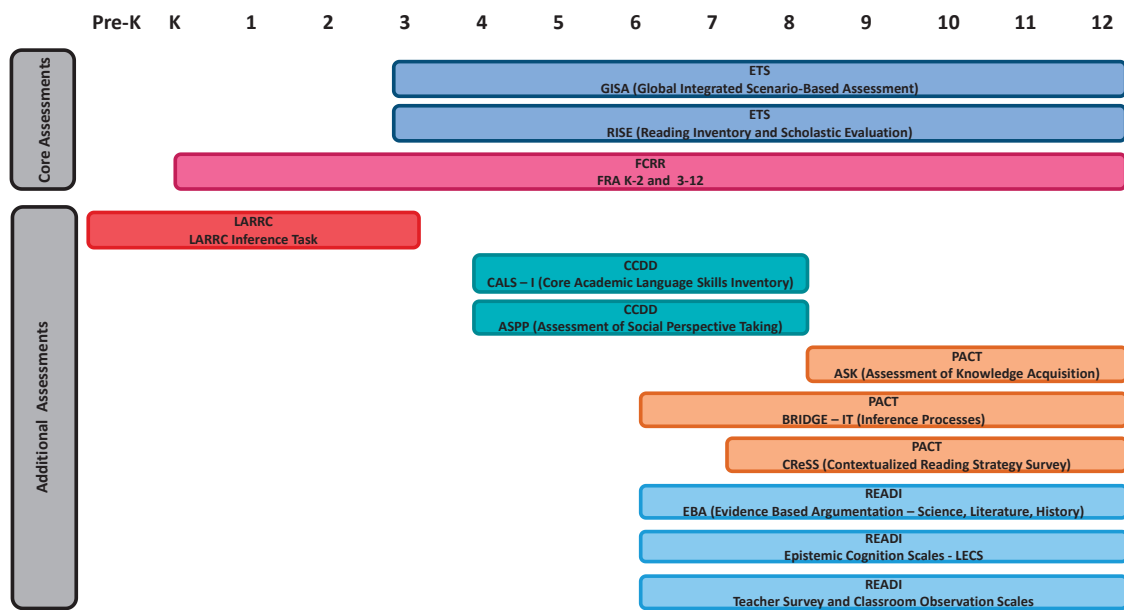


FIGURE 3-1 Assessments and classroom surveys reviewed.

## CONTRIBUTIONS OF THE RFU RESEARCH INITIATIVE AND A FUTURE RESEARCH AGENDA

To evaluate the contributions of the RfU research initiative on assessment, we followed an integrative approach that focused on the minimum criteria put forth by the RAND Research Study Group (2002), current trends in reading comprehension research, and an in-depth review of each assessment. The review of each assessment focused on the conceptual framework guiding development, content and sample items, administration and scoring guidelines, and evidence for technical quality focusing specifically on validity, reliability/precision, fairness in testing, and intended use of scores (AERA, APA, & NCME, 2014).

It is important to keep in mind the distinction between the assessments that emerged from the **core assessment mission** versus those developed to allow researchers to measure key facets of their interventions. For the assessments involved in the core mission, it was absolutely imperative to adhere to the highest psychometric standards; this meant that the ETS-FCRR team needed to engage in extensive and iterative large-scale, validity studies of the assessments. The other five teams were not funded to engage in extensive psychometric analyses, but they did engage in standard procedures for establishing the reliability, validity, and utility of their measures for the populations of students with whom they carried out their interventions. Nonetheless, we applied the same standards to all of the assessments introduced in this chapter and elaborated in the more detailed accounts in Appendix 3-1. Our hope in doing so was that readers of this report might understand the comprehensiveness of assessment tools that the RfU has made available to the worldwide community of researchers and educators.

Through this integrative, evaluative process, nine themes emerged that helped summarize the contributions of the RfU assessment research. What follows is the discussion of each of those theme contributions. The discussion of each theme concludes, where appropriate, with suggestions for more research that may be needed.

### **Authenticity: Complicating the Reading Comprehension Construct**

Reading comprehension is among the most complex of human activities. It involves processing words, connecting words using rules of syntax to understand sentences (Perfetti & Stafura, 2014), integrating meaning across sentences, drawing on relevant knowledge, generating inferences, identifying the structure of the text, and taking into consideration the authors' goals and motives (Graesser, 2015). The end product is a mental representation, what has been termed the "situation model" (Kintsch & van Dijk, 1978), which reflects the overall meaning of the text. For all of these processes to be successful, many interacting factors are playing a role, such as reader characteristics, text properties, context, and the demands of the reading task (Kintsch, 1998; RRSg, 2002).

The assessment consortium embraced the complexity of reading comprehension and expanded the construct definition. The construct was identified as that of *reading literacy*, defined as

the deployment of a constellation of cognitive, language, and social reasoning skills, knowledge, strategies, and dispositions, directed towards achieving specific reading purposes. (Sabatini, O'Reilly, & Deane, 2013, p. 7)

The decision to define and assess a broad construct such as reading literacy was innovative and contemporary. The decision was driven by recent policy efforts, including the Common Core State Standards for K–12 education in the United States (NGA & CCSSO, 2010), new social studies (NCSS, 2013) and science standards (NRC, 2012), the Partnership for 21st Century Skills (2008), frameworks for international assessments of reading such as the Programme for International Student Assessment (PISA; OECD, 2009a), the Programme for the International Assessment of Adult Competencies (PIAAC; OECD, 2009b), and the Progress in International Reading Literacy Study (PIRLS; Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009), and other assessment efforts and reforms (Bennett, 2011; Bennett & Gitomer, 2009; Gordon Commission, 2013).

Adopting a broad construct of reading comprehension embraces its complexity and allows for a focus on the entire *range* of reading processes, from foundational to higher-order processes (Fletcher, 2009; Goldman & Snow, 2015; Snow, 2018). Indeed, the focus in RISE and FRA is mostly on foundational reading skills, whereas the focus in GISA is on higher-level and goal-directed reading comprehension. Targeting this broad range of processes and embracing the complexity of reading also necessitates the integration of important variables that are expected to influence performance. These variables—(a) prior knowledge, (b) metacognitive and self-regulatory strategies, (c) reading strategies, and (d) student motivation and engagement—can affect the interpretation of reading comprehension scores (O’Reilly & Sabatini, 2013). For this reason, these variables were either directly assessed in the context of the assessment (this was the case for prior knowledge) or integrated in the assessment design (this was the case for all four). This approach is a considerable strength of GISA.

Expanding the reading comprehension construct enabled focus not only on higher-level processes during assessment, but also on deeper comprehension (Graesser, 2015; O’Reilly, Sabatini, & Wang, 2018), and thus deeper learning (Goldman & Pellegrino, 2015). As a result, and consistent with the recommendations made by the RRSO (2002), the focus shifted from comprehension to the consequences of reading with comprehension, such as acquiring and applying knowledge. This was accomplished by using a scenario-based assessment design (Bennett & Gitomer, 2009; O’Reilly & Sheehan, 2009), which approaches reading comprehension assessment as learning: it focuses on the consequences of comprehension rather than comprehension itself. The shift in focus from comprehension to learning is the main difference between GISA and most traditional reading comprehension assessments. This shift already has been embraced by several international literacy assessments (e.g., PISA, PIAAC, and PIRLS) and, once widely adopted, will present both a challenge and an opportunity for theory and practice in reading comprehension. *In other words, theory and practice also need to shift in focus from comprehension to learning, an issue that needs to be addressed in the future research agenda.*

### **Theoretically Based: Component and Process Theories of Reading Comprehension**

It has been argued repeatedly that reading comprehension models and theories have not directly informed past assessment efforts, and that new assessments should be based on an elaborated theory of reading comprehension (RRSO, 2002). The ETS-FCRR consortium drew on multiple theoretical frameworks and models to inform their

The primary goal of the ETS assessment project was to build a theoretically-driven, developmentally sensitive assessment system that spanned pre-K to grade 12. Our subgoal was to design assessments that address an expanding 21st-century reading construct, incorporate reading and learning science in the designs, and enhance instructional relevance, while still maintaining feasibility of implementation and psychometric quality.

—John Sabatini, Steering Committee Representative from ETS

assessment efforts. The use of multiple theories (as opposed to a single theory or model) is consistent with the inherent complexity of reading comprehension that makes it challenging for a single theory to describe the full range of cognitive, social, and linguistic processes involved (Perfetti & Stafura, 2014) or to make precise, testable predictions (Kendeou & O'Brien, 2018). Specifically, the consortium drew on both *component* and *process* models of reading comprehension, integrating different theoretical perspectives and views.

Component models focus on the identification of *component skills* that explain reading comprehension performance.

Reading component skills are subskills that can be isolated and assessed independently from higher-level reading comprehension (Perfetti & Adlof, 2012). Relevant to the RfU, component models of reading comprehension have been particularly influential for the development of the core assessments FRA and RISE as well as additional assessments, such as CALS-I, ASPP, and ASK (see Figure 3-1). These assessments include several of the component skills known to predict reading comprehension, such as word decoding and its precursors (Ehri, 2014), reading fluency (Fuchs, Fuchs, Hosp, & Jenkins, 2001), syntactic awareness (Cain & Nash, 2011; Crosson & Lesaux, 2013), vocabulary knowledge (Quinn, Wagner, Petscher, & Lopez, 2015), academic language (Snow, Lawrence, & White, 2009; Uccelli et al., 2015a, 2015b), language comprehension (Connor et al., 2014, 2018; Kendeou, van den Broek, White, & Lynch, 2009; Kim, 2016; Storch & Whitehurst, 2002), and perspective taking (LaRusso et al., 2016). Several of these components have been termed “pressure points” (Compton & Pearson, 2016), defined as skills that can result in robust variations in reading comprehension performance (Perfetti & Adlof, 2012). Among the component models in the extant literature, the Simple View of Reading (SVR; Hoover & Gough, 1990), which describes reading comprehension as the product of decoding and language comprehension, has been very influential for the development of both RISE and FRA. In the context of the SVR, decoding includes processes needed to decipher written code, such as phonological processing, orthographic processing, and word recognition, whereas language comprehension includes processes needed to build a coherent mental representation, such as vocabulary, academic language, and inference generation.

Process models focus on the identification of various *processes* involved in the construction of a mental text representation during reading (see McNamara & Magliano, 2009, for a review). An important assumption in most process models is that reading is a purposeful or goal-driven activity (Britt, Rouet, & Durik, 2018; McCrudden, Magliano, & Schraw, 2011). These purposes or goals influence readers’ desired level of comprehension or standards of coherence (van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011) and thus comprehension and learning from text. Relevant

to the RFU, several process models of reading comprehension have been particularly influential for the development of the core assessment GISA as well as additional assessments, such as BRIDGE-IT, the LARRC inference task, and EBA (see Figure 3-1). Among these models, the Construction-Integration Model (Kintsch & van Dijk, 1978) describes reading comprehension as the activation and integration of text information and relevant background knowledge into a coherent mental representation (i.e., a situation model) (Kintsch, 1988; van den Broek et al., 2005). The Landscape Model (van den Broek, Young, Tzeng, & Linderholm, 1999) specifies how the construction and integration processes are influenced by readers' standards of coherence or criteria for comprehension. The Documents Model Framework (Perfetti, Rouet, & Britt, 1999) and the Multiple-Document Task-based Relevance Assessment and Content Extraction model (MD-TRACE; Rouet & Britt, 2011) describe reading comprehension of multiple documents and texts and identify additional processes that are relevant in this context, including the evaluation and integration of information across sources (Goldman, Greenleaf, et al., 2016).

To the extent that theories of reading comprehension inform the development of reading comprehension assessments, evidence from the use of these assessments can also inform further development of reading comprehension theories. Indeed, the development of theoretically-based assessments has already begun to facilitate this reciprocal relation between theory and assessment. For example, ongoing work by the assessment consortium has produced new insights with respect to the relation of core component skills, such as decoding and reading comprehension. Wang, Sabatini, O'Reilly, and Weeks (2019) provided evidence for the nonlinear relation between decoding and reading comprehension by identifying a *decoding threshold* in grades 5–10 using RISE. Decoding below this threshold was only weakly related to reading comprehension and reading comprehension performance was limited. Decoding above this threshold positively predicted performance in reading. Wang et al. (2019) argued that the Decoding Threshold Hypothesis has the potential to explain differences in prominent reading theories in terms of the role of decoding in reading comprehension across development. *Thus, using evidence from the use of these assessments to further develop current theories of reading comprehension is an important goal in the future research agenda.*

### **Developmental Sensitivity: A Dynamic Construct**

Reading comprehension is a dynamic construct that changes across development (Weeks, 2018). That is because reader characteristics change with age and experience. As a result, the relative contribution of these characteristics to reading comprehension varies across development (van den Broek & Kendeou, 2017). For example, in the early elementary school grades decoding skills (the “reading” in reading comprehension) are a major contributor to reading comprehension, but in later elementary school grades and onward comprehension skills (the “comprehension” in reading comprehension), such as inference generation and oral language, are stronger predictors (Catts, Hogan, & Fey, 2003; Ehri, Nunes, Stahl, & Willows, 2001). This shift coincides with a transition from *learning to read* to *reading to learn* as complex informational texts become more common in the curriculum (Chall, Jacobs, & Baldwin, 1990; Goldman, Snow, & Vaughn, 2016; Snow & Sweet, 2003), and fits with conceptualizations of reading development as



a dynamic system (van Geert, 1991). This dynamic nature of the construct itself presents a challenge when the goal is to develop an assessment system that can span stages of development (e.g., K–12). The ETS-FCRR consortium addressed this challenge by taking into account the main determinants of the construct across development.

Specifically, in the FRA assessment there is a clear differentiation between kindergarten through grade 2 and grades 3–12, such that basic decoding skills are initially assessed with tasks that provide direct, specific measurement of letter-sound knowledge, phonological awareness, and spelling, whereas they are later assessed with tasks that provide measurement of the application of decoding skills, such as word recognition and vocabulary knowledge (Fitzgerald et al., 2014; Foorman, Francis, Davidson, Harm, & Griffin, 2004). Similarly, comprehension assessment also shifts from *listening* to *reading*, requiring students to either listen to or read passages (depending on their decoding proficiency). Furthermore, the system is designed to be administered in fall, winter, and spring to effectively track development of these “dynamic” skills in shorter periods.

RISE evaluates components of reading comprehension in grades 3–12, one dimension of the broader construct of reading literacy (O’Reilly et al., 2012; Sabatini, Bruce, & Steinberg, 2013; Sabatini, Weeks, et al., 2019). The range of these component skills—beginning with the recognition or decoding of words, to understanding the meanings of words and sentences, to building meaning from a text—is consistent with the developmental progression of reading comprehension (RRSG, 2002; Snow, 2018). Sabatini et al. (2015) provided initial evidence for unidimensionality of each component/subscale and examined all of the grade-level means and standard deviations, noting that, in general, the means incrementally increased across grades 5–10. Recently, Sabatini, Weeks, et al. (2019) provided further evidence for the unidimensionality of these components in grades 3–12.

GISA evaluates the expanded construct of reading comprehension following the same design across grades 3–12. The available GISA forms are scenario based, framed around different literacy goals, include either science or history topics, and have various numbers (range from 27 to 59) and types (constructed response, graphic organizer, and multiple choice) of questions. Despite these variations, Sabatini and colleagues (Sabatini, O’Reilly, et al., 2019) provided adequate evidence for the scale’s unidimensionality. Sabatini, O’Reilly, et al. (2016) also examined all of the grade-level means and standard deviations and noted that they reflected developmental differences in ability across grades. In general, the means increased across all grades. The one exception was in grade 12, where the mean was slightly lower than that in grade 11.

Sabatini, Halderman, O’Reilly, and Weeks (2016) also developed and tested a GISA form for K–3 in a small-scale study to evaluate the feasibility of the scenario-based assessment for younger students. The results showed ability differences across grade levels, even though third graders read silently, second graders read aloud, and kindergarten and first graders listened to the texts. Technical adequacy indices, though, were susceptible to the changes in delivery/modality, making it challenging to further develop GISA forms for K–2; for this reason, GISA begins at grade 3.

Thus, FRA, RISE, and GISA collectively assess a dynamic construct of reading comprehension across grades and demonstrate adequate developmental sensitivity. It is important to note that, with respect to component skills, the assessments extend from kindergarten through grade 12 (RISE is in grades 3–12 and FRA is in kindergarten

through grade 12) but with a validation sample only up to grade 10 for FRA. With respect to global reading literacy (GISA), the assessment extends from grades 3–12 but with less developmental sensitivity in grades 11 and 12 (and has no forms in K–2). This pattern of results suggests that at the ends of the developmental spectrum (pre-K, K–2, and grades 11 and 12), the broader construct of reading comprehension is not yet adequately captured by these assessment systems. *Thus, further developing measures of global reading literacy for younger readers while also refining those for older readers is an important goal in the future research agenda.*

The changing nature of the reading comprehension construct for older readers is also reflected in the approaches adopted by the three RfU teams that focused specifically on the development of interventions for adolescent students (Goldman, Snow, & Vaughn, 2016). Specifically, the READI team approached reading comprehension in grades 6–12 as a discipline-specific task (Shanahan & Shanahan, 2008) that requires readers to analyze, synthesize, and evaluate information within and across sources (Goldman, Greenleaf, et al., 2016, 2019; Lee & Goldman, 2015). The focus on sources also expanded the traditional notion of “text” to include print-based texts, images, audio, and video texts. This approach provided new insights on the higher-order and discipline-specific aspects of reading comprehension that are important for readers in the 21st century. Likewise, the PACT team approached reading comprehension in grades 7–12 through the lenses of content learning (Gersten, Baker, Smith-Johnson, Dimino, & Peterson, 2006; Vaughn et al., 2009), focusing primarily on prior knowledge activation, vocabulary building, text-based learning, and team-based learning (Vaughn et al., 2017). Finally, the CCDD team approached reading comprehension in grades 4–7 through discussion and debate, focusing on identifying different perspectives expressed in texts, learning academic vocabulary, and practicing academic language structures orally and in writing (Jones et al., 2019). The team’s work was motivated by an analysis of the tasks adolescents are meant to be accomplishing through reading, and how they differ from the tasks typically embedded in traditional comprehension assessments, which often require only relatively shallow inferences. Importantly, these three teams (READI, PACT, and CCDD) also developed measures to evaluate core constructs related to their intervention research, such as EBA (Goldman et al., 2019), knowledge acquisition (ASK knowledge measure; Vaughn et al., 2013), academic language (CALSI; Phillips Galloway & Uccelli, 2019; Uccelli et al., 2015a, 2015b), and social perspective taking (ASPP; Kim et al., 2018). These measures reflect a few of the additional aspects of the broader reading comprehension construct that are developmentally appropriate for middle and high school readers. *An important goal in the future research agenda would be to continue to identify aspects of the broader reading comprehension construct that are developmentally appropriate for different populations and disciplines.*

### **Instructional Sensitivity: Reflect the Effects of Intervention**

Instructional sensitivity, namely, an assessment’s capacity to reflect the effect of instruction or intervention, has been set as a core assessment criterion by the RRSB (2002) that has been realized in several RfU assessments. This is an important goal for any assessment since, historically, traditional measures of reading comprehension rarely show such sensitivity (Denton, Wexler, Vaughn, & Bryan, 2008). This is due, in

part, to the fact that different reading comprehension assessments often measure different aspects of the construct (Keenan, Betjemann, & Olson, 2008), and evidence for the impact of an intervention depends on whether the aspects of the construct being assessed are the same as those being trained (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014). This lack of sensitivity is also a consequence of transfer failure. Transfer (Barnett & Ceci, 2002; Day & Goldstone, 2012) is very difficult to evaluate and achieve in education settings in general, and in reading in particular (Gick & Holyoak, 1980, 1983). As a result, researchers have often distinguished between “proximal” and “distal” measures of reading comprehension in intervention studies (Connor et al., 2014). *Proximal* measures are closely tied to the intervention/instruction (require *near* or no transfer), whereas *distal* measures are generalized outcomes we expect to be influenced by the intervention/instruction (require *far* transfer). What is particularly promising for this new generation of assessments is that the three core assessments—GISA, RISE, and FRA—have shown some sensitivity to instruction, even though they would be considered distal measures.

Specifically, O'Reilly et al. (2014) demonstrated GISA's use as a summative assessment designed to provide evidence for the efficacy of the Reading Apprenticeship intervention (Monte-Sano, 2010). It is important to note that the underlying approach to assessment design in GISA (see O'Reilly & Sabatini, 2013; Sabatini & O'Reilly, 2013; Sabatini, O'Reilly, & Deane, 2013) had several elements in common with the Reading Apprenticeship program, making it more a proximal rather than a distal measure. The program was designed to train disciplinary reading in high school history, science, and literature and three GISA forms were specifically designed to evaluate the outcomes of the intervention in grades 9–12. O'Reilly et al. (2014) concluded that the GISA assessments were promising for use as outcomes in the intervention and sensitive to intervention effects.

Similarly, Goldman et al. (2019) also used GISA as a distal measure in a randomized controlled trial evaluating the efficacy of the READI Science intervention when compared to a business-as-usual control in grade 9. The READI Science intervention aims to improve reading comprehension by training evidence-based argumentation across multiple sources in science. The results showed that GISA was sensitive to the READI intervention effects with the treatment condition scoring significantly higher on GISA than the control condition. Notably, this effect held even after controlling for RISE at pretest.

Kim et al. (2017) included both RISE and GISA with the goal to evaluate the efficacy of the Strategic Adolescent Reading Intervention (STARI). STARI aims to improve reading comprehension by using reciprocal teaching strategies (Palincsar & Brown, 1984) and student discussion and debate within thematic text units in grades 6–8, while also building fluency through carefully selected leveled texts. Kim et al. (2017) reported that the program demonstrated significantly positive effects on the RISE word recognition, morphological awareness, and efficiency subtests.

Foorman, Herrera, Dombek, Schatschneider, and Petscher (2017) also demonstrated the utility of FRA K–2 as a summative measure to provide evidence for the efficacy of interventions. The study consisted of a randomized controlled trial that compared two early literacy interventions—one using standalone materials and one using materials embedded in the existing core reading program. The findings showed that the FRA

K–2 system was sensitive to intervention effects by demonstrating that the standalone intervention significantly improved spelling outcomes relative to the embedded intervention; other student outcomes were similar for the two interventions.

As noted earlier, the RfU also resulted in a set of additional measures developed specifically to evaluate the efficacy of various interventions and other reading-related constructs. These measures cover a range of different constructs and age groups. Specifically, the CCDD team developed the CALS-I (Uccelli et al., 2015a, 2015b) to evaluate the efficacy of the STARI and Word Generation interventions on improving students' *academic language* in grades 4–8 (LaRusso et al., 2016). The team also developed the ASPP measure (Kim et al., 2018) to assess students' social perspective taking since their intervention program hypothesized *social perspective-taking* performance as a core mechanism of learning. Both CALS-I and ASPP have shown evidence for instructional sensitivity (Kim et al., 2017; Phillips Galloway & Uccelli, 2019; Uccelli et al., 2015a, 2015b) and predictive relationships to comprehension (LaRusso et al., 2016).

The LARRC team developed an inference measure to assess global and local inference-making skills during listening comprehension in children pre-K through grade 3 (LARRC & Muijselaar, 2018) in the context of the language-based comprehension instruction Let's Know! (LARRC, Jiang, & Logan, 2019). LARRC (2015) provided evidence for the validity of a discourse skills factor that included this inference task along with four other measures of discourse skills. LARRC (2017) also provided evidence for the validity of a listening comprehension factor that included this inference task along with two other listening comprehension tests. Even though the team suggests that the measure could be used to evaluate the effects of language comprehension intervention or instruction, such evidence has not been provided yet.

The PACT team developed the ASK measure to evaluate middle and high school students' learning of U.S. history. ASK includes two subtests, one that measures content knowledge relevant to the intervention and one that measures reading comprehension. ASK has been used successfully to evaluate the efficacy of the PACT intervention in improving students' social studies content knowledge in grade 8 (Vaughn et al., 2013, 2015). The measure was also used to evaluate PACT's efficacy for English learners in grade 8 (Vaughn et al., 2017; Wanzek, Swanson, Vaughn, Roberts, & Fall, 2016), students with disabilities in grade 8 (Swanson et al., 2016; Wanzek, Swanson, Vaughn, Roberts, & Fall, 2016), and students in grade 11 (Wanzek et al., 2015).

The READI team developed the EBA measure to evaluate adolescents' ability to make evidence-based arguments from multiple sources in science. The EBA measure was used to evaluate the efficacy of the READI intervention that was designed to engage students in evidence-based argumentation from multiple text-based sources in grade 9 life sciences (Goldman, Greenleaf, et al., 2016). Goldman et al. (2019) showed that the multiple-choice component of the EBA measure was sensitive to instruction, with the intervention group performing significantly higher compared to the control group. EBAs were also developed for history and literature, along with rubrics for evaluating them. These were used in the context of the iterative design-based research conducted with a small number of teachers in each discipline. The EBAs in history and literature remain to be validated with larger samples of students.

The READI team also developed epistemic cognition scales in history, science, and literature. The science and history scales emphasized two dimensions of epistemic

cognition for multiple sources in history and science: the importance of corroborating across documents (history) and data sets and experiments (science), and the complexity and uncertainty of historical and scientific knowledge. The Literature Epistemic Cognition Scale (LECS; Yukhymenko-Lescroart et al., 2016) emphasized three dimensions: the multiple meanings of any literary work, the relevance of literature to life, and the importance of multiple readings of a literary work. The READI literature intervention (Goldman, Greenleaf, et al., 2016) used the LECS in the context of a 2-year longitudinal study of adolescents during their sophomore and junior years. Two of the subscales (multiple meanings and relevance to life) were significantly correlated with students' perceptions of the instructional context for literature (e.g., encouraging them to consider readings from multiple perspectives and to think about why writers and characters they create do what they do) as well as their self-reports of how frequently they analyzed their readings from different perspectives and considered how others interpreted readings.

With respect to the classroom survey measures, there is also increasing evidence for their instructional sensitivity. The PACT team developed the CReSS to evaluate four constructs related to students' reading strategy use in grades 7–12 (Denton, Wolters, et al., 2015). These constructs included *evaluation and integration strategies* (integrating current text information with previous text information and prior knowledge), *note-taking strategies* (identification of important text information), *regulation strategies* (adjustment of reading in response to difficulty), and *help-seeking strategies* (asking for help in response to difficulty). The survey is designed so that students respond to items targeting the use of comprehension strategies in four imagined reading situations. The first scenario involves reading a social studies textbook to prepare for a small group discussion and class presentation. The second scenario involves reading a story from an English language arts book to prepare for a quiz. The third scenario involves reading a self-selected nonfiction book in social studies in preparation for a written short report. The fourth scenario involves reading two articles from the internet to prepare for a class report. Denton, Wolters, et al. (2015) reported higher use of evaluation/integration and regulation strategies by adequate than struggling comprehenders, while the use of help seeking and note taking did not differ between these groups. Students at higher grade levels also reported greater use of evaluation/integration and regulation strategies than those in lower grades.

Finally, the READI team developed a self-report survey to assess teachers' *attitudes, self-efficacy, and argument/multiple source practices* (Goldman et al., 2019) in an effort to evaluate the impact of teacher professional development activities. Although developed and piloted in all three content areas (history, literature, and science), it was only validated in life sciences. Goldman et al. (2019) provided evidence that READI science intervention teachers scored significantly higher than those in the control condition on argument/multiple source practices at the conclusion of the intervention although there were no differences between the groups prior to the intervention. Even though there were no significant differences in attitude and self-efficacy from pre- to post-intervention, intervention teachers consistently scored higher than control teachers on the post-intervention administration. The READI team also developed a classroom observation protocol for the life sciences efficacy study and used it to evaluate teacher and student activities. Goldman et al. (2019) reported that, of the six constructs on that

protocol, READI intervention teachers improved from the first to the second observation but control teachers did not. Furthermore, at the end of the intervention, READI intervention teachers scored higher than control teachers on all six constructs, with significant differences and large effect sizes on two of the constructs (support for reading and collaboration).

Collectively, the measures developed in the context of the RfU show increased instructional sensitivity. This is true for the measures that by design were well aligned with the outcomes of the respective interventions, but also for the core measures (GISA, RISE, and FRA), as well as the classroom survey measures. Thus, the RfU has contributed to the literature a set of instructionally sensitive measures of knowledge, skills, and processes that contribute to using information obtained through reading single and multiple texts to address important questions. These include more generic skills such as use of academic language and perspective taking, as well as discipline-specific knowledge and skills. This “toolbox” of measures enriches the range of possibilities available to researchers in the field of reading and enables measurement of aspects of reading comprehension that are contemporary and innovative—aspects that were not possible to adequately measure before (e.g., global literacy, evidence-based argumentation, academic language, and social perspective taking). *An important goal in a future research agenda would be further development, calibration, and scale-up of these measures to evaluate their practical utility. In doing so, access to these measures by the scientific community would be necessary.*

### **Instructional Value:**

#### **Identify Student Strengths and Weaknesses to Inform Instruction**

Teachers need information with respect to students’ strengths and weaknesses in reading, as well as specific instructional recommendations that can address these weaknesses (Denton, Enos, et al., 2015; Pellegrino, DiBello, & Goldman, 2016). Indeed, effective teachers systematically collect and share student assessment data to make instructional decisions that improve student performance (Lipson, Mosenthal, Mekkelsen, & Russ, 2004) by as much as 0.20 standard deviations (Kingston & Nash, 2011). Effective evaluation of students’ reading skills and instruction planning, however, requires high-quality formative assessments that assess both comprehension processes and their products (Kendeou, McMaster, & Christ, 2016; van den Broek & Kendeou, 2014). Until recently, there were only a few high-quality formative assessments of reading comprehension (Afflerbach, Cho, & Kim, 2015), a need that has also been highlighted in the research agenda by the RRSF (2002).

The three core assessments produced in the context of the RfU partly address this need, particularly with respect to *component skills* of reading comprehension. Sabatini et al. (2015) suggested that RISE could be used to identify students’ strengths and weaknesses in conjunction with GISA. For example, RISE can be used to detect whether foundational reading skills are possible barriers to achieving higher levels of reading comprehension performance as reflected in GISA performance. Sabatini et al. (2014a) provided “proof of concept” of this approach in a small-scale study where they used RISE to create four subgroups of students (proficient, high basic, low basic, and below basic) and subsequently explored the extent to which each RISE subtest predicted

unique variance in GISA across these four ability groups. In this sample, each RISE subtest added significant unique variance predicting GISA scores, and together accounted for approximately 69 percent of the variance. Part of the residual variance unaccounted for presumably comprises the complex, deep comprehension required in GISA that cannot be captured by the individual subtests themselves. The results from this proof-of-concept study suggest that scores for each RISE subtest provide evidence for readers' strengths and weaknesses, and the combination of RISE and GISA assessments can provide useful insights on understanding students' reading ability. *It remains an open question whether training of these specific component skills improves reading comprehension. That would be an important next step in examining RISE's diagnostic accuracy.*

Foorman et al. (2015a, 2015b) provided strong evidence that the FRA assessment is an effective screening and diagnostic system of foundational reading comprehension skills. *Screening* in kindergarten through grade 2 is accomplished by evaluating foundational skills, such as phonological awareness, letter sounds, word reading, spelling, vocabulary, and following directions. In grades 3–12, screening is accomplished by evaluating word recognition, vocabulary knowledge, and reading comprehension. In each system, these tasks produce the Probability of Literacy Success (PLS) score following a weighted formula. The PLS score indicates the likelihood that a student will reach end-of-year expectations in literacy. For the purposes of the FRA, reaching expectations is defined as performing at or above the 50th percentile on the Stanford Achievement Test, Tenth Edition. The PLS is also color coded, providing the teacher with actionable information: red indicates the student is at high risk and needs targeted intervention ( $PLS < .50$ ), yellow indicates the student may be at risk and needs supplemental instruction ( $PLS > .50$  and  $\leq .70$ ), and green indicates the student is likely not at risk ( $PLS > .70$ ). Foorman et al. (2015a, 2015b) provided strong evidence for the predictive power of the PLS cutoff score in kindergarten through grade 10. The FRA team indicated that even though in the initial studies they also included grades 11 and 12, the sample was skewed toward lower-performing students in Florida, so they described it as having a K–10 proficiency range.

It is important to note that, despite the progress made with RISE and FRA, there are currently no formative assessments that evaluate the *actual processes* during reading comprehension. As outlined earlier in this report, current models of reading comprehension assume that comprehension involves the construction of a coherent mental representation of a text or "situation model" (Kintsch & van Dijk, 1978). These models differentiate between the actual processes that give rise to a mental product. An important next step in the development of assessments with instructional value is measures that can provide insights into the cognitive processes "in the moment." For example, the BRIDGE-IT measure (Barth et al., 2015) developed by the PACT team, a computerized inference measure for students in grades 6–12, is a good example of how one core comprehension process—inference making—can be evaluated in the moment. The test evaluates inference making by asking students to judge whether a continuation sentence is consistent or inconsistent with prior text; both accuracy and response times are considered as evidence for inference making. It is during these moment-by-moment processes that comprehension succeeds or fails (e.g., Kintsch, 1998). *Thus, the development, calibration, and scale-up of process assessment measures should be an important goal in the future research agenda. Technological advancements (e.g., eye-tracking methodologies) and trace or log data recorded in digital environments can be particularly helpful in this context.*

### Increased Complexity: Texts and Tasks

Most published, standardized reading comprehension assessments in the United States include a set of independent texts each with related literal and/or inferential multiple-choice questions with a task or goal to simply perform in the context of the assessment (Rupp, Ferne, & Choi, 2006). The majority of these assessments are also paper-based and do not include aspects of online reading or digital literacy (Kiili et al., 2018; Sabatini et al., 2015). The assessment consortium took a contemporary approach that expanded the types of tasks, texts, and questions associated with these texts.

With respect to text types, GISA shifted from the traditional set of independent texts to a *set of interrelated texts* that includes different sources and interactive communications (Sabatini et al., 2015). This was enabled with the adoption of a scenario-based design (Bennett, 2011; Bennett & Gitomer, 2009; O'Reilly & Sheehan, 2009). A scenario-based design provides test takers with a specific purpose for reading, a set of materials, and relevant questions. With respect to types of questions, GISA depends heavily on *multiple-choice questions*, but also incorporates two other types, *constructed-response items* summarizing text content and *graphic organizer items* organizing text content. These additional item types require strategies such as integration, synthesis, and application. With respect to task or context, GISA includes aspects of technology and digital environments by design (e.g., simulated peers, multiple sources), making the students' experience akin to learning (rather than testing). The tasks call for students to analyze, evaluate, synthesize, and report information and ideas.

The inclusion of various texts, questions, and tasks begins to address the RRSB (2002) call for assessments to evaluate the performance of an individual across activities with varying tasks and text types. In GISA, this was accomplished by using a scenario-based design. This was also accomplished in additional measures, such as EBA. EBA aligned tasks and texts with their disciplinary reading context (Lee & Goldman, 2015). Even though these are important steps forward, more work is needed to better understand how increased difficulty or complexity can be accomplished by taking into account various combinations of tasks and texts, and how to best utilize the affordances of digital environments in doing so. *Thus, exploring the extent to which scenario-based assessments can be used to introduce increased complexity in the assessment of reading comprehension is an important question in the future research agenda.*

### Prior Knowledge: An Integral Component

The inherent influence of prior knowledge on reading comprehension has always been a challenge for reading comprehension assessments. Traditionally, reading comprehension assessments aimed to eliminate rather than integrate prior knowledge, by including content that reduced knowledge demands (Francis et al., 2009; RRSB, 2002). This approach is less than optimal because prior knowledge is not only an integral component of reading comprehension, it is also one of the factors that carries the largest variability (Ahmed et al., 2016; Kendeou et al., 2016; McNamara & Kintsch, 1996) in middle and high school students (Goldman, Snow, & Vaughn, 2016). Prior knowledge is an integral component because at various points during reading, the reader draws on *different sources of knowledge*, including linguistic knowledge and general world knowledge (Perfetti & Stafura, 2014). The accuracy of that knowledge is also important:



accurate knowledge can facilitate reading comprehension, whereas inaccurate knowledge can severely disrupt it (Kendeou & O'Brien, 2015).

Rather than controlling prior knowledge, GISA included it as one of the important moderators in the assessment design (O'Reilly & Sabatini, 2013). This was accomplished by (a) measuring prior knowledge directly, (b) providing access to additional content during the assessment (e.g., videos, audio, definitions, diagrams) that supported students' prior knowledge, and (c) structuring the sequence of sources to facilitate knowledge acquisition. To measure prior knowledge, students were presented with a list of words/terms from a natural language processing database that provided a topical-association index for each word to a topic (Deane, 2012), and were asked to decide whether a term was related to the topic of the text. Students responded "Yes," "No," or "I don't know." O'Reilly et al. (2014) showed that this task was a quick and valid indicator of topic knowledge. By integrating a measure of prior knowledge into the assessment, one can investigate directly how student proficiency might interact with prior knowledge and whether students learn new content after taking the assessment. Indeed, GISA included prior knowledge measurement in a selected number of forms that functioned as a proof of concept for this approach (McCarthy et al., 2018; O'Reilly, Sabatini, & Wang, 2019).

In an elegant analysis, O'Reilly, Wang, and Sabatini (2019) used scores in this prior knowledge assessment to identify a *knowledge threshold*. Below the threshold, the relation between knowledge and performance on GISA was weak ( $\beta = 0.18$ ), whereas above the threshold, the relation between knowledge and performance was strong ( $\beta = 0.81$ ). These results show that integrating prior knowledge assessment into reading comprehension not only is feasible, but may also help identify what is the minimum knowledge required to comprehend information on a topic. *An important goal in the future research agenda is to evaluate at a larger scale the utility of integrating this type of prior knowledge test into assessment, and to understand better the implications for score interpretation.*

### Technical Adequacy

Following measurement theory and sound testing practices are key criteria for the construction of new assessments. In this report, the evaluation of the technical quality of the RfU assessments focused specifically on validity, reliability/precision, fairness in testing, and intended use of scores as outlined by the *Standards for Educational and Psychological Testing* (AERA et al., 2014). To meet these standards, the assessment consortium used sophisticated methodologies associated with test development and statistical analyses (e.g., measurement theory, classical test theory, and item response theory).

The calibration and validation studies for the three core assessment systems (GISA, RISE, and FRA) were *extensive*. More than 100,000 students in grades 3–12 participated from the Midwest, Northeast, Southern, and Western United States for RISE and GISA (Sabatini, 2017; Sabatini, Weeks, et al., 2019), and more than 70,000 students participated from kindergarten through grade 10 for FRA from the southern United States (Foorman et al., 2015a, 2015b). These studies not only included large national samples but were also *iterative* in item design and sample selection, resulting in significant improvements over the course of 5 years of development. Detailed technical reports have been produced for each assessment system that allow researchers and teachers to

evaluate whether each assessment fits their assessment needs from a technical adequacy perspective and help understand the scales, scores, and samples used to create them.

Across these assessments, the validity argument (Kane, 2013a, 2013b) integrated three types of evidence: (1) *evidence based on internal structure*, namely, the extent to which the relations among the test components conform to the hypothesized construct (in all instances, unidimensionality was hypothesized and tested); (2) *evidence based on test content*, namely, alignment of content to student learning standards; and (3) *evidence based on relations with other variables*, and specifically the extent to which test scores and other measures intended to assess similar constructs provided convergent and predictive evidence. Reliability/precision evidence was based primarily on *internal-consistency coefficients*. Finally, fairness in testing was based primarily on evidence for *lack of measurement bias* using differential item functioning. The latter was in line with the RRSB (2002) call for assessments that would *not* reflect social, linguistic, or cultural variation in reading comprehension performance.

Taken together, the results of the iterative and extensive calibration and validation studies suggest that the GISA, RISE, and FRA assessments have defensible psychometric properties. Given the large number of features that were novel in the design of these assessment (e.g., expanded construct, item types, being web based, automated scoring, and being computer adaptive), this is no small feat. This is particularly important for GISA, which uses a scenario-based assessment design in a digital environment, and various themed texts and types of items across forms.

It is also important to note that the additional assessments developed by the other RfU teams (LARRC, CCDD, PACT, and READI) and reviewed in this report also met basic technical adequacy standards. These assessments were developed primarily to evaluate reading-related constructs and intervention effects, so the validation and calibration studies were not extensive.

### Standardization and Efficiency

The three core assessments are characterized by standardization and efficiency. Specifically, both GISA and RISE (grades 3–12) are web administered and automatically scored (including selected constructed-response items). GISA takes 45 minutes to complete, whereas RISE takes 45–60 minutes. Both assessments make reference to reporting support that has the potential to be scalable at the classroom, school, or district level. It remains unclear, however, how researchers and practitioners can gain access to each of these assessments.

The FRA system consists of a K–2 system and a grades 3–12 system administered at three periods (fall, winter, and spring). Each system takes 45 minutes to complete. The K–2 system consists of screening, comprehension, and diagnostic tests that the teacher administers to students individually. The grades 3–12 system consists of screening and comprehension tests using web-based administration. The systems include reporting support that is scalable at the classroom, school, or district level. Notably, the FRA system is a computer-adaptive system; namely, the selection, order, and number of items administered depend on a student's response to the first item and each subsequent item of the assessment. Students receive harder or easier items based on their performance, and the system stops administering items once it has enough information

about the student's ability (i.e., a small enough amount of error or uncertainty associated with a student's score). Thus, this adaptive assessment maximizes precision efficiency—the maximum of precision of information with a minimum of time spent gaining it (Mitchell, Truckenmiller, & Petscher, 2015).

The efficiency of these new assessments with respect to administration and scoring “raises the bar” for current testing practices. *The future research agenda needs to continue to explore approaches, methodologies, and technologies that will increase further standardization and efficiency of reading comprehension assessments.*

## CONCLUSION

Historically, the assessment of reading comprehension is one of the most important outcomes of reform movements (Pearson & Hamm, 2005). Our evaluation is that the RfU research initiative had a profound impact on assessment, akin to that of reform movements. Collectively, the three core assessments developed—RISE, GISA, and FRA—can be characterized as a *new generation* of reading assessments. These assessments have a strong theoretical basis, reflect a broader and more authentic conceptualization of reading comprehension, are developmentally sensitive, emphasize instructional sensitivity and value, and have defensible psychometric properties. The calibration and validation studies were extensive, iterative, and undertaken across the United States. The result is a set of *forward-thinking assessments* that not only meet the standards of educational and psychological testing, but also promise to advance both research and practice in reading comprehension for years to come.

What the RfU has also contributed to the literature is a set of *additional measures of reading-related constructs* that are sensitive to high-quality instruction designed to improve different aspects of reading comprehension. These assessments also have a strong theoretical basis, reflect various aspects of reading comprehension or reading-related constructs, emphasize instructional sensitivity, and have defensible psychometric properties. This toolbox of measures enriches the range of possibilities available to researchers in the field of reading comprehension and enables measurement of aspects of reading comprehension that are contemporary and innovative (e.g., evidence-based argumentation, academic language, social perspective taking, online inference making). An important goal in a future research agenda would be further development, calibration, and scale-up of these measures to evaluate their practical utility.

The multiyear iterative efforts to develop these assessments also produced an incredible volume of empirical research that has used these assessments in small-scale, proof-of-concept studies; intervention studies; and large-scale calibration studies. The findings from the use of these assessments in various populations and contexts can inform further development and refinement of reading comprehension theories and models, help evaluate with better precision additional aspects of reading comprehension in younger and older readers, and help understand more deeply the implications of integrating important moderators (such as prior knowledge) into assessment design. An important goal in the future research agenda would be to use these assessments in place of more traditional standardized reading comprehension measures.

Advances in assessment influence instruction, and this new generation of assessments has the potential to transform current assessment practices and, thus, significantly

influence instructional practices in reading comprehension. Finally, because these new assessments reflect some of the inherent complexities of the comprehension process that only now have been realized in assessment, they open new possibilities in the future research agenda that can significantly advance the field of reading comprehension.

## REFERENCES

- Ackerman, B. P. (1984). The effects of storage and processing complexity on comprehension repair in children and adults. *Journal of Experimental Child Psychology, 37*, 303–334.
- Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading include a fluency component? *Reading and Writing, 19*, 933–958. doi: 10.1007/s11145-006-9024-z.
- AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Afflerbach, P., Cho, B. Y., & Kim, J. Y. (2015). Conceptualizing and assessing higher-order thinking in reading. *Theory into Practice, 54*(3), 203–212.
- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44–45*, 68–82.
- Albrecht, J. E., & O'Brien, E. J. (1991). Effects of centrality on retrieval of text-based concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 932–939.
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1061–1070.
- Armbruster, B. B., Anderson, T. H., & Meyer, J. L. (1991). Improving content-area reading using instructional graphics. *Reading Research Quarterly, 26*, 394–416. doi: 10.2307/747895.
- Barnes, M. A., Faulkner, H., Wilkinson, M., & Dennis, M. (2004). Meaning construction and integration in children with hydrocephalus. *Brain and Language, 89*, 47–56.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.
- Barr, C. D., Uccelli, P., & Phillips Galloway, E. (2019). Specifying the academic language skills that support text understanding in the middle grades: The design and validation of core academic language skills construct and instrument. *Language Learning, 69*, 978–1021.
- Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing, 28*(5), 587–609.
- Bean, T. W., Singer, H., Sorter, J., & Frazee, C. (1986). The effect of metacognitive instruction in outlining and graphic organizer construction on students' comprehension in a tenth-grade world history class. *Journal of Reading Behavior, 18*(2), 153–169. doi: 10.1080/10862968609547562.
- Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior, 16*(4), 297–306. doi: 10.1080/10862968409547523.
- Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 789–814). New York: Longman.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Beck, I. L., McKeown, M.G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples*. New York: Guilford.
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*, 70–91. doi: 10.1080/15366367.2010.508686.
- Bennett, R. E. (2011). *CBAL: Results from piloting innovative K-12 assessment* (ETS Research Report No. RR-11-23). Princeton, NJ: Educational Testing Service.

- Bennett, R. E., & Gitomer, D. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt Smith & J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61). Dordrecht: Springer. doi: 10.1007/978-1-4020-9964-9\_3.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Zimowski, M. F. (1997). Multi-group IRT. In W. J. Van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 433–448). New York: Springer-Verlag.
- Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, 44(1), 6–28.
- Britt, M. A., Rouet, J.-F., & Durik, A. M. (2018). *Literacy beyond text comprehension: A theory of purposeful reading*. New York: Routledge.
- Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, 25, 313–339. doi: 10.1080/02702710490522658.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429–441. doi: 10.1037/a0022824.
- Cain, K., & Oakhill, J. V. (1999). Inference making and its relation to comprehension failure. *Reading and Writing: An Interdisciplinary Journal*, 11, 489–503. doi: 10.1023/A:1008084120205.
- Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, 114, 647–662.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31–42. doi: 10.1037/0022-0663.96.1.31.
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96(4), 671–681.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12(3), 169–190.
- Carrow-Woolfolk, E. (1995). *Oral and written language survey (OWLS)*. Circle River, MN: American Guidance Service.
- Carrow-Woolfolk, E. (2008). *Comprehensive assessment of spoken language*. Torrance, CA: Western Psychological Services.
- Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36, 151–164. doi: 10.1177/002221940303600208.
- Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard.
- Coiro, J. (2009). Rethinking reading assessment in a digital age: How is reading comprehension different and where do we turn now? *Educational Leadership*, 66, 59–63.
- Coleman, D., & Pimentel, S. (2011). *Revised publishers' criteria for the Common Core State Standards in English language arts and literacy, Grades 3–12*. Washington, DC: National Governors Association and Council of Chief State School Officers. Retrieved from [http://www.corestandards.org/assets/Publishers\\_Criteria\\_for\\_3-12.pdf](http://www.corestandards.org/assets/Publishers_Criteria_for_3-12.pdf).
- Compton, D. L., & Pearson, P. D. (2016). Identifying robust variations associated with reading comprehension skill: The search for pressure points. *Journal of Research on Educational Effectiveness*, 9, 223–231.
- Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, 29(2), 186–204.
- Connor, C., Phillips, B., Kaschak, M., Apel, K., Kim, Y.-S., Al Otaiba, S., et al. (2014). Comprehension tools for teachers: Reading for understanding from prekindergarten through fourth grade. *Educational Psychology Review*, 26(3), 379–401. doi: 10.1007/s10648-014-9267-1.

- Connor, C., Phillips, B., Kim, Y.-S., Lonigan, C., Kaschak, M., 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*, 462–484.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.
- Crosson, A. C., & Lesaux, N. K. (2013). Connectives: Fitting another piece of the vocabulary instruction puzzle. *The Reading Teacher, 67*(3), 193–200. doi: 10.1002/TRTR.1197.
- Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57–75. doi: 10.1016/j.jecp.2015.03.005.
- Dahlberg, L. L., Toal, S. B., & Behrens, C. B. (1998). *Measuring violence-related attitudes, beliefs, and behaviors among youths: A compendium of assessment tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.
- Dawson, T. L. (2002). A comparison of three development stage scoring systems. *Journal of Applied Measurement, 3*, 146–189.
- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176.
- Deane, P. (2012). NLP methods for supporting vocabulary analysis. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Researching an understanding: Innovations in how we view reading assessment* (pp. 117–144). Lanham, MD: Rowman & Littlefield Education.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., et al. (2015). Text-processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and informational text. *Reading Research Quarterly, 50*, 393–416.
- Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research & Practice, 23*, 79–89.
- Denton, C. A., Wolters, C. A., York, M. J., Swanson, E., Kulesz, P. A., & Francis, D. J. (2015). Adolescents' use of reading comprehension strategies: Differences related to reading proficiency, grade level, and gender. *Learning and Individual Differences, 37*, 81–95.
- Diazgranados, S., Dionne, M. O., & Selman, R. L. (2011). *The social perspective taking acts measure (SPTAM)*. Unpublished manuscript, Graduate School of Education, Harvard University, Cambridge, MA.
- Diazgranados, S., Selman, R. L., & Dionne, M. (2015). Acts of social perspective taking: A functional construct and the validation of a performance measure for early adolescents. *Social Development, 25*, 572–601. doi: 10.1111/sode.12157.
- Dorans, N. J., & Kulick, E. (2006). Differential item functioning on the mini-mental state examination: An application of the Mantel Haenszel and standardization procedures. *Medical Care, 44*, S107–S114. doi: 10.1097/01.mlr.0000245182.36914.4a.
- Douglas, K. M., & Albro, E. R. (2014). The progress and promise of the reading for understanding research initiative. *Educational Psychology Review, 26*(3), 341–355.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test-IV*. Circle Pines, MN: American Guidance Service.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*, 5–21. doi: 10.1080/10888438.2013.819356.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*, 393–447. doi: 10.3102/00346543071003393.
- Ehri, L. C., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.
- Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology, 109*, 761–781.
- Ferguson, C. (1994). Dialect, register, and genre: Working assumptions about conventionalism. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 15–30). New York: Oxford University Press.

- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought and emotion. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development: Handbook of child psychology* (6th ed., Vol. 1, pp. 313–399). New York: Wiley.
- Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E. H., Bowen, K., Sanford-Moore, E., & Stenner, A. J. (2014). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology, 93*, 3–22. doi: 10.1037/a0037289.
- Fletcher, J. M. (2009). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323–330.
- Foorman, B., Espinosa, A., Wood, C., & Wu, Y. (2016). *Using computer-adaptive literacy assessments to monitor the progress of English language learner students*. (REL 2016-149). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.
- Foorman, B. R., Francis, D. J., Davidson, K. C., Harm, M., & Griffin, J. (2004). Variability in text features in sex grade 1 basal reading programs. *Scientific Studies of Reading, 82*(2), 167–197.
- Foorman, B., Herrera, S., Dombek, J., Schatschneider, C., & Petscher, Y. (2017). *The relative effectiveness of two approaches to early literacy intervention in grades K–12*. Retrieved from [https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL\\_2017251.pdf](https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2017251.pdf).
- Foorman, B. R., Petscher, Y., & Schatschneider, C. (2015a). *Florida Center for Reading Research (FCRR) Reading Assessment (FRA): Grades 3 through 12 technical manual*. Retrieved from <http://www.fcrr.org/for-researchers/fra.asp>.
- Foorman, B. R., Petscher, Y., & Schatschneider, C. (2015b). *Florida Center for Reading Research (FCRR) Reading Assessment (FRA): Kindergarten to grade 2 technical manual*. Retrieved from <http://www.fcrr.org/for-researchers/fra.asp>.
- Foorman, B. R., Petscher, Y., Stanley, C., & Truckenmiller, A. (2017). Latent profiles of reading and language and their association with standardized reading outcomes in kindergarten through tenth grade. *Journal of Research on Educational Effectiveness, 10*(3), 619–645. doi: 1080/19345747.2016.1237597.
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2009). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*, 301–322.
- Freed, J., & Cain, K. (2017). Assessing school-aged children's inference making: The effect of test story format in listening comprehension. *International Journal of Language and Communications Disorders, 52*, 95–105. doi: 10.1111/1460-6984.12260.
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45–58.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239–256.
- Gersten, R., Baker, S. K., Smith-Johnson, J., Dimino, J., & Peterson, A. (2006). Eyes on the prize: Teaching complex historical content to middle school students with learning disabilities. *Exceptional Children, 72*(3), 264–280.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.
- Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language–Receptive*. Austin, TX: Pro-Ed.
- Goldman, S. (2012). Adolescent literacy: Learning and understanding content. *Future of Children, 22*, 89–116.
- Goldman, S., & Snow, C. E. (2015). Adolescent literacy: Development and instruction. In A. Pollatsek & R. Treiman (Eds.), *Handbook on reading* (pp. 463–478). New York: Oxford University Press.
- Goldman, S., Snow, C. E., & Vaughn, S. (2016). Common themes in teaching reading for understanding: Lessons from three projects. *Journal of Adolescent and Adult Literacy, 60*, 255–264. doi: 10.1002/jaal.586.
- Goldman, S. R., Greenleaf, C., & Yukhymenko-Lescroart, M. (with Brown, W., Ko, M., Emig, J., George, M., Wallace, P., Blum, D., Britt, M. A., & Project READI). (2019). Explanatory modeling in science through text-based investigation: Testing the efficacy of the READI intervention approach. *American Educational Research Journal, 56*(4), 1148–1216. doi: 10.3102/0002831219831041.
- Goldman, S. R., McCarthy, K. S., & Burkett, C. (2014). Interpretive inferences in literature. In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 386–415). New York: Cambridge University Press.

- Goldman, S. R., & Pellegrino, J. W. (2015). Research on learning and instruction implications for curriculum, instruction, and assessment. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 33–41.
- Goldman, S. R., & Rakestraw, J. S. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Lawrence Erlbaum.
- Gordon Commission. (2013). *To assess, to teach, to learn: A vision for the future of assessment*. Princeton, NJ: Author. Retrieved from [http://www.gordoncommission.org/rsc/pdfs/gordon\\_commission\\_technical\\_report.pdf](http://www.gordoncommission.org/rsc/pdfs/gordon_commission_technical_report.pdf).
- Graesser, A. C. (2015). Deeper learning with advances in discourse science and technology. *Policy Insights from the Behavioral and Brain Sciences*, 2, 42–50. doi: 10.1177/2372732215600888.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395.
- Griffin, C. C., Malone, L. D., & Kameenui, E. J. (1995). Effects of graphic organizer instruction on fifth-grade students. *Journal of Educational Research*, 89, 98–107. doi: 10.1080/00220671.1995.9941200.
- Haberman, S. J. (2005). *When can subscores have value?* (Research Report No. RR-05-08). Princeton, NJ: Educational Testing Service.
- Halliday, M. A. K. (2004). *The language of science: Collected works of M.A.K. Halliday* (Vol. 5). New York: Continuum.
- Hill, M. (1991). Writing summaries promotes thinking and learning across the curriculum—but why are they so difficult to write? *Journal of Reading*, 34(7), 536–639.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Hogan, T., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children*, 44(3), 1–20.
- Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.
- Hsin, L., & Snow, C. (2017). From theory of mind to social perspective-taking: A benefit of bilingualism in academic writing. *Reading and Writing: An Interdisciplinary Journal*, 30(6), 1193–1214.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi: 10.1080/10705519909540118.
- IES (Institute of Education Sciences). (2010). *Reading for Understanding initiative*. Washington, DC: U.S. Department of Education. Retrieved from <https://ies.ed.gov/ncer/projects/program.asp?ProgID=62>.
- Jones, S. M., LaRusso, M., Kim, J., Kim, H. Y., Selman, R., Uccelli, P., et al. (2019). Experimental effects of word generation on vocabulary, academic language, perspective taking, and reading comprehension in high poverty schools. *Journal of Research on Educational Effectiveness*, 12(3), 448–483.
- Kane, M. T. (2013a). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Kane, M. T. (2013b). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1–73.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281–300.
- Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights from the Behavioral and Brain Sciences*, 3, 62–69.
- Kendeou, P., & O'Brien, E. J. (2015). Prior knowledge: Acquisition and revision. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Text and context* (pp. 151–163). New York: Routledge.
- Kendeou, P., & O'Brien, E. J. (2018). Theories of text processing: A view from the top-down. In M. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Handbook of discourse processes* (2nd ed.). New York: Routledge.
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101(4), 765–778. doi: 10.1037/a0015956.
- Kiili, C., Leu, D. J., Utriainen, J., Coiro, J., Kannianen, L., Tolvanen, A., Lohvansuu, K., & Leppanen, P. H. T. (2018). Reading to learn from online information: Modeling the factor structure. *Journal of Literacy Research*, 50, 304–334.



- Kim, H. Y., LaRusso, M., Hsin, L., Selman, R., & Snow, C. (2018). Social perspective-taking performance: Construct, measurement, and relations with academic performance and engagement. *Journal of Applied Developmental Psychology, 57*, 24–41.
- Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly, 52*(3), 357–382. doi: 10.1002/rrq.171.
- Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101–120. doi: 10.1016/j.jecp.2015.08.003.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L. & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch Report 8-75. Millington, TN: Naval Technical Training, U.S. Naval Air Station, Memphis, TN.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 164–182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Klingner, J. K. (2004). Assessing reading comprehension. *Assessment for Effective Intervention, 29*(4), 59–70.
- Kolen, M. J., & Brennan, R. L. (2004). *Testing equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.
- LARRC (Language and Reading Research Consortium). (2015). The dimensionality of language ability in young children. *Child Development, 86*, 1948–1965. doi: 10.1111/cdev.12450.
- LARRC. (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*, 1273–1284. doi: 10.1044/2017\_JSLHR-L-16-0039.
- LARRC & Muijselaar, M. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading, 22*, 117–136.
- 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*(8), 2812–2828.
- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., et al. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness, 9*(2), 201–222.
- Lee, C. D., & Goldman, S. R. (2015). Assessing literary reasoning: Text and task complexities. *Theory into Practice, 54*(3), 213–227.
- Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenge of adolescent literacy*. New York: Carnegie Corporation.
- Leslie, L., & Caldwell, J. S. (2011). *Qualitative reading inventory* (5th ed.). Boston: Pearson.
- Lipson, M. Y., Mosenthal, J. H., Mekkelsen, J., & Russ, B. (2004). Building knowledge and fashioning success one school at a time. *The Reading Teacher, 53*–542.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- MacGinitie, W. H., & MacGinitie, R. K. (1988). *Gates MacGinitie Reading Test*. Itasca, IL: Riverside.
- Madnani, N., Burstein, J., Sabatini, J., & O'Reilly, T. (2013). Automated scoring of a summary writing task designed to measure reading comprehension. In J. Tetreault, J. Burstein, & C. Leacock (Eds.), *Proceedings of the eighth workshop on innovative use of NLP for building educational applications* (pp. 163–168). Atlanta, GA: Association for Computational Linguistics.
- Martin, J., Sokol, B., & Elfers, T. (2008). Taking and coordination perspectives: From pre-reflective interactivity, through reflective inter-subjectivity, to meta-reflective sociality. *Human Development, 51*, 294–317. doi: 10.1159/000170892.

- McCarthy, K., Guerrero, T., Kent, K., Allen, L., McNamara, D., Chao, S., et al. (2018). Comprehension in a scenario-based assessment: Domain and topic-specific background knowledge. *Discourse Processes*, 55(5–6), 510–524. doi: 10.1080/0163853X.2018.1460159.
- McCrudden, M. T., Magliano, J., & Schraw, G. (Eds.). (2011). *Text relevance and learning from text*. Greenwich, CT: Information Age Publishing.
- McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19, 113–139. doi: 10.1007/s10648-006-9010-7.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288. doi: 10.1080/0163853X.1996.10648006-9010-7.
- McNamara, D. S., & Magliano, J. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297–384). New York: Academic Press.
- Mead, G. H. (1938). *Mind, self and society: From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257–306). Westport, CT: American Council on Education/Praeger.
- Mislevy, R. J. (2008). How cognitive science challenges the educational measurement tradition. *Measurement: Interdisciplinary Research and Perspectives*, 6, 124.
- Mislevy, R. J., & Sabatini, J. P. (2012). How research on reading and research on assessment are transforming reading assessment (or if they aren't, how they ought to). In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 119–134). Lanham, MD: Rowman & Littlefield Education.
- Mitchell, A. M., Truckenmiller, A., & Petscher, Y. (2015). Understanding computer adaptive assessments: Fundamentals and considerations for school psychologists. *Communique*, 43, 8.
- Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences*, 19, 539–568. doi: 10.1080/10508406.2010.481014.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from <https://files.eric.ed.gov/fulltext/ED512410.pdf>.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah, NJ: Lawrence Erlbaum.
- NCSS (National Council for the Social Studies). (2013). *The College, Career, and Civic life (C3) framework for social studies state standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history*. Silver Spring, MD: NCSS.
- NELP (National Early Literacy Panel). (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Retrieved from <https://lincs.ed.gov/publications/pdf/NELPReport09.pdf>.
- NGA & CCSSO (National Governors Association Center for Best Practices and Council of Chief State School Officers). (2010). *Common Core State Standards: College and career readiness standards for reading, writing and communication*. Washington, DC: Author. Retrieved from <http://www.corestandards.org/ELA-Literacy>.
- NICHD (National Institute of Child Health and Human Development). (2000). *Report of the National Reading Panel. Teaching children to read: Report of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Department of Health and Human Services.
- NRC (National Research Council). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- NRC. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Oakhill, J., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16, 91–121, doi: 10.1080/10888438.2010.529219.
- OECD (Organisation for Economic Co-operation and Development). (2009a). *PISA 2009 assessment framework—Key competencies in reading, mathematics, and science*. Paris: Organisation for Economic Co-operation and Development. Retrieved from [http://www.oecd.org/document/44/0,3746,en\\_2649\\_35845621\\_44455276\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/44/0,3746,en_2649_35845621_44455276_1_1_1_1,00.html).

- OECD. (2009b). *PIAAC literacy: A conceptual framework*. Paris: Organisation for Economic Co-operation and Development. Retrieved from <http://www.oecdilibrary.org/content/workingpaper/220348414075>.
- O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (ETS Research Report No. RR-13-31). Retrieved from [https://www.ets.org/research/topics/reading\\_for\\_understanding/publications](https://www.ets.org/research/topics/reading_for_understanding/publications).
- O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a RTI framework. *Reading Psychology, 33*(1–2), 162–189. doi: 10.1080/02702711.2012.631865.
- O'Reilly, T., Sabatini, J., & Wang, Z. (2018). Using scenario-based assessments to measure deep learning. In K. Millis, D. Long, J. Magliano, & K. Weimer (Eds.), *Deep comprehension: Multi-disciplinary approaches to understanding, enhancing, and measuring comprehension* (pp. 197–208). New York: Routledge.
- O'Reilly, T., Sabatini, J., & Wang, Z. (2019). What you don't know won't hurt you, unless you don't know you're wrong. *Reading Psychology, 40*, 638–677.
- O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively based assessment of, for and as learning: A framework for assessing reading competency* (RM-09-26). Princeton, NJ: Educational Testing Service.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science, 30*, 1344–1351.
- O'Reilly, T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading comprehension assessments for reading interventions: How a theoretically motivated assessment can serve as an outcome measure. *Educational Psychology Review, 26*, 403–424. doi: 10.1007/s10648-014-9269-z.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566. doi: 10.1037/0022-0663.98.3.554.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175.
- Partnership for 21st Century Skills. (2008). *21st century skills map*. Washington, DC: Author. Retrieved from [http://www.p21.org/storage/documents/21st\\_century\\_skills\\_english\\_map.pdf](http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf).
- Pearson, P. D., & Hamm, D. N. (2005). The history of reading comprehension assessment. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–70). Mahwah, NJ: Erlbaum.
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81.
- Perfetti, C., & Adlof, S. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how to assess reading ability* (pp. 3–20). Lanham, MD: Rowman & Littlefield.
- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum Associates.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22–37. doi: 10.1080/10888438.2013.827687.
- Petscher, Y., & Foorman, B. R. (2011). *Summary of the predictive relationship between the FAIR and the FCAT in grades 3–10: 2010–2011*. Tallahassee, FL: Florida Center for Reading Research.
- Phillips Galloway, E., & Uccelli, P. (2019). Examining developmental relations between core academic language skills and reading comprehension for English learners and their peers. *Journal of Educational Psychology, 111*(1), 15–31.
- Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology, 105*, 342–355.
- Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86*(1), 159–175. doi: 10.1111/cdev.12292.
- Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language, 29*, 419–448.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74.

- Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Relevance instructions and goal-focusing in text learning* (pp. 19–52). Greenwich, CT: Information Age Publishing.
- RRSG (RAND Reading Study Group). (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Corporation.
- Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23*, 441–474.
- Sabatini, J., Bruce, K., & Steinberg, J. (2013). *SARA reading components tests, RISE form: Test design and technical adequacy* (ETS Research Report No. RR-13-08). Princeton, NJ: Educational Testing Service.
- Sabatini, J., Bruce, K., Steinberg, J., & Weeks, J. (2015). *SARA reading components tests, RISE forms: Technical adequacy and test design, 2nd edition* (ETS Research Report No. RR-15-32). Princeton, NJ: Educational Testing Service.
- Sabatini, J., O'Reilly, T., & Deane, P. (2013). *Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design* (ETS Research Report No. RR-13-30). Princeton, NJ: Educational Testing Service.
- Sabatini, J., O'Reilly, T., Weeks, J., & Steinberg, J. (2016, April). *The validity of scenario-based assessment: Empirical results*. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2019). Engineering a 21st century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*. doi: 10.1080/15305058.2018.1551224.
- Sabatini, J., Weeks, J., O'Reilly, T., Bruce, K., Steinberg, J., & Chao, Z. (2019). *SARA reading components tests, RISE forms: Technical adequacy and test design, 3rd edition* (Research Report No. RR-15-32). Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12076.
- Sabatini, J. P. (2017). *Reading for Understanding assessment network year 7 annual report*. Princeton, NJ: ETS.
- Sabatini, J. P., & Bruce, K. (2009). *PIAAC reading component: A conceptual framework* (OECD Education Working Papers No. 33). Paris: OECD Publishing.
- Sabatini, J. P., Halderman, L. K., O'Reilly, T., & Weeks, J. P. (2016). Assessing comprehension in kindergarten through third grade. *Topics in Language Disorders, 36*, 334–355.
- Sabatini, J. P., & O'Reilly, T. (2013). Rationale for a new generation of reading comprehension assessments. In B. Miller, L. E. Cutting, & P. McCardle (Eds.), *Unraveling reading comprehension: Behavioral, neurobiological and genetic components* (pp. 100–111). Baltimore: Paul H. Brookes.
- Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014a). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disability Research & Practice, 29*(1), 36–43. doi: 10.1111/ldrp.12028.
- Sabatini, J. P., O'Reilly, T., Halderman, L., & Bruce, K. (2014b). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Année psychologique, 114*, 693–723. doi: 10.4071/S0003503314004059.
- Schatschneider, C., Petscher, Y., & Williams, K. M. (2008). How to evaluate a screening process: The vocabulary of screening and what educators need to know. In L. Justice & C. Vukelic (Eds.), *Every moment counts: Achieving excellence in preschool language and literacy instruction* (pp. 304–317). New York: Guilford Press.
- Schraw, G. (2000). Reader beliefs and meaning construction in narrative text. *Journal of Educational Psychology, 90*, 705–715. doi: 10.1037/0022-0663.92.1.96.
- Semel, E., Wiig, E., & Secord, W. (2003). *Clinical evaluation of language fundamentals—4 (CELF-4)*. San Antonio, TX: PsychCorp.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content area literacy. *Harvard Educational Review, 78*(1), 40–59.
- Shin, J., Deno, S. L., & Espin, C. A. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *Journal of Special Education, 34*, 164–172. doi: 10.1177/002246690003400305.
- Sinatra, G. M., Kienhues, D., & Hofer, B. K. (2014). Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist, 49*(2), 123–138.

- Sinharay, S., Haberman, S., & Puhan, G. (2007). Subscores based on classical test theory: To report or not to report. *Educational Measurement: Issues and Practice*, 26, 21–28. doi: 10.1111/j.1745-3992.2007.00105.x.
- Snow, C., Lawrence, J., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), 325–344.
- Snow, C. E. (2018). Simple and not-so-simple views of reading. *Remedial and Special Education*, 39, 313–316. doi: 10.1177/0741932518770288.
- Snow, C. E., & Sweet, A. P. (2003). Reading for comprehension. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 1–11). New York: Guilford Press.
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York: Cambridge University Press.
- Snyder, L., Caccamise, D., & Wise, B. (2005). The assessment of reading comprehension: Considerations and cautions. *Topics in Language Disorders*, 25, 33–50.
- Stanford Achievement Test Series Tenth Edition (SAT-10). *Pearson Assessment*. Pearson.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. doi: 10.1037//0012-1649.38.6.934.
- Swanson, E. A., Wanzek, J., McCulley, L., Stillman-Spisak, S., Vaughn, S., Simmons, D., et al. (2016). Literacy and text reading in middle and high school social studies and English language arts classrooms. *Reading & Writing Quarterly*, 32(3), 199–222.
- Swanson, E. A., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A.-M. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children*, 81, 426–442.
- Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository material. *Journal of Educational Psychology*, 74, 323–340.
- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency—second edition (TOWRE-2)*. Austin, TX: Pro-Ed.
- Uccelli, P., Barr, C. D., Dobbs, C. L., Phillips Galloway, E., Meneses, A., & Sanchez, E. (2015a). Core Academic Language Skills (CALs): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics*, 5, 1075–1107. doi: 10.1017/S0142716400006X.
- Uccelli, P., Phillips Galloway, E., Aguilar, G., & Allen, M. (2019). Amplifying and affirming students' voices through CALs-informed instruction. *Theory into Practice*, 59, 75–88.
- Uccelli, P., Phillips Galloway, E., Barr, C., Meneses, A., & Dobbs, C. (2015b). Beyond vocabulary: Core Academic Language Skills (CALs) that support text comprehension. *Reading Research Quarterly*, 50(3), 261–359. doi: 10.1002/rrq.10.
- van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–140). Charlotte, NC: Information Age Publishing.
- van den Broek, P., & Kendeou, P. (2014). Special issue guest editors' preface problems in reading comprehension: Connecting theory and practice. *Learning Disabilities Research & Practice*, 29(1), 2.
- van den Broek, P., & Kendeou, P. (2017). Development of reading comprehension: Change and continuity in the ability to construct coherent representations. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 283–308). Amsterdam: John Benjamins.
- van den Broek, P., Kendeou, P., Kremer, K., Lynch, J. S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Erlbaum.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates.

- van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98, 3–53. doi: 10.1037/0033-295X.98.1.3.
- Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, 2(4), 297–324.
- Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E. A., & Fall, A.-M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology*, 109, 22–34.
- Vaughn, S., Roberts, G., Wexler, J., Vaughn, M., Fall, A.-M., & Schnakenberg, J. (2015). High school students with reading comprehension difficulties: Results of a randomized control trial of a two-year reading intervention. *Journal of Learning Disabilities*, 48(5), 546–558.
- Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, 48, 77–93.
- Vellutino, F., Tunmer, W., Jaccard, J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11, 3–32. doi: 10.1207/s1532799xssr1101\_2.
- von Davier, H., & Xu, X. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika*, 76(2), 318–336. doi: 10.1007/s11336-011-9202-z.
- Wagner, R.K., Torgesen, J., Rashotte, C.A., & Pearson N. (2010). *Test of Sentence Reading Efficiency and Comprehension*. Austin, TX: Pro-Ed.
- Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology*, 111, 387–401.
- Wanzek, J., Swanson, E., Vaughn, S., Roberts, G., & Fall, M-A. (2016). English learner and non-English learner students with disabilities: Content acquisition and comprehension. *Exceptional Children*, 82, 428–442.
- Wanzek, J., Swanson, E. A., Vaughn, S., Roberts, G., & Kent, S. C. (2015). Promoting acceleration of comprehension and content through text in high school social studies classes. *Journal of Research on Educational Effectiveness*, 8, 169–188.
- Weeks, J. P. (2018). An application of multidimensional vertical scaling. *Measurement: Interdisciplinary Research and Perspectives*, 16, 139–154. doi: 10.1080/15366367.2018.1502005.
- Wellborn, J. G., & Connell, J. P. (1987). *Manual for the Rochester Assessment Package for Schools*. Rochester, NY: University of Rochester.
- 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. doi: 10.1002/pits.20307.
- Wood, P., & Kardash, C. (2002). Critical elements in the design and analysis of studies of epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 231–260). Mahwah, NJ: Lawrence Erlbaum Associates.
- Woodcock, R. W., McGraw, K. S., & Mather, N. (2001). *Woodcock-Johnson III Test*. Itasca, IL: Riverside Publishing.
- Yukhymenko-Lescroart, M. A., Briner, S. W., Magliano, J. P., Lawless, K., Burkett, C., McCarthy, K. S., et al. (2016). Development and initial validation of the Literature Epistemic Cognition Scale (LECS). *Learning and Individual Differences*, 51, 242–248.



# Appendix 3-1<sup>1</sup>

## Brief Reviews of Reading for Understanding Assessments

### CORE ASSESSMENTS

- Global Integrated Scenario-Based Assessment (GISA)
- Reading Inventory and Scholastic Evaluation (RISE)
- Florida Center for Reading Research Reading Assessment (FRA)
- FRA K–2 System
- FRA Grades 3–12 System

### ADDITIONAL ASSESSMENTS

- LARRC Inference Task
- Core Academic Language Skills Instrument (CALSI)
- Assessment of Social Perspective-Taking Performance (ASPP)
- ASK Knowledge Acquisition Measure
- BRIDGE-IT Measure
- READI Literature Epistemic Cognition Scale (LECS)
- READI Evidence-Based Argument (EBA) Measure

### GLOBAL INTEGRATED SCENARIO-BASED ASSESSMENT (GISA)

#### Conceptual Framework

GISA is designed to measure *global reading literacy*, the second dimension of the broader construct of *reading literacy* (O'Reilly et al., 2012; Sabatini & Bruce, 2009; Sabatini, Bruce, & Steinberg, 2013). Global reading literacy is defined as “the deployment of a constellation of cognitive, language, and social reasoning skills, knowledge, strategies, and dispositions, directed towards achieving specific reading purposes” (Sabatini, O'Reilly, & Deane, 2013, p. 7).

This assessment system uses a scenario-based design (Bennett, 2011; Bennett & Gitomer, 2009) that measures various levels of reading comprehension in a range of reading situations. Specifically, a scenario-based design provides test takers with a specific purpose for reading, a set of materials, and relevant questions. The scenario-based design is consistent with that of Cognitively-Based Assessment for, of, and as Learning (CBAL), a large Educational Testing Service (ETS) initiative that has been focusing on building innovative assessments in English language arts, math, and science (Bennett, 2010).

---

<sup>1</sup> These brief reviews were based on technical reports and other publications for each measure provided by the research teams at the time of writing this publication. The primary sources should be consulted for complete and detailed information.



GISA builds heavily on process models of reading comprehension that attempt to identify core processes that explain reading comprehension performance. In this context, the assessment team identified three principles to guide the assessment design (principles 4, 5, and 6; Sabatini, O'Reilly, & Deane, 2013). These principles state that reading is viewed as a purposeful activity (McCrudden, Magliano, & Schraw, 2011) that involves the construction of meaning at multiple levels, from literal to text base and situation models (Kintsch, 1998; McNamara & Kintsch, 1996); skilled reading includes proficiency in evaluating and synthesizing information across multiple texts in a digital environment (Britt & Sommer, 2004; Rouet & Britt, 2011); and reading growth involves the expansion of both knowledge and skills (RRSG, 2002). The focus in GISA is on higher-level reading comprehension rather than foundational reading skills. In this context, a set of moderators is expected to influence performance. These moderators—(1) background knowledge, (2) metacognitive and self-regulatory strategies, (3) reading strategies, and (4) student motivation and engagement—affect the interpretation of reading comprehension scores (O'Reilly & Sabatini, 2013). For this reason, the moderators are either directly assessed in the context of the assessment (this is the case for background knowledge and reading strategies) or integrated in the assessment design (this is the case for all four moderators).

*Background knowledge* provides an indicator of students' knowledge on the topic of the texts in the assessment. This is an important moderator because it can be used as an indicator of students' ability to learn, update, and apply information. In the assessment design this is accomplished by (a) measuring background knowledge directly, (b) providing access to additional content during the assessment (e.g., videos, audio, definitions, diagrams) that supports the test taker's background knowledge, and (c) structuring the sequence of sources (from general to specific) to facilitate knowledge building.

*Metacognitive and self-regulatory strategies* and behavior provide an indicator of students' ability to monitor their understanding and their ability to repair gaps, errors, and misconceptions. This is an important moderator because it can be used as an indicator of the accuracy of students' judgments of learning and ability to use available resources to solve problems and correct mistakes. In the assessment design this is accomplished by (a) setting goals for reading, (b) sequencing sources, (c) providing feedback/hints after an error, (d) evaluating peer responses in a simulated peer collaboration, and (e) accessing and using supplemental resources.

*Reading strategies* provide an indicator of students' strategic use of text. This is an important moderator because it can be used as an indicator of students' ability to use strategies such as paraphrasing and summarization. In the assessment design this is accomplished by including items that require students to (a) paraphrase, (b) summarize, and (c) graphically organize information.

*Motivation and engagement* provide an indicator of students' willingness to expend sufficient effort to understand text. This is an important moderator because it can be used as an indicator of students' interest on the topics and texts and engagement with specific tasks. In the assessment design this is accomplished by including goal-directed, authentic scenarios.

GISA includes a number of different item types, designed to integrate in the design information about the aforementioned moderators. These include constructed-response (CR), graphic organizer (GO), and multiple-choice (MC) item types. The CR items involve

constructing summaries according to a rubric that requires the following: the first sentence must be about the entire text; the next three sentences must be about each of the paragraphs in the text; and students must use their own words and exclude their own opinions (Madnani, Burstein, Sabatini, & O'Reilly, 2013). The choice to include a summary type of CR item was motivated by evidence demonstrating that summarization enhances both comprehension (Bean & Steenwyk, 1984; Hill, 1991; Taylor, 1982) and metacognition (Thiede & Anderson, 2003). The GO items involve visualizing and understanding the organizational structure of a text. These items are always partially completed to help students understand the structure of each text (e.g., a 3 × 4 cell). The choice to include GO items was motivated by evidence suggesting that these tasks help construct coherent models of text content (Armbruster, Anderson, & Meyer, 1991; Bean, Singer, Sorter, & Frazee, 1986; Griffin, Malone, & Kameenui, 1995). The MC items involve higher-order processes such as evaluating sources, questions, perspectives, and quality of information.

In some test forms, it is possible to assess relevant *background knowledge* before, after, or before and after to evaluate learning from the assessment. For the background knowledge test students are asked to decide whether a term is *related* to the topic of the text (e.g., farming). Students can choose "Yes," "No," or "I don't know" and the instructions make clear that the section will not count toward their total reading score. Previous work has shown that this task is a quick and valid indicator of students' prior knowledge of the topic (Deane, 2012; O'Reilly et al., 2014).

### Description

GISA measures a broader conception of reading literacy ability, consistent with cognitive models of reading comprehension. GISA is a 45-minute, web-administered, scenario-based assessment that includes authentic reading situations. Specifically, test takers are provided with a specific purpose for reading (e.g., studying for a test, preparing for a class presentation, etc.), a set of materials (e.g., websites, blogs, newspaper articles, etc.), and progress through the materials in a structured and scaffolded way. GISA examines the test taker's proficiencies in (a) constructing different levels of mental model representations (Kintsch, 1998), (b) familiarity with text structure and genre differences (Goldman & Rakestraw, 2000), (c) deployment of executive/metacognitive processes (Schraw, 2000), and (d) application of strategies for attaining a literacy goal (McCrudden & Schraw, 2007; van den Broek et al., 2011). Currently, there are 19 test forms available for grades 3–12 that include either science or history/language topics. The number of items varies across forms and ranges from 27 to 59. The types of items also vary across forms and include CR, GO, and MC. Each test form follows the same structure as described below.

*GISA Forms (Sabatini, O'Reilly, Weeks, & Steinberg, 2016)*

First, prior to reading the texts, students are presented with a scenario. For example (from the *Organic Farming* test form; Sabatini et al., 2014a):

Your class has decided to create a website about organic farming to help members of the community become more familiar with the subject. The website will provide information to answer the following questions: What are the natural methods used in organic

farming? How are these methods different from the methods used on non-organic, or conventional, farms? What are the pros and cons of organic farming? You will work with three classmates on the project.

The sources provided in this test form were texts on techniques used in organic farming, simulated results of a web search, advantages/disadvantages of organic farming, a simulated web discussion, cartoons, charts, and graphic organizers. Text readability ranged from grades 4–9 based on the Flesch-Kincaid readability formula (Kincaid, Fishburne, Rogers, & Chissom, 1975). This specific test form includes a total of 35 items as follows: 2 CR items summarizing text content; 7 GO items organizing text content; 3 MC items demonstrating detailed text understanding; 3 MC items demonstrating web source evaluation; 2 MC items and 4 GO items demonstrating evaluation of the advantages and disadvantages of organic farming; 5 MC items demonstrating word and sentence understanding by choosing synonyms in sentence context; and 9 MC items evaluating perspective taking and information quality in a simulated web discussion.

In this test form, relevant *background knowledge* is also being assessed. Students are asked to decide whether a term is *related* to the topic of farming. Students can choose “Yes,” “No,” or “I don’t know” and the instructions make clear that the section will not count toward their total reading score.

#### *GISA Administration and Scoring*

GISA includes a computerized administration, automated scoring, and reporting support scalable at the classroom, school, or district level. A single score is produced that is subsequently scaled using 2PL/GPCM (Bock & Zimowski, 1997) and then rescaled. The scores are rescaled to have a mean of 1000 and a standard deviation of 100.

#### **Sample**

Evidence reported next (unless otherwise noted) is based on a large-scale field study that recruited students from all four regions of the United States: Midwest, Northeast, South, and West (Sabatini, O’Reilly, Weeks, & Steinberg, 2016). In this study, a total of 12,317 students in grades 3–12 participated. Specifically, there were 1,107 students in grade 3, 1,089 in grade 4, 1,178 in grade 5, 1,355 in grade 6, 1,403 in grade 7, 1,231 in grade 8, 1,401 in grade 9, 1,388 in grade 10, 1,153 in grade 11, and 1,012 in grade 12. In terms of ethnicity, 31.8 percent were Hispanic/Latino. In terms of race, 1.1 percent were American Indian/Native Alaskan, 2.9 percent Asian, 11.9 percent Black, 0.6 percent Native Hawaiian/Pacific Islander, 33.8 percent White, and 17.2 percent other/not reported. Also, 51.4 percent were female, 48.5 percent male, and 1 percent not reported. No other demographic information was reported other than the sample median of students receiving English-language learners’ services (5 percent). Tests were administered in school computer labs and proctored by trained school staff members.

#### **Reliability/Precision**

Reliability/precision evidence for GISA was based on *internal-consistency coefficients*. Sabatini, O’Reilly, et al. (2016) computed Cronbach’s alpha coefficients (Cronbach, 1951)

for each test form across all grades (grades 3–12). The range of reliabilities were generally within acceptable range (.72. to .89).

Additional reliability coefficients have also been reported in several papers that evaluated specific forms of GISA. For example, Sabatini, O'Reilly, Halderman, and Bruce (2014b) performed a component reliability analysis for the Organic Farming GISA form, which included 35 items. A sample of 426 grade 6 students completed this form. Cronbach's alpha coefficient was  $\alpha(426) = .89$ . The split-half reliability was  $r(426) = .76$ , with each half of the test also showing adequate alpha reliability ( $\alpha = .80$  and  $\alpha = .82$ , respectively). Also, a subsample of 283 students was administered the same form again at the beginning of the next school year. Test-retest reliability was  $r(283) = .87$ . Reliabilities for items related to reader mental models (16 items;  $\alpha = .78$ ), digital literacy (13 items;  $\alpha = .78$ ), and other/vocabulary (8 items;  $\alpha = .64$ ) were within acceptable range.

### Validity

The validity argument for GISA integrates *evidence based on internal structure*, namely, the extent to which the relations among the test components conform to the hypothesized construct, and *evidence based on relations with other variables*, specifically the extent to which test scores and other measures intended to assess similar constructs provide convergent evidence, and the extent to which there was test-criterion predictive evidence.

#### *Evidence Based on Internal Structure*

Sabatini, O'Reilly, Weeks, and Steinberg (2016) theorized that the underlying literacy construct assessed by GISA is unidimensional. They evaluated unidimensionality in a large-scale study. To enable the creation of a vertical scale, a nonequivalent groups common item design (Kolen & Brennan, 2004) was used, which included at least two parallel forms in each grade (to be used as alternate forms in subsequent test administrations) and a linking form. Unidimensionality was evaluated in two ways: (1) factor analysis, and (2) item response theory (IRT) analysis. To evaluate unidimensionality using factor analysis, Sabatini et al. fit and compared three theoretically driven models: a unidimensional model, a two-factor exploratory model, and a two-factor simple-structure model where items associated with science passages loaded on one factor and items associated with history/language arts passages loaded on the second factor. Results from this comparison were mixed. The analysis showed that the unidimensional model fit better than either of the multidimensional models but only when the Bayesian information criterion (BIC) was used. Differences between the three models, however, were small overall. Although the indices provide mixed information, the penalty term is greater in the BIC compared to the Akaike information criterion (AIC). Due to the penalty difference, the BIC is a more conservative estimate and was deemed more appropriate for model selection. Subsequently, the unidimensional model was retained. Thus, the construct measured by the GISA across grades appears to be unidimensional.

On this basis, a unidimensional vertical scale was created using IRT analysis. To evaluate unidimensionality using IRT analysis, the item response curve for the two-parameter logistic (2PL) model (Birnbaum, 1968) was used to create a common scale.

The end result was a set of unidimensional vertical scales spanning grades 3–12. The item parameters for each scale were estimated using marginal maximum likelihood via a multigroup extension of the 2PL model (Bock & Zimowski, 1997), whereas the ability parameters were estimated using expected a posteriori. The item and ability parameters were estimated using the software program MDLTM (von Davier & Xu, 2011). As a final step, scores were rescaled to have a mean of 1,000 and a standard deviation of 100. Sabatini, O'Reilly, et al. (2016) examined all of the grade-level means and standard deviations and noted that they reflected developmental differences in ability across grades. In general, the means increased across all grades. The one exception was in grade 12, where the mean was slightly lower than the grade 11 mean.

#### *Evidence Based on Relations to Other Variables*

In the year 7 annual report, Sabatini (2017) reported preliminary findings from an integrated study design on the convergent evidence among RISE, GISA, CBAL, and the Gates-MacGinitie Reading Test (GMRT). The design was very complex and involved approximately 8,000 students. Preliminary results indicated that the correlation between the GISA and the GMRT was in the expected range. Specifically, the average correlation across different GISA forms and GMRT was  $r = .69$  with the correlations ranging from  $.54$  to  $.75$ . In the same report, Sabatini et al. also reported preliminary findings from an integrated study (*aka* the Mississippi Study) on the relations of GISA, FRA Reading Comprehension, and GMRT in elementary (grades K–2), middle (grades 3–5), and high school (grades 6–10) students. The correlations between GISA and FRA Reading Comprehension were in the moderate to high range ( $.483$  to  $.777$ ), whereas the correlations between GISA and GMRT were in the low to high range ( $.399$  to  $.770$ ). These moderately high correlations are very encouraging because they indicate that these reading comprehension tests are measuring a similar (but not identical) construct: reading.

Sabatini, O'Reilly, Halderman, and Bruce (2014a) provided preliminary evidence for predictive evidence in a small-scale study. In this study,  $n = 237$  students in grade 6 were given the RISE battery which measured core reading skills such as word recognition, decoding, vocabulary, and morphology, as well as a pilot GISA form test (i.e., Organic Farming). The pattern of correlations among measures showed relatively strong relations (range  $r = .704$  to  $.773$ ), suggesting that all the component skills measured by the RISE are related to comprehension on GISA. As expected, the highest correlation was between the RISE Reading Comprehension subtest and GISA ( $r = .773$ ), with RISE Reading Efficiency a close second ( $r = .762$ ). A hierarchical, multiple regression analysis predicting GISA total scores from RISE subtest scores showed that each subtest added significant unique variance with an adjusted total of 69 percent of the variance in GISA accounted for by all the RISE subtests. Overall, the RISE and GISA robust correlations suggested that both batteries measure some overlapping aspects of reading comprehension across the ability range. However, there was also evidence that adequate lower-level skills may be necessary, but not sufficient prerequisites, to higher levels of reading performance as indicated by the regression analysis, providing further evidence that GISA measures complex comprehension.

## Fairness

Sabatini, O'Reilly, et al. (2016) followed ETS Standards for fairness. In this context, every item was independently reviewed by ETS staff specifically trained in ensuring the fairness of test items. Evidence for fairness was also based on *lack of measurement bias*. Sabatini et al. (2016) examined effects of potential differential item functioning (DIF) as a function of gender across grades and forms. The DIF procedure determines whether any differential item performance exists between two groups matched for ability above and beyond expectations. The criteria for assessing the presence of DIF are based on Dorans and Kulick (2006) and have three levels based on values of the Mantel-Haenszel chi square statistic: A (negligible), B (moderate), and C (significant). The analysis showed very little presence of significant DIF (7 out of 708 items).

## Proposed Intended Use of Scores

The review of GISA demonstrated evidence of careful test construction consistent with current conceptual frameworks of reading comprehension processes, appropriate administration and scoring, adequate score reliability, adequate evidence for validity based on test content, internal structure, and on relations with other variables, and attention to fairness with an emphasis on minimizing measurement bias.

GISA forms have some features that are different from existing standardized assessments. Among the most striking differences in design are the following. First, all assessments are contextualized within a scenario that provides a purpose for integrating multiple sources. Second, all assessments are delivered on computer, which allows for the assessment of "digital literacy" (Coiro, 2009). Third, the assessment uses simulated peers that provide instruction and guidance in "collaborating" with the test taker, making it a more authentic reading situation. Fourth, the assessments taps higher-level skills such as integration, evaluation, and application.

### *GISA Domain-Specific Assessments for Intervention Studies (O'Reilly et al., 2014)*

With respect to its intended purposes, O'Reilly et al. (2014) demonstrated GISA's use as a summative assessment designed to provide evidence for the efficacy of reading interventions. While GISA forms have been developed and evaluated for different grade bands, topics, and skill foci, the GISA forms reported in this study were specifically designed to evaluate the outcomes of a specific intervention in mind: Reading Apprenticeship. The Reading Apprenticeship intervention views reading as an inquiry-based, problem-solving activity that builds knowledge about text content. For instance, reading in history involves evaluating facts and interpretations, the quality of sources (e.g., primary versus secondary), the corroboration of evidence, and an evaluation of the context in which information was collected (Monte-Sano, 2010). Reading in science involves using representations, models, and principles to reason and express key relationships among variables (Goldman, 2012). Reading in literature involves understanding human experience (Lee & Spratley, 2010). The underlying approach to assessment design in GISA (see O'Reilly & Sabatini, 2013; Sabatini & O'Reilly, 2013; Sabatini, O'Reilly, & Deane, 2013) had several elements in common with the Reading Apprenticeship program. Despite the similarities, the biggest difference between the

Reading Apprenticeship intervention and the GISA designs was the strong focus on content and disciplinary reading in high school history, science, and literature. Thus, three GISA summative forms were developed, intended to measure students' abilities to read and understand in each domain. Texts and tasks were sourced from topics in each domain. Each form also included an integrated background knowledge assessment following Deane (2012), in which students are asked to decide whether a term is *related* or *unrelated* to the topic of the text. Students can choose "Yes," "No," or "I don't know."

O'Reilly et al. (2014) analyzed data from a sample of 12,715 high school students in grades 9–12 from 43 schools in California and Pennsylvania. The three domain-specific forms exhibited good reliability (Cronbach's alpha range .84 to .88), adequate score variation, and positive correlations with measures of background knowledge. Furthermore, the results of a bifactor model suggested that there was a general reading comprehension factor underlying the domain-specific tests. Scores on the specific factors under the bifactor model were correlated at around .70 with the scores on the simple structure model. Thus, from a pure measurement standpoint, O'Reilly et al. (2014) concluded that the GISA assessments were adequate for use as outcomes in equations comparing treatment versus control students in the intervention.

Goldman et al. (2019) also used GISA as a distal measure in a randomized controlled trial evaluating the efficacy of the READI Science intervention to improve reading comprehension in grade 9 science when compared to a business-as-usual control. The results showed that GISA was sensitive to the intervention effects.

## READING INVENTORY AND SCHOLASTIC EVALUATION (RISE)

### Conceptual Framework

RISE is designed to measure foundational *components of reading*, one dimension of the broader construct of *reading literacy* (O'Reilly et al., 2012; Sabatini & Bruce, 2009; Sabatini, Bruce, & Steinberg, 2013; Sabatini, Weeks, et al., 2019). Reading component skills are subskills of reading that can be isolated and assessed independently from higher-level reading comprehension (Perfetti & Adlof, 2012; Sabatini, Bruce, & Steinberg, 2013). RISE builds heavily on component models of reading comprehension that attempt to identify core linguistic and cognitive skills to explain reading comprehension performance. In this context, the assessment team identified three principles to guide the assessment design (principles 1, 2, and 3; Sabatini, O'Reilly, & Deane, 2013). The first principle states that print skills and language comprehension are each considered necessary components of reading proficiency, though neither individually is sufficient to ensure proficiency (Adlof, Catts, & Little, 2006; Vellutino, Tunmer, Jaccard, & Chen, 2007). The second principle states that both breadth and depth of vocabulary knowledge are essential for understanding (Nagy & Scott, 2000; Ouellette, 2006). The third principle states that readers construct mental models of text meaning at multiple levels, from literal to gist to complex situation models (Kintsch, 1998; McNamara & Kintsch, 1996).

Consistent with these principles, components of the RISE inventory include foundational skills such as word recognition and decoding (Ehri, 2014), reading fluency (Fuchs et al., 2001), vocabulary knowledge (Beck & McKeown, 1991; Quinn et al., 2015), morphology (Carlisle, 2000; Hogan, Bridges, Justice, & Cain, 2011), syntax (Perfetti &

Adlof, 2012), and lower-level reading comprehension—at the sentence and single text level (Kintsch, 1998; McNamara & Kintsch, 1996). The range of these foundational skills—beginning with the recognition or decoding of words, to understanding the meanings of words and sentences, to building meaning from a passage—is consistent with the developmental progression of reading comprehension (RRSG, 2002).

This assessment system was designed for a finite set of purposes, including screening (i.e., identify students at risk of meeting grade-level expectations), diagnosis (i.e., identify students' strengths and weaknesses), formative assessment (i.e., provide actionable information for teachers), and summative assessment (i.e., provide accountability or outcome information) (Sabatini et al., 2015).

### Description

RISE (Sabatini et al., 2015) is a 45- to 60-minute web-administered assessment of foundational reading skills in grades 3–12. The RISE is part of a larger reading assessment system called Study Aid and Reading Assistant (SARA). It contains six subtests, each of which targets a specific component of reading that may be affecting a student's progress toward higher levels of reading comprehension proficiency. Reading components are defined here as foundational subskills related to reading comprehension performance. Specifically, the RISE subtests target (a) decoding and recognizing words in isolation; (b) recognizing meaning or semantic relationships of individual words; (c) using knowledge of word parts to identify which word fits the meaning and syntax of a sentence; (d) building meaning from sentences by understanding causal connectors, pronouns, and relationships among terms; (e) reading for basic understanding with fluency; and (f) comprehending the basic meaning of passages. The initial scaling of RISE (Sabatini et al., 2015) had four forms in grade 5, six forms in grades 6–9, and three forms in grade 10. In the final scaling (Sabatini, Weeks, et al., 2019) these forms were reused with a national sample of students in grades 3–12; an additional form for grade 3 students was also developed. Thus, the final scaling includes one form in grade 3, four forms in grades 3–5, six forms in grades 6–9, and three forms in grades 9–12. Each subtest is described in more detail next.

The RISE Word Recognition and Decoding subtest uses three item types to measure a student's ability both to recognize sight words and to decode nonwords:

1. Real words, including content-area words that middle school students will encounter in their school curricula;
2. Nonwords, including a range of spelling and morphological patterns; and
3. Pseudohomophones, including nonwords that sound exactly like real English words.

Students are presented with one item on the screen at a time and are asked to decide if what they see (a) is a real word, (b) is not a real word, or (c) sounds exactly like a real word.

The RISE Vocabulary subtest includes both tier 2 and tier 3 words. Tier 2 words are general academic words, whereas tier 3 words are domain-specific, less frequently used words (Beck, McKeown, & Kucan, 2002, 2008; Coleman & Pimentel, 2011).



Students are presented with a target word on the screen and are asked to select either a synonym or a meaning associate of the target from three choices:

- An example of a synonym item is *data* (information, schedule, star).
- An example of a meaning associate item is *thermal* (heat, bridge, evil).

The RISE Morphology subtest focuses on derivational morphology—those words that have prefixes and/or suffixes attached to a root. The test uses a cloze (fill-in-the-blank) item type. Each item is a sentence. The sentences are designed with straightforward syntactic structures and relatively easy vocabulary so that students focus on the derived words:

- That man treats everyone with respect and \_\_\_\_\_. (civility, civilization, civilian)

The RISE Sentence Processing subtest focuses on single-sentence semantic and syntactic processing. The focus is on the student's ability to construct basic meaning from print at the sentence level. The cloze (fill-in-the-blank) items in the subtest require the student to process all parts of the sentence to select the correct answer among three choices:

- The dog that chased the cat around the yard spent all night \_\_\_\_\_. (barking, meowing, writing)

The RISE Efficiency of Basic Reading Comprehension subtest uses the maze selection technique (Fuchs & Fuchs, 1992; Shin, Deno, & Espin, 2000); that is, in each sentence within a passage, one of the words is replaced with three choices, only one of which makes sense in the sentence. Accurately selecting the correct response for each item does require that the reader is comprehending each sentence and likely building a cross-sentence general model of passage gist. Because the task is timed, the simultaneous demand that students read quickly also captures an indicator of silent reading fluency or efficiency. The subtest comprises informational texts. Students have 3 minutes to complete each passage.

- Passage excerpt: During the Neolithic Age, humans developed agriculture—what we think of as farming. Agriculture meant that people stayed in one place to grow their crops/baskets/rings. They stopped moving from place to place to follow herds of animals or to find new wild plants to eat/win/cry. And because they were settling down, people built permanent shelters/planets/secrets.

The RISE Reading Comprehension subtest assesses discourse-level comprehension. Students read a text and answer related question items. The items show a range of difficulties (from a verbatim understanding of the words and phrases to the “gist” understanding of what is being read and low-level inference making):

- Question (Locate/Paraphrase): What did people use to heat water in Neolithic houses? (*hot rocks, burning sticks, the sun, mud*)

- Question (Low-Level Inference): In the sentence “They gave people more protection from the weather and from wild animals.” the word “they” refers to: (*permanent shelters, caves, herds, agriculture*)

### *RISE Administration and Scoring*

RISE includes a computerized administration, automated scoring, and reporting support that is scalable at the classroom, school, or district level. Scores for each subtest provide evidence of instructionally malleable targets of readers’ strengths and weaknesses. The item parameters for each scale are estimated using marginal maximum likelihood via a multigroup extension of the 2PL model (Bock & Zimowski, 1997). The scores for each scale are then rescaled to have a mean of 250 and a standard deviation of 15.

### **Sample**

The technical quality of the RISE system was initially evaluated based on the findings reported by Sabatini et al. (2015) and subsequently Sabatini, Weeks, et al. (2019). Data were collected in a large, urban school district in the mid-Atlantic region of the United States. A total of  $n = 17,383$  students in grades 5–10 participated. Specifically, there were  $n = 2,947$  in grade 5,  $n = 3,540$  in grade 6,  $n = 3,477$  in grade 7,  $n = 3,114$  in grade 8,  $n = 2,885$  in grade 9, and  $n = 1,420$  in grade 10. In terms of ethnicity, 3.6 percent were Hispanic/Latino. In terms of race, 0.3 percent were American Indian/Native Alaskan, 1.1 percent Asian, 87.7 percent Black, 0.2 percent Native Hawaiian/Pacific Islander, 10.7 percent White, and 0.2 percent other/not reported. Also, 51.4 percent were female, 48.5 percent male, and 1 percent not reported. No exclusions were mandated. In fact, 15.5 percent of the sample was receiving special education services and 1.3 percent English language learner services. Tests were administered in school computer labs and proctored by school staff members who were trained to the protocol. In their year 7 annual report, Sabatini (2017) noted that they conducted a large-scale field study in which they recruited students from all four regions of the United States: Midwest, Northeast, South, and West. Sample size increased to  $n = 51,391$  and grade levels expanded from 4 to 12. Sabatini et al. stated that no meaningful differences were observed compared to those reported previously (Sabatini et al., 2015) and they planned on updating information about the sample as well as the assessment psychometric properties. Indeed, the most recent technical report (Sabatini, Weeks, et al., 2019) includes the updated information. Note that the results of the analyses reported next are based on both the 2015 and 2019 RISE technical reports.

### **Validity**

The validity argument for RISE integrates *evidence based on test content*, namely, the relations between the content of the test and the construct it is intended to measure; *evidence based on internal structure*, namely, the extent to which the relations among the test components conform to the hypothesized construct; and *evidence based on relations with other variables*, and specifically the extent to which test scores and other measures intended to assess similar constructs provide convergent evidence.

### *Evidence Based on Test Content*

Sabatini et al. (2015) theorized that each subtest construct represents a somewhat distinct component or subskill. Drawing on the extant reading literature (RRSG, 2002), it would be predicted that the various subtests would be moderately to strongly related (Mislevy & Sabatini, 2012). Indeed, the analysis showed moderate to strong correlations (Pearson's  $r$ ) between the subtests within each grade level (grade 5 range .450 to .679; grade 6 range .504 to .718; grade 7 range .535 to .699; grade 8 range .522 to .699; grade 9 range .497 to .667; and grade 10 range .570 to .711).

### *Evidence Based on Internal Structure*

Sabatini et al. (2015) used IRT (Lord & Novick, 1968), specifically the item response curve for the 2PL model (Birnbaum, 1968), to create a common scale for each subtest. The result was a set of six unidimensional vertical scales spanning grades 5–10. The item parameters for each scale were estimated using marginal maximum likelihood via a multigroup extension of the 2PL model (Bock & Zimowski, 1997), where the item parameters for the common items were constrained to be equal across groups. As a final step, the scores for all six scales were rescaled to have a mean of 250 and a standard deviation of 15. This analysis provided evidence for the hypothesized unidimensionality of each subscale/construct.

Sabatini, Weeks, et al. (2019) evaluated further the separation between the components across grades 3–12; three factor structures were considered. The first was a unidimensional structure where all the items loaded on a single factor. The second was a six-factor simple structure where the items associated with each component skill loaded only on the respective factor. The third was a two-factor simple view structure where the word reading, vocabulary, and morphology items loaded on one factor (*decoding*) and sentence comprehension, efficiency, and reading comprehension items loaded on the other factor (*comprehension*). The results suggested that both the two-factor and six-factor multidimensional structures had good fit to the data.

### *Evidence Based on Relations to Other Variables*

Sabatini (2017) reported preliminary findings from a large-scale integrated study design on the relations among RISE, GISA, CBAL, and GMRT. Preliminary results indicated that the correlations between the GMRT and the RISE subtests were in the expected range. For instance, the correlation between the Gates-MacGinitie Vocabulary and the RISE Vocabulary subtest was  $r(626) = .70, p < .01$ . Similarly, the correlation between the Gates-MacGinitie Reading Comprehension test and the RISE Reading Comprehension subtest was  $r(706) = .61, p < .01$ . These correlations suggest that the assessments measure related, but not identical, constructs.

Sabatini, Weeks et al. (2019) also reported that the RISE vocabulary and morphology tests were correlated with the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 2012)  $r = .36$  to  $.56$ , the Peabody Picture Vocabulary Test (PPVT)  $r = .52$  to  $.57$ , and the Clinical Evaluation of Language Fundamentals (CELF) language measures  $r = .38$  to  $.51$ . Also, RISE Reading Comprehension and GMRT correlation is  $.77$ , whereas RISE Reading Comprehension and GISA correlation is  $.65$ .

### Reliability/Precision

Reliability/precision evidence was based on *internal-consistency coefficients*. Sabatini et al. (2015) reported Cronbach's alpha coefficients (Cronbach, 1951) for each subtest within each administration, form, and grade. The reliabilities represented as median values within a grade across forms were generally within acceptable range, specifically, for word recognition and decoding (range .899 to .921), for vocabulary (range .830 to .900), for morphology (range .871 to .920), for sentence comprehension (range .826 to .873), for reading efficiency (range .922 to .948), and for reading comprehension (range .604 to .833).

Sabatini et al. (2015) also evaluated subtest scores for consistency (versus the use of a total score) following the approach advocated by Haberman (2005) and Sinharay, Haberman, and Puhan (2007). The input information included Cronbach's alpha reliability values, average raw scores and standard deviations for each subtest, and the correlation between the subtest score and the total score. For purposes of this analysis, the total score was computed as the sum of the six subtest raw scores, and the total reliability coefficient was computed based on all item-level data across subtests merged together by unique student identifier. The analysis provided some evidence for subscore utility. Specifically, across 19 comparisons, 15 (79 percent) met the criteria for subscore utility. The four comparisons that did not meet the criteria involved grades 5 or 6 and three the reading comprehension subtest.

### Fairness

Sabatini et al. (2015) followed ETS Standards for fairness. In this context, every item was independently reviewed by ETS staff specifically trained in ensuring the fairness of test items. Evidence for fairness was also based on *lack of measurement bias*. Specifically, Sabatini et al. (2015) examined effects of potential DIF by comparing item-level data for gender and race across grades and forms. The DIF procedure determines whether any differential item performance exists between two groups matched for ability above and beyond expectations. The criteria for assessing the presence of DIF were based on Dorans and Kulick (2006) and had three levels based on values of the Mantel-Haenszel chi square statistic: A (negligible), B (moderate), and C (significant). The analysis showed very little presence of significant DIF, suggesting no differential item performance as a function of gender or race. The updated analysis reported by Sabatini et al. (2019) using the national sample also showed very little presence of significant DIF.

### Proposed Intended Use of Scores

The review of RISE demonstrated evidence of careful test construction consistent with current conceptual frameworks of reading comprehension components; appropriate administration and scoring; adequate score reliability; adequate evidence for validity based on test content, on internal structure, and on relations with other variables; and attention to fairness with an emphasis on minimizing measurement bias. With respect to its intended purposes, Sabatini, Weeks, et al. (2019) suggested that RISE could be used for diagnosis (i.e., identify students' strengths and weaknesses),

formative assessment (i.e., provide actionable information for teachers), and summative assessment (i.e., provide accountability or outcome information). For example, with respect to diagnosis, RISE can detect whether foundational reading skills are barriers to achieving higher levels of reading comprehension performance. If foundational skills are lacking, then teachers should take this information into account when designing instruction to address student needs. Wang et al. (2019) identified a decoding threshold in RISE word recognition and demonstrated that students who initially fell below the threshold in the early grades showed little to no growth in reading comprehension over time.

With respect to evaluating instructional outcomes, RISE has been used in several large-scale intervention studies, demonstrating its instructional sensitivity. Specifically, Kim et al. (2017) used RISE to evaluate the efficacy of the Strategic Adolescent Reading Intervention (STARI) with low-achieving middle school students. The results showed that students who participated in STARI scored higher than control students on RISE efficiency of basic reading comprehension (Cohen's  $d = 0.21$ ). In other words, the RISE was sensitive to the effects of the reading intervention.

## FLORIDA CENTER FOR READING RESEARCH READING ASSESSMENT (FRA)

### Conceptual Framework

FRA draws on decades of research about what predicts reading comprehension success in the English language system (NELP, 2008; NICHD, 2000; NRC, 1998; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; RRSG, 2002). Specifically, in an alphabetic orthography such as English, *mastering the alphabetic principle*, namely, acquiring basic decoding skills, is a necessary skill that needs to be explicitly and systematically taught (Ehri, Nunes, Willows, et al., 2001). However, mastering the alphabetic principle is not a sufficient condition for understanding written text. Understanding written text (i.e., reading comprehension) also requires knowledge of *word meaning or lexical quality* (Perfetti & Stafura, 2014), namely, knowledge of pronunciation, spelling, multiple meanings in a variety of contexts, synonyms, and morphological structure. Understanding written text (i.e., reading comprehension) also requires *syntactic awareness*, namely, understanding of the rules that govern how words are ordered to make meaningful sentences.

The emphasis on achieving the alphabetic principle, lexical quality, and syntactic awareness ensures adequate reading comprehension. However, individual differences in readers' background knowledge, motivation, memory, and attention will also create variability in reading comprehension. Furthermore, because reading comprehension is affected by the interactions of variables related to reader and text characteristics (RRSG, 2002), text genre is also expected to influence performance.

In FRA K–2 the *alphabetic principle* is assessed with tasks that measure letter-sound knowledge, phonological awareness, ability to link sounds to letters, word reading, word building, and spelling. *Knowledge of word meanings* or lexical quality is measured by a word matching task. *Syntactic awareness* is assessed using a following directions

task and a sentence comprehension task. *Reading comprehension* is assessed with a listening or reading passage comprehension task that includes both literary and informational passages (Fitzgerald et al., 2014; Foorman et al., 2004).

In FRA Grades 3–12 the *alphabetic principle* is assessed with a word recognition task. *Knowledge of word meanings* or lexical quality is measured by a vocabulary knowledge task that taps morphological awareness and includes words that signal inferential or decontextualized language. *Syntactic awareness* is assessed with syntactic knowledge tasks that taps the meaning and use of connectives (Cain & Nash, 2011; Crosson & Lesaux, 2013). *Reading comprehension* is assessed efficiently with a computer-adaptive reading passage comprehension task that includes both literary and informational passages (Fitzgerald et al., 2014; Foorman et al., 2004).

The FRA system consists of a K–2 system and a Grades 3–12 system administered at three periods (fall, winter, and spring). Each system consists of a series of tasks for which students receive five items at grade level and then additional tasks that the system adapts up or down in grade level based on performance to reach a precise estimate of a student’s ability. The K–2 system consists of screening, comprehension, and diagnostic tasks that the teacher administers to students individually. The Grades 3–12 system consists of screening and comprehension tests that students complete online.

FRA is a computer-adaptive assessment system; namely, the selection, order, and number of items administered depend on a student’s ability at the time of the assessment. Students receive harder or easier items based on their performance, and the system stops administering items once it has enough information about the student’s ability. Thus, adaptive assessments maximize precision of information while minimizing time spent gaining it (Mitchell et al., 2015).

## FRA K–2 SYSTEM

### Description

The FRA K–2 system (Foorman et al., 2015a) is a 45-minute web-administered assessment of foundational reading skills. In the K–2 system the teacher scores the responses as correct or incorrect. The system is computer adaptive; namely, the selection, order, and number of items administered depend on a student’s ability. FRA consists of six computer-adaptive tests, which evaluate students’ phonological awareness, letter sounds, word reading, spelling, vocabulary, and following directions. These tasks collectively function as screening and produce the Probability of Literacy Success (PLS) score following a weighted formula. Students whose PLS score predicts that they are at risk of meeting grade-level expectations go on to take *Diagnostic* tasks. These computer-administered tasks are criterion referenced to developmental expectations for beginning readers and are scored for mastery (i.e., 80 percent correct). FRA also assesses comprehension using a listening/reading comprehension task and a sentence comprehension task.

### Screening

There are six screening tasks. The *Phonological Awareness* task (K only) requires students to listen to a word that has been broken into parts and then blend them together to reproduce the full word. The *Letter Sounds* task (K only) presents students with a letter and asks them to provide the sound that the letter represents. The *Word Reading* task (grades 1 and 2) displays a word on a screen and students respond by reading the word out loud. The *Spelling Task* (grade 2) aurally presents a word and uses it in a sentence. Students respond by typing the word. The *Vocabulary Pairs* task (K–2) presents and pronounces three words. The student then selects the two words that go together best (e.g., dark, night, swim). The *Following Directions* task (K–2) requires students to listen and attend as they hear directions. Students respond to the directions by clicking on or moving the specified objects on the computer monitor (e.g., put the square in front of the chair and then put the circle behind the chair).

There are two comprehension tasks. The *Listening and Reading Comprehension* task (K–2) requires students to either listen to or read one passage and answer comprehension questions. Students are placed into listening or reading comprehension passages based on their performance on the *Screening* (and specifically the *Word Reading* task). Each passage has five multiple-choice questions. For each passage, the number of questions answered correctly, the number of words read correctly, and the words read correctly per minute are used in conjunction with the student's classroom performance to descriptively inform classroom instruction. The *Sentence Comprehension* task (K–2) requires students to select the one picture out of the four presented that depicts the sentence given by the computer (e.g., click on the picture of the bird flying toward the nest).

### Administration and Scoring

In K–2, each task has four stop rules that determine when administration of each task is complete. Specifically: (a) a reliable estimate of the student's abilities is reached (i.e., standard error is less than 0.316); (b) the student has responded to 30 items (29 items in Letter Sounds); (c) the student responds correctly to all of the first eight items; and (d) the student responds incorrectly to all of the first eight items.

FRA produces three different scores. An *ability score* and a *percentile rank score* are provided for each computer adaptive task (Letter Sounds, Phonological Awareness, Word Reading, Vocabulary Pairs, Following Directions, Spelling, and Sentence Comprehension in K) at each time point. A PLS score is provided at each assessment period, which is an aggregate of the individual student's scores. In K the aggregate is based on Letter Sounds, Phonological Awareness, Vocabulary Pairs, and Following Directions. In grade 1 the aggregate is based on Word Reading, Vocabulary Pairs, and Following Directions. In grade 2, the aggregate is based on Word Reading, Vocabulary Pairs, Spelling, and Following Directions.

The PLS score indicates the likelihood that a student will reach end-of-year expectations in literacy. For the purposes of FRA, reaching expectations is defined as performing at or above the 40th percentile on the Stanford Achievement Test, Tenth Edition (SAT-10). The PLS is also color coded: red indicates the student is at high risk and needs targeted intervention, yellow indicates the student may be at risk and needs supplemental instruction, and green indicates the student is likely not at risk.

*Ability scores* provide an estimate of a student's development in a particular skill. The range is approximately 200 to 1,000, with a mean of 500 and standard deviation of 100. This score has an equal interval scale and is used to determine the degree of growth in a skill for individual students.

*Percentile ranks* vary from 1 to 99. The median percentile rank on FRA is 50. The percentile rank is an ordinal variable and is used to compare a student's performance to other students within a grade level.

### Sample

Evidence reported next (unless otherwise noted) is based on a large-scale field study that recruited students in Florida. A total of 27,862 students in kindergarten through grade 2 across multiple districts in Florida participated in the calibration and validation studies (Foorman et al., 2015a). These studies involved students being administered subsets of items from each task depending on their grade level. Demographic information for the sample approximated that of the state of Florida: 40 percent White, 31 percent Hispanic, 23 percent Black, and 6 percent other; 65 percent eligible for free or reduced-price lunch; and 18 percent limited English proficient.

### Validity

The validity argument for FRA K–2 integrates *evidence based on test content*, namely, the relations between the content of the test and the construct is intended to measure; *evidence based on internal structure*, namely, the extent to which the relations among the test components conform to the hypothesized construct; and *evidence based on relations with other variables*, and specifically the extent to which test scores provide convergent evidence and predict criterion performance.

#### *Evidence Based on Test Content*

The expectation was that oral language and reading measures would be moderately correlated with higher intercorrelations within each cluster. Indeed, FRA scores in K–2 were moderately interrelated ( $r = .20$  to  $.78$ ) with the highest correlations observed within oral language measures (e.g., Following Directions and Sentence Comprehension in K,  $r = .61$ ) and reading measures (e.g., Spelling and Word Reading in grade 2,  $r = .78$ ).

#### *Evidence Based on Relations to Other Variables*

Convergent evidence was provided by correlating performance on the FRA screening tasks with well-known clinical measures. Specifically, the FRA Phonological Awareness task scores in a low-performing sample of 100 English learners correlated  $r = .36$  with the Letter-Word Identification task of the Woodcock-Johnson III Test of Achievement (Woodcock, McGrew, & Mather, 2001). FRA Letter Sounds correlated  $r = .52$  with the Phonemic Awareness task of the Woodcock-Johnson III Test of Achievement (Woodcock et al., 2001). FRA Sentence Comprehension scores correlated  $r = .48$  in K,  $r = .44$  in grade 1, and  $r = .40$  in grade 2 with the Sentence Structure subtest from the



CELF-4 (Semel, Wigg, & Secord, 2003). FRA Vocabulary Pairs scores correlated  $r = .46$  in K,  $r = .59$  in grade 1, and  $r = .50$  in grade 2 with the PPVT-4 (Dunn & Dunn, 2007). FRA Following Directions scores correlated  $r = .58$  in K,  $r = .58$  in grade 1, and  $r = .64$  in grade 2 with the CELF-4 Concepts and Following Directions (Semel et al., 2003).

Test-criterion predictive evidence was obtained in two ways. First, multiple regression analysis was used to estimate the total amount of variance that the linear combination of the FRA predictors explained in SAT-10 Word Reading in K and SAT Reading Comprehension in grades 1 and 2 (Foorman et al., 2015a). The analysis showed that FRA predicted a significant amount of variance at each grade (.46, .43, and .51, respectively). Sabatini (2017) also reported preliminary findings from an integrated study (*aka* the Mississippi Study) on the relations of FRA (K–2) with GISA and GMRT. The correlations between FRA (K–2), GISA, and GMRT were low to moderate (range .291 to .640). A series of regression analyses showed that FRA (K–2) accounted for 24.9 percent of variance in GISA and 54.8 percent of variance in GMRT.

Second, logistic regression analysis was used to estimate the predictive power of the PLS cutoff score. Recall that the PLS score is used to estimate the probability that a student is at risk of meeting grade-level expectations. This analysis focused on negative predictive power (Schatschneider, Petscher, & Williams, 2008), namely, the percentage of students who are identified as “not at risk” on the FRA screening but performing below benchmark on the outcome tests (< 40th percentile on SAT Word Reading and Reading Comprehension). The analysis evaluated PLS cutoff scores of .85 and .70, following previous work (Petscher & Foorman, 2011), and showed that a PLS score of .70 not only reduces false positives (range from .83 to .94), but also increases positive predictive power (range from .52 to .82) and the overall correct classification (range from .66 to .82).

### Reliability/Precision

Across all grades and assessment periods, Foorman et al. (2015a) reported *marginal reliability coefficients* for the computer-adaptive tasks ranging from .85 to .96. *Test-retest reliability* was evaluated at three testing points: fall, winter, and spring. Across tasks and grade levels, correlations ranged between .42 to .80 in fall-winter, .44 to .72 in winter-spring, and .23 to .65 in fall-spring. The lowest correlations were consistently for the Vocabulary Pairs task.

### Fairness

Evidence for fairness was based on *lack of measurement bias*. Specifically, the PLS cutoff score was evaluated for differential accuracy across different demographic groups. This procedure involved a series of logistic regressions predicting success on the SAT-10 tests (i.e., at or above the 50th percentile). The independent variables included a variable that represented whether students were identified as not at risk (PLS  $\geq .70$ ; coded as “1”) or at risk (PLS  $< .70$ ; coded as “0”), a variable that represented a selected demographic group, as well as an interaction term between the two variables. A statistically significant interaction term would suggest differential accuracy. For the combination of FRA screening task scores, differential accuracy was separately tested

for Black and Latino students as well as for students identified as English language learners and students who were eligible for free or reduced-price lunch. These analyses showed no significant interactions and thus no differential effects.

### **Proposed Intended Use of Scores**

The review of FRA K–2 demonstrated evidence of careful test construction consistent with current conceptual frameworks of reading comprehension, appropriate administration and scoring, adequate score reliability, adequate evidence for validity based on test content, internal structure, and on relations with other variables, and attention to fairness with an emphasis on minimizing measurement bias. With respect to intended use, Foorman and colleagues provided evidence for score appropriateness in evaluating the efficacy of interventions and identifying profiles of readers with instructional utility.

#### *Evaluating Intervention Effects (Foorman, Herrera, et al., 2017)*

In this study, the utility of FRA K–2 as a pre- and postintervention measure was evaluated in a randomized controlled trial in 55 low-performing schools across Florida that compared two pull-out early literacy interventions—one using standalone materials and one using materials embedded in the existing core reading program. The interventions were delivered daily for 45 minutes for 27 weeks in small groups of students at risk of literacy failure in K–2 for 2 consecutive years. A three-level hierarchical linear model with students nested in small groups, nested in schools, was used to estimate treatment effects by grade. The findings showed that the standalone intervention significantly improved grade 2 spelling outcomes relative to the embedded intervention, but impacts on other student outcomes were similar for the two interventions. On average, students in schools that used the standalone intervention and students in schools that used the embedded intervention showed similar improvement in reading and language outcomes. The two interventions also had similar impacts on reading and language outcomes among English learner students.

#### *Identifying Latent Profiles with Instructional Utility (Foorman, Petscher, Stanley, & Truckenmiller, 2017)*

This investigation had several aims, one of which was to determine the latent profiles of reading and language skills as measured by FRA and the extent to which these latent profiles were related to important reading outcomes, namely, SAT-10 Reading Comprehension (SESAT Word Reading for K). A total of 7,752 students in kindergarten through grade 10 across multiple districts in Florida participated in this study. Demographic information for the sample approximated that of the state of Florida: 42.18 percent White, 29.10 percent Hispanic, 22.5 percent Black, and 3.59 percent other; 60 percent eligible for free or reduced-price lunch; and 10.39 percent limited English proficient. There were 2,295 students in K–2. Latent profile analysis (LPA) identified five to six classes in the elementary grades. Profiles revealed high and low patterns in addition to interesting heterogeneous patterns (e.g., vocabulary deficit in K; vocabulary

and word reading deficit in grade 1; word reading and spelling deficit in grade 2). These profiles have implications for differentiating instruction.

## FRA GRADES 3–12 SYSTEM

### Description

The FRA Grades 3–12 system is a 45-minute web-administered assessment. It includes four computer-adaptive tests, which evaluate students' word recognition, vocabulary knowledge, syntactic knowledge, and reading comprehension. *Screening* measures are the Word Recognition and the Vocabulary Knowledge tasks. *Diagnostic* measure is the Syntactic Knowledge task. Students are placed on a comprehension passage in the Reading Comprehension task based on their scores on the Word Recognition and Vocabulary Knowledge tasks.

#### *Screening*

There are two screening measures. The Word Recognition task presents a word to the student aurally and the student selects the correctly spelled word from three options. The Vocabulary Knowledge task presents one sentence at a time with a word missing. The missing word is replaced with a choice of three morphologically related words. The student selects the word that best completes the sentence.

#### *Diagnostic*

The Syntactic Knowledge task presents to the student one sentence (or sentences) aurally. Each sentence is missing one word. The computer also displays the sentence(s) for the student to read along. The student selects the missing word from a dropdown menu of three choices.

#### *Reading Comprehension*

The Reading Comprehension task presents students with a sample of one to three passages that are between 200 and 1,300 words in length. Each passage has seven to nine multiple-choice questions. All questions associated with the passage are displayed at the same time and the passage is also available during question answering.

#### *Administration and Scoring*

In grades 3–12 each task (except for Reading Comprehension) has four stop rules that determine when administration of each task is complete; specifically, (a) a reliable estimate of the student's abilities is reached (i.e., standard error is less than .50), (b) the student has responded to 30 items, (c) the student responds correctly to all of the first eight items, and (d) the student responds incorrectly to all of the first eight items.

FRA produces three different scores. An *ability score* and a *percentile rank score* are provided for each computer adaptive task (Word Recognition, Vocabulary Knowledge, Syntactic Knowledge, and Reading Comprehension) at each time point. A *probability*

*of literacy success score* is provided at each assessment period, which is an aggregate of the individual student's scores. In grades 3–12 the aggregate is based on Word Recognition, Vocabulary Knowledge, and Reading Comprehension. The PLS score indicates the likelihood that a student will reach end-of-year expectations in literacy. For the purposes of FRA, reaching expectations is defined as performing at or above the 40th percentile on the SAT-10. The PLS is also color coded: red indicates the student is at high risk and needs targeted intervention, yellow indicates the student may be at risk and needs supplemental instruction, and green indicates the student is likely not at risk. The *ability score* provides an estimate of a student's development in a particular skill. The range is approximately 200 to 1,000, with a mean of 500 and standard deviation of 100. This score has an equal interval scale and is used to determine the degree of growth in a skill for individual students. Finally, *percentile ranks* vary from 1 to 99. The median percentile rank on FRA is 50. The percentile rank is an ordinal variable and is used to compare a student's performance to other students within a grade level.

### Sample

Evidence reported next (unless otherwise noted) is based on a large-scale field study that recruited students in Florida. A total of 44,780 students in grades 3–10<sup>2</sup> across multiple districts in Florida participated in the calibration and validation studies (Foorman et al., 2015b). These studies involved students being administered subsets of items from each task depending on their grade level. Demographic information for the sample approximated that of the state of Florida: 41 percent White, 30 percent Hispanic, 23 percent Black, and 6 percent other; 60 percent eligible for free or reduced-price lunch; and 8 percent limited English proficient.

### Validity

The validity argument for FRA Grades 3–12 integrates *evidence based on test content*, namely, the relations between the content of the test and the construct is intended to measure; *evidence based on internal structure*, namely, the extent to which the relations among the test components conform to the hypothesized construct; and *evidence based on relations to other variables*, and specifically the extent to which test scores provide convergent evidence and predict criterion performance.

#### *Evidence Based on Test Content*

FRA tasks were expected to be moderately correlated. Indeed, across grades FRA scores were moderately interrelated (range  $r = .29$  to  $.63$ ) with the highest correlations observed between reading comprehension and the three other measures (Vocabulary Knowledge, Word Recognition, and Syntactic Knowledge).

---

<sup>2</sup> The FRA team indicated that even though in their initial studies they also included grades 11 and 12, the sample is skewed toward lower-performing students. As a result they describe the sample as having a grade 3–10 proficiency range.

### *Evidence Based on Internal Structure*

A series of parametric factor analyses by grade within each task were conducted. The comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA) were used to evaluate model fit for the Vocabulary Knowledge, Word Recognition, and Syntax Knowledge tasks. CFI and TLI values of at least .90 are considered acceptable as are RMSEA values less than .10. With respect to the Vocabulary Knowledge, Word Recognition, and Syntax Knowledge tasks the results provided support for a unidimensional construct in each case. RMSEA values ranged between 0.000 and 0.028, CFI between 0.89 and 1.00, and TLI between 0.88 and 1.00 across grades. For the Reading Comprehension task, a unidimensional model was compared to a testlet model using the AIC and BIC indices. Results from this comparison were mixed. The AIC suggested that the testlet model should be used while the BIC and adjusted BIC values were smaller for the unidimensional model. Although the indices provided mixed information, the penalty term was greater in the BIC compared to the AIC. Due to the penalty difference, the BIC is a more conservative estimate and was deemed more appropriate for model selection. Subsequently, the unidimensional model was retained.

### *Evidence Based on Relations to Other Variables*

A study that involved  $n = 1,825$  students in grades 3–10 was used to provide convergent evidence. Students were administered the FRA tasks and well-known clinical measures. These measures included the TOWRE (Torgesen, Wagner, & Rashotte, 2012), the PPVT-4 (Dunn & Dunn, 2007), and the Grammaticality Judgment Test of the Comprehensive Assessment of Spoken Language (GJT; Carrow-Woolfolk, 2008). The analyses showed that the average correlation between the FRA Vocabulary Knowledge task and the PPVT-4 was  $r = .52$  (range of .47 to .67); that of the FRA Word Recognition task and the TOWRE Real Word test was  $r = .33$  (range of .24 to .49); that of the FRA Word Recognition task and the TOWRE Non-Word test was  $r = .38$  (range of .30 to .47); and that of the FRA Syntax Knowledge task and the GJT was  $r = .49$  (range of .37 to .61). Convergent evidence was also reported for students with low (< 40th quantile), average (40th to 60th quantile), and high (> 60th quantile) scores using quantile correlation analysis. The quantile correlations demonstrated a trend that higher correlations between the measures were observed for students who scored low or average on each measure.

Discriminant evidence was provided by estimating correlations between the FRA tasks and variables such as sex and birth date. The results showed overall weak relations across grades for both sex (range  $-.26$  to  $.22$ ) and birthdate (range  $.01$  to  $.28$ ).

Test-criterion predictive evidence was obtained in two ways (Foorman et al., 2015b). First, multiple regression analysis was used to estimate the total amount of variance that the linear combination of the FRA predictors explained in SAT-10 Reading Comprehension. The analysis showed that FRA predicted a significant amount of variance at each grade (range from  $.39$  to  $.62$ ). Sabatini (2017) reported preliminary findings from an integrated study (*aka* the Mississippi Study) on the relations of FRA (3–12) with GISA and GMRT. The correlations between FRA (3–12), GISA, and GMRT were moderate to high at middle (range  $.475$  to  $.716$ ) and high school (range  $.419$  to  $.777$ ) levels. A series of regression analyses showed that FRA accounted for 52.5 percent and 57.3 percent

of variance in GISA middle and high school levels, respectively. Also, FRA accounted for 62.9 percent and 60.2 percent of variance in GMRT middle and high school levels, respectively.

Second, logistic regression analysis was used to estimate the predictive power of the PLS cutoff score. Recall that the PLS score is used to estimate the probability that a student is at risk of meeting grade-level expectations. This analysis focused on negative predictive power (Schatschneider et al., 2008), namely, the percentage of students who are identified as “not at risk” on the screening assessment (FRA) but performing below benchmark on the outcome tests (< 40th percentile on SAT Reading Comprehension). The analysis evaluated cutoff scores of .85 and .70 following previous work (Petscher & Foorman, 2011) and showed that a PLS score of .70 not only reduces false positives (range from .84 to .91), but also increases positive predictive power (range from .45 to .68) and overall correct classification (range from .71 to .86).

### Reliability/Precision

Across all grades and assessment periods, Foorman et al. (2015b) reported average *marginal reliabilities* for the computer-adaptive tasks ranging from .86 to .93. *Test-retest reliability* was evaluated at three testing points: fall, winter, and spring. Across tasks and grade levels, correlations ranged between .46 to .85 in fall-winter, .51 to .80 in winter-spring, and .31 to .80 in fall-spring. The lowest correlations were consistently for fall-spring, which was expected as students’ performance differentially changes from the beginning to the end of the year.

### Fairness

Evidence for fairness was based on *lack of measurement bias*. Specifically, the PLS cutoff score was evaluated for differential accuracy across different demographic groups. This procedure involved a series of logistic regressions predicting success on the SAT-10 test (i.e., at or above the 50th percentile). The independent variables included a variable that represented whether students were identified as not at risk (PLS  $\geq$  .70; coded as “1”) or at risk (PLS < .70; coded as “0”), a variable that represented a selected demographic group, as well as an interaction term between the two variables. A statistically significant interaction term would suggest differential accuracy. For the combination of FRA screening task scores, differential accuracy was separately tested for Black and Latino students as well as for students identified as English language learners and students who were eligible for free or reduced-price lunch. These analyses showed only one significant interaction between the PLS cut point and minority status in grade 4 ( $p = .005$ ) such that White students with a PLS above the cut point had a greater chance of being at or above the 50th percentile on the SAT-10 compared to Black students above the cut point on the PLS. The researchers noted the need to replicate this effect before definitive conclusions can be drawn.

In a subsequent study, Foorman, Espinosa, Wood, and Wu (2016) examined the appropriateness of FRA for English learner students. A sample of  $n = 102$  English learner students in grades 3–5 participated. The students were classified as English levels 1 and 2 based on district-determined ranges of ability scores on the Comprehensive English

Language Learning Assessment for grades 3–5. The study showed that it was feasible for teachers to use FRA score reports and graphs to note students' strengths and weaknesses in oral language and reading and to differentiate instruction. They also used scores to monitor student progress, make instructional adjustments as needed, and report progress to parents.

### **Proposed Intended Use of Scores**

The review of FRA Grades 3–10 demonstrated evidence of careful test construction consistent with current conceptual frameworks of reading comprehension; appropriate administration and scoring; adequate score reliability; adequate evidence for validity based on test content, on internal structure, and on relations to other variables; and attention to fairness with an emphasis on minimizing measurement bias. With respect to intended use, Foorman and colleagues provided evidence for score appropriateness in identifying profiles of readers with instructional utility.

#### *Identifying Latent Profiles with Instructional Utility (Foorman, Petscher, et al., 2017)*

This investigation had several aims, one of which was to determine the latent profiles of reading and language skills as measured by FRA and the extent to which these latent profiles were related to important reading outcomes, namely, SAT-10 Reading Comprehension (SESAT Word Reading for K). A total of 7,752 students in kindergarten through grade 10 across multiple districts in Florida participated in this study. Demographic information for the sample approximated that of the state of Florida: 42.18 percent White, 29.10 percent Hispanic, 22.5 percent Black, 3.59 percent other; 60 percent eligible for free or reduced-price lunch; and 10.39 percent limited English proficient. There were 5,457 students in grades 3–10. LPA identified three classes. Profiles in grades 3–10 followed a high, medium, and low pattern. In all grades, the latent profiles were significantly related to the reading outcome scores, explaining from 24 percent to 61 percent of the variance, with the mode being 42 percent. These profiles have possible implications for differentiating instruction.

## **LARRC INFERENCE TASK**

### **Conceptual Framework**

The LARRC team was particularly interested in the dimensionality of language (LARRC, 2015) and aimed to assess different levels of receptive and expressing language (i.e., single word, sentence, and discourse levels). In this context, the team developed the Inference Making task to evaluate discourse-level language comprehension following the work of Cain and Oakhill (1999) and Oakhill and Cain (2012). Thus, the LARRC Inference Making task builds heavily on process models of reading comprehension. Inference making is necessary to establish both local and global coherence (Graesser, Singer, & Trabasso, 1994) during comprehension of written or spoken text (Kintsch & van Dijk, 1978). Local coherence inferences are necessary in order to integrate information from adjacent pieces of text, whereas global coherence inferences are used

to fill in details not explicitly stated that are needed to construct a globally coherent representation of text meaning (Cain & Oakhill, 1999, 2014; Currie & Cain, 2015; Freed & Cain, 2017). Inference making in general is a critical skill to successful reading and listening comprehension both concurrently and longitudinally, over and above cognitive factors such as general ability and memory (Cain, Oakhill, & Bryant, 2004; Elleman, 2017; Kim, 2016; Oakhill & Cain, 2012).

### *Description*

The Inference Making task was developed to assess global and local inference-making skills during listening comprehension in children in pre-kindergarten (pre-K) through grade 3. The task includes two stories at each grade level, each one followed by eight questions to assess the ability to generate local and global coherence inferences (four questions each for local and global coherence inferences per text). The stories and questions were based on the work of Cain and Oakhill (1999) and Oakhill and Cain (2012). The second story at each grade level was repeated at the subsequent grade, such that there was one unique story at each grade. Students were read each story and then asked inferential questions.

#### **Story Excerpt:**

Today was Grandma's birthday. The family was getting ready for the party. Dad and Josh were putting up the party tent in the back lawn. Mom told them to put on some sunscreen, so that they didn't burn. Mom drove over to pick up Grandma, who lived an hour away. Mom told Linzie to keep an eye on the cake in the oven and to make some fruit punch.

#### **Sample Questions:**

- What were the family getting ready for?\*
- Answer: Grandma's (birthday) party (2 points); a party (1 point); to go out (0 points)
- What was the weather like?
- Answer: (hot and) sunny (2 points); warm (1 point); rainy (0 points)

### *Administration and Scoring*

The task is individually administered and scored. In this task, children listen to two narrative passages read aloud and are asked a series of inference-based questions. Children's responses are audio-recorded and postscored. Questions are scored as either



correct (2 points), partially correct (1 point), or incorrect (0 points) using a rubric. The total score is the average score on all questions.<sup>3</sup>

### Sample

The LARRC Inference Making task was evaluated using a sample of participants from the larger longitudinal study on listening and reading comprehension conducted in the context of the Reading for Understanding (RfU) research initiative (LARRC, 2017; LARRC & Muijselaar, 2018). Participants were 416 pre-kindergartners (241 boys,  $M = 5$  years and 1 month,  $SD = 4.33$  months), 520 kindergartners (289 boys,  $M = 6$  years and 1 month,  $SD = 3.93$  months), 620 first graders (324 boys,  $M = 7$  years and 1 month,  $SD = 4.10$  months), 724 second graders (380 boys,  $M = 8$  years and 1 month,  $SD = 4.19$  months), and 783 third graders (400 boys,  $M = 9$  years and 1 month,  $SD = 4.10$  months). Children in each grade level were selected from research sites in Arizona, Kansas, Nebraska, and Ohio. It is important to note that the sample was predominantly White (83–94 percent across grades); most with high income level (12.8 percent < \$30,000, 25.3 percent \$31,000–\$60,000, 61.9 percent > \$60,000); and 14.6 percent on free or reduced-price lunch.

### Validity

#### *Evidence Based on Internal Structure*

Several confirmatory factor models that assumed unidimensionality of inference making but accounted for text and coherence factors to various degrees were tested. Three models were directly compared: (1) a one-factor model in which all items loaded on a general inference-making factor; (2) a bifactor model in which all items loaded on a general inference making factor, and in addition, on the text to which they belonged; and (3) multitrait, multimethod (MTMM) model in which each item loaded on a local or global inference factor, in addition to the loadings on the general factor and one of the text factors. The latent factors in all models were specified to be uncorrelated. The fit of the models was evaluated with inspection of three indices: the chi-square goodness-of-fit test statistic, the RMSEA, and the CFI (Kline, 2011). A nonsignificant chi square indicated good overall model fit, whereas a significant chi square showed poor fit. The ratio  $\chi^2/df$  was also used to evaluate model fit. A  $\chi^2/df$  ratio < 2 confirmed a good fit. A model with an RMSEA below .05 has a good approximate fit, an RMSEA between .05 and .08 was taken as satisfactory approximate fit, and values above .10 indicated poor approximate model fit (Browne & Cudeck, 1993). A model with a CFI larger than .95 had a good incremental fit to the data, and a CFI larger than .90 was taken as acceptable (Hu & Bentler, 1999). Differences between nested models were tested with the corrected chi-square difference test (with Satorra-Bentler correction) (Kline, 2011). These analyses showed that, across grades, the MTMM model had the best fit to the data. The general factor explained most of the variance in the items, whereas the latent text and inference factors explained little additional variance. This suggests that the

---

<sup>3</sup> The local and global subscores were also evaluated for reliability and validity but were not deemed adequate (LARRC & Muijselaar, 2018).

construct of inference making is broadly unidimensional. Furthermore, even though it is important to account for text and type of inference, the additional explanatory power of these factors is limited.

#### *Evidence Based on Relations to Other Variables*

Convergent evidence was evaluated with a series of correlation analysis between the Inference Making task scores and scores on the Listening Comprehension Measure (LCM) from the Qualitative Reading Inventory–Fifth Edition (QRI-5; Leslie & Caldwell, 2011), the CELF-4 Subtest Understanding Spoken Paragraphs (USP; Semel, Wiig, & Secord, 2003), and the Test of Narrative Language–Receptive (TNL; Gillam & Pearson, 2004). Across grades, correlations of the Inference Making task with LCM from QRI-5 ranged from .48 to .69; with the USP from CELF-4 ranged from .37 to .61; and with the TNL ranged from .40 to .72. These moderate to high correlations suggest that the Inference Making task is a valid measure of listening comprehension.

#### **Reliability/Precision**

The *internal consistency coefficients* of the test at each grade level were acceptable. Cronbach's alphas for pre-K = .78, kindergarten = .64, grade 1 = .71, grade 2 = .74, and grade 3 = .69. Test-retest reliability was evaluated with correlations for consecutive years. These were consistently moderate (pre-K to kindergarten,  $r = .63$ ; kindergarten to grade 1,  $r = .58$ ; grade 1 to grade 2,  $r = .56$ ; and grade 2 to grade 3,  $r = .54$ ).

#### **Proposed Intended Use of Scores**

This experimenter-developed Inference Making task is a reliable and valid measure to assess discourse listening comprehension in pre-K through grade 3. LARRC and Muijselaar (2018) suggest that the Inference Making task could be used as a measure for general listening comprehension or as a measure of discourse narrative comprehension with a focus on inference making. Indeed, the LARRC team included this measure as one of the main dependent variables in a randomized controlled trial designed to evaluate the efficacy of a language-based comprehension instruction in pre-K through grade 3.

### **CORE ACADEMIC LANGUAGE SKILLS INSTRUMENT (CALSI)**

#### **Conceptual Framework**

The Catalyzing Comprehension Through Discussion and Debate (CCDD) team proposed an expanded operationalization of academic language skills, namely, skills that involve understanding the meanings of words and the syntactic and discourse constructions in which they are embedded (Halliday, 2004; Snow & Uccelli, 2009). The focus on academic language proficiency was driven by evidence that it may be one key source of difficulty in accessing the meaning of texts, particularly in preadolescents and adolescents. The construct Core Academic Language Skills or CALS was defined as “knowledge and deployment of a repertoire of language forms and functions that

co-occur with oral and written school learning tasks across disciplines” (Uccelli et al., 2015a, p. 1). Instead of focusing on discipline-specific language proficiency, CALS focus on the high-utility language skills hypothesized to support reading comprehension *across content areas*. Instead of focusing only on English learners, as most prior research on academic language proficiency had, CALS were hypothesized to be significant contributors of reading comprehension also for English-proficient students.

The work on CALS is situated in a sociocultural pragmatics-based view of language, which views language as inseparable from its social context and posits that language continues to develop throughout adolescence and even adulthood as people continue to learn new ways of using language to navigate more social contexts (Uccelli et al., 2015a, 2015b). During adolescence, language development entails developing “rhetorical flexibility” (Ferguson, 1994; Ravid & Tolchinsky, 2002), defined as the ability to use lexicogrammatical and discourse resources appropriately and flexibly in a variety of social contexts.

### Description

The CALS Instrument (CALS-I) is a researcher-designed group-administered instrument for students in grades 4–8 that measures CALS. CALS are operationalized as a set of skills that correspond to linguistic features prevalent in academic texts across content areas yet are rare in colloquial conversations. This set of skills was hypothesized to support academic reading across school content areas and to encompass the following nonexhaustive domains:

- *Unpacking dense information*: skill in unpacking dense information in academic texts at the word and sentence levels:
  - decomposing complex words (e.g., decomposing nominalizations: *invasion* > *invade*); and
  - understanding complex sentences (e.g., extended noun phrases, embedded clauses).
- *Connecting ideas*: skill in comprehending connectives used to signal relations between ideas in academic texts (e.g., *consequently*, *in contrast*, *in other words*).
- *Tracking themes*: skill in identifying terms or phrases used to refer to the same participants or themes throughout an academic text, specifically tracking conceptual anaphors, those that refer to a complex concept mentioned in a different part of the text (e.g., *Water evaporates at 100 degrees Celsius. This process . . .*).
- *Organizing argumentative texts*: skill in organizing argumentative texts according to conventional academic structures, especially argumentative texts (e.g., thesis, argument, example, and conclusion).
- *Understanding metalinguistic vocabulary*: skill in understanding metalinguistic vocabulary or words that refer to—or qualify—thinking and reasoning processes (e.g., *hypothesis*, *generalization*, *contradictory*).
- *Understanding a writer’s viewpoint*: skill in understanding markers that signal a writer’s viewpoint, especially epistemic stance markers, those that signal a writer’s degree of certainty in relation to a claim (e.g., *certainly*; *it is unlikely that*).

- *Recognizing academic language*: skill in recognizing more academic language when contrasted with more colloquial language in communicative contexts where academic language is expected (e.g., more colloquial versus more academic dictionary-like noun definitions).

### *Administration and Scoring*

CALS-I is a group-administered 45-minute test that has two vertically equated forms: Form 1 (for grades 4–6) contains 49 items, and Form 2 (for grades 7 and 8) contains 46 items. A total of 29 items are common across both forms. Most items are scored dichotomously as correct (1) or incorrect (0), except for those of one task (i.e., organizing argumentative texts), which can receive partial credit. All partial-credit items are rescaled to be between 0 and 1. Scores include raw scores, percent correct scores, factor scores, and extended CALS-I (or ECALS) scores. The ECALS scores are the original factor scores (which are on a z score metric) rescaled to a new scale that has a mean of 500 and a standard deviation of 50 (Barr, Uccelli, & Phillips Galloway, 2019).

### **Sample**

To date, three main studies have provided technical quality evidence for the CALS-I with samples of participants that included English-proficient and bilingual students designated as English learners across grades 4–8. A total of 7,152 students across grades 4–8 from 6 districts and 36 urban public schools in the Northeast and Middle Atlantic regions of the United States participated in the final norming study. The sample was balanced by gender (50.1 percent female), with a majority of students classified as English proficient and 12 percent identified as English learners according to official school records. Students were predominantly from low-income backgrounds (81 percent) as indexed by their eligibility for free or reduced-price lunch.

### **Validity**

#### *Evidence Based on Test Content*

CALS-I tasks were expected to be moderately correlated (Uccelli et al., 2015a). Indeed, across grades CALS-I scores were moderately interrelated (range  $r = .23$  to  $.64$ ), with the lowest correlations observed for the academic register task. Findings revealed also that CALS-I captured developmental trends with upward trends in higher grades, yet considerable individual differences within grade.

#### *Evidence Based on Internal Structure*

This was evaluated using confirmatory factor analysis (CFA) and Rasch IRT. The authors assumed unidimensionality of CALS-I. In these analyses, the CALS-I task-specific scores (Unpacking Complex Words, Comprehending Complex Sentences, Connecting Ideas, Tracking Themes, Structuring Argumentative Texts, Identifying Academic definitions, and Producing Academic Definitions) were used. In two separate studies with students in grades 4–8 (Uccelli et al., 2015a, 2015b) the CFA results supported a

single factor solution (CFI = .93 and .95, TLI = .92 and .94, RMSEA < .05 and = .06) and offered evidence for unidimensionality. In a third study (Barr, Phillips Galloway, & Uccelli, 2019), several competing models were tested to investigate the dimensionality of the construct assessed and the Rasch unidimensional measurement model was selected as the best model due to theoretical, empirical, and practical considerations.

#### *Evidence Based on Relation to Other Variables*

Validity evidence was provided in two studies that included Gates McGinitie (Uccelli, Phillips Galloway, Aguilar, & Allen, 2020) and GISA (Barr, Uccelli, & Phillips Galloway, 2019), respectively. Relations with Gates-MacGinitie Passage Comprehension as indexed by the zero-order correlations were .70 for Form 1 and .75 for Form 2. Relations between the CALS-I and the GISA reading comprehension scores were .69 for Form 1 and .71 for Form 2.

#### **Reliability/Precision**

For Form 1 reliability was .90 and for Form 2 it was .86 as indexed by coefficient alpha. Reliability of the CALS-I was also assessed by comparing the test information function and standard error of measurement for each of the two forms. Both forms had adequate test information function to standard error of measurement ratios (Form 1, -2.8 to 2.6; Form 2, -2.3 to 2.8), indicating that both forms offered adequate estimates of student ability across the expected range, with Form 1 scores having a higher reliability for low-performing students and Form 2 scores for students at higher ability levels (Barr et al., 2019).

#### **Proposed Intended Use of Scores**

This researcher-developed instrument is a reliable and valid measure to assess academic language skills in grades 4–8. The CCDD team used the CALS-I to model the relation between academic language proficiency and reading comprehension (LaRusso et al., 2016), to track concurrent longitudinal development in academic language and reading comprehension (Phillips Galloway & Uccelli, 2019), and to evaluate the impact of interventions (developed in the RfU initiative) on improving students' academic language proficiency and reading comprehension (Jones et al., 2019). The CALS-I is presently available for use as a research instrument upon request. Results of the CALS-I have been used effectively in teachers' professional development to raise awareness of the importance of paying attention to core academic language skills during instruction. Additional uses of the CALS-I to inform pedagogical practice are being investigated (Uccelli et al., 2020).

## THE ASSESSMENT OF SOCIAL PERSPECTIVE-TAKING PERFORMANCE (ASPP) MEASURE

### Conceptual Framework

Diazgranados, Selman, and Dionne (2015) identified the functional dimensions of social perspective taking (SPT) in the context of social-relational frameworks (Martin, Sokol, & Elfers, 2008; Mead, 1934). Social-relational frameworks differ from the cognitive-representational approaches (e.g., theory of mind, executive functions) that have largely dominated the literature on perspective taking. They followed a grounded-theory approach to develop a framework that resulted in the development of the Social Perspective Taking Acts Measure (SPTAM; Diazgranados et al., 2015) initially, and a revised version subsequently (ASPP; Kim et al., 2018). This work identified SPT as acts that serve different functions. Specifically, when students were challenged to resolve social situations presented in a scenario, they produced responses that (1) *acknowledged* the existence of different actors; (2) *articulated* the thoughts, feelings, and orientations to action of those actors; and (3) *positioned* these actors according to their characteristics, social roles, or circumstances in the scenario. Responses varied in their levels of *integration*, as participants demonstrated different abilities to acknowledge, articulate, and position the perspectives of multiple actors in the scenario. Kim et al. suggest that the ability to consider multiple perspectives is a critical skill for learning in 21st-century classrooms, facilitating the processing and integration of information from multiple sources.

### Description

ASPP is a revised and extended version of SPTAM (Diazgranados et al., 2015), a scenario-based assessment of students' ability to perform SPT acts in response to written texts about specific social situations. ASPP is designed to assess children's ability to *acknowledge*, *articulate*, and *position* the perspectives of multiple stakeholders in a given social conflict and to provide solutions that consider and integrate their different positions. The measure puts students in the shoes of an advisor, who needs to make a recommendation to address social conflicts that occur at the interpersonal, group, and institutional levels. Specifically, students are presented with a subset (typically three) of four scenarios. In each scenario, an actor who is observing a social problem (i.e., a witness to teasing, mockery, or breaking school rules) does not know what to do and is asking different people for advice. Students are prompted to think about the recommendations this observer might receive from the following two types of advisors: (1) someone who was recently teased, whose privacy was recently violated, or is otherwise oriented in opposition to the perpetrator(s); and (2) someone who often socializes with the teasers or rule violators, or is otherwise in sympathy with the perpetrator(s). Then, students answer three questions: (1) What would (the prompted actor) recommend to the observer? (2) Why would (the prompted actor) make that recommendation? and (3) What might go wrong with this recommendation? This structure (four scenarios × two advisors) provides participants with the opportunity to produce open-ended responses to these sets of questions. Answers to all three questions provided by each advisor constitute one unit of analysis, which receives one score for each of the three

subscales: *acknowledgment*, *articulation*, and *positioning*. Note that the revised version of ASPP excluded acknowledgment as a core component and focused primarily on *articulation* and *positioning* since both of these depend on actors being acknowledged (Kim et al., 2018). These subscales refer to the *function* of the SPT act, with acknowledgment serving the basic function of introducing a potential actor; articulating that actor's perspective is a more advanced act, while positioning that actor's perspective in light of her social role represents the pinnacle of SPT skill in ASPP.

### *Administration and Scoring*

ASPP is a group-administered 30-minute test identified for students as "The Advice on Making Social Choices Measure." An experimenter reads the instructions and walks participants through the scenarios and questions, providing them with 4 minutes to answer each prompt. If participants complete a section before others, they are allowed to move on at their own pace. Coding follows detailed guidelines with examples that can be found in the coding manual (Diazgranados et al., 2011). The coding system was deemed stable when interrater reliability reached .90 (which reflected the proportion of units on which raters agreed out of the total number of units coded). This coding system results in three subscale scores (acknowledgment, articulation, and positioning). ASPP includes two forms (i.e., with the addition of new social scenarios and changes in elicited perspectives of the two advisor roles; Form A and Form B), and scoring excludes the *acknowledgment* dimension (even though it is initially coded).

*Acknowledgment* is the act of identifying the various actors involved. It can be determined by counting, only once per unit of analysis, the names and pronouns that refer to any particular actor that is included in the unit of analysis, irrespective of whether anything further is said about that actor.

*Articulation* is the act of describing the thoughts, feelings, or orientations to action of distinct actors involved. It can be determined by counting, only once per unit of analysis, the actors whose feelings, opinions, beliefs, preferences, and orientations to action are described in the scenarios.

*Positioning* is the act of identifying the roles, circumstances, or attributes that qualify the position distinct actors hold in a social scenario. It can be determined by counting, only once per unit of analysis, the actors whose roles, attributes, experiences, or circumstances are identified in the scenario as motivations for their beliefs, thoughts, actions, or potential actions.

Separate scores are assigned to each of the dimensions: for acknowledgment, 1 point per (potential) actor named; for articulation, 1 point per perspective described; and for positioning, 1 point per perspective explicitly positioned. In some past research, scores for each dimension have been scaled separately using item response theory to facilitate analysis.

### **Sample**

Diazgranados et al. (2015) evaluated the initial SPTAM measure using a sample of participants from the larger study conducted in the context of the RfU Research Initiative. Participants were  $n = 459$  students in grades 4–8 (50 percent boys), 25 percent

in grade 4, 21 percent in grade 5, 18 percent in grade 6, 16 percent in grade 7, and 16 percent in grade 8. Subsequently, Kim et al. (2018) evaluated the ASPP measure in the same context, drawing on  $n = 1,299$  students in grades 4–7. The current participants include 52 percent female students, 14 percent Black, 39 percent White, 4 percent Asian, 39 percent Latino, 3 percent mixed race/other, 79 percent eligible for free or reduced-price lunch, 12 percent English language learners, and 14 percent with special education classification.

## Validity

### *Evidence Based on Internal Structure*

Diazgranados et al. (2015) evaluated the SPTAM factor structure using a CFA, which provided support for a three-dimensional model in which SPT is a factor comprising acknowledgment, articulation, and positioning. The results showed that all parameter estimates were positive, statistically significant, and exhibited loadings in the range .62–.71. The three subscales exhibited positive, moderate, and statistically significant correlations with each other ( $r$  range .40 to .46).

Kim et al. (2018) tested the two-factor structure of the ASPP, articulation and positioning, using multigroup categorical confirmatory factor analysis (CCFA). The standardized factor loadings of the articulation items on the articulation factor ranged from .55 to .77, and factor loadings of the positioning items on the positioning factor ranged from .49 to .68. This model with two dimensions had a good fit,  $\chi^2(135) = 174.72$  (Form A = 67.44, Form B = 107.28),  $p = .01$ , RMSEA = .02, 90% CI [.01, .03], CFI = .99, and TLI = .99. These multidimensional models fit the data significantly better than unidimensional models,  $\Delta\chi^2(1) = 231.91$ ,  $p < .001$ .

### *Evidence Based on Relations to Other Variables*

Diazgranados et al. (2015) examined hypothesized relations between SPT and several other constructs, while controlling for the rest. Specifically, they expected and confirmed that children in higher grades would perform better (for every additional grade level, students scored .37 points more on SPTAM,  $p < .001$ ) and that girls would perform better (girls scored 1.37 points higher on SPTAM than boys,  $p < .001$ ). It was also expected that SPTAM would have a negative and moderate association with the Aggressive Interpersonal Strategies (AINS) measure (Dalhberg, Toal, & Behrens, 1998). Indeed, for every additional unit in the AINS measure, students obtained .33 fewer points on SPTAM ( $p < .10$ ). Finally, SPTAM was expected to have a moderate positive association with the Written Language Scale of the Oral and Written Language Scale (OWLS-II; Carrow-Woolfolk, 1995) because of its high language production demands. Indeed, for every additional point in the OWLS writing test, students scored 11.54 more points on SPTAM ( $p < .001$ ). SPTAM was not related to measures of complex reasoning (LAS; Dawson, 2002; Fischer & Bidell, 2006), academic language (CALs-I; Uccelli et al., 2015a, 2015b), and reading (GMRT; MacGinitie & MacGinitie, 1988).

Kim et al. (2018) evaluated the relation of the two ASPP factors to several academic and engagement outcomes. The results showed that the overall ASPP model explained



52 percent to 54 percent of the variance in Reading Engagement (Wigfield et al., 2008), 33 percent to 34 percent of the variance in Classroom Engagement (Wellborn & Connell, 1987), 62 percent to 67 percent of the variance in ELA, and 62 percent to 64 percent of the variance in the Mathematics state test scores. The model had adequate goodness of fit,  $\chi^2(913) = 1138.10$  (Form A = 501.37, Form B = 636.73),  $p < .001$ , RMSEA = .02, 90% CI [.02, .02], CFI = .97, and TLI = .96. When demographic variables were considered, the results also showed that students who scored higher on ASPP were more likely to be in higher grades and female. English language learners and students eligible for special education were likely to score lower on the ASPP.

### Reliability/Precision

Diazgranados et al. (2015) reported Cronbach's alpha for each subscale of SPTAM:  $\alpha_{\text{acknowledgment}} = .80$ ,  $\alpha_{\text{articulation}} = .83$ , and  $\alpha_{\text{positioning}} = .70$ . The latent factor of SPT exhibited excellent internal consistency (.90).

For each of the two forms of ASPP, Kim et al. (2018) reported both Cronbach alpha coefficients,  $\alpha_A = .82$  and  $\alpha_B = .78$  for articulation, and  $\alpha_A = .67$  and  $\alpha_B = .66$  for positioning, and omega reliabilities,  $\Omega_A = .86$  and  $\Omega_B = .83$  for the articulation scale and  $\Omega_A = .74$  and  $\Omega_B = .76$  for the positioning scale. In the CCFA context, omega reliabilities should be interpreted as more representative, and for both articulation and positioning they were acceptably high.

### Proposed Intended Use of Scores

Diazgranados et al. (2015) and Kim et al. (2018) suggest that SPTAM/ASPP provides researchers with a tool to assess early adolescents' ability to produce SPT acts in an innovative way. This instrument can be particularly useful in the context of intervention programs whose theory of change includes SPT performance as a mechanism of change or outcome. For example, Hsin and Snow (2017) used a modification of the SPTAM coding scheme to examine the incidence of SPT acts in the argumentative essays of language-minority and English-only students in grades 4–6, and then associated the SPT found in students' writing with their ASPP scores. The results showed that language-minority students matched or surpassed the English-only students on perspective taking, and that there was a significant relationship between essay SPT and ASPP scores among language-minority students but not among English-only students.

## ASK KNOWLEDGE ACQUISITION MEASURE

### Conceptual Framework

Promoting Acceleration of Comprehension and Content through Text (PACT) is a multicomponent treatment aimed at improving content-area knowledge acquisition in social studies/history and also improving reading comprehension, consistent with the Common Core State Standards (CCSS). The CCSS requires teachers to emphasize students' understanding and learning from complex reading materials. Existing research shows that middle school teachers must make adjustments to current instructional practices to provide the reading opportunities and instruction necessary to ensure that

students meet the CCSS expectations. To highlight the problem, students engaged in reading texts in only 38 percent of middle and secondary social studies classes and fewer than 20 percent of middle school social studies classes. Text reading consumed only 10.4 percent of social studies instructional time (Swanson, Wanzek, Vaughn, Roberts, & Fall, 2016). Vaughn et al. (2013) identified five components of the PACT intervention informed by the content learning model (Gersten et al., 2006; Vaughn et al., 2009) that focus on improving understanding while reading text, and provide opportunities for students to connect new, text-based learning to previous learning. They also infused the intervention with motivational aspects to further bolster its effectiveness among adolescent learners. These components are (1) a comprehension canopy that contains a motivational springboard and an overarching issue or question, (2) essential words or key vocabulary related to the unit, (3) knowledge acquisition (appropriate text-based instruction and reading), (4) team-based learning (TBL) comprehension checks, and (5) TBL knowledge application. In this context, it was deemed important to develop an appropriate knowledge test, the ASK Knowledge Acquisition measure.

### Description

The ASK Knowledge Acquisition measure is one of the two subtests of the ASK assessment (Vaughn et al., 2015). The other subtest is a passage comprehension measure. The assessment is a researcher-developed measure. This knowledge subtest is a 42-item, four-option, untimed multiple-choice test that measures content knowledge in the three units that comprised the intervention (Colonial America, Road to Revolution, and Revolutionary War). The items comprising the test were collected (with permission) from released Texas state social studies tests (Texas Assessment of Knowledge and Skills), released Massachusetts state social studies tests (Massachusetts Comprehensive Assessment System), and released advanced placement tests in social studies from the College Board. Researcher-developed vocabulary items were also included in the item set. The item pool from these released items was further narrowed to align with the content of the Texas and Florida state content standards for the units covered in the PACT intervention. Following the first year of implementation of PACT the psychometric properties of the ASK content items were evaluated. Poor-performing items were removed from the assessment and a final version was created.

### *Administration and Scoring*

The ASK Knowledge Acquisition measure is dichotomously scored (1 for correct, 0 for incorrect responses). The test was administered at pretest, posttest, 4 weeks following intervention, and again 12 weeks following intervention (Wanzek et al., 2015).

### Sample

Participants were 1,487 students (male = 712), 39 percent qualified for free or reduced-price lunch, 4.8 percent were classified as limited English proficient, and 7.9 percent of students qualified for special education services. Students' average age was 13.16 in the treatment condition and 13.16 in the comparison condition (Vaughn et al., 2015).

## Validity

### *Evidence Based on Internal Structure*

Item response theory was used to analyze initial data from the validation process. IRT parameters for the 42 items reflect a sizable range of underlying knowledge acquisition ( $-2.12$  to  $+2.67$ ) and good item discrimination ( $0.05$  to  $2.13$ ). Vaughn et al. (2013) used confirmatory factor analysis on pretest data to evaluate the degree to which the hypothesized models represented their observed data. Model fit was very good for the ASK Knowledge Acquisition test:  $\chi^2 = 1,022.69$ ,  $df = 989$ ,  $p = .22$ , CFI = .97, RMSEA = .009.

### Reliability/Precision

Reliability information from IRT analyses was above .80 from  $-1.6$  to  $+1.2$  thetas. Alpha coefficients for the ASK knowledge acquisition measure was .89 (Vaughn et al., 2013). Vaughn et al. (2017) reported Cronbach's alpha of .93. Wanzek, Swanson, Vaughn, Roberts, and Kent (2015) reported alpha of .90.

### Proposed Intended Use of Scores

This experimenter-developed measure is a valid and reliable indicator of middle and secondary school students' U.S. history knowledge and reading comprehension ability in the social studies domain. ASK has been used to evaluate the efficacy of the PACT intervention (developed as part of the RfU initiative) for improving students' social studies content knowledge and text comprehension among typical grade 8 students (Vaughn et al., 2013, 2015), grade 8 English learners (Vaughn et al., 2017; Wanzek et al., 2016), grade 8 students with disabilities (Swanson et al., 2016; Wanzek et al., 2016), and grade 11 students (Wanzek et al., 2015).

## BRIDGE-IT MEASURE

### Conceptual Framework

Barth, Barnes, Francis, Vaughn, and York (2015) aimed to develop a computerized inference measure drawing on the extant literature in discourse processes. Inferential processes support the integration of text-derived information and general world knowledge (Graesser, Singer, & Trabasso, 1994). These inferential processes involve maintaining local and global coherence during reading. Local and global coherence has been examined by studies that manipulate textual features, including distance in the text that separates two sentences or ideas that need to be integrated (Albrecht & O'Brien, 1991). Shorter distances between sentences may draw on local-coherence processes such as accessing and retrieving information from working memory (Albrecht & O'Brien, 1993). Larger distances are more likely to tap global coherence processes such as integration of information and bridging inferences. Coherence breaks become easier to detect with age (e.g., Ackerman, 1984) and seem to be more difficult to detect over longer distances (Pike, Barnes, & Barron, 2010). Moreover, skilled comprehenders detect coherence breaks more easily than less-skilled comprehenders, especially with

larger distances between information units (e.g., Barnes, Faulkner, Wilkinson, & Dennis, 2004; Cain, Oakhill, & Lemmon, 2004).

### Description

Bridge-IT was designed to measure the effects of textual distance (i.e., near versus far) on students' ability to generate inferences by judging the consistency or inconsistency of a continuation sentence with prior text. The Bridge-IT consists of 32 five-sentence narrative passages, presented on a computer monitor. Each story consists of five sentences and contains a key sentence important to making the consistency judgment. In the near condition, the key sentence is the final sentence in the story. In the far condition, the first sentence of the story serves as the key sentence. In both conditions, the additional sentences of the story are compatible with either the consistent or inconsistent continuation sentence. Correct judgments in the near condition require that readers evaluate information presented earlier in the text as well as the critical information presented in the final sentence. This information is likely still accessible by the reader. Correct judgments in the far condition require that readers evaluate information they just read as well as critical information in the first sentence of the story, which likely needs to be reactivated from episodic memory.

### *Administration and Scoring*

A "Ready" prompt appears on the computer monitor for one second, followed by a five-sentence story. Instructions prompt students to press the spacebar after they finish the story. The spacebar removes the story and presents an asterisk in the center of the screen to signal the presentation of the test sentence. Participants receive instructions to read the test sentence and then to press a green button if they judge the sentence as a good continuation (i.e., consistent) or a red button if they judge that the sentence is not a good continuation (i.e., inconsistent). Judgments are to be made as quickly and accurately as possible. Students are provided two practice items to ensure familiarity with good and poor continuations and the task procedure.

Students receive a testlet that consists of eight items in each condition (i.e., near-consistent, far-consistent, near-inconsistent, and far-inconsistent). Items are counter-balanced across conditions. For each item, reading time is measured for the passage, and accuracy and response time are measured for continuation sentence judgments. Continuation sentences range from 3 to 12 words in length across items. Word length is consistent across consistent and inconsistent versions of the continuation sentence for each passage.

In terms of scoring, a total accuracy score and condition accuracy scores in all four conditions (i.e., near-consistent, far-consistent, near-inconsistent, and far-inconsistent) are calculated. Accuracy scores represent the proportion of items answered correctly after trimming for outliers.

### Sample

Barth et al. (2015) evaluated the Bridge-IT measure using a sample of 1,203 students ( $n = 531$  struggling comprehenders, 11 percent in grade 6, 14 percent in grade 7, 19 percent in grade 8, 15 percent in grade 9, 17 percent in grade 10, 15 percent in grade 11, and 10 percent in grade 12; and  $n = 675$  adequate comprehenders, 9 percent in grade 6, 14 percent in grade 7, 13 percent in grade 8, 17 percent in grade 9, 17 percent in grade 10, 16 percent in grade 11, and 13 percent in grade 12). Adequate comprehenders were students attaining scale above 2,150 on the Texas Assessment of Knowledge and Skills (TAKS) Reading Test.

### Validity

#### *Evidence Based on Test Content*

Barth et al. (2015) hypothesized grade-level changes in inferential processes across grades 6–12 for both adequate and struggling comprehenders, especially in the far condition. With regard to accuracy, results indicated that in the near condition, the effect of grade within distance was significant ( $p < .001$ ). Students in grades 6 and 7 were less accurate than students in grades 10–12; students in grades 8 and 9 were less accurate than students in grade 10 ( $p < .007$ ). In the far condition, students in grade 10 were more accurate than students in grades 6–9; and students in grade 12 were more accurate than students in grade 9 ( $p < .007$ ). With regard to response time, students in grade 6 were slower at sentence continuation judgments than students in grades 8–12; students in grade 7 were slower than students in grades 9–12; students in grade 8 were slower than students in grades 10–12; and students in grade 9 were slower than students in grades 11 and 12 ( $p < .007$ ).

#### *Evidence Based on Relation to Other Variables*

Barth et al. (2015) also hypothesized that inferential processes would account for unique variance in passage-level comprehension but not single-sentence comprehension, after controlling for working memory and a host of other reading-related variables. Predictive validity evidence was assessed with a series of hierarchical regression analyses. Bridge-IT-near explained 0.7 percent of the variance in the Test of Sentence Reading Efficiency and Comprehension (TOSREC) standard scores, and 3 percent of the variance in Gates MacGinitie reading test-Lexile Score over and above grade level, WJ-III letter word identification, TOWRE, WJ-III numbers reversed, KBIT-2 verbal knowledge, and reader group status.

Bridge-IT-far explained 0.3 percent of the variance in TOSREC (Wagner, Torgesen, & Rashotte, 2010) standard score and 2 percent of the variance in the Gates MacGinitie test-Lexile Score over and above grade level and other linguistic and cognitive measures (Barth et al., 2015).

### Reliability/Precision

Average reliability coefficients (Kuder-Richardson 20) were .85 for near-consistent, .87 for near-inconsistent, .83 for far-consistent, and .87 for far-inconsistent continuations.

### Proposed Intended Use of Scores

Barth et al. (2015) suggest that Bridge-IT adequately discriminates inference making, local and global, across grade levels 6–12 and comprehension skill (skilled versus less-skilled comprehenders). Thus, Bridge-IT can be used as a process measure of inference making.

## READI LITERATURE EPISTEMIC COGNITION SCALE (LECS)

### Conceptual Framework

Epistemic cognition has been broadly defined as the knowledge and beliefs people draw from in order to understand particular phenomena (Hofer & Pintrich, 1997; Yukhymenko et al., 2016). Epistemic cognition has been found to be related to students' problem solving, learning, and reasoning about topics in the natural and social sciences (Bråten, Strømsø, & Britt, 2009; Conley, Pintrich, Vekiri, & Harrison, 2004; Sinatra, Kienhues, & Hofer, 2014). Whether epistemic cognition also relates to students' understanding of literature remains less clear.

Literary reading (i.e., understanding literature, response to literature) can be conceptualized as a complex problem-solving task that requires readers to go beyond basic comprehension of the explicit content in literary texts. Readers must make deeper interpretative inferences about the literary text content, such as inferences about the moral and theme of the text (Goldman, McCarthy, & Burkett, 2014). Literary reading is also an ill-defined problem-solving task, as readers do not come to the same interpretations even after reading the same literary text. Thus, literary reading adopts some of the problem-solving characteristics found in the natural and social sciences. In this regard, epistemic cognition may also play an important role in literary reading. The Literature Epistemic Cognition Scale (LECS; Yukhymenko-Lescroart et al., 2016) was developed to measure epistemic cognition in literature.

Note that the READI team also developed epistemic cognition scales in history and science. The science and history scales emphasized two dimensions of epistemic cognition for multiple sources in history and science: the importance of corroborating across documents (history) and data sets and experiments (science), and the complexity and uncertainty of historical/scientific knowledge. These scales have not been validated to the same extent as LECS yet, and thus are not reviewed here.

### Description

LECS measures three epistemic constructs for literature in adolescents (grades 6–12): relevance to life, multiple meanings, and multiple readings. *Relevance to life* measures the degree to which readers believe that reading literature can help them understand the human condition. Reading literature in order to understand the human condition is a

fundamental assumption in the field of literary studies. *Multiple meanings* refer to readers' tendency to view literary texts as amenable to multiple interpretations. *Multiple readings* reflect readers' belief in the benefit of multiple readings in understanding a literary text. These constructs are thought to be central to adolescents' understanding of literature.

LECS consists of 16 items (the prevalidation version had 29 items) across the three different subscales: 5 items for multiple meanings, 6 items for multiple reading, and 5 items for relevance to life.

#### *Administration and Scoring*

LECS is individually administered and scored. Higher scores on the multiple meanings and relevance to life subscales reflect more sophisticated beliefs. Higher scores on the multiple readings subscale reflect less sophisticated beliefs. Prior to analysis, items corresponding to the multiple readings subscale were recoded so that higher scores reflected more sophisticated beliefs.

### **Sample**

LECS was evaluated using a sample of 798 students. Of the total students, 455 were in middle school ( $M = 13.2$  years,  $SD = 0.95$  years) and 343 were in high school ( $M = 16.1$  years,  $SD = 1.27$  years). Of the total sample, 53.5 percent were female, and gender was evenly distributed across grades. Students were chosen from 47 classrooms across four middle schools and four high schools in a district located near a large urban Midwestern area. In regards to race, 33.4 percent of participants self-reported as Hispanic or Latino, 24.1 percent as White, 21.4 percent as Asian, 6.8 percent as Black, 1.6 percent as American Indian or Alaska Native, 0.9 percent as Native Hawaiian or Pacific Islander, and 11.7 percent as other.

### **Validity**

#### *Evidence Based on Internal Structure*

The 798 surveys were divided into split-half samples after stratifying across gender and grade. The first sample ( $n = 399$ ) was used to perform a confirmatory factor analysis to test the three-factor structure of the original 29-item scale. The model fit was evaluated with inspection of several indices: chi-square index, RMSEA, standardized root mean square residual (SRMSR), CFI, Tucker-Lewis index (TLI), and chi-square to degrees of freedom ratio. The results did not indicate good model fit:  $\chi^2(374, n = 399) = 822.2$ ,  $p < .001$  ( $\chi^2/df = 2.20$ ), CFI = .892, TLI = .882, RMSEA = .055, 90% CI [0.050, 0.060], SRMSR = .068.

The second sample ( $n = 399$ ) was also used to perform a confirmatory factor analysis to test the model fit of an adjusted LECS with 16 items. Results indicated a good model fit:  $\chi^2(101, n = 399) = 124.3$ ,  $p = .058$  ( $\chi^2/df = 1.23$ ), CFI = .987, TLI = .985, RMSEA = .024, 90% CI [0, 0.037], SRMSR = .035. Model fit indices also did not change significantly by grade and gender for models measuring invariance of factor pattern, loadings, and variances, indicating that the model is valid for all genders in middle and high school.

### *Evidence Based on Relations to Other Variables*

Criterion validity was evaluated with correlational analyses between the subscales of LECS, the Speed of Knowledge Acquisition subscale from the Wood and Kardash (2002) epistemology scale, and students' reading habits. The Speed of Knowledge Acquisition subscale measured students' beliefs about the speed of learning that ranged from learning is quick and straightforward to learning is complex and gradual. Speed of knowledge acquisition was predicted to correlate with the multiple meaning and multiple reading subscales. Students' reading habits were assessed by their response on two questions about their reading habit outside of school. Students who read more outside a school setting were thought to find reading more enjoyable, which would be associated with positive ratings on all three of the epistemic cognition constructs. Speed of knowledge acquisition was positively correlated with multiple reading,  $r(397) = .49$ ,  $p < .001$ , and multiple meaning,  $r(397) = .50$ ,  $p < .001$ . Liking of reading was positively correlated with multiple reading,  $r(397) = .36$ ,  $p < .001$ , relevance to life,  $r(397) = .21$ ,  $p < .001$ , and multiple meaning,  $r(397) = .17$ ,  $p = .006$ .

### **Reliability/Precision**

The omega reliability for each subscale was acceptable: .78 for multiple meaning, .85 for relevance to life, and .89 for multiple reading.

### **Proposed Intended Use of Scores**

LECS is a reliable and valid measure to assess epistemic cognition for literature in adolescents (grades 6–12). As the first measure of epistemic cognition for literature, LECS can be used to explore the relationship between epistemic cognition and literary reading. The READI Literature intervention (Goldman, Greenleaf, et al., 2016) incorporated LECS as a pre- and posttest and the analysis showed that pretest scores on the multiple meaning, multiple reading, and relevance to life subscales predicted posttest scores. Importantly, the multiple meaning and relevance to life beliefs changed as a result of the intervention.

## **READI EVIDENCE-BASED ARGUMENT (EBA) MEASURE**

### **Conceptual Framework**

The ability to identify, evaluate, and synthesize information across multiple sources is a very important literacy skill in the 21st century. The READI team has focused on developing an instructional and curricular intervention that can help adolescents develop evidence-based argumentative skills from multiple sources across academic disciplines (Goldman, Greenleaf, et al., 2016; Goldman et al., 2019). However, what constitutes an evidence-based argument differs according to discipline. As a result, students must learn to engage in different reading practices that reflect different disciplinary epistemologies. This is challenging because students are rarely taught the discipline-specific skills and knowledge required to do so. The READI Evidence-Based Argument (EBA) assessments were designed to evaluate adolescents' ability to make



evidence-based arguments from multiple sources in each of three disciplines, one in science, one in literature, and one in history. The science EBA was most extensively tested and was used as a proximal outcome measure in the randomized controlled trial efficacy study. The literature EBA was developed in the context of the design-based classroom research and was administered in these classrooms as well as in the 2-year longitudinal study. The history EBA was likewise developed and tested in the context of the design-based research classrooms. Because the technical qualities of the literature and history EBAs need further testing in larger samples of students, this review focuses primarily on the science EBA.

### *Science EBA*

The READI science EBA was aligned with the learning goals of the science intervention. These include the following: Students need an understanding of what knowledge and knowledge building in science means. They must understand how claims and evidence are established or justified in science as well as the reasoning principles used to connect evidence to claims. Students must be able to understand different types of scientific texts and graphics that present scientific information. Finally, they must be able to understand the technical expressions and language conventions used in the texts and graphics. When students are equipped with this knowledge, they will be able to engage in evidence-based argumentation from multiple sources in science. More specifically, they will be able to use information from scientific texts to construct their own explanations of science phenomena, support their explanations, and critique explanations. These are the skills that are specifically assessed by the READI science EBA measure.

### **Description**

The READI science EBA measure consists of five different tasks that tap evidence-based argumentative skills from multiple sources in science:

- *Reading*: closely reading and annotating scientific texts;
- *Essay*: reading and synthesizing task-relevant information within and across scientific texts;
- *Multiple-choice (9 items)*: reading and synthesizing task relevant information within and across scientific texts;
- *Graphical model comparison*: analyzing two graphic explanatory models related to the topic, selecting the better of the two, critiquing explanatory graphic model; and
- *“Peer” essay evaluation*: critiquing explanatory graphic models.

### *Administration and Scoring*

Each student was provided with a text set on skin cancer or coral bleaching. A text set consisted of one text that provided background information about one of the two topics, two additional texts providing more information about the topic, and two

graphics that portrayed an explanation of a phenomenon associated with the topic. The texts in a given set and tasks were chosen so that students would have to read and synthesize information across multiple sources.

The science EBA was administered over 2 school days. On the first day, students were administered a six-question survey that measured their prior knowledge of skin cancer or coral bleaching. Students were then told they were going to read about one of two topics. They were explicitly told that they would have to read and use information across multiple sources to explain a phenomenon related to the topic. Students received a text set and were asked to *read and annotate the texts*. Texts could be read in any order, although students were encouraged to read the background text first. On the second day, students were given the same text set and a booklet in which they would complete the other tasks.

The *essays* were scored according to the number of concepts and connections that students provided. Connections were indicated by the students' use of causal language. Essays were scored sentence by sentence.

The *graphical model comparison* task was given a score of 1 or 0 based on a rubric of acceptable answers. The justification that students provided for the model they selected had to include a variation of the following acceptable language conventions: steps, step-by-step, order, cause and effect, the way it is organized, process, chain reaction, and how they connect to each other.

The *peer evaluation essays* were scored based on the inclusion of six variables of interest that were present in the two peer essays: relevance, coherence, completeness, importance of sourcing, mentioning the graph, and mentioning a concept tied to the graph. A score of 1 was given for each variable if students wrote about the variable in at least one of their two evaluations. The acceptable language conventions for each variable were provided in a rubric.

### Sample

Participants were  $n = 964$  students in grade 9 (567 READI) from 95 classrooms (48 READI) in 24 schools (12 READI) and were present for all 4 days of the EBA assessment (two pre and two post).

### Validity

The science EBA assessment consisted of several tasks that were designed to assess the skills outlined in the READI science intervention designed to help adolescents develop evidence-based argumentative skills from multiple sources (Goldman et al., 2019). Students need an understanding of what knowledge, knowledge building, reasoning, and knowledge expression (in text and graphics) in science means. When students are equipped with this knowledge, they will be able to engage in evidence-based argumentation from multiple sources in science. Thus, the assessment has a solid theoretical basis with respect to the dimensions of the construct being measured.

### **Reliability/Precision**

Interrater reliability for the scoring of essays was determined by two coders who were trained to code essays on one topic and another coder who was trained to code essays on both topics. The two single-topic coders scored six subsets of essays that made up the total set, while the double topic coder randomly coded 20 percent of each subset. The kappa scores for the coral bleaching essays were .75, .89, .85, .86, .86, and .93 while the kappa scores for the skin cancer essays were .64, .92, .88, .89, .85, and .93.

Interrater reliability for the scoring of model evaluation responses was determined by one coder who scored all the responses in three different subsets and another coder who scored 20 percent of responses within each subset. The kappa scores were .90, .92, and .91.

Interrater reliability for the scoring of the peer evaluation task was determined by one coder who scored all the essays and another coder who scored a small set of evaluations over a period of time. Kappa scores were .86, .80, and .84.

### **Proposed Intended Use of Scores**

The READI science EBA is a reliable assessment of evidence-based argumentation from multiple sources in science for students in grade 9. The science EBA measure was used to evaluate the efficacy of the READI intervention and showed sensitivity to intervention effects. Specifically, on average, the intervention group had 5.7 percent higher scores on the multiple-choice task than the control group (Goldman et al., 2019). In regard to essay task performance, students in the intervention and control groups did not differ significantly in the percentage of nodes and connections included in their essays, although the intervention group's scores were generally higher than those of the control group.

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

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

## CONTENTS

INTRODUCTION .....	143
THE RATIONALE FOR THE PEDAGOGICAL EMPHASIS IN THE RFU.....	145
Instruction as First Among Equals.....	145
The Pedagogical Charge.....	146
Previewing the Curriculum and Instruction Portfolio of Work.....	147
EXAMINING THE RFU TEAM PORTFOLIOS.....	148
Language and Reading Research Consortium.....	148
Florida Center for Reading Research.....	155
Catalyzing Comprehension Through Discussion and Debate.....	171
Promoting Adolescents' Comprehension of Text.....	180
Reading, Evidence, and Argumentation in Disciplinary Instruction.....	188
Looking Across the Array.....	198
REFERENCES.....	198

## INTRODUCTION

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

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

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

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

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

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

---

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

<sup>2</sup> Appendix 4-2 summarizes the demographic information, by team and RCT, of the students involved in the RCTs.

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

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

### THE RATIONALE FOR THE PEDAGOGICAL EMPHASIS IN THE RFU

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

#### Instruction as First Among Equals

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

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

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

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

### **The Pedagogical Charge**

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

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

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

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

### **Previewing the Curriculum and Instruction Portfolio of Work**

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

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

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

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

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



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

## EXAMINING THE RFU TEAM PORTFOLIOS

### Language and Reading Research Consortium

#### *Overview*

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

#### *Developing the Let's Know! Curriculum*

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

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

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

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

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

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

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

#### *The Randomized Controlled Trial*

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

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

#### *Follow-Up to the Randomized Controlled Trial*

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

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

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

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

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

### Summary

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

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

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

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

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

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

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

## Florida Center for Reading Research

### *Overview*

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



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

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

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

### *Developing the CTT Portfolio*

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

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

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

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

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

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

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

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

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

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

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

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

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

#### *The Design of the CTT Intervention Studies*

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

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

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

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

rather than by component intervention, and dissemination of results is ongoing. As of when this synthesis was ready for publication, we had access to results only from grades 3 and 4 for CE<sub>1</sub> (Connor et al., 2018). In addition, some of the component interventions underwent earlier, smaller-scale RCTs prior to CE<sub>1</sub> and CE<sub>2</sub>. Thus, where efficacy results were not yet published for a component intervention, we summarized the earlier, smaller RCTs (i.e., for LIM and MAT) whenever they were available. DAWS, WKeB, and CALI, as we said, were evaluated in free-standing RCTs.

CE<sub>2</sub>. In the second comparative efficacy study (CE<sub>2</sub>), FCRR researchers (C. Lonigan, personal communication, July 29, 2019) used the results of CE<sub>1</sub> to judiciously combine treatments into curricular approaches that bear a stronger resemblance to multicomponent interventions, hypothesizing that the combined approaches might overcome the somewhat sporadic pattern of mainly proximal effects observed for single-component interventions in CE<sub>1</sub> and yield, in their stead, more robust and consistent effects. The implementation design of CE<sub>2</sub> was limited to DR, LIM, and COMPASS in pre-K and kindergarten, with three combination interventions (DR/LIM, DR/COMPASS, and LIM/COMPASS) each compared to BAU. In grade 4, two versions of TEXTS (the CE<sub>1</sub> version and a newly created adaptive version [TEXTS<sup>A</sup>], with provision for individualized journeys through the curriculum) were compared to BAU. As of the date this synthesis was ready for publication, no results from CE<sub>2</sub> were available for summarization. And as we intimated in the earlier overview, the trends for the paired interventions in CE<sub>2</sub> appear to be more promising than the results of individual components in CE<sub>1</sub>.

**CE<sub>1</sub> for grades 3 and 4.** The main results available for CE<sub>1</sub> are for grades 3 and 4 (Connor et al., 2018). The sample consisted of 338 grade 3 students and 307 grade 4 students who qualified for reading comprehension intervention (meaning that they scored below the 45th percentile on the EOWPVT). Children came from 33 and 31 schools and 135 and 115 classrooms in grades 3 and 4, respectively. Students within schools were assigned to conditions (COMPASS, LIM, TEXTS, ERC, or BAU) using an incomplete-random-blocks design. The interventions were delivered not by classroom teachers but by members of the research team. BAU typically included reliance on core literacy curricula approved by the state of Florida: *Treasures*, *Wonders*, *Open Court Imagine*, or *Journeys*. Instruction in each of these curricula focused on reading comprehension, strategies, discussion, vocabulary, writing, decoding, and spelling, and researchers deemed it unlikely that this instruction included any intensive focus on the same components as the interventions under investigation.

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

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

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

### Results

We summarize RCT results here for the most recent results available at the time of this writing for each of the CTT interventions. This approach is possible because FCRR analyzed all data by grade level and compared interventions only to BAU. There have been no direct comparisons among the interventions themselves in the work published thus far. Most of these results come from CE<sub>1</sub>, but some from free-standing RCTs.

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

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

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

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

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

**LIM.** RCT results are available for LIM in pre-K and grade 3. The analytic sample, measures, design, and analyses for the grade 3 version of LIM were derived from CE<sub>1</sub> (Connor et al., 2018). An additional early RCT conducted in pre-K (Phillips et al., 2016) involved 82 children randomized to either pull-out LIM instruction or BAU. Children were drawn from Title I public schools with pre-K programs where 77 percent or more of students received free or reduced-price meals. The racially and ethnically diverse sample was drawn from 10 classrooms in 5 schools. To qualify for intervention, students had to perform below the 35th percentile on either a spoken language syntax measure or a listening comprehension measure, or both. In addition to the screening measure, five measures were administered at pretest and posttest. Students completed three standardized measures of expressive and receptive language and two researcher-developed, intervention-aligned measures of listening comprehension and of language targeted by LIM. Some of these measures were also administered mid-intervention. As with the grade 3 trial, moderation effects based on pretest scores were also examined.

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

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

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

Target Constructs		Listening Comprehension			Vocabulary			Syntax		Intervention-aligned	
Grade	CELFL	OWLS	TNLS	Intervention-aligned	EOWPVT	CELFL	WJ-III	CASL	CELFL	Intervention-aligned	
Pre-K	NA	0.31	NA	<b>0.79</b>	NA	NA	-0.14	NA	0.08	<b>0.88</b>	
3	-0.14	0.08	0.04	NA	0.15	0.02	NA	0.09	NA	NA	
Additional Constructs											
Reading Comprehension				Word Recognition				Knowledge			
TOSREC	NA	GMRT	Comprehension Monitoring	Inconsistency Detection	WJ-III	TOWRE-SWE	TOWRE-PDE	WJ-III	NA	NA	
Pre-K	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
3	0.04	-0.10	-0.18	-0.18	-0.06	0.18	-0.02	0.16	0.16	0.16	

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



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

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

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

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

Grade	Reading Comprehension	Morphology				Word Recognition	
	TOSREC	Nonsense Morphemic Blending	Deriva- tional Awareness	Multi- morphemic Spelling	Affix Identifi- cation	TOWRE- SWE	TOWRE- PDE
K	NA	<b>1.26</b>	<b>0.82</b>	NA	NA	0.00	0.00
1	0.26	<b>0.67</b>	0.41	<b>0.82</b>	<b>2.54</b>	0.11	-0.39
2	0.14	<b>0.86</b>	<b>1.07</b>	-0.03	<b>1.52</b>	0.12	0.28

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

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

**TEXTS.** Teaching Expository Text Structures results (see Table 4-6) were available in grade 4 only (Connor et al., 2017) for CE<sub>1</sub>. Although no longer significant after correcting for multiple comparisons, two positive main effects were observed prior to correction. The first was for listening comprehension (on the Oral and Written Language Scales [OWLS]) and the second, which was only marginally significant to begin with, was for academic knowledge. TEXTS also demonstrated three significant moderation effects based on incoming student characteristics, only one of which remained significant after correcting for multiple comparisons. The latter effect was for academic knowledge such that TEXTS students with poorer academic knowledge at pretest outperformed BAU students at posttest. While students with average academic knowledge at pretest only marginally outperformed BAU students in the initial analysis, this effect was significant after the multiple comparisons correction. By contrast, an effect on listening comprehension for students with lower listening comprehension at pretest was not maintained after correcting for multiple comparisons.

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

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

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

**ERC.** Enacted Reading Comprehension was studied in both grades 3 and 4 (see Table 4-7) under the umbrella of CE<sub>1</sub> (Connor et al., 2018). ERC did not show significantly better performance at posttest compared to BAU for any outcome except one measure of expressive vocabulary, which demonstrated a small positive effect for ERC relative to BAU in grade 3. However, once a statistical correction was applied to control for error due to multiple comparisons, this effect was no longer significant. Moderator analyses for ERC based on student characteristics at pretest revealed significant effects for two additional outcomes in grade 3 and one additional outcome in grade 4. In grade 3, students with poorer expressive vocabulary at pretest (on the EOWPVT) showed positive, significant differences compared to BAU students on two measures of expressive vocabulary, the CELF-4 and the EOWPVT. However, once a statistical correction was applied to control for error due to multiple comparisons, only the effect on EOWPVT was still significant. In addition, students with average pretest expressive vocabulary (on the EOWPVT) also showed a significant positive effect on the EOWPVT, but this effect was also no longer significant after controlling for multiple comparisons.

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

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

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

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

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

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

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

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

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

Study	Applications		Academic Language	Morphology
	Narrative Writing	Editing		
RCT <sub>1</sub>	0.28	<b>0.69</b>	<b>0.44</b>	NA
RCT <sub>2</sub>	<b>0.21</b>	<b>1.48</b>	NA	<b>0.33</b>

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

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

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

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

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

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

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

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

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

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

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

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

Grade	Reading Comprehension		Listening Comprehension	Vocabulary	Knowledge	
	WJ-III	Reading-2-Comprehension	WJ-III	WJ-III	Social Studies	Science
K	NR		NR	NR		
1	NR		NR	NR		
2	NR		NR	NR		
3	NR	NR	NR	NR		
4	<b>0.22</b>	<b>NR</b>	<b>0.47</b>	<b>1.20</b>		
K-4					<b>2.27</b>	<b>2.10</b>

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

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

### Summary

The sheer volume of interventions developed and the complexity of some of the RCT designs make it difficult to do justice to FCRR work. Across the interventions, results were most positive for CALI, the one FCRR multicomponent intervention, and for the more targeted DAWS, LIM, and MAT interventions. Results were largely null for COMPASS, ERC, and TEXTS. However, in personal communications (C. Lonigan, personal communication, July 29, 2019), FCRR researchers shared the observation that CE<sub>1</sub> yielded positive results for LIM and COMPASS in pre-K through grade 2, as well as for a modified version of DR in pre-K. As a result, researchers explored combinations of pairs of three pre-K and kindergarten interventions: DR, LIM, and COMPASS in CE<sub>2</sub>. In addition, FCRR tested two different version of TEXTS in grade 4. More details on the results of both comparative efficacy studies will be forthcoming from FCRR.

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

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

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

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

### Catalyzing Comprehension Through Discussion and Debate

#### *Overview*

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

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



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

#### *Word Generation (WG)*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

#### *Strategic Adolescent Reading Intervention (STARI)*

STARI (Kim et al., 2016), though not entirely new, had been much less fully developed at the start of the CCDD project. STARI was designed as a multicomponent, Tier 2, small-group intervention for students identified as specifically struggling with reading. As a supplemental program, STARI focused instruction on a wide swath of students' requisite skills—word reading, fluency, vocabulary, and comprehension—all situated within a peer discussion framework designed to promote comprehension and engagement with reading. STARI used thematic units (e.g., “How can we find a place where we

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

—RfU Participating Teacher

really belong?”) that combined disciplinary learning with reading instruction. Instructional materials included a range of texts, from poems to autobiographies to first-person accounts of events, and novels or full-length works of nonfiction. Reading materials were chosen using two primary criteria: (1) relevance to unit themes, and (2) accessibility and cognitive challenge for target students. Researchers hypothesized that “challenging text characteristics would promote classroom talk about text and help move struggling readers beyond very literal and limited responses to text” (Kim et al., 2016, p. 366), with discussion serving as a learning opportunity (and motivating factor) for students. Given the reader profiles of participating

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

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

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

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

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

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

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

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

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

Grade	Reading Comprehension		Vocabulary	Sentence Processing	Morphology	Word Recognition
	Multiple Choice	Maze				
6-8	0.08	<b>0.21</b>	0.16	0.15	<b>0.18</b>	<b>0.20</b>

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

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

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

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

#### *Looking Across the Two CCDD Interventions*

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

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



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

### Promoting Adolescents' Comprehension of Text

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

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

Over the course of the 5-year RfU initiative, the work related to these interventions included design studies to devise, revise, and refine key instructional tools; pilot

We were committed to identifying feasible comprehension practices that content area reading teachers could integrate into their teaching routines that would both promote content learning and comprehension. We think that the PACT intervention practices are on the right path to promote both content learning and reading comprehension in secondary settings.

—Sharon Vaughn, Steering Committee Representative from PACT

studies and smaller-scale efficacy studies to evaluate the contribution of particular facets of comprehension instruction; and, most important to our synthesis, RCTs to assess the magnitude of the effects of these multicomponent interventions on the comprehension and knowledge acquisition of key demographic groups (e.g., a general population of learners, students with learning disabilities, and, for some but not others, ELs). We examine the results of each and then discuss patterns and distinctions among the three.

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

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

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

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

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

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

Study	Reading Comprehension		Knowledge
	GMRT	MASK	ASK
RCT <sub>1</sub>	<b>0.20</b>	<b>0.29</b>	<b>0.17</b>
RCT <sub>2</sub>	0.01	0.02	<b>0.32</b>
RCT <sub>3</sub>	0.12	<b>0.20</b>	<b>0.40</b>

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

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

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

**RCT<sub>1</sub>.** The first RCT (Vaughn et al., 2013) was the smallest scale ( $N = 416$ ), involving five teachers teaching three units to 27 (16 PACT) sections of grade 8 American history. For the three immediate outcomes, effect sizes favoring PACT over BAU were found for the three major outcomes: knowledge (ASK), intervention-aligned reading comprehension (MASK), and distal reading comprehension (the latent GMRT distal measure). Results for the follow-up content ASK measure indicated that a continued advantage for PACT was still present 4 weeks later.<sup>4</sup>

**RCT<sub>2</sub>.** In 2015, Vaughn and colleagues (2015) published the results of a much larger replication (19 teachers teaching 1,487 students in 85 sections, 47 of which implemented PACT) of the protocol used in the 2013 RCT<sub>1</sub>. The intervention-aligned measure of knowledge (ASK and its follow-up versions) revealed a reliable advantage for PACT over BAU immediately after the treatment, after 4 weeks, and after 8 weeks. Moreover, the effect of PACT was found to be fully mediated by implementation fidelity. However, neither the intervention-aligned MASK nor the distal GMRT measure of comprehension revealed significant differences between PACT and BAU, nor did implementation fidelity mediate observed differences. One can view the lack of a significant effect on

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

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

*RCT<sub>3</sub>*. A third RCT (Vaughn et al., 2017) involved 19 teachers teaching 94 sections, 49 of which were PACT. It focused intentionally on the performance of ELs ( $N = 1,629$ ) by sampling from schools and classes with sizable EL populations (ranging from 42 to 52 percent ELs). What differed from the previous PACT RCTs is that *RCT<sub>3</sub>* was supplemented with tools (e.g., Baker et al., 2014; Francis, Rivera, Lesaux, Keiffer, & Rivera, 2006) designed "to enhance the features of instruction and promote best practice for teaching ELs" (Vaughn et al., 2017, p. 24). In a departure from the previous RCTs, the researchers fit hierarchical linear models to their data, including not only a main effect for the PACT treatment, but also main and interaction effects for EL status and the percentage of EL students in a class. As a result, where interactions with the PACT intervention were significant, effects must be interpreted in light of those interactions.

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

In contrast to the ASK findings in the third RCT, ELs and non-ELs equally outperformed students in BAU classes on MASK regardless of the percentage of EL students in a class. As in the *RCT<sub>2</sub>*, the PACT effects did not generalize to the distal GMRT measure of reading comprehension. Thus, effects for the modified PACT intervention were significant for both intervention-aligned measures, with the effect on reading comprehension extending to all students and classes; by contrast, the effect on intervention-aligned content learning (ASK) depended on student EL status and the percentage of ELs in a class.

*Secondary analyses of the major PACT RCTs.* The PACT team conducted secondary analyses of data from these RCTs to tease out more complex accounts of the impact of PACT on specific populations of learners, most commonly students with learning disabilities (e.g., Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015; Wanzek, Swanson, Vaughn, Roberts, & Fall, 2016). In general, the pattern of results for the overall population was replicated in that effects favoring PACT on knowledge were stronger and more consistent than those on content-based or general reading comprehension. In the Swanson et al. (2015) reanalysis of the Vaughn et al. (2013, 2015) RCT<sub>1</sub> and RCT<sub>2</sub> data sets, PACT students with learning disabilities (LDs) outperformed BAU students with LDs on the intervention-aligned ASK content measure and the intervention-aligned MASK comprehension measure but not on the distal GMRT measure of comprehension. In the Wanzek et al. (2016) reanalysis of the Vaughn et al. (2017) RCT<sub>3</sub> with ELs, PACT students with LDs scored higher than their BAU counterparts on the ASK but not on either of the comprehension measures—MASK or GMRT. An additional analysis corroborated the fact that the effect of PACT on the outcome measures was similar for both EL and non-EL students with learning disabilities, which led the PACT team to conclude that the curriculum was both accessible to and beneficial for all students, including those who had been diagnosed with a learning disability and were coping with a curriculum presented in a second language.

The PACT team has also conducted a follow-up analysis to examine the moderating effects of other individual difference variables. Wanzek, Roberts, Vaughn, Swanson, and Sargent (2019) reexamined the data from the Vaughn et al. (2015) RCT<sub>2</sub> replication to determine whether the typical PACT effect on content acquisition and content-related comprehension was moderated by the incoming class mean scores on prior knowledge of American history or incoming general reading achievement (the GMRT). They found no hint of any interaction effects. Students in classes with higher or lower levels of knowledge or achievement benefited equally from PACT instruction.

### *Team-Based Learning*

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

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

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

### *Comprehension Circuit Training*

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

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

Study	Reading Comprehension		Knowledge	
	GMRT	ASK	ASK-Comprehension Items	ASK-Vocabulary Items
1	0.03	<b>0.19</b>	NA	NA
2	NA	0.50	0.38	<b>1.01</b>

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

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

Study	Grade	Reading Comprehension						Vocabulary	ORF	SWE
		Narrative	Expository	GMRT	Latent	TOSREC	STAAR			
RCT <sub>1</sub>	7–10	0.01	0.03	–0.01	NA	NA	NA	NA	NA	NA
RCT <sub>2</sub>	6–8	0.06	NA	0.16	NA	NA	NA	NA	NA	NA
RCT <sub>3</sub>	6–8	NA	NA	NA	<b>0.14</b>	<b>0.28</b>	0.10	<b>0.43</b>	–0.08	–0.04

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

**RCT<sub>1</sub>.** The first study, RCT<sub>1</sub>, conducted even before they had settled on the CCT moniker (Simmons et al., 2014), was more or less the proof of concept for the intervention, although no statistically significant main effects were found (see Table 4-14). A follow-up moderator analysis based on pretest performance yielded two small but provocative findings. First, when they compared students with GMRT scores below the 15th percentile with the rest of the sample, they found that the lower group made significantly more pre- to posttest progress on GMRT. On the Adolescent Literacy Inventory (ALI)-adapted passages, there were no differential effects attributable to pretest comprehension on the more narrative-like of the passages, but in a reversal of the GMRT findings, students who scored above the 15th percentile exhibited greater statistically reliable gains than those below the 15th percentile.

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

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

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

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

### *Summary*

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

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

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



Regarding language diversity, the results for PACT RCT<sub>3</sub> suggest that as language diversity in PACT classrooms increased, ELs' gains with PACT diminished, which the researchers hypothesize may be due to discourse patterns with a decreasing incidence of English academic language in these classrooms. As a result, PACT researchers suggest that for a discourse-based treatment like PACT to sustain positive impact on students' learning, additional supports are needed as the percentage of ELs increases. Results also indicate that PACT benefits students with LDs in similar ways to students without LDs, and neither prior class level of content knowledge nor reading achievement predicted responsiveness to the PACT intervention. With TBL work at the high school level, ASK pretest performance moderated ASK posttest scores, with the relative advantage over BAU students accruing to those who started with the most knowledge. For LD students, greater content learning growth attributable to TBL was evident for items focused on vocabulary compared to recognition of key content.

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

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

## **Reading, Evidence, and Argumentation in Disciplinary Instruction**

### *Overview*

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

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

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

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

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

—Susan Goldman, Steering Committee Representative from Project READI

### *The Development Process*

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

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

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

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

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

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

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

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

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

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

#### *The Randomized Controlled Trial*

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

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

---

<sup>5</sup> Available through the Project READI case library at <https://www.projectreadi.org/case-library>.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

	Reading Comprehension		Application: Evidence-Based Argumentation Essay	
	GISA	Multiple-Choice Evidence-Based Argumentation	Concepts	Connections
ES	<b>0.32</b>	<b>0.26</b>	0.11	0.08

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



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

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

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

Survey						
	CCSS	Attitude	Self-efficacy	Teaching Philosophy	Science Reading	Higher-Order Teaching
ES	0.45	0.53	0.41	0.46	<b>1.36</b>	<b>2.21</b>
Practices						
	Argumentation Practices	Content Reading	Metacognitive Modeling	Metacognitive Practice	Negotiation Instruction	
ES	<b>1.73</b>	<b>1.60</b>	<b>1.34</b>	<b>2.24</b>	<b>1.89</b>	
Higher-Order Teaching						
	Opportunities	Support	Inquiry	Strategies	Argumentation	Collaboration
ES	<b>1.28</b>	<b>1.49</b>	<b>1.09</b>	<b>1.37</b>	<b>1.07</b>	<b>0.83</b>

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

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

### *Summary*

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

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

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

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

### Looking Across the Array

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

### REFERENCES

- Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system*. Eugene, OR: Center for Educational Assessment Accountability.
- Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education, National Institute of Education.
- Apel, K., & Diehm, E. (2013). Morphological awareness intervention with kindergarteners and first and second grade students from low SES homes: A small efficacy study. *Journal of Learning Disabilities, 47*, 65–75.
- Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from [http://ies.ed.gov/ncee/wwc/publications\\_reviews.aspx](http://ies.ed.gov/ncee/wwc/publications_reviews.aspx).
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289–300.
- Blaum, D., Griffin, T. D., Wiley, J., & Britt, M. A. (2017). Thinking about global warming: The effect of policy-related documents and prompts on learning about causes of climate change. *Discourse Processes, 54*, 303–316.

- Brozo, W., & Afflerbach, P. (2011). *Adolescent Literacy Inventory, Grades 6–12*. New York: Pearson.
- Burkett, C., & Goldman, S. R. (2016). "Getting the point" of literature: Relationships between processing and interpretation. *Discourse Processes*, 53(5–6), 457–487.
- Capin, P., & Vaughn, S. (2017). Improving reading and social studies learning for secondary students with reading disabilities. *TEACHING Exceptional Children*, 49(4), 249–261.
- Chinn, C., & Sandoval, W. (2018). Epistemic cognition and epistemic development. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 24–33). New York: Routledge.
- Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, 38(1), 35–70.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge E-Book. *Computers & Education*, 128, 284–311.
- Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zaegar, E., Wood, T., & Petscher, Y. (2017). Acquiring science and social studies knowledge in kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of content-area literacy instruction (CALI). *Journal of Educational Psychology*, 109(3), 301–320.
- Connor, C. M., Phillips, B. M., Kaschak, M., Apel, K., Kim, Y. S., Al Otaiba, S., Crowe, E. C., Thomas-Tate, S., Johnson, L. C., & Lonigan, C. J. (2014). Comprehension tools for teachers: Reading for understanding from prekindergarten through fourth grade. *Educational Psychology Review*, 26(3), 379–401. doi: 10.1007/s10648-014-9267-1.
- Connor, C. M., Phillips, B. M., Kim, Y. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & Otaiba, S. A. (2018). Examining the efficacy of targeted component interventions on language and literacy for third and fourth graders who are at risk of comprehension difficulties. *Scientific Studies of Reading*, 22(6), 462–484.
- Cribb, G., Maglio, C., & Greenleaf, C. (2018). Collaborative argumentation: 10th graders read modern Iranian history. *The History Teacher*, 51, 477–526.
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99(2), 311–325.
- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34, 3–14.
- Duhaylongsod, L., Snow, C. E., Selman, R. L., & Donovan, M. S. (2015). Toward disciplinary literacy: Dilemmas and challenges in designing history curriculum to support middle school students. *Harvard Educational Review*, 85(4), 587–608.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10, 323–330. doi: 10.1207/s1532799xssr1003\_7.
- Fogarty, M., Clemens, N., Simmons, D., Anderson, L., Davis, J., Smith, A., Wang, H., Kwok, O., Simmons, L. E., & Oslund, E. (2017). Impact of a technology-mediated reading intervention on adolescents' reading comprehension. *Journal of Research on Educational Effectiveness*, 10(2), 326–353.
- Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N. H., & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review*, 26, 425–449.
- Francis, D. J., Rivera, M., Lesaux, N., Keiffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Portsmouth, NH: Center on Instruction. Retrieved from <http://www.centeroninstruction.org/files/ELL1-Interventions.pdf>.
- Freebody, P., & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: An Australian Journal of TESOL*, 5(7), 7–16.
- Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, 25, 99–125.
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japunitich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Research*, 96(3), 424–436.

- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children*, 22(2), 89–116.
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Shanahan, C., & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246.
- Goldman, S. R., Greenleaf, C., Yukhymenko-Lescroart, M., Brown, W., Ko, M. L. M., Emig, J. M., George, M., Wallace, P., Blaum, D., & Britt, M. A. (2019). Explanatory modeling in science through text-based investigation: Testing the efficacy of the Project READI intervention approach. *American Educational Research Journal*, 56(4), 1148–1216.
- Goldman, S. R., McCarthy, K., & Burkett, C. (2015). Interpretative inferences in literature. In E. O'Brien, A. Cook, & R. Lorch (Eds.), *Inferences during reading* (pp. 386–415). Boston: Cambridge University Press.
- Goldman, S., & Snow, C. E. (2015). Adolescent literacy: Development and instruction. In A. Pollatsek & R. Treiman (Eds.), *Handbook on reading* (pp. 463–478). New York: Oxford University Press.
- Goldman, S. R., Snow, C., & Vaughn, S. (2016). Common themes in teaching reading for understanding: Lessons from three projects. *Journal of Adolescent & Adult Literacy*, 60(3), 255–264. doi: 10.1002/jaal.586.
- Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., Schneider, S. A., Madden, S., & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, 48(3), 647–717.
- Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, 15(1), 75–90.
- Henbest, V. S., & Apel, K. (2017). Effective word reading instruction: What does the evidence tell us? *Communication Disorders Quarterly*, 39(1), 303–311.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177.
- Hsin, L. B., Phillips Galloway, E., & Snow, C. E. (n.d.). *Social strengths amid linguistic limitations in the argumentative writing of English-language learners*. Unpublished manuscript.
- Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism*, 19(3), 314–332.
- IES (Institute of Education Sciences). (2009). *Request for applications: Reading for Understanding research initiative, CFDA Number: 84.305F*. Retrieved from [https://ies.ed.gov/funding/pdf/2010\\_84305F.pdf](https://ies.ed.gov/funding/pdf/2010_84305F.pdf).
- Johanson, M., & Arthur, A. M. (2016). Improving the language skills of pre-kindergarten students: Preliminary impacts of the Let's Know! experimental curriculum. *Child & Youth Care Forum*, 45(3), 367–392.
- Johnson, L., Terry, N. P., Connor, C. M., & Thomas-Tate, S. (2017). The effects of dialect awareness instruction on nonmainstream American English speakers. *Reading and Writing*, 30(9), 2009–2038.
- Jones, S. M., LaRusso, M., Kim, J., Yeon Kim, H., Selman, R., Uccelli, P., Barnes, S. P., Donovan, S., & Snow, C. (2019). Experimental effects of Word Generation on vocabulary, academic language, perspective taking, and reading comprehension in high-poverty schools. *Journal of Research on Educational Effectiveness*, 12(3), 448–483.
- Kent, S., Wanzek, J., Swanson, E. A., & Vaughn, S. (2015). Team-based learning for students with high-incidence disabilities in high school social studies classrooms. *Learning Disabilities Research & Practice*, 30(1), 3–14.
- Kim, H. Y., Hsin, L. B., & Snow, C. E. (2018). Reducing academic inequalities for English language learners: Variation in experimental effects of Word Generation in high-poverty schools. *International Journal of Bilingual Education and Bilingualism*, 1–19. doi: 10.1080/13670050.2018.1535574.
- Kim, H. Y., LaRusso, M. D., Hsin, L. B., Harbaugh, A. G., Selman, R. L., & Snow, C. E. (2018). Social perspective-taking performance: Construct, measurement, and relations with academic performance and engagement. *Journal of Applied Developmental Psychology*, 57, 24–41.
- Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2016). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly*, 52(3), 357–382.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

- LARRC (Language and Reading Research Consortium). (2016). Use of the Curriculum Research Framework (CRF) for developing a reading-comprehension curricular supplement for the primary grades. *The Elementary School Journal*, 116(3), 459–486.
- LARRC, Arthur, A. M., & Davis, D. L. (2016). A pilot study of the impact of double-dose robust vocabulary instruction on children's vocabulary growth. *Journal of Research on Educational Effectiveness*, 9(2), 173–200.
- LARRC, Jiang, H., & Davis, D. (2017). Let's Know! proximal impacts on prekindergarten through grade 3 students' comprehension-related skills. *The Elementary School Journal*, 118(2), 177–206.
- LARRC, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*, 62, 2812–2828.
- LARCC, Johanson, M., & Arthur, A. M. (2016). Improving the language skills of pre-kindergarten students: Preliminary impacts of the Let's Know! experimental curriculum. *Child Youth Care Forum*, 45, 367–392. doi: 10.1007/s10566-015-9332-z.
- LARRC, Pratt, A., & Logan, J. (2014). Improving language-focused comprehension instruction in primary-grade classrooms: Impacts of the "Let's Know!" experimental curriculum. *Educational Psychology Review*, 26(3), 357–377. doi: 10.1007/s10648-014-9275-1.
- LaRusso, M. D., Donovan, S., & Snow, C. (2016). Implementation challenges for tier one and tier two school-based programs for early adolescents. *New Directions for Child and Adolescent Development*, 154, 11–30.
- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, C., Donovan, S., & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, 9(2), 201–222.
- Lawrence, J., Capotosto, L., Branum-Martin, L., & Snow, C. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition*, 15, 437–451.
- Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, 52(4), 750–786.
- Lawrence, J. F., Rolland, R. G., Branum-Martin, L., & Snow, C. E. (2014). Generating vocabulary knowledge for at-risk middle school readers: Contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk*, 19(2), 76–97.
- Lee, C. D. (2007). *Culture, literacy and learning: Taking blooming in the midst of the whirlwind*. New York: Teachers College Press.
- Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York: Carnegie Corporation.
- Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87–113). London: Routledge.
- Leslie, L., & Caldwell, J. (2011). *Qualitative Reading Inventory, 5th edition*. New York: Pearson.
- Levine, S., & Horton, W. (2015). Helping high school students read like experts: Affective evaluation, salience, and literary interpretation. *Cognition and Instruction*, 33(2), 125–153.
- Litman, C., & Greenleaf, C. (2018). Argumentation tasks in secondary English language arts, history, and science: Variations in instructional focus and inquiry space. *Reading Research Quarterly*, 53, 107–126.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests—directions for administration (levels 7/9 & 10/12)* (4th ed.). Rolling Meadows, IL: Riverside.
- MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2006). *Gates-MacGinitie Reading Tests*. Itasca, IL: Riverside.
- Martin, N., & Brownell, R. (2010). *EOWPVT-4: Expressive One-Word Picture Vocabulary Test—Fourth Edition*. Indianapolis, IN: Pro-Ed.
- McCarthy, K. S., & Goldman, S. R. (2015). Comprehension of short stories: Effects of task instructions on literary interpretation. *Discourse Processes*, 52, 585–608.
- Miller, J. F., & Chapman, R. S. (2008). *Systematic analysis of language transcripts (computer software)*. Madison: University of Wisconsin–Madison, Waisman Center, Language Analysis Laboratory.
- NAEP (National Assessment of Educational Progress). (2017). *Reading framework for the 2017 National Assessment of Educational Progress*. National Assessment Governing Board, U.S. Department of Education.

- NAEP. (2019a). *NAEP reading stat facts*. Retrieved from <https://nces.ed.gov/nationsreportcard/reading>.
- NAEP. (2019b). *NAEP report card: Reading*. Retrieved from <https://www.nationsreportcard.gov/reading/?grade=4>.
- NELP (National Early Literacy Panel). (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Retrieved from <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>.
- NGA & CCSSO (National Governors Association Center for Best Practices and Council of Chief State School Officers). (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Author.
- NICHHD (National Institute of Child Health and Human Development). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- NRC (National Research Council). (2014). *Developing assessments for the next generation science standards*. Washington, DC: The National Academies Press.
- Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education*, 32(2), 203–218.
- Park, M., Anderson, E., & Yoon, S. (2017). Learning biology coherently through complex systems: A model supported with scientific practices and agent-based simulations. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Proceedings of the 12th International Conference for Computer Supported Collaborative Learning* (pp. 199–206). Philadelphia: International Society of the Learning Sciences.
- Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Science and literacy: Each in the service of the other. *Science*, 328, 459–463.
- Petersen, D. B., & Spencer, T. D. (2012). The Narrative Language Measures: Tools for language screening, progress monitoring, and intervention planning. *Perspectives on Language Learning and Education*, 19(4), 119–129.
- Phillips, B. M., Tabulda, G., Ingrole, S. A., Burriss, P. W., Sedgwick, T. K., & Chen, S. (2016). Literate language intervention with high-need prekindergarten children: A randomized trial. *Journal of Speech, Language, and Hearing Research*, 59(6), 1409–1420.
- Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System [CLASS] manual*. Baltimore: Brookes.
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. M. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 23(4), 431–451.
- PISA (Programme for International Student Assessment). (2018). *PISA 2018 results*. Retrieved from <https://www.oecd.org/pisa/publications/pisa-2018-results.htm>.
- Pratt, A., & Logan, J. (2014). Improving language-focused comprehension instruction in primary-grade classrooms: Impacts of the Let's Know! experimental curriculum. *Educational Psychology Review*, 26(3), 357–377.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum Associates.
- Purcell-Gates, V., Perry, K., & Briseño, A. (2011). Analyzing literacy practice: Grounded theory to model. *Research in the Teaching of English*, 45, 439–458.
- Schoenbach, R., & Greenleaf, C. (2009). Fostering adolescents' engaged academic literacy. In L. Christenburg, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 98–112). New York: Guilford Press.
- Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). *Reading for understanding: How Reading Apprenticeship improves disciplinary learning in secondary and college classrooms* (2nd ed.). San Francisco: Jossey-Bass.
- Schoenbach, R., Greenleaf, C., & Murphy, L. (2016). *Leading for literacy: A Reading Apprenticeship approach*. San Francisco: Wiley Press.
- Seymour, H. N., Roeper, T. W., & deVilliers, J. (2003). *Diagnostic evaluation of language variation, screening test examiner's manual*. San Antonio, TX: The Psychological Corporation.
- Shanahan, C., Heppler, J., Manderino, M., Bolz, M., Cribb, G., & Goldman, S. R. (2016). Deepening what it means to read (and write) like a historian: Progressions of instruction across a school year in an eleventh grade U.S. history class. *The History Teacher*, 49, 241–270.

- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.
- Simmons, D., Fogarty, M., Oslund, E. L., Simmons, L., Hairrell, A., Davis, J., Anderson, L., Clemens, N., Vaughn, S., Roberts, G., Stillman, S., & Fall, A. M. (2014). Integrating content knowledge-building and student-regulated comprehension practices in secondary English language arts classes. *Journal of Research on Educational Effectiveness, 7*(4), 309–330.
- Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325–344.
- Sosa, T., Hall, A. H., Goldman, S. R., & Lee, C. D. (2016). Developing symbolic interpretation through literary argumentation. *Journal of the Learning Sciences, 25*(1), 93–132.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.
- Swanson, E., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children, 81*(4), 426–442.
- Texas Education Agency. (2013). *State of Texas assessments of academic readiness*. Austin, TX: Texas Education Agency.
- Torgesen, J., Wagner, R., & Rashotte, C. (2010). *Test of word reading efficiency, 2nd ed.* New York: Pearson.
- Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *TOWRE-2 test of word reading efficiency, 2nd ed.* San Antonio, TX: Pearson.
- Troyer, M. (2017). Teachers' adaptations to and orientations toward an adolescent literacy curriculum. *Journal of Curriculum Studies, 51*(2), 202–228. doi: 10.1080/00220272.2017.1407458.
- Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A.-M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology, 109*(1), 22–34.
- Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A. M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review, 27*(1), 31–50.
- Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly, 48*(1), 77–93.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of silent reading efficiency and comprehension*. Austin, TX: Pro-Ed.
- Wanzek, J., Roberts, G., Vaughn, S., Swanson, E., & Sargent, K. (2019). Examining the role of pre-instruction academic performance within a text-based approach to improving student content knowledge and understanding. *Exceptional Children, 85*(2), 212–228.
- Wanzek, J., Swanson, E., Vaughn, S., Roberts, G., & Fall, A.-M. (2016). English learner and non-English learner students with disabilities: Content acquisition and comprehension. *Exceptional Children, 82*, 428–442.
- Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., Fall, A.-M., Stillman-Spisak, S. J., & Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness, 7*(2), 183–204. doi: 10.1080/19345747.2013.836765.
- Wiederholt, J., & Bryant, B. (2012). *Gray oral reading tests-Fifth edition (GORT-5)*. Austin, TX: Pro-Ed.
- Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating the effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools, 45*, 432–445.
- Wiley, J., Jaeger, A. J., & Griffin, T. D. (2018). Effects of task instructions on comprehension of multiple documents in history and science. In J. L. G. Braasch, I. Braten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 341–361). New York: Routledge.
- Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. American Guidance Service. New York: Pearson.
- Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia: Temple University Press.



- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of achievement*. Rolling Meadows, IL: Riverside.
- Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly, 52*(2), 203–226.

## Appendix 4-1

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

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

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

TABLE APPENDIX 4-1 Continued

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

TABLE APPENDIX 4-1 Continued

Construct	Assessment	Description	RfU Approach
Vocabulary (continued)	Expressive One Word Picture Vocabulary Test, 4th Edition (EOWPVT)	Norm-referenced, individually administered, ungraded, untimed test in which students name objects, actions, and concepts based on illustrations in increasing order of difficulty	COMPASS ERC LIM TEXTS
	RISE vocabulary subtest	See Chapter 3	STARI
	WJ-III expressive vocabulary subtest	Norm-referenced, individually administered, ungraded, untimed test in which students name pictured objects	LIM
Syntax	Clinical Evaluation of Language Fundamentals Preschool, 2nd Edition (CELF2) sentence structure subtest	Norm-referenced, individually administered, ungraded, untimed test in which students point to a picture that matches verbal prompts	LIM
	Comprehensive Assessment of Spoken Language (CASL) syntax construction subtest	Norm-referenced, individually administered, grade-leveled, untimed test in which students respond orally to a verbal prompt and picture with a grammatically and semantically appropriate word, phrase, or sentence	COMPASS ERC TEXTS
Morphology	RISE morphology subtest	See Chapter 3	STARI
Word recognition	easyCBM passage reading fluency subtest	Norm-referenced, individually administered, grade-leveled, timed test in which students read a passage aloud for 1 minute and are scored based on the number of words read aloud correctly	CCT
	RISE word recognition and decoding subtest	See Chapter 3	STARI
	Test of Word Reading Efficiency-2nd Edition (TOWRE2) phonemic decoding efficiency subtest	Norm-referenced, individually administered, ungraded, timed test in which students read nonsense words listed in order of increasing difficulty, reading as many as possible in 45 seconds	COMPASS ERC LIM TEXTS
	TOWRE2 sight word efficiency subtest	Norm-referenced, individually administered, ungraded, timed test in which students read real words listed in order of increasing difficulty, reading as many as possible in 45 seconds	CCT COMPASS ERC LIM TEXTS MAT
	WJ-III letter word identification subtest	Norm-referenced, individually administered, ungraded, untimed test in which students read letters and words in increasing order of difficulty	COMPASS ERC LIM MAT TEXTS



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

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

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

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

*continued*



TABLE APPENDIX 4-2 Continued

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

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



# 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 .....	216
PEDAGOGICAL EFFECTS ACROSS THE RFU PORTFOLIO .....	216
Judging the Magnitude of Effects.....	216
Our Decision .....	217
Effect Size Patterns .....	217
Measures of Comprehension and Beyond .....	221
Moderating and Mediating Effects.....	223
Moderation Across the Entire RfU Portfolio.....	226
Mediation Across the Entire RfU Portfolio .....	227
Moving the Needle on Reading Comprehension.....	227
EXAMINING THE PEDAGOGICAL FEATURES OF THE RFU PORTFOLIO .....	228
The Relationships Among Enabling Skills, Knowledge, Language, and Reading Comprehension Are Dynamic and Synergistic .....	228
Many Kinds of Knowledge Play a Role in Reading Comprehension.....	230
Learning to Read and Reading to Learn Are Better Regarded as Complementary Processes Than Separate Stages of Development .....	233
Comprehension Can Be Conceptualized as a Waypoint, or the End State, of a Journey.....	235
Metacognitive Processes Play a Role in the Comprehension Instruction Repertoire.....	236
Collaboration Is Often a Key Element of Effective Interventions .....	237
Engagement with Texts and Tasks Supports Comprehension.....	240
LESSONS LEARNED .....	241
A CODA FOR THE PEDAGOGICAL PORTFOLIO OF THE RFU.....	247
REFERENCES.....	247

## 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			<b>DAWS</b> ••••	<b>DAWS</b> ••••	<b>DAWS</b> ••••	<b>WG</b> •	<b>WG</b> •	<b>WG</b> •	<b>WG</b> •	<b>READI</b> •
Knowledge and learning	P			<b>CALI</b> ••••	<b>CALI</b> ••••	<b>CALI</b> ••••	<b>CALI</b> ••••			<b>PACT</b> ••	<b>TBL</b> •
	R			<b>CALI</b> ••••	<b>CALI</b> ••••	<b>CALI</b> ••••	<b>CALI</b> ••••			<b>PACT</b> ••	<b>TBL</b> •
	P			<b>COMPASS</b> <sup>n</sup>	<b>ERC</b> •	<b>TEXTS</b> ••	<b>LIM</b> •				
Reading comprehension	R							<b>CCT</b> <sup>n</sup>	<b>CCT</b> <sup>n</sup>	<b>CCT</b> <sup>n</sup>	<b>READI</b> ••
	P			<b>MAT</b> ••	<b>MAT</b> •	<b>COMPASS</b> <sup>n</sup>	<b>ERC</b> <sup>n</sup>	<b>WG</b> •	<b>CCT</b> ••	<b>CCT</b> ••	<b>READI</b> ••
	R			<b>MAT</b> ••	<b>MAT</b> •	<b>COMPASS</b> <sup>n</sup>	<b>ERC</b> <sup>n</sup>	<b>WG</b> •	<b>CCT</b> ••	<b>CCT</b> ••	<b>READI</b> ••
	P			<b>MAT</b> ••	<b>MAT</b> •	<b>COMPASS</b> <sup>n</sup>	<b>ERC</b> <sup>n</sup>	<b>WG</b> •	<b>STARI</b> ••	<b>STARI</b> ••	<b>READI</b> ••
	R			<b>LIM</b> ••	<b>LK</b> <sup>B,n</sup>	<b>LK</b> <sup>B,••</sup>	<b>LK</b> <sup>B,••</sup>		<b>WG</b> •	<b>WG</b> •	<b>TBL</b> <sup>n</sup>
	P			<b>LK</b> <sup>D,••••</sup>	<b>LK</b> <sup>D,n</sup>	<b>LK</b> <sup>D,••••</sup>	<b>LK</b> <sup>D,••••</sup>		<b>WG</b> •	<b>WG</b> •	<b>TBL</b> <sup>n</sup>
	R			<b>LK</b> <sup>D,••••</sup>	<b>LK</b> <sup>n</sup>	<b>LK</b> <sup>n</sup>	<b>LK</b> <sup>n</sup>		<b>WG</b> •	<b>WG</b> •	<b>TBL</b> <sup>n</sup>
	P			<b>LIM</b> ••	<b>COMPASS</b> •	<b>CALI</b> ••	<b>ERC</b> <sup>n</sup>	<b>TEXTS</b> ••			
	R			<b>LIM</b> ••	<b>COMPASS</b> •	<b>CALI</b> ••	<b>ERC</b> <sup>n</sup>	<b>TEXTS</b> ••			
	P			<b>LIM</b> ••	<b>COMPASS</b> •	<b>CALI</b> ••	<b>ERC</b> <sup>n</sup>	<b>TEXTS</b> ••			

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; <sup>n</sup> = 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 <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LIM <sup>n</sup> LK <sup>B•••••</sup> LK <sup>D•••••</sup> LK <sup>•••••</sup>	ERC <sup>n</sup> TEXTS <sup>n</sup>					
	P										
	R				DAWS <sup>••</sup>	DAWS <sup>••</sup> WG <sup>n</sup>	DAWS <sup>••</sup> WG <sup>n</sup>	WG <sup>n</sup>	WG <sup>n</sup>	WG <sup>n</sup>	
Vocabulary	R	LK <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LK <sup>D•••••</sup>	LK <sup>B•••••</sup> LK <sup>D•••••</sup> LK <sup>•••••</sup>	WG <sup>••</sup>	WG <sup>••</sup>	CCT <sup>••</sup> WG <sup>•</sup>	CCT <sup>••</sup> WG <sup>•</sup>	CCT <sup>••</sup>	
	P										
	R	LIM <sup>n</sup>			COMPASS <sup>•</sup> ERC <sup>••</sup> LIM <sup>•</sup>	CALI <sup>•••••</sup> ERC <sup>•</sup> TEXTS <sup>•</sup>		STARI <sup>•</sup>	STARI <sup>•</sup>	STARI <sup>•</sup>	STARI <sup>•</sup>
Syntax	R	LIM <sup>•••••</sup>									
	P				COMPASS <sup>•</sup> ERC <sup>•</sup> LIM <sup>•</sup>						
Morphology	R				DAWS <sup>••</sup> MAT <sup>•••••</sup>	DAWS <sup>••</sup>					
	P									STARI <sup>•</sup>	STARI <sup>•</sup>
Word recognition	R										
	P				MAT <sup>n</sup>	MAT <sup>•</sup>	MAT <sup>••</sup>	COMPASS <sup>n</sup> ERC <sup>n</sup> LIM <sup>•</sup>	ERC <sup>•</sup> TEXTS <sup>n</sup>	CCT <sup>n</sup> STARI <sup>••</sup>	CCT <sup>n</sup> STARI <sup>••</sup>
	P										

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; <sup>n</sup> = 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<sup>1</sup> (•••) following the acronym for the intervention. Thus, the notation of LK<sup>D••••</sup> 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,

<sup>1</sup> 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 multi-component 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.<sup>2</sup> 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

---

<sup>2</sup> WG and STARI results were positively mediated by levels of student engagement, and PACT outcomes in RCT<sub>3</sub> 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<sub>3</sub> 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.

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

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

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<sup>D</sup> and LK<sup>B</sup>) 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 CALL, 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.

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

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

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 engage-

ment. 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<sub>3</sub>; 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<sub>3</sub>. 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 ( $RC = DEC \times LC$ ; 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 instruction* (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](http://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.
- Sanson, 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](https://nces.ed.gov/nationsreportcard/blog/mathematics_reading_2019.aspx).

# Taking Stock of the Reading for Understanding Initiative

P. David Pearson, *University of California, Berkeley*  
 Annemarie Sullivan Palincsar, *University of Michigan*  
 Peter Afflerbach, *University of Maryland*  
 Gina N. Cervetti, *University of Michigan*  
 Panayiota Kendeou, *University of Minnesota*  
 Gina Biancarosa, *University of Oregon*  
 Jennifer Higgs, *University of California, Davis*  
 Miranda Fitzgerald, *University of North Carolina at Charlotte*  
 Amy I. Berman, *National Academy of Education*

## CONTENTS

INTRODUCTION .....	252
QUESTION 1: SUMMARIZING THE CONTRIBUTIONS OF THE RFU INITIATIVE .....	252
Nature and Development of Comprehension .....	252
Assessment.....	253
Curriculum and Instruction.....	255
QUESTION 2: ADDITIONAL LESSONS LEARNED FROM THE RFU WORK ...	256
The Affordances of the RfU Consortium Model .....	256
Teacher Professional Learning .....	258
Insights About “Moving the Needle” .....	262
Where Does the Analysis Leave Us? .....	267
QUESTION 3: BENCHMARKS FOR GAUGING PROGRESS OF THE RFU INITIATIVE .....	268
The Simple View of Reading .....	268
The RAND Reading Study Group .....	271
Adolescent and Disciplinary Literacy .....	273
QUESTION 4: WHAT MIGHT THE RFA FOR RFU 2.0 LOOK LIKE? .....	275
Research Priorities for Nature and Development, Assessment, and Curriculum and Instruction .....	275
Overarching Research Priorities for Reading Comprehension .....	277
THE GRAND QUESTION .....	285
REFERENCES.....	287



## INTRODUCTION

In this chapter, we address the central issue of this volume, “What rewards did we reap from the substantial investment made by the Institute of Education Sciences in this focused effort to better understand and improve reading comprehension on the pathway to improved student achievement?” We try to answer this grand question by deconstructing it into more specific questions that, taken together, move us toward a grand answer:

1. Looking across the work of the six teams within the three domains detailed in the earlier chapters of this volume (the strands of nature and development, assessment, and curriculum and instruction), what findings and common features of the work stand out as markers of progress?
2. Recognizing that not all of the salutary features of the Reading for Understanding (RfU) initiative fit neatly into our three-strand structure, what other noteworthy outcomes of the initiative contribute to the narrative about what was learned?
3. Using as benchmarks the contextual issues and movements that were influential when this work began almost 10 years ago, how do we think differently about reading comprehension as a result of the RfU?
4. What legacy, in terms of an agenda for future reading comprehension research, has the RfU left for the field?

Our approach to answering the grand question and its four facets is straightforward. First, we look across Chapters 2–5 to offer a concise summary of what we have already covered in greater detail. Second, we add a section on additional lessons learned, mainly about the affordances of the process employed in carrying out this work. Third, we look back at the influential contextual factors (policies, theories about reading development, or pedagogical movements) we introduced in Chapter 1 and use them as benchmarks for assessing how we think differently about reading comprehension a decade after the RfU began. Finally, we look beyond the RfU to assess its legacy in terms of future work by responding to “What would the Request for Application (RfA) for the next RfU look like?”

### QUESTION 1: SUMMARIZING THE CONTRIBUTIONS OF THE RFU INITIATIVE

#### Nature and Development of Comprehension

With respect to the *nature and development of comprehension*, the RfU studies revealed the number and range of skills, knowledge, and dispositional characteristics that support reading comprehension, as well as the relative importance of these skills and knowledge as students matriculate through the grades. These studies informed the field’s understanding of the *linguistic* and *cognitive* skills associated with successful listening and reading comprehension and called our attention to the role of both *word* and *world* knowledge in comprehension activity.

At the preschool and elementary levels, linguistic, cognitive, and behavioral skills were particularly significant predictors of comprehension, while at the adolescent level,

other reader-specific factors (e.g., background knowledge, vocabulary knowledge, discourse expertise, strategy use, and inferencing) emerged as explanatory factors regarding performance and development. The factors included in the adolescent portfolio are largely discipline-specific language and reasoning skills: academic language (both vocabulary and discourse) that characterizes advanced discussion about text and language, as well as epistemological (how we know what we know and do not know) and perspective-taking dispositions (how we learn to decenter in examining text-based narratives, arguments, and explanations).

Specific to language skills, the evidence from the developmental studies suggested that language might most productively be regarded as a single entity, or perhaps as a cluster or assemblage of closely related skills, thus calling into question both assessments and interventions that privilege discrete language skills. Specific to cognitive skills, the RfU development studies revealed that some skills, namely, attentional control and self-regulation, made small but significant contributions to comprehension, while comprehension monitoring and inferencing made more substantial contributions to comprehension, especially in distinguishing between stronger and weaker comprehension. The RfU studies, while confirming the role of word and world knowledge in listening and reading comprehension, extended our understanding of this relationship by illustrating the role that word and world knowledge play in supporting readers' ability to engage in inferencing and comprehension monitoring.

Finally, research on the nature and development of comprehension raised new questions for future research regarding the significance and malleability of different knowledge sources and skills at different points in development and in relation to different text genres and text characteristics. These questions constitute one of the research legacies of the RfU initiative.

### Assessment

The RfU initiative had a profound influence on the development and validation of *reading comprehension assessments*, giving rise to a new generation of relevant measures of the construct.

First, we learned that authenticity, complexity, and psychometric adequacy can all be achieved, even in a single assessment. The Global Integrated Scenario-Based Assessment (GISA) system demonstrated the feasibility of assessing the broad, multidimensional comprehension construct in an authentic way while also achieving technical adequacy.

Second, assessment does not have to be distinct from learning; in fact, the approach to scenario-based assessment utilized in GISA entails learning as a fundamental premise of the assessment process.

Third, to reflect and capture the dynamic nature of reading comprehension, we need multiple assessment systems that vary in construct coverage and tasks. Within the RfU portfolio, both the Reading Inventory and Scholastic Evaluation (RISE) and Florida Center for Reading Research (FCRR) Reading Assessment (FRA) assessment systems ensure broad coverage of the components that are either a part of reading comprehension or on the developmental pathway to it. GISA, by contrast, represents the orchestration of reading comprehension and other variables in using the fruits of reading comprehension to perform related comprehension tasks.

Fourth, knowledge is an integral component of reading comprehension, and as such it should be integrated, not simply treated as a nuisance variable and controlled. GISA has integrated knowledge in the design of the assessment as an integral component and provided evidence for the feasibility and efficacy of this approach.

Fifth, looking across the entire range of the RfU assessments, including those developed for particular studies by the RfU teams, the consortium addressed not only prior knowledge, but also metacognitive and self-regulatory strategies, reading strategies, and motivation and engagement. The integration of these variables represented a significant advance in comprehension assessment design and provided a set of tools sensitive enough to inform and evaluate the effects of high-quality instruction. A number of these assessments were theoretically robust and reflected a reconceptualization of comprehension consistent with advances in theory and research as well as numerous national and international standards and movements. The stage is set for future work in which these tools can replace traditional standardized reading comprehension assessments. Furthermore, these assessments can be investigated for their potential to support new approaches to curriculum and pedagogy, particularly those that privilege differentiated instruction and the application of text-supported knowledge to new problems and situations.

Sixth, the connection between assessment and instruction is underscored by an inherent reciprocity that has been realized by the full scope of assessment efforts of the RfU. This reciprocity is identified by two important characteristics of the assessment itself, instructional value and instructional sensitivity. Instructional value refers to an assessment's capacity to provide information about a student's strengths and weaknesses in particular skills or processes that might become candidates for instruction. Instructional sensitivity refers to an assessment's capacity to reflect the effects of instruction or intervention; this is a key attribute in pedagogical research designed to promote particular comprehension processes. Because the RfU assessments acknowledged and reflected a broader and more authentic conceptualization of reading comprehension that emphasized instructional value and sensitivity, significant progress has been made in this area. Specifically, interventions that moved the needle on GISA, the broader index of reading comprehension, were primarily multicomponent, suggesting good coverage between the aspects of the construct being assessed and those being trained. Even though these effects were generally small, they underscore the importance of reflecting the multicomponent nature of reading comprehension, including the capacity to orchestrate those components to produce new knowledge or learning, in a summative assessment. Similarly, reflecting the multicomponent nature of reading comprehension also increased instructional value. Both the FRA and RISE assessment systems evaluate multiple reading components and thus provide information regarding students' strengths and weaknesses that can inform instructional decisions.

Compared to other recent initiatives, such as the Partnership for Assessment of Readiness for College and Careers (PARRC) and Smarter Balanced Assessment Consortium efforts promoted by various state consortia to be responsive to curricula consistent with the Common Core State Standards (CCSS), GISA bears a stronger resemblance to these other assessments than does either RISE or FRA. Both the GISA and CCSS-aligned assessment privilege some approximation to authentic literacy activities. However, neither the PARRC nor the Smarter Balanced assessments come close to GISA in privileging

purpose (knowing from the beginning what the culminating product will be) or context (completing the assessment in a virtual community setting).

In conclusion, the assessment systems developed by the RfU are broader and more authentic than those typically available in the educational marketplace. They are also developmentally sensitive, and emphasize instructional sensitivity and value. These assessments have a strong theoretical basis and defensible psychometric properties. The calibration and validation studies were extensive and iterative, and were undertaken across the United States. The result is a set of forward-thinking assessments that not only meet the standards of educational and psychological testing, but also promise to advance both research and practice in reading comprehension for years to come.

### Curriculum and Instruction

With respect to *curriculum and instruction*, the RfU initiative produced a range of positive, but often inconsistent, results on a wide range of measures across the K–12 continuum. Effects were greater and more consistent for assessments that represented curriculum-aligned and researcher-developed measures than for those that were curriculum-independent and published measures of key outcomes. The strongest effects were observed for measures of vocabulary, morphology, comprehension monitoring, and knowledge acquisition. Those interventions that most consistently “moved the needle” on reading comprehension and a host of related measures (such as vocabulary, knowledge acquisition, application, and enabling skills) were characterized by well-orchestrated, multicomponent instruction.

Consistent with those studies that focused on development, instructional studies revealed that the relationships among enabling skills, knowledge, and reading comprehension are dynamic (changing across grades) and synergistic (improvement in one process can enhance performance in another). Across the RfU teams, researchers catalogued the roles of different types of knowledge in reading comprehension: declarative, procedural, conditional, disciplinary, and epistemic. This portfolio represents an *advancement in our conceptualization of the role of knowledge in comprehension*. In previous eras, we would have suggested that declarative knowledge of the topics of the texts was the most important resource for comprehension, followed closely by procedural knowledge of how to engage in strategies or practices to demonstrate comprehension, with a nod to conditional knowledge (when, why, and where to use it). Twenty years ago, that trio would have told the story about the role of knowledge in comprehension. The insights emerging from the RfU, hard on the heels of developments in adolescent literacy in the 2010s, demonstrates that the knowledge repertoire needs to be extended to include how disciplinary knowledge (how we talk, write, think, explain, and argue about key ideas in the major academic fields of study) and epistemic knowledge (how knowledge is generated and evaluated within the disciplines) change the relationship between knowledge and comprehension in substantive ways. The RfU work suggests that these five types of knowledge are variously cause, consequence, and covariate of reading comprehension.

*Learning to read* and *reading to learn* surfaced in the RfU portfolio as complementary goals rather than separate stages of development. These two complex processes were revealed to be interwoven fruitfully across students’ school careers, and the RfU research

The strategies enabled the students to see different nuances, discuss, and pull information together.

—*RfU Participating Teacher*

described their ongoing development and interaction with one another. Granted, we learned, particularly from the adolescent teams (Catalyzing Comprehension through Discussion and Debate [CCDD], Promoting Adolescents' Comprehension of Text [PACT], and Reading, Evidence, and Argumentation in Disciplinary

Instruction [READI]), that both learning to read and reading to learn are different enterprises in history, literature, and science; however, within each discipline, we also learned that reading to learn and learning to read were complementary. Conversely, we learned that even in the primary grades, as early as pre-kindergarten (pre-K), there is also a complementary relationship between reading to learn and learning to read (see Let's Know! [LK] and Content Area Literacy Instruction [CALI] for examples in Chapter 4).

The instructional research provided further evidence that student *engagement* with texts and tasks supports comprehension. The RfU researchers provided details on the instruction and classroom environments that contribute to students' engagement, and on the particular aspects of reading comprehension that benefit from engagement.

## QUESTION 2: ADDITIONAL LESSONS LEARNED FROM THE RFU WORK

As suggested earlier, not all of the lessons learned emerged from the substance of the three strands of development, assessment, and pedagogy. Some were implicit in the processes used to carry out the work and in the constellation of findings across teams. Three such lessons stand out: affordances of the consortium model, methodology to account for professional learning issues, and insights about why it is hard to “move the needle” on (particularly distal) measures of reading comprehension.

### The Affordances of the RfU Consortium Model

The research model enacted in the RfU consortium provides a demonstration of what is possible in the design, implementation, and analysis of lines of inquiry with the affordances of adequate funding, extended time frames, and a diverse array of expertise to carry out the work. When there is a sufficiently long runway, scholars have the opportunity to exploit the complementarity of research methods, scholarly traditions, and academic disciplines. As a result, the RfU scholars were able to engage in:

- \* Fine-grained, longitudinal, largely theory-driven examinations of the infrastructure of processes and knowledge developed by readers over time and in response to instruction;
- \* Examinations of exemplary and/or typical practice in school settings to establish the contexts in which pedagogical work might be situated;
- \* Design-based research and development in which teachers and schools are partners, not subjects, and, through interaction and consensus building with researchers, contribute to the creation and implementation of curriculum, assessments, and teaching practices;

- \* Extensive professional development, often in sustained, long-term professional learning communities;
- \* Well-designed, pedagogically complex, adequately powered efficacy studies and randomized controlled trials (RCTs) in which both learning and teaching are measured;
- \* Efforts that built upon each team's prior, less well-funded, efforts to design curricular and instructional interventions in the same pedagogical family;
- \* Secondary analyses of their own data, with an eye toward determining whether particular approaches were effective with particular subgroups (e.g., low-performing readers, emergent bilingual learners (EBs), or students from low-income schools); and
- \* Regular interaction with the researchers from the other teams in meetings brokered by the Institute of Education Sciences (IES) staff overseeing the RfU.

Even though the six projects were very different from one another, there were productive opportunities to interact with one another about the entire spectrum of research concerns—including models of reading, useful measures, sampling designs, new statistical tools, professional development formats, and tools for promoting or monitoring treatment fidelity.

At its best, the RfU research reflected a broad view of the science of reading, drawing from diverse traditions, paradigms, and theories. The research was also informed by the relevant scientific perspectives of researchers from affiliated fields, including psychology, sociology, learning sciences, linguistics, and literary criticism. This broad perspective was evident in how the RfU research was designed, the nature of data collection and analysis, and the breadth of examined outcomes. In terms of the “science” of reading, the RfU teams were informed by scientific literature that ran the gamut from phonemic awareness to phonics to morphological awareness to vocabulary to metacognition to motivation, and to comprehension itself.

Related science also made important contributions to the conceptualization and execution of the RfU research. For instance, FCRR employed a common set of measures across a wide range of language-related interventions to allow for the possibility of highly precise comparisons across the entire family of interventions. The Language and Reading Research Consortium (LARRC) used a design study to build understanding of local school needs and dovetail them with research intent, to enlist and gain support of key participants in the research work, and to determine the “place” of research in existing school contexts. READI conducted theoretical analyses of history, science, and literary criticism to determine common and unique features of the culture of each discipline, which then informed comprehension curriculum and instruction, and the development of evidence-based argumentation and of extensive scaffolds—heuristics for sense making, evaluating the validity of information, developing arguments, and reflecting on personal progress along the way. PACT and CCDD designed comprehension instruction that attempted to boost students' motivation and engagement, while LARRC sought to enhance young children's metacognition related to language comprehension.

In short, the affordances of the RfU grant mechanism enabled multiple, complementary strands of research (on the development, assessment, and instruction of reading

comprehension) to be conducted simultaneously. The coordination of multiple research goals, along with time to allow for iterations of design and refinement and of longitudinal data, helped to increase the “yield” of the RfU. Moreover, the RfU researchers will continue to publish results from this investment for years to come because of the richness of the data gathered. To continue to foster this kind of breadth, depth, and coordination of research inquiries, funders such as IES should reconsider whether the separation of project types (i.e., project types 1 through 4 are labeled exploration, development and innovation, initial efficacy and follow-up, and measurement) and investment in narrowly focused inquiries is ever likely to provide the kind of nuanced results that the RfU has produced. In fact, the very structure of the RfU acted as something of a guarantee that, even if an intervention or hypothesis was not fruitful, researchers were still left with data that improved our understanding of reading comprehension.

### Teacher Professional Learning

One objective of the RfU initiative was to encourage changes in reading comprehension instruction as a means of improving student reading comprehension performance. The RfU studies explicitly set out to disrupt “business as usual” by enacting

One of the largest shifts in my practice was due to unpacking my own moves, first with “Flowers” [used in the professional development session] and then in my classroom texts. Shifting the purpose of classroom discussion—from asking questions about the text that students would answer or asking students to only make personal connections—to apprenticing students into the kind of thinking that I do as an expert reader in the discipline. [It] shifted the way I planned for discussions.

—*RfU Participating Teacher*

a reconceptualization of comprehension teaching routines, reading comprehension curricula, expectations of student performance, and learning outcomes. Underlying this goal for student learning was an expectation that teacher expertise, as indexed by teacher knowledge of how to enact more challenging classroom practices, would be the focus of professional development activities, in a form that was developmentally appropriate for a given grade-level band. In such instances, teaching basic reading strategies (e.g., using context to help determine word meaning, or identifying an explicit main idea statement) is important and necessary, but not sufficient, for developing students’ capacity to engage with

texts and tasks in more complex and challenging ways, enacting what we have recently come to call deeper learning (NRC, 2012). If the practices that students engage in are to be disrupted, so too must the practices that teachers are asked to enact in prescribed commercially available or mandated curricula. The centrality of teacher professional development in support of helping students move to enhanced levels of comprehension and performance is evident across many of the RfU teams. The nature of this professional development—ongoing, detailed, collaborative, and tailored to individual teachers’ needs—is critical.

Many of the resulting RfU approaches to reading comprehension instruction necessitated new approaches to teaching that placed new demands on teachers. The

RfU comprehension curricula required the creation of means to educate and evaluate teachers, which was accomplished through professional development (for the education) and measuring fidelity of implementation (for the evaluation). A common medium for professional development in the RfU consortium was the design study, which, according to the NRC (2002), focuses on “the evolution of learning. The learning might be that of a student, a teacher, or an organization” (p. 28). This approach to research and development demanded collaboration among varied stakeholders, particularly researchers and teachers, and the iterative construction of their goals and needs along the pathway to new and challenging curricula. The approach also encouraged broad participation in framing research questions, designing interventions, and professional development. Here, teachers’ professional development was tied to learning the particulars of comprehension instruction within a specific research project, and was a necessary means of promoting growth in teachers’ declarative, procedural, conditional, disciplinary and epistemic knowledge for a particular curriculum, such as those reviewed in Chapter 4 (i.e., LK, Comprehension Tools for Teachers [CTT], Word Generation [WG], Strategic Adolescent Reading Intervention [STARI], PACT, Comprehension Circuit Training, or READI).

In terms of innovative RfU comprehension curricula, professional development was necessary to ensure that teachers were able to teach new curricula that variously promoted young children’s metacognitive strategies (LARRC, Johansson, & Arthur, 2016), middle schoolers’ debate, discussion and academic language (Kim et al., 2016; LaRusso, Kim, et al., 2016), and high schoolers’ epistemic development (Lee et al., 2016; Shanahan et al., 2016). Accordingly, the RfU research featured teachers’ professional development in the ways and means of innovative reading comprehension instruction. The nature of this professional development—ongoing (LARRC, Farquharson, & Murphy, 2016), detailed (Wanzek & Vaughn, 2016), and sometimes on call and responsive to particular needs in particular situations (Connor et al., 2018)—was critical. Teaching strategies to students at an unprecedented young age (e.g., comprehension monitoring to pre-K students; LARRC, Johansson, & Arthur, 2016), and teaching to the twin targets of learning to read while reading to learn (LaRusso, Kim, et al., 2016) demanded that teachers learn new instructional approaches. As well, providing instruction that fostered students’ ability to identify and then engage in evidence-based argument (Goldman et al., 2019) was preceded by teachers’ learning the requisite strategies and mindsets that they would eventually scaffold for their students. Professional development was accomplished through attention to both theory and practice. For example, at a theory level, READI devised professional development that helped high school teachers understand the nature and workings of disciplinary and epistemic knowledge, which in turn informed their disciplinary teaching practices (Goldman et al., 2019). While grounded in theory, all of the RfU studies eventually focused on the practical aspects of enacting these curricula. Teachers’ capacity to learn and adopt new practices related to new

Usually I am very lecture based rather than text based. It was neat to see students pull things from text that I hadn’t necessarily considered or chosen. They turned into little experts because of the text.

—RfU Participating Teacher



curricula, strategies, and mindsets was expected to take time and involve challenges. As students in the RfU studies learned new strategies and stances related to reading comprehension, so teachers learned new practices for teaching and supporting this comprehension. This model of teacher professional development in which university and K–12 educators collaborate to translate theory into practice and improve that practice in the context of classroom instruction is reminiscent of the National Writing Project model (Wood & Lieberman, 2000).

As described in the RfU studies, there are necessary areas of growth in teachers' declarative, procedural, conditional, disciplinary, and epistemic knowledge. For instance, fostering students' ability to read like a historian differs from helping students read and learn the facts of history, or from reading like a scientist or literary critic. Modeling strategies for students so that they might question authors' use of claim and evidence, or reconcile competing accounts of scientific cause-and-effect phenomena, differs from teaching summarization strategies. Teachers must also help students learn to be comfortable with the uncertainty of the knowledge they are acquiring. Students may find that they cannot determine the truthfulness of two opposing eyewitness accounts, but they might be able to determine which account is more credible. This stands in contrast to many students' school experience of reading to determine facts and the "right" answer.

Fidelity of implementation is a perennial challenge (Foorman, Dombek, & Smith, 2016), and involves both "logistics and establishing an infrastructure for ensuring adequate implementation" (Gersten, 2016, p. 113). Fidelity of implementation was a goal across the RfU teams, and the sum of the RfU studies provides a tutorial in conceptualizing, working toward, and analyzing fidelity of implementation. The research designs used in the RfU studies required considerable a priori efforts, including resources, to establish guidelines for building and maintaining fidelity. Looking across the continuum of RfU fidelity of implementation efforts one encounters a range of efforts to account for fidelity. Some studies focused on measuring teacher adherence to researcher-created instructional scripts and routines (Connor et al., 2018), while other studies saw professional development as a means of establishing fidelity (Goldman et al., 2019). A related continuum illustrates that professional development might focus on the practical (i.e., did teachers follow the script of a particular lesson?) and the theoretical (i.e., was teachers' understanding of epistemic knowledge evident in a particular lesson?). In the final analysis, few studies used metrics of professional development (e.g., high versus low knowledge or implementation) to predict student performance.

As many RfU fidelity efforts were linked to professional development, they simultaneously measured fidelity and enhanced the probability that teachers in treatment classrooms were adept at teaching with new curricula. FCRR provided its already experienced instructional assistants (mainly former teachers and/or graduate assistants) with professional development that ranged from 6 to 12 hours of initial training, followed by 3 to 6 hours of "booster" professional development (Connor et al., 2018). Teachers also utilized electronic bulletin boards to post teaching-related questions and to submit responses to weekly implementation quizzes. Accompanying these efforts was fidelity monitoring, by which teachers were observed in person and provided immediate feedback. Formative ratings provided a means for amending instruction, while summative ratings told the overall story of fidelity of implementation.

LARRC, Johansson, and Arthur (2016) investigated the effectiveness of teacher professional development in relation to fidelity using classroom teachers' online surveys following completion of the professional development module, three observations of lessons, lesson logs for every lesson, and an end-of-unit teacher survey and guided interview. CCDD (LaRusso, Donovan, & Snow, 2016) used summary coaching reports, implementation challenge checklists, and semistructured interviews, including open-ended questions about implementation, along with teacher case summaries that described implementation progress and barriers and student completion of instructional materials. These measures both reported current implementation fidelity and shaped future fidelity efforts. Furthermore, CCDD scholars (LaRusso, Donovan, & Snow, 2016), in both the STARI and WG interventions, documented structural challenges to fidelity, which included lack of sufficient teaching time and difficulties with new programs.

To summarize, the learning involved in professional development that supports teaching comprehension parallels students' ongoing learning related to reading comprehension. In essence, teachers and students involved in many of the RfU projects were working on parallel learning tracks. While students were learning new strategies, routines, and stances to become better learners, teachers were acquiring knowledge and routines to become better at scaffolding that student learning. This is not to say that the pathway to deeper learning on the part of teachers or students was always clear, well marked, or free of obstacles. For example, as we reported in Chapter 5, teachers in the PACT intervention were able to learn and implement some teaching routines better than others. In particular, routines involving scaffolding more basic skills and knowledge (engaging prior knowledge or teaching key unit vocabulary) exhibited greater uptake from teachers than routines intended to promote higher-level practices such as close reading, interpretation, and critical reading (Wanzek & Vaughn, 2016). When queried about barriers to the uptake of new routines, teachers using the CCDD's WG and STARI (LaRusso, Donovan, & Snow, 2016) interventions cited interference with other, often more high-stakes, initiatives, such as state test preparation or covering the adopted curriculum or the normal variations of student behavior in classroom cultures. Even so, working closely with teachers in supportive teacher learning community settings, researchers were able to promote changes in teacher practices in the direction of key intervention principles (see, in particular, Goldman, Britt, et al., 2016; Goldman, Snow, & Vaughn, 2016) among intervention teachers in comparison to business-as-usual teachers. Moreover, in some instances, changes in teacher practice were associated with gains in student achievement (Lawrence, Rolland, Branum-Martin, & Snow, 2014, Wanzek & Vaughn, 2016). Hard work, supportive settings for teachers to try new perspectives, and staying the course on everyone's part seem to be consistent threads in the more successful ventures.

The approaches to professional development and measuring teaching by the RfU teams built upon rich traditions established in decades of work in enacting (e.g., Coburn, 2003) and measuring (Rowan & Correnti, 2009) changes in teachers' knowledge, beliefs, and practices in reform-motivated curriculum projects. Thus, these approaches were not entirely new or ground breaking, but the professional development they embodied was exceptional in its durability (many of the teacher learning communities lasted over several years), focus (enacting particular curricula and measuring uptake of their key

components and principles), and engagement (finding a setting in which teachers could develop ownership of the reform curriculum).

### Insights About “Moving the Needle”

Several scholars within the RfU consortium (e.g., Catts, 2018; Lonigan, Burgess, & Schatschneider, 2018; Phillips, Kim, Lonigan, & Connor, 2015; Piasta, LARRC, & Jiang, 2016; Wanzek, Swanson, Vaughn, Roberts, & Fall, 2016) as well as those outside the RfU initiative (e.g., Elleman, Lindo, Morphy, & Compton, 2009; Fuchs et al., 2018; Lesaux, Kieffer, Faller, & Kelley, 2010) have noted how hard it is to move the needle on reading comprehension measures, especially distal measures, with even the most comprehensive, well-designed, and faithfully implemented of interventions. Earlier, in Chapter 5, we raised the same concern, adding to the mix the findings of Hill, Bloom, Black, and Lipsey (2008) and Lortie-Forgues and Inglis (2019).

The vexing question is: “Why?” Informed by the RfU teams, other scholars in the field, and our own experience as researchers in the same community, we offer several plausible explanations, ever cognizant of the perils involved in asserting or even implying causation.

#### *Insensitive Measures*

Perhaps the most common argument for the sparse and small reading comprehension effects in the pedagogical literature is the lack of instructional sensitivity of the assessments that have traditionally been used (see Chapter 3 in this volume; Pearson, Valencia, & Wixson, 2014). We just have not had or used measures of reading comprehension that are sufficiently sensitive to the types of interventions implemented in the RfU initiative or a host of other instructional programs that have surfaced over the past several decades, roughly since comprehension rose to prominence in reading pedagogy in the early phases of the cognitive revolution (Pearson & Cervetti, 2017).

There is substantial evidence documenting the difficulty of moving the needle for the very sort of distal measures we tend to demand as evidence of far transfer: intervention-unrelated standardized measures such as the Gates-MacGinitie Reading Test (GMRT) or the Woodcock-Johnson IV Tests of Achievement (WJ-IV). For example, Scammacca, Roberts, Vaughn, and Stuebing (2015), in examining intervention studies from two time periods (1980–2004 and 2005–2011), noted an average effect size for distal measures of comprehension of .24 across both time periods. By contrast, the comparable effect size for ALL measures across both sample time periods was .49. Interestingly, they also found across-the-board decreases in all effect sizes for the 2005–2011 sample; a finding that they attributed “at least in part to increased use of standardized measures, more rigorous and complex research designs, differences in participant characteristics, and improvements in the school’s ‘business-as-usual’ instruction that often serves as the comparison condition in intervention studies” (p. 369). Similar findings (greater effect sizes for proximal investigator-designed assessments over standardized measures) have been reported by Bloom, Hill, Black, and Lipsey (2008) and Moran, Ferdig, Pearson, Wardrop, and Blomeyer (2008), in which the proximal-distal comparison was .56 versus .30.

To the credit of the RfU initiative, a major goal of the Educational Testing Service (ETS)–FCRR portfolio of the RfU work (Chapter 3 in this volume; Sabatini & O’Reilly, 2013) was to develop measures of reading comprehension that reflected more ambitious goals, such as applying the knowledge gained from text comprehension to novel problems and projects and assessing a full array of contributing skills and knowledge. In fact, GISA, the RfU assessment system designed to measure the former sort of application of reading comprehension, did reflect a modest effect size and a modicum of instructional sensitivity for CCDD’s WG intervention and the READI science intervention, both of which placed a premium on using text understandings as a source of evidence to warrant arguments. Additionally, CCDD’s STARI intervention, with its emphasis on ensuring that struggling readers have an opportunity to bolster their entire repertoire of skills within the context of carefully designed thematic units, found an instructionally sensitive set of specific component measures in the RISE assessment.

In contrast, and no doubt partly due to the belief that neither RISE nor GISA was a good fit for their interventions, the other three teams (LARRC, FCRR, and PACT) for both narrative and expository texts chose either more widely used distal measures of reading comprehension (for PACT, the long-standing GMRT and for FCRR, the WJ-IV) or a specially crafted measure (for LARRC, it was their own listening comprehension measure, roughly modeled on the Qualitative Reading Inventory; Leslie & Caldwell, 2006). For PACT, even though they used the GMRT as their distal measure, they developed the Assessment of Social Studies Knowledge (ASK) measure of knowledge acquisition as their most proximal and the Modified Assessment of Social Studies Knowledge and Reading Comprehension (MASK) measure (comprehension items collected from the released items of many state assessments) as anchoring a spot on the continuum between proximal and distal. As reviewers, we were puzzled that neither of the early grade teams availed themselves of the extensive battery of well-designed, carefully validated measures within the FRA assessment system.

One conclusion that can be drawn from the RfU pedagogical portfolio (see the effect size analysis at the outset of Chapter 5) is that the instructional (in)sensitivity of distal outcomes did play a role in shaping our conclusions about how far an intervention might travel from decidedly proximal to increasingly distal measures. Above and beyond the need for assessments with instructional value and sensitivity, it is important to highlight that, like with any other construct, moving the needle on reading comprehension is also a function of transfer. Transfer (Barnett & Ceci, 2002; Day & Goldstone, 2012) is very difficult to achieve and evaluate in education in general, and in reading in particular (Gick & Holyoak, 1980, 1983; Pearson et al., 2014). Moving forward, we need to consider how transfer interacts with the purpose and grain size of reading assessments. For example, if students make progress on assessments of component skills, does that progress also emerge in a more global assessment of reading comprehension, such as GISA? Transfer—being able to apply new skills and ideas in settings and on measures that differ from the instructional context—is, and always has been, a worthy goal (perhaps the gold standard) of curriculum and instruction enterprises. But it is—and likely always will be—a challenge to achieve.

*Design and Implementation Issues*

A second commonly offered explanation is faulty design and implementation of the interventions themselves. Historically, pedagogical experiments in reading have often been inadequately designed and implemented. They have been guilty of one or more of the following flaws. The studies:

- were underpowered (especially to detect small effects) due largely to inadequate sampling;
- employed samples of convenience rather than intentional or random samples;
- failed to last long enough for treatments to take effect;
- were implemented in schools that did not really want or value the interventions in order to satisfy the demands of random assignment;
- failed to assess fidelity of implementation; and
- until very recently, ignored modern, sophisticated statistical analyses.

Thus, we have not, as a field, been able to say much about either the validity of any effects obtained or, more germane to the current concern, the trustworthiness of null or weak results. Instead, when we did not find significant and/or robust results, we have typically argued that some set of situational issues (usually design, duration, or measurement issues) conspired to sabotage our attempt.

However, as we suggested in our reporting of the curriculum and instruction portfolio in Chapter 5, poor design and implementation were not issues in the RfU initiative. The intervention portfolio is expansive and complex, with a wide range of independent variables, including many of the potentially malleable factors discussed extensively in Chapter 2. These were well-designed and well-implemented RCTs. All of the interventions emanated from a theoretical base (but not always the same theoretical base) about the nature and development of reading comprehension. Hence, it is reasonable to assume that they possessed a kind of *prima facie* construct validity. They detailed explicit models (or theories of action) of how particular facets of the reading comprehension puzzle can be shaped in instructional settings to elicit changes in performance. The details of the actual interventions were, in general, as well informed by the wisdom of practice as they were by the theories on which they were built. Teachers were involved as co-designers or critics along the way, often in extensive design research efforts. They employed a range of outcome variables, both proximal measures of whether students learned what was taught and distal measures of how “far” the interventions traveled to more general indices of comprehension or learning. They measured teaching as well as learning, always documenting what actually occurred in the intervention classrooms and, most often, in the business-as-usual control groups. In contrast to many prior efforts in pedagogical research, these were well-powered efforts, with samples sufficiently large and well defined to detect even small effects. About the only implementation standard they may not have met is the length of the treatment; most study directors would have wished for more time, more even than a single school year, for their interventions to take root in classroom ecologies. Even so, in comparison to most intervention studies, these were substantial periods of enactment, ranging from 8 to more than 20 weeks.

In short, given the care with which the interventions were designed and implemented, there was every reason to believe, going into the RCT phase of the RfU

initiative, that if there were effective interventions to be found, they would be found in this initiative. Conversely, if the effects proved to be null, weak, or small, that might be the truth of the matter; such results might be all that should be expected from this sort of intensive effort to move the needle. In short, improvements in reading comprehension itself might be more difficult to achieve than previously thought.

### *Unrealistic Expectations*

At the outset of this section, we suggested that many who conduct and review experimental work on curriculum and instruction bemoan the difficulty of finding large and statistically reliable effects. We also noted the empirical reviews suggesting that the typical effect sizes on distal measures in reading hover in the lower regions of the small (.20 to .49) range. We all want large effect sizes, but they are seldom forthcoming and may, in fact, be flatly unrealistic.

We think the same perspective can be applied to the RfU portfolio. As we note in Chapter 5, it is all too easy to look across the results presented in Chapters 4 and 5 with a glass-half-empty perspective. The effects could have been stronger, significant results more plentiful, and results more consistent across groups and measures. We suggest an alternative “glass a bit more than half full” interpretation, namely, that the aggregate RfU results provide grounds for cautious optimism and guidance for future reading comprehension instruction. We argue that the perspective of disappointment, to mix our metaphors, misses the forest for the trees. If, as we believe, the RCTs and efficacy studies within the RfU possessed reasonably robust designs, psychometrically and conceptually trustworthy measures, adequate power, sufficient dosage/duration, and sensitive statistical analyses, then perhaps we have collectively set our sights too high for achievable effect sizes.

Another interpretation is possible, even plausible: 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 during the RfU initiative. The RfU results suggest that carefully developed and orchestrated multicomponent (and intersectional) instruction, when implemented with fidelity by teachers who are supported by robust professional development, can yield effects that are strong enough to move the needle on reading comprehension and a host of related measures, such as vocabulary, knowledge acquisition, application, and many enabling skills. The needle might not move as radically as we desire, but it most certainly has 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 see significant progress in instruction and resultant reading comprehension.

### *Challenge*

The RfU studies in which teachers were observed teaching comprehension (Wanzek & Vaughn, 2016) or queried about implementation barriers (LaRusso, Kim, et al., 2016) offer some additional clues about why it is hard to move the needle. First, there is a lot to get in the way of barriers to serious implementation, as LaRusso, Donovan, and Snow (2016) learned. At the top of the list is test prep, which annually disrupts instruction

in the spring, followed closely by the time demands associated with adhering to the school or district curriculum and dealing with student behavioral disruptions. Second, as Wanzenk and Vaughn (2016) discovered, teachers are much more likely to possess the knowledge and material resources to implement instruction that targets the “low-hanging fruit” of the curriculum (invoking or providing relevant prior knowledge and teaching key words) than the harder-to-reach fruit (engaging in close reading, interpretation, application, or critique).

Even so, we know from the RfU work that these higher-order activities can be implemented, as we are reminded from the changes in teachers’ practices in the READI work on evidence-based argument (Goldman, Britt, et al., 2016), and that when implemented with fidelity (as in CCDD’s WG intervention) they can mediate student learning (Lawrence, Crosson, Paré-Blagoev, & Snow, 2015). These contrastive findings suggest, first and foremost, that this kind of teaching and learning is hard, for teachers as well as students. It further suggests (see R. Anderson, personal communication, September 17, 2019; Sun et al., 2020), for work going forward, that careful monitoring to ensure fidelity not only to the treatment but also to goals of both cognitive and affective engagement is required. Many of the RfU multicomponent interventions put a premium on collaboration and conversation. It seems reasonable to conclude, from the successes that the RfU did achieve, that engagement in higher-order talk within collaborative discussions about interesting and even controversial texts might be the most plausible pathway toward more successful outcomes for students and teachers.

This bundle of requirements (collaboration, talk, and thought-provoking texts) might also explain why we struggle to achieve even modest effects. At the very least, this is an important endeavor for next steps in unpacking the pedagogical puzzle around reading comprehension and learning in the presence of texts. But it is equally as important to provide considerable support for both teachers and students when we ask them to stay the course in these collaborative endeavors. This is as true for learning communities that support teachers’ focused, ongoing efforts as it is for students when we ask them to collaborate with their peers in challenging comprehension, critique, and composing tasks.

### *Moving the Needle Differentially*

An explicit goal of educational reform in the United States (and surely in much of the rest of the world) is to close the achievement gap between the educational haves and have nots. This goal often comes couched in a moral imperative such as, “America’s problem is not the overall low achievement of its students, it’s the unconscionable gap between x and y,” where we can fill in the x and y blanks with any of several pairs: majority and minority, rich and poor, native English speakers and English learners (ELs). Students’ percentile rank within the overall achievement distribution is remarkably stable from grade to grade. As long ago as 1988, Juel documented this phenomenon in a longitudinal study of students who struggled with reading (Juel, 1988). Within the RfU portfolio, Lonigan (2016) found a similar resistance to differential change in percentile rank in the longitudinal analyses of student growth over time. The story is that, left to the natural ebb and flow in curricular and pedagogical forces, students are likely to maintain their place in the achievement distribution.

But neither the Juel nor the Lonigan analysis involved interventions that intentionally try to disrupt this stable pattern of achievement across the years. Evidence that we could close the achievement gap would mean that we found student characteristic-by-treatment interactions that benefited lower achievers more than high achievers. That is, in a perfect world, all students would make growth over time, but those who started out low would make differentially greater growth than their initially higher-performing peers. Such a pattern was found only occasionally in the RfU work, as with the increased growth of ELs and other language-minority learners in vocabulary and in perspective taking, compared to English-only students, within the CCDD WG curriculum (see Chapter 4). As we suggested there, the news on student characteristic-by-treatment interactions is mixed and complicated. For some interventions, such interactions did not surface; where PACT worked, for example, it worked equally to the benefit of all the identified subgroups. For other interventions (e.g., the FCRR collection), the patterns of interactions were so complex and inconsistent that they defy explanation: in some grades, for some groups, the intervention outpaced business as usual (BAU), but then the pattern flipped at other grades, with BAU (very occasionally) exhibiting greater growth. Likewise, in some studies, these interaction effects could not be examined because students were selected into the study based on low-level skill in reading comprehension or one of its component skills. Nonetheless, those studies often demonstrated main effects that suggested, at the very least, that students selected into the study outperformed students with similar pre-intervention skills.

To study the differential impact of an intervention on students coming into the study with differing characteristics or performance profiles, researchers must make deliberate choices about sampling, design, and analysis. Without an even more substantial investment than was made within the RfU initiative, it is unrealistic to expect that any single efficacy trial could simultaneously account for both student characteristics and pre-intervention performance profiles in determining what works for whom. Researchers need to make these sampling, design, and analytic choices informed by either theory-based or policy-driven priorities. This means that related lines of inquiry, conducted systematically over time, are needed to establish a complete picture of what moves the needle differentially for our most underserved learners.

### **Where Does This Analysis Leave Us?**

One might be tempted, after encountering all of the factors that deter us from our goal, to retreat altogether from the effort to move the needle. To the contrary, we think that the intractability of the problem, and the signs of promise unearthed by the RfU, should motivate us to redouble our efforts to solve it. As we have suggested, the glass-half-full perspective on what we did learn gives us a foothold to resume the quest. More realistic expectations, coupled with building more multicomponent curricula, reframing instruction as a cultural pursuit, accepting the challenge of the diligence it takes to sustain interventions inside classrooms and within teacher learning communities, and expanding our portfolio of innovative assessments of reading comprehension may be the basis—and the best hope—of future efforts to address the goal of moving the needle.



### QUESTION 3: BENCHMARKS FOR GAUGING PROGRESS OF THE RFU INITIATIVE

In Chapter 1, we grounded the RfU initiative in a number of contextual factors—theories, practices, policy initiatives, and trends—that had risen to a level of influence in the first decade of the 21st century such that they necessarily influenced what could or should be done in the name of the RfU initiative. In return, these contextual factors were likely to be influenced by the RfU work; in that sense, they provide convenient benchmarks for assessing the influence of the RfU. Here, we address two consequential theories influencing the development of the RfU initiative, the Simple View of Reading (SVR) and the RAND model of reading comprehension. Then we address adolescent literacy and its sibling construct of disciplinary literacy, both influential developments in the 2010s.

#### The Simple View of Reading

The SVR served as an explicit jumping-off point for the RfU initiative (IES, 2009, p. 5). The SVR was originally intended to provide a broad model for understanding the role of decoding in reading comprehension and potential sources of reading disabilities (Gough & Tunmer, 1986). The SVR describes reading comprehension as the product of decoding and listening comprehension. In doing so, it specifies that, in general, readers who have underdeveloped skill in quickly and accurately recognizing words (decoding) or in constructing meaning from discourse (listening comprehension) will struggle with reading comprehension. Although each of the contributors is actually quite complex, involving an array of skills and knowledge (see, for example, Francis, Kulesz, & Benoit, 2018), framing comprehension as the product of these two broad contributors has long been viewed as a useful heuristic for understanding sources of reading success and difficulty and shaping the purpose and goals of reading pedagogy. However, the relative simplicity of the SVR also invites scrutiny of its explanatory power.

In assessing the overall contribution of the RfU work in advancing our knowledge about the SVR, we conclude that the RfU effort complicated the SVR substantially by adding to our knowledge about (1) the subcomponents that comprise the key components of listening comprehension and, to a lesser degree, decoding; (2) how those components shift in relation to one another and to the ultimate reading comprehension outcome across the development span for pre-K through grade 12; and (3) what other, including exogeneous, factors need to be considered to allow us to explain more of the variance in reading comprehension, for both the youngest and older readers. Now to the evidence that warrants this conclusion.

Several RfU studies examined the validity of the SVR. This research confirmed previous research that had established the validity and credibility of the model: the vast majority of the variance in reading comprehension is accounted for by readers' skill in decoding and language comprehension, at least among elementary-age students (e.g., LARRC, 2015a, 2015b, 2015c; LARRC & Chiu, 2018). Moreover, early language and code-related skills predict the components of the SVR later in school (LARRC & Chiu, 2018). Wang, Sabatini, O'Reilly, and Weeks (2019) provided evidence for a nonlinear relation between decoding and reading comprehension and the identification of a threshold; below this threshold decoding was only weakly related to reading

comprehension and reading comprehension performance was limited. Decoding above this threshold positively predicted performance in reading. RfU findings like these further enhance the credibility of the model and provide plausible hypotheses about malleable factors that can inform pedagogical interventions.

The SVR was also the backbone of the pedagogical portfolio of both of the primary grade RfU teams—FCRR and LARRC. Many of the single-component interventions of FCRR (e.g., Language in Motion, Comprehension Monitoring and Providing Awareness of Story Structure, Morphological Awareness Training, or Enacted Reading Comprehension) or the constituent components of the multicomponent LARRC intervention (language comprehension, comprehension, monitoring, vocabulary, or text structure awareness) can be viewed as an attempt to expand the infrastructure of the listening comprehension factor in the basic SVR formula (reading comprehension [RC] = decoding [DEC] × listening comprehension [LC]).

Similarly, two of the three major assessment efforts, the FRA of FCRR and the RISE of ETS (see Chapter 3 of this volume), could be viewed as attempting to provide at least a partial answer to the question, “What would you need to assess, if you wanted to assess the major internal components of the three variables (reading comprehension, listening comprehension, and decoding) in the SVR?”

The RfU research also sheds light on limitations of the model. For example, the conceptualization and representation of decoding and language comprehension as “necessary, and thus, of equal importance, for reading comprehension” (Hoover & Tunmer, 2018, p. 304) serves the broad conceptual model, but it may obscure the complex dynamic relations among key variables when applied to understandings about reading development and reading instruction. As the RfU studies attest, in practical terms, the role of components and subcomponents shifts across age and comprehension skill level; in particular, the explanatory power of the decoding component attenuates across grades (e.g., LARRC, 2015b; Lonigan et al., 2018).

Several additional issues regarding the clarity and utility of SVR were raised or left unresolved by the RfU research. For example, it is still unclear what subcomponents belong in each of the two broad SVR constructs. For example, does vocabulary adequately index listening comprehension (LARRC, 2015a; Wagner, Herrera, Spencer, & Quinn, 2015)? Should additional components be explicitly acknowledged in the model (e.g., where should one place the powerhouse factor of declarative world knowledge)? Are there underlying factors (such as fundamental cognitive components like memory or attention) that explain the substantial shared variance between decoding and listening comprehension found in many empirical studies of the model (Catts, 2018; LARRC & Chiu, 2018; Lonigan et al., 2018)?

Although the model accounts for most of the variance in reading comprehension in the primary grades, it may not provide sufficient guidance for the development and application of interventions. Indeed, as Gough, Hoover, and Peterson (1996) declared: “Only a fool would deny that reading is complex. Reading clearly involves many subprocesses, and those subprocesses must be skillfully coordinated” (p. 1). In focusing on two broad predictors of comprehension that are underspecified and difficult to distinguish in the earliest grades (Lonigan & Burgess, 2017), the model offers less guidance about the particular underlying factors that will affect some students’ reading comprehension later in school.

Similarly, explaining comprehension for older students may involve unpacking the infrastructure of the SVR (e.g., what is entailed in the listening comprehension component?) or augmenting it with additional facets, such as those investigated in other models. For example, the FCRR team subscribed to a longitudinal elaboration of the SVR called the “lattice model” that accounts for the reciprocal relations between decoding and listening comprehension, as well as other cognitive processes, over time (Connor et al., 2014). Ahmed, Francis, York, Fletcher, Barnes, and Kulesz (2016) validated the Direct and Inferential Mediation (DIME) model (Cromley & Azevedo, 2007) in which background knowledge, vocabulary knowledge, reading comprehension, word reading skill, inference making, and reading strategy use all make significant direct contributions to comprehension in adolescence. Using the RfU data, Francis, Kulesz, and Benoit (2018) also examined an alternative model, one they dub the Complete View of Reading (CVR*i*), that accounts for idiosyncratic variation based not only on readers but also on texts by unifying discourse-based cognitive models of reading comprehension with the SVR. They found evidence of variation in rates of reading growth over time that reflect not only variation between readers in reading skills, but also between texts, which shows evidence of differential impact on readers of differing levels of achievement. For example, expository texts and more difficult texts have a negative impact on fluency (i.e., causing students to read more slowly), but especially so for better readers who adjust their reading rate more than poorer readers as they encounter more challenging texts. According to Francis et al. (2018), these findings suggest that models like the SVR that attribute comprehension entirely to component skills may overlook important variation in how individuals approach the task of reading comprehension across different situations and texts (reflecting the task/activity dimension of the RAND model). As a result, they may thus overlook potential pathways for intervention (see Valencia, Wixson, and Pearson [2014] for examples of what these pathways might look like).

In fact, the CCDD and READI work was based on the hypothesis that the SVR declined in relevance to middle grades reading because it obscured or ignored key elements that are crucial to success in reading literature, history, and science in the upper grades. For CCDD, these elements were academic language skills, perspective-taking skills, and reasoning skills (LaRusso, Kim, et al., 2016). For READI, they included the discourse conventions that render oral and written texts discipline specific and the complex set of reasoning skills that define evidence-based argumentation within disciplines (Goldman, 2018). One might argue that the listening comprehension component of the SVR covers academic language, but that interpretation obscures the fact that we are more likely to see than *hear* academic language; the major site for exposure to it is in literate contexts. Similarly, social perspective taking, which starts early with the development of Theory of Mind (Brown-Schmidt, 2009), is not fully accounted for in the broad label of listening comprehension because it involves the ability to infer and project the likely different perspectives of multiple participants in a social scenario using more than just linguistic cues. Finally, given that many texts in literature, science, and social studies require following a multistep, often probabilistically or conditionally stipulated cascade of sequential and/or causally related claims, the ability to follow the logic of complex arguments comes into play as a determinant of successful comprehension (Goldman, 2018; Snow, 2018).

As we suggested at the outset of this analysis, the RfU work advanced our understanding of the SVR by complicating the range of subcomponents that influence listening comprehension and decoding, the shift in influence of within-word and language factors across the developmental span, and the gradual entry of exogenous variables as explanatory factors.

In a recent article (Hoover & Tunmer, 2018), two of the developers of the SVR note that the original intent of the model was to suggest that, “at the broadest level of analysis,” reading comprehension is determined by decoding (or word recognition) and language comprehension (p. 304). It is at the broadest level of analysis that the SVR continues to be most useful. It still provides a useful heuristic for conceptualizing and discussing the major “clusters” of factors that account for reading comprehension. The work that remains to be completed is to better understand, and ultimately validate, the key components that constitute the components, particularly the listening comprehension component, across levels of development. However, the collective RfU findings suggest several promising avenues, not only for a better elaborated and more global theory of reading comprehension, but also one that better specifies promising pathways for intervention.

### **The RAND Reading Study Group**

In 2002, the RAND study group posed a set of challenges that could serve as a blueprint for guiding a research agenda specific to reading comprehension. For example, the RAND group heuristic suggested that future research should focus on the independent and joint influence of the reader, task or activity, and text, all of which are nested within sociocultural contexts, on comprehension.

Among the reader factors that received significant attention in the RfU research was reader knowledge, including the quality of that knowledge (see CALI, PACT, and READI) as well as its range. In particular, the RfU (see Chapter 5) portfolio moved us beyond the familiar triad of declarative, procedural, and conditional to include both disciplinary and epistemological knowledge.

The RAND study group proposed that there was a lot to be learned about the influence of text features on comprehension. The RfU teams extended the range of text features to include unfamiliar content using complex language forms (Shanahan et al., 2016), novel syntactic constructions, discourse organization, linguistic markers, multisyllabic words (academic language), and metalinguistic terms. In addition, the RfU researchers studied the role of sequencing texts to build vocabulary and knowledge. The texts included ambiguous story characters, unexpected plot developments, and the representation of contrasting positions. Text features were considered in designing all of the intervention and curriculum materials; for example, the SoGen units developed by CCDD intentionally offered relatively small chunks of text (providing information in lists of facts to be sorted into “pro and con” for the debates, for example), rather than longer, denser paragraphs. Furthermore, alternative text types (videos, cartoons, etc.) were used both to present information and as targets of analysis in both WG and READI units. Texts were an especially important influence on the instructional routines and settings within READI, with a special emphasis on scaffolding and supporting students as they grappled with complex, challenging texts;

in particular, disciplinary tasks and social supports were critical in helping students feel comfortable with challenging texts, even leading to cognitive and affective reflection about their encounters with texts.

With respect to readers' tasks and activities related to comprehension, the RfU teams augmented the more traditional classroom practices of recalling and summarizing by having students determine the meanings of unfamiliar words and constructs, analyze text structures, recognize intertextual references, integrate information across texts, and transform text-based information into knowledge that could be used to construct arguments, explanations, and even reports and projects. This work was especially prominent in the work of the adolescent teams (PACT, CCDD, READI), but it was also present in FCRR's CALI. The adolescent teams also expanded study of the purposes for which readers read to include solving problems using text-based information, critiquing arguments, and building arguments (PACT, CCDD, and READI).

The RAND study group raised questions about the role of direct comprehension instruction versus instruction that was embedded in inquiry and authentic reading. Supporting ways to embed comprehension in inquiry and reading for authentic purposes was prominent in the RfU work, especially in the teams that focused on adolescents. For example, PACT attended to cognitive and motivational aspects of the reading process in the design of their interventions; CCDD examined academic language, perspective taking, and reasoning skills; and READI focused on reading and reasoning in different disciplines by attending to oral discourse frames, text genres, and academic language that distinguish disciplines of history, science, and literature. A number of the interventions attended to establishing an explicit purpose for reading that went beyond answering questions or passing a test; for example, the interventions used essential questions (e.g., PACT) or explicit unit goals connected to students' lives and experiences (e.g., PACT and CALI), juxtaposed texts (e.g., READI), conducted highly focused debates (e.g., WG), and used peer participation structures (pair-share, team-based learning; e.g., PACT, READI). In both WG and STARI, students read texts that were chosen to be of interest and relevance to readers (e.g., immigration, nontraditional families), that were organized in thematic units, and that posed discussable questions. Both programs made efforts to align texts and topics to curriculum standards. They emphasized concepts and vocabulary that were specific to the disciplines, and they selected texts to deepen knowledge.

The RAND study group wondered about the relative power of various instructional delivery systems. There were multiple modes of delivery systems investigated across the RfU sites. Most of FCRR's CTT curricula used scripted approaches, but the CALI curriculum used semiscripted lessons, and Word Knowledge e-Book produced technology for independent but guided practice in reading for meaning. PACT produced its own modules for history and also developed technological curriculum assets. CCDD produced supplementary curricula for both WG and STARI. READI did not produce curricula, but instead engaged in close design collaboration with teachers planning around district and school curricula, and often, as suggested earlier, involving non-textbook texts. Most teams used professional development, in-class coaching, and professional learning communities to build extended, not single-shot, teacher learning opportunities. Above all, all teams concerned themselves with both supporting and/or measuring the quality and fidelity of implementation.

The contextual factors identified in the RAND document, such as economic resources, ethnicity, neighborhood, and school culture, did not figure prominently in the RFU portfolio, most likely because of the focus in the request for applications on development, assessment, and pedagogy. In the recommendations for future research that we identify in the next section, we suggest how these contextual factors might figure more prominently in future comprehension research.

### Adolescent and Disciplinary Literacy

As we suggested in Chapter 1, adolescent literacy had gained traction in the language arts field in the first decade of the 21st century due in no small part to a systemic effort on the part of the Carnegie Corporation to highlight an emerging groundswell of theoretical and practical work on the reading problems facing adolescents in content-laden secondary classes (Biancarosa & Snow, 2006; Snow & Moje, 2010). With attention focused on adolescents, it seemed a natural step to consider the question of whether the practices of these subject-matter classes were general (applying to all subjects) or subject specific. This distinction led many scholars to the idea of disciplinary literacy as a construct we could use to characterize the goals and challenges of reading, writing, and thinking in what we have traditionally labeled subject-matter or content-area classes (science, history, mathematics, and sometimes even literature). Chief among the perspectives arising from this work was the idea that while there might be general reading, writing, and learning practices, there were also likely to be subject- or discipline-specific practices—or at the very least discipline-specific instantiations of more general practices (Lee & Spratley, 2010; Shanahan & Shanahan, 2008). Also prominent in the disciplinary literacy perspective was the idea that the language of texts and talk about key ideas varied across disciplines—that there were indeed discipline-specific vocabulary and discourse patterns, and even ways of thinking and epistemologies. Phrases like “thinking like a historian” or “reasoning like a scientist” became more common. So how did the RfU initiative influence the ways we think about adolescent and disciplinary literacy?

Combined, the RfU teams that focused on older students demonstrated that the notion of transitioning from learning to read to reading to learn is a false dichotomy. Students in grades 5–12 must learn new strategies, stances, and forms of knowledge to fully comprehend school texts. Recall that three RfU projects (PACT, CCDD, and READI) were funded to attend specifically to older students (grades 5–12). Even though the projects differed in important ways, they were united by their interest in addressing the unique challenges experienced by students as they move from the intermediate grades of elementary school into middle and secondary school. Specifically, these challenges include (1) the amount of unfamiliar content presented in texts, rendering less effectual the typical strategy of encouraging students to use their prior knowledge to make connections and draw inferences; (2) the complexity of academic language encountered in text (including unfamiliar, multisyllabic words and less familiar (and seldom used) syntactic constructions); and (3) the task demands associated with, for example, integrating information from multiple texts, critiquing arguments for claims made in texts, and building one’s own arguments from text-based evidence.

Goldman, Snow, and Vaughn (2016) summarized the similar practices that emerged across their three projects, given these challenges. The first of these is

*active, purposeful, engaged reading*, which entailed identifying explicit goals for reading that were connected to students' lives—for example, by posing a controversy to which students could relate that would be addressed in the text. Another example included the use of essential questions to which the students returned as they read. To support engaged reading, all three projects used nontextbook texts, often replacing

My practice moved from attention to plot and asking students to make surface connections to characters (“I know men like Rasheed.”) to attention to language and the way in which it helps the reader to understand characters, theme, etc. This shift in my planning made a difference in the way students talked about literature.

—RfU Participating Teacher

textbooks with shorter texts that were sequenced in increasing difficulty and contained information germane to the essential question, or that supported the construction of arguments and or explanations. The second common practice was *social support for reading*. Working in pairs or small groups, students prepared for debates, jointly wrestled with the ideas in the text, and shared common challenges and successes in interpreting and learning from text. Whole-class discussions were used as occasions to model repair strategies and as occasions

for teachers to teach disciplinary-specific uses of language and reasoning. The third feature that was characteristic across the projects was *promoting deeper learning* by activating prior knowledge and positioning readers to apply the information they were acquiring to solve a novel problem or articulate an explanation. These three features are critical for understanding the importance of a more “cultural” understanding of comprehension practices as a way of helping students come to terms, as they move through their schooling career, with increasingly challenging and complex literacy activities.

These RfU teams also helped to refine what it means to take a disciplinary stance toward language and learning within secondary classrooms, helping us come to understand “disciplinarity” in several manifestations:

- Representing knowledge, including grappling with the epistemology question of how we know what we know (or do not know);
- Deploying specialized reading comprehension strategies that can help to crack open the puzzles of opaque language, both vocabulary and syntax;
- Engaging in discourse practices that define how we explain phenomena and argue about the validity of competing explanations; and
- Pursuing goals (often taking the form of projects or solutions to problems) representative of the discipline (see Goldman et al., 2019).

This disciplinary knowledge complements the declarative and procedural knowledge that is necessary for literal and inferential interpretation of text; it allows student readers to move beyond literal and inferential comprehension to forms of understanding that include analysis, critique, evaluation, and, above all, integration (Goldman et al., 2019; Shanahan et al., 2016). A clear finding across these adolescent teams is that curriculum and instruction in upper grades must attend to students' ongoing need to learn

to read texts and to participate in tasks of increasing complexity and challenge. That is the essence of our new understanding of disciplinary literacy.

In 2011, Wilkinson and Son proposed that we were on the verge of taking a dialogic turn in comprehension instruction, emphasizing dialogue (talk!) as a medium for paying more attention to discourse and collaboration as a means of improving comprehension and learning in our schools. It seems clear that the RfU teams that focused on adolescents shifted the emphasis of comprehension instruction to just such an agenda; in the work of the adolescent teams, students actively and collaboratively constructed and extracted meaning from texts, used language in the form of discourse to sharpen and deepen their understanding, and applied the knowledge gained from reading, thinking, and talking to solve problems and explain how and why things work the way they do.

#### QUESTION 4: WHAT MIGHT THE RFA FOR RFU 2.0 LOOK LIKE?

We want to close our stocktaking by looking toward the future and proposing what we think are the absolutely essential initiatives for the literacy research field to undertake in order to “write the next chapter on reading comprehension.” To prepare for such a proposal, we begin by summarizing the future research agendas implicit if not explicit in the core chapters on development, assessment, and pedagogy.

#### Research Priorities for Nature and Development, Assessment, and Curriculum and Instruction

In Table 6-1, we remind readers, in highly truncated and interpreted form, of the future research priorities identified in more elaborated form in Chapters 2, 3, and 5.

**TABLE 6-1** Recommendations for Future Research Across the Three Strands

Strand	Issue	Recommendation
Nature and Development	Discrete versus interactive language development	Determine whether rich and broad language experiences develop multiple aspects of language concurrently compared with the independent development of discrete language skills.
	Individual differences in cognitive and attentional skills	Examine whether metacognitive, cognitive, and attentional skills have a critical role in comprehension for particular groups of students, suggesting different pedagogical pathways to improvement.
	Knowledge as a broader mediator	We know a great deal about the mediating role of knowledge for comprehension but not for attentional and retrieval processes on the way toward more facile inferencing or monitoring.
	Linking vocabulary and knowledge	If the semantic and conceptual facets of word knowledge are emphasized over definitional knowledge, there might be synergistic growth in both vocabulary learning and knowledge acquisition.

*continued*



**TABLE 6-1** Continued

Strand	Issue	Recommendation
Assessment	Authenticity	Evaluate the validity and utility of using knowledge gained from comprehension as a deep index of comprehension.
	Theory: process versus componential measures	Use evidence from assessments to evaluate competing theories, such as assembly versus orchestration of key components.
	Instructional sensitivity	Develop measures of global reading literacy for younger readers while also refining those for older readers.  Determine elements of global literacy appropriate at different age levels, populations, and disciplines.  Evaluate the everyday utility of global measures.
	Instructional value	Determine if training test-identified specific skills improves more general comprehension performance.  Evaluate the feasibility of formative measures of global comprehension to complement summative measures.
	Complexity	Explore how to increase the depth of learning required to complete tasks in order to expand the ceiling for comprehension measures.
	Prior knowledge	Examine the increase in explanatory power of assessments by including prior knowledge probes as possible mediators.
Curriculum and Instruction	Emergent bilingual learners (EBs)	Additional research focused on all underserved, but especially EB, populations to help teachers develop more effective practices and close gaps.
	Pedagogical theory (assembly versus orchestration)	Compare the relative merits of assembly versus orchestration models of acquiring skills and knowledge.
	Engagement	Evaluate ways to embed engagement and motivation as inputs (malleable factors), outcomes (measuring the constructs), and mediators (catalysts for enhancing comprehension and learning).
	Text	Examine ways of positioning text in a more central role in our pedagogical research, as a malleable factor rather than simply a medium for discussion.
	Critique	Evaluate the role that a mindset for critique plays in shaping a purpose for close reading and comprehension (see Recommendation 1 below).
	Measuring teaching	Find ways to describe “business-as-usual” conditions with the same care and detail with which interventions are described.
	Transfer	Develop better approaches to scaling and describing the degree of alignment between assessments and interventions.
	Metacognition	Determine optimal approaches to teaching metacognition as the natural counterpart to comprehension and comprehension-related tasks.
	Knowledge	Determine the breadth and depth of prior knowledge that are necessary for comprehension and knowledge application.

### Overarching Research Priorities for Reading Comprehension

In addition to these highly specific recommendations (some of which, such as the assembly-versus-orchestration issue, emerge in all three strands), there are several recommendations for a future RfU agenda that are more overarching in character and, as such, are elaborated on below.

**Recommendation 1: We need to incorporate the relatively new perspectives of new literacies, digital literacies, and multiliteracies into our comprehension research portfolio.** As we suggested in Chapter 1, it was never the expectation that the RfU portfolio would necessarily extend to the recently minted cluster of theories and practices that includes perspectives, practices, and affordances that are relative newcomers to the literacy scene. In fact, terms like “new” and “digital” literacies are not explicitly mentioned in the RfA for the RfU. Whereas the RfU initiative is clearly grounded in a long tradition of cognitive and, to a lesser extent, social perspectives on how we understand and use what we read in the service of learning, these new traditions are grounded more in the epistemological and theoretical traditions of sociocultural or critical perspectives on literacy. As such, they represent opportunities for cross-fertilization across these currently independent research efforts. We think that the work informing the nature, development, assessment, and instruction of reading comprehension, as instantiated by the RfU initiative, has as much to learn from these new traditions (especially in situating comprehension research squarely in the contexts in which its purpose is instantiated) as the new traditions have to learn from the more cognitively grounded work exemplified by the RfU (especially when it comes to research methods that can be used to warrant explanatory and causal accounts of key relationships).

In this brief review of research on technology-related reading comprehension research, as well as research related to multimodal meaning making (both digital and nondigital) and reading comprehension in out-of-school contexts (see Fitzgerald, Higgs, & Palincsar [2020], a white paper available on the National Academy of Education website, for a more extensive treatment of these developments), our goal is to highlight future directions for reading comprehension research that complement those conducted by the RfU teams. As reading increasingly shifts from traditional print to screens, online platforms, and other digital representations in school, work, and community spaces, readers need increasing facility using search engines to locate information, critically evaluate online information to determine the reliability of the text(s) identified, and use online communication tools, such as email, blogs, or infographics, to communicate information. Across the past decade, researchers have investigated a number of questions about reading comprehension with digital text, particularly in the context of the Internet (Leu et al., 2015). Examples include investigations of readers’ use of strategies during online reading, pointing to the interplay of new and traditional reading strategies (e.g., Cho & Afflerbach, 2015; Goldman, Braash, Wiley, Graesser, & Brodowinska, 2012); facilitative and detrimental cognitive and social processes during online inquiry (Coiro, Sekeres, Castek, & Guzniczak, 2014); contextual factors that may influence online research and comprehension (e.g., Kennedy, Rhoads, & Leu, 2016; Leu et al., 2015); and how learners evaluate the quality of online information (Coiro, Coscarelli, Maykel, & Forzani, 2015). We suspect that these will continue to be important lines of inquiry.

In a related area of research, a number of studies have focused on identifying facilitative and detrimental cognitive and social processes during K–12 students' online inquiry. While some cognitive and social processes appear to facilitate students' performance on online research and comprehension tasks, others inhibit performance (e.g., Castek, Coiro, Guzniczak, & Bradshaw, 2012; Cho, Woodward, & Li, 2017; Coiro, Sekeres, et al., 2014; Delgado, Vargas, Ackerman, & Salmerón, 2018). While studies are emerging that differentiate more and less successful online reading, there is limited research that speaks to instructional practices and tools that K–12 teachers might adopt in order to foster student's performance on online research and comprehension tasks, and how instructional practices and tools might differ across grade levels; this is another area ripe for inquiry.

While the Internet has the potential to democratize access to vast quantities of information, it also places unprecedented responsibility on readers to evaluate the quality and reliability of the information they encounter online (McGrew, Breakstone, Ortega, Smith, & Wineburg, 2018). Future research in this area could productively include observational research in classrooms to understand curriculum and practices teachers are already using to support students to interpret and evaluate digital text in online spaces, as well as design-based implementation research using multiple qualitative and quantitative research methods and conducted in collaboration with teachers, schools, and districts to design, test, and iterate upon the design and enactment of curriculum materials to support digital literacy. In addition, while some research in this area has focused on students in the elementary grades, the vast majority of the research on online reading comprehension has been conducted in secondary and postsecondary contexts, suggesting that the field needs to expand research efforts to earlier grades in order to better understand how students develop strategies and skills for online reading over time.

There is increasing interest in multimodal literacy, most germane to this report being multimodal composition as a form of comprehension assessment. For example, Kesler and colleagues (2016) studied the digital stories created by fifth graders to share their interpretations of historical fiction novels. Analyses suggested that students' multimodal designs showed inferential skills, metaphorical thinking, and their understandings of character motivation. The digital stories also made visible to researchers and teachers the limits of students' understandings, such as misconceptions about plot sequence and gaps in background knowledge of historical context. These studies underscore a synergistic relationship between reading and writing.

A number of possible future directions for multimodal comprehension research are warranted. First, the field would benefit from longitudinal research that follows students over time to determine the effects of scaffolded learning in designed digital environments and the implications for students' achievement in and beyond the classroom. Second, although findings suggest that multimodal nondigital texts are useful tools that can support comprehension in K–16 settings and among diverse learners, it would be helpful to know how teachers might support students in learning how to interpret and synthesize communicative modes as they read. Studies suggest that interpretation of even mundane multimodal texts such as picture books or textbooks is a complex process, and one that requires thoughtful guidance and ongoing opportunities to practice. Finally, there is still relatively little comprehension research that focuses on critical literacy and multimodality. While some researchers have explored how

young people understand and evaluate multimodal texts using explicitly sociocritical lenses (e.g., Ajayi, 2015; Begoray, Higgins, Harrison, & Collins-Emery, 2013), studies that consider how integration of modal resources can support learners' inferential and sociocritical understandings of texts are still uncommon.

Finally, studies of comprehension in out-of-school settings have attended to varied contexts, disciplines, and age groups, including a STEM program for nondominant middle school girls (Pinkard, Erete, Martin, & McKinney de Royston, 2017), an after-school literacy program for recent-arrival immigrant teenagers (Park, 2016), and a summer science and data literacy camp for high school students (Sommer, Hinojosa, Traut, Polman, & Weidler-Lewis, 2017). While there is clear interest in what young people read and how they make meaning of texts outside of school in digital and non-digital environments (e.g., Hutchinson, Woodward, & Colwell, 2016; Jiménez & Meyer, 2016), a sustained line of inquiry related to reading comprehension in out-of-school spaces is not yet clear. Indeed, many of the questions raised by Hull and Schultz (2001) in their review of research related to out-of-school literacy learning remain salient directions for future research almost two decades later. For example, more research is needed to understand reading comprehension in out-of-school spaces and its relationship to in-school learning, including how to bridge students' out-of-school worlds and lived experiences with classroom practice, how to leverage learning in afterschool and other "school-like" spaces in the classroom, and how to support teachers to view and leverage students' out-of-school meaning-making practices as assets for classroom learning.

Recently, Ito et al. (2020), summarized a decade of research, conducted by the Connected Learning consortium to address gaps between in-school and out-of-school learning. Ito et al. (2020) advocate for research that asks, for example, (a) how the field can optimally use "the growing abundance of free and open learning resources to support the learning and interests of diverse young people"; (b) how "new media [can] be mobilized to forge shared rather than divergent interests and literacies between young people, parents, and teachers"; (c) what "new literacies [are] required by the new media ecosystem"; (d) "what forms of measurement, documentation, and evaluation can capture learning across settings"; and (e) how "factors such as social connection, affinity, and belonging influence learning" (Ito et al., 2020, p. 66).

**Recommendation 2: We need to develop more precise tools for evaluating the implementation of interventions by incorporating insights from the relatively new field of improvement science.** Like most educational researchers, reading researchers are more prone to be guided in their work by developments within rather than outside their own fields of study. But in light of what the RfU teams learned, particularly about just how hard it is to maintain the momentum needed to sustain implementation fidelity—and even more particularly for sustaining collaborative deeper learning practices—perhaps the time has come for scholars who do efficacy studies and RCTs within curricular settings to incorporate even more principles and tools from other fields. In particular, we think that pedagogical researchers have much to learn from the relatively new but rapidly exploding field of improvement science (Bryk, Gomez, Grunow, & LeMahieu, 2015; LeMahieu, Grunow, Baker, Nordstrom, & Gomez, 2017). Important in the field of improvement science is moving toward metrics that assess not only what individual players are learning (e.g., measures of *student* learning or *teacher* fidelity) but also indicators of *system*

learning, where the degree to which entities like schools, districts, and collaboratives are also assessed for the enhancements or barriers they construct in reform efforts. In the process, some constructs change. So, for example, implementation fidelity gets replaced by implementation integrity (LeMahieu, 2011), where the consequential index is not how closely the implementation of the reform matches the “ideal” but how well it is situated in a particular context of implementation. We think research efforts, even tightly controlled RCTs, would benefit from a more ecologically sensitive approach to examining the constraints and affordances of implementation, especially when we have compelling evidence of their consequential influence on research outcomes.

**Recommendation 3: We need to add both breadth and depth to our study of the knowledge-comprehension relationship.** The aphorism that we learn what is new in terms of what we already know has been with us probably from the onset of human cognition—as a matter of folk wisdom—and from the early days of educational and psychological studies of human cognition (e.g., Thorndike, 1917)—as a matter of empirical documentation. And the RfU scholars often referred to it as a key factor in both research design (e.g., Vaughn et al., 2015) and interpreting results (Goldman, Snow, & Vaughn, 2016). In particular, we see two directions for this expansion, one that looks inward to the RfU work and the other more outward looking.

*Testing the power of the RfU-expanded view of knowledge to help us understand and improve comprehension.* We acknowledge the importance of the RfU initiative’s contribution of highlighting the role that disciplinary and epistemic knowledge—over and above the traditional triad of declarative, procedural, and conditional knowledge—play in describing and improving what students must do to read complex content with deep understanding. At the same time, we assert that we have much to learn about the potential value added of these newer forms of knowledge. At a basic level, we do not know how independent these allegedly distinct forms of knowledge are. Do they develop independently or in concert with one another? Are these two new categories important only for older students, beginning perhaps in middle school, or are they equally important for younger readers? In what ways do students really learn to read like historians or apply knowledge like scientists as they advance through school? If one looks at the Common Core State Standards (and other related state standards documents) or even the Reading Framework for the National Assessment of Educational Progress, our current sources of guidance assume a march toward disciplinarity in thinking about pedagogy and assessment in the service of reading for understanding; indeed, the disciplinary grounding of the RfU work in curriculum and instruction reinforces that perspective. But there are basic and applied research efforts that should be undertaken before we dismiss the idea that there might also be some value in more generic constructs and instructional practices. We think the expansion from the RfU work is important and influential; however, prudence suggests that we continue to examine and refine the power of this expansion.

*Expand the scope of the work on the relationship between knowledge and comprehension.* In addition to incorporating research on newer categories of knowledge championed in the RfU portfolio, there is still a great deal of unfinished business on the knowledge-comprehension relationship within the realm of more conventional categories of knowledge, such as the familiar triad of declarative, procedural, and conditional knowledge.

*Building knowledge within language arts instruction.* The RfU research has added to the substantial body of research documenting the significant, positive impact of topic knowledge on reading comprehension, particularly among adolescent readers. Children and adolescents are asked to read texts on a wide range of topics in their lives as students. Starting early in building knowledge of the topics they are likely to encounter is one of the most promising ways to ensure they will successfully comprehend the increasingly complex texts they encounter. As many literacy researchers have pointed out, we have often viewed reading primarily as an opportunity for strategy and skill development, even when the texts are content rich (e.g., Neuman & Celano, 2006; Norris et al., 2008; Palincsar & Duke, 2004), and we have reduced time devoted to content area instruction in the early grades, often without considering the consequences for students' continued literacy development. Even so, the teaching profession is faced with the strong likelihood that the English Language Arts (ELA) block will continue to dominate curricular space (over science and social studies) at the elementary level. We should evaluate opportunities for students to engage with content-rich reading and learning from the earliest years of schools, and ask how ELA instruction can be put to work in building students' knowledge of the natural and social world. A starting point would be to think of literature, as some RfU efforts did, as rich in the content of understanding the human experience—with an emphasis on the big themes of love, friendship, conflict, betrayal, empathy, interacting with the environment, and the like. Literature may prove to be as rich a source of knowledge as science and history.

*Leveraging knowledge for other facets of literacy development.* The RfU research—and much research on knowledge and comprehension—has focused on the role of topic knowledge in helping students comprehend text by filling gaps and establishing conceptual coherence. A small, but intriguing body of research suggests that knowledge may have a broader role to play in literacy development, supporting students' incidental acquisition of word knowledge as they read (e.g., Barnes, Ginther, & Cochran, 1989; Cervetti, Wright, & Hwang, 2016; Kaefer, Neuman, & Pinkham, 2015; Pulido, 2004), and supporting their acquisition of comprehension strategies (Gaultney, 1995). Future research might well focus on this sort of reciprocity. A common heuristic among practicing teachers goes something like this: if you are presenting a new process for students, situate it in familiar content; and if you are presenting new content, situate it in a familiar process. Investigating the efficacy of such a heuristic could add valuable insights to how we think about the knowledge–literacy relationship.

*Leveraging student interests and cultural knowledge.* An essential and complex question for future research is how to leverage students' experiential and cultural knowledge in the interest of their literacy development. Studies have demonstrated that cultural knowledge supports students' text comprehension (e.g., Bell & Clark, 1998; Kelley, Siwatu, Tost, & Martinez, 2015; McCullough, 2013; Pritchard, 1990; Pulido, 2004). Although there is promising research demonstrating that cultural knowledge impacts text comprehension, this type of knowledge has yet to be used purposefully in classroom instruction with the goal of supporting students' reading comprehension. There remains substantial work ahead in bringing together two rich research traditions: (1) research on instructional programs that build reading comprehension, and (2) research documenting the efficacy of sociocultural funds of knowledge (Moll,

Amanti, Neff, & Gonzalez, 1992) and both culturally sustaining (Paris, 2012) and culturally relevant (Ladson-Billings, 2014) pedagogy.

**Recommendation 4: More of our work on comprehension needs to be directed toward populations currently underserved in U.S. schools.** The list of currently marginalized populations is long because it includes cultural and minoritized groups and children of poverty irrespective of race, ethnicity, or home language. But at the top of the list should be emergent bilingual or translingual learners. The particular irony of EBs is that, even though they bring rich language experiences to the classroom, we seem unable to exploit their first language or interlingual/translanguage (first- to second-language connections) resources to craft effective programs for deep reading experiences in English as a second language. Developing curriculum, as well as assessments, that exploit their linguistic resources is a special challenge that scholars of comprehension need to embrace.

Two loosely coupled but separate issues complicate this recommendation. First, we need an explicit effort to include underserved populations in experimental studies

Close reading skills became a foundation of the class. We stopped reading just to read, but we began dissecting while we read. My ESL students were able to easily and readily admit when they came across words they struggled with. They began developing tools to deal with words they struggled with. They were able to visualize their reading throughout a text. They were able to set goals before reading a text.

—RfU Participating Teacher

for equity purposes. If they are included just incidentally (as a part of a random sample of the entire population, for instance), they will be underrepresented and inappropriate conclusions and recommendations about what works for particular populations will be made. Second, there are important theoretical issues about the relationship of language to knowledge and comprehension that can be addressed only if they are included. This is doubly important for EBs because, for them, knowledge is constant, but proficiency in the two languages in which they operate will vary. To fail to target this population is a missed opportunity

to better our understanding of the relationship between language, knowledge, and comprehension.

**Recommendation 5: Writing, especially writing in response to reading and learning from text, is a likely candidate for improving reading comprehension.** Writing as the natural complement to and outcome of reading comprehension (Collins, Lee, Fox, & Madigan, 2017; Graham & Hebert, 2011) was implicit in all of the middle and high school interventions—CCDD, PACT, and READI. Sometimes it took the form of group work that required students to collaborate on a joint project (PACT), sometimes the development of arguments about key issues in the text (READI), and sometimes short syntheses and perspective taking on key issues across a set of texts (WG). But in all cases, the writing tasks served the function of promoting integration of key ideas unpacked in one or more texts. Much remains to be examined vis-à-vis the role that writing plays in promoting integration and analysis of key textual ideas. An under-explored area is the role that writing can play in units in which writing is deployed

systematically as students encounter multiple texts along the way to producing some sort of culminating product (an argument, an essay, or even a website design) that documents what students have understood and learned across texts they have read.

**Recommendation 6: We need to redouble our efforts to understand, measure, and organize instructional experiences to promote students' language skill and knowledge.** The RfU teams embraced language as a major component of and/or contributor to reading comprehension in all three strands of their efforts—nature and development, assessment, and pedagogy. Language is reflected most broadly in the logic of the SVR ( $RC = LC \times DEC$ ). In this regard, one of the RfU teams (LARRC, 2017) has presented the case that the LC component in the SVR model can more profitably be thought of as oral language comprehension than just listening comprehension, to drive home the point that it is language, not just listening, that is necessary to have a better understanding of the nature and development of reading comprehension. Because this recommendation overlaps considerably with Recommendation 7 regarding the relative merits of assembly versus orchestration, we defer our list of language-related possibilities to the next recommendation.

**Recommendation 7: Given the prevalent tension within the RfU initiative between the assembly and orchestration models of skill acquisition, the field (perhaps with the leadership of IES) should undertake a major national initiative, including meta-analyses of existing research and new research studies, to evaluate the relative merits of competing theories of the process and pedagogical models of delivery.** Albeit with different terminology, the issue of which metaphor—assembly or orchestration—better captures the character of reading (and reading comprehension) development arose in each strand. Chapter 2 referred to the discrete versus connected development of skills. Chapter 3 contrasted process versus componential assessments of reading comprehension. Chapter 5 discussed the tension between assembled versus orchestrated approaches to individual skill instruction, acknowledging that this tension revealed, at its core, a pedagogical grain size issue.

As we suggest in a similar recommendation for pedagogy in Chapter 5, the RfU teams varied considerably in their theoretical position on this tension. 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 LC factor in the SVR formula ( $RC = LC \times DEC$ ) with a curated collection of language structures and routines that might ultimately drive reading comprehension. At the orchestrated activity end of the continuum stood 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 work of CCDD and LARRC leaned toward the READI end of the continuum, and PACT seems best positioned squarely in the middle. Much work needs to be completed on this important but enormously complex issue.

*Conduct close examinations of the skill infrastructure of older readers.* We have very elaborate analyses of the changing interrelationships among subword level, vocabulary, and



comprehension skills from pre-K through grades 4 or 5. Save for DIME and the recent work from the RfU (e.g., Francis et al., 2018; Jones et al., 2019), we do not possess a rich database for older readers. We need to expand our understanding of these interactive developments during adolescence.

*Follow through on the logic of the FCRR approach.* As we have suggested, one can conceptualize the FCRR approach as answering the question, “What might the infrastructure of the LC term in the SVR look like if we unpacked it with the same care and fervor as has been accomplished for the DEC term over the last 30 years?” Even though the initial attempt to accomplish that goal was only partially achieved (the mixed results and unfinished analyses reported in Chapter 4)—and even though the evidence for orchestration is stronger than for componential assembly—we think there is merit in staying the course to determine whether and what key malleable facets of LC might look like. In particular, moving beyond simple indicators of vocabulary acquisition (i.e., selecting definitions or words to fill a sentence slot) to consider more nuanced aspects of vocabulary and syntactic and pragmatic aspects of language is necessary before closing down such a line of inquiry.

*Consider the possibility of middle-ground approaches.* Here we suggest that there may be some middle ground between the “assembly” assumption that students only learn what we teach (so we make sure to teach everything separately and explicitly to some level of mastery) and the “orchestration” assumption that some combination of close reading, rich discussion, and collaboration in applying the fruits of comprehension (i.e., the knowledge and insight one acquires from such routines) to authentic real-world tasks will naturally improve skill infrastructure without the cumbersome baggage of heavy-duty skill and strategy instruction. Middle-ground positions might include:

- \* Emphasizing some “mini-assemblages” or skill clusters (e.g., causal reasoning, predicting, and inferring), and
- \* On-demand excursions into explicit instruction for components only when formative assessments suggest a mini-intervention.

We have too many convictions and too little empirical evidence to resolve or manage this tension. It is wise, we think, to devote more resources and conceptual energy to understanding and managing, if not resolving, these tensions.

**Recommendation 8: In future initiatives in which a separate team is charged with the responsibility of developing relevant assessments for the entire network, employ a different model of assessment development and utilization.** We identify two issues regarding the relationship between assessment and its use in evaluating matters of reading development and pedagogy. One is focused on timing, and a second on common measures.

*Ensure lead time for assessment development.* If the network involves a separate team devoted to the assessment of the core construct under study, along with the enabling skills that feed into it, provide the assessment group a substantial head start (3 years at a minimum) if the core teams examining development and pedagogy are expected to use these measures in their work.

*Require common measures across teams.* Regardless of the source of the assessments, IES should insist on a core of common measures across projects that focus on similar populations (e.g., kindergarten through grade 3 or grades 6–8). While in the RfU initiative three of the teams (CCDD, READI, and FCRR) used GISA (although FCRR did not use it for their core efficacy studies and CCDD used it only for WG) and both READI and CCDD used RISE (as a pretest control variable for READI and as a progress indicator for STARI within CCDD), there was no common measure across all teams and even when the same measure was used, as noted, it was used differentially. At the very least, common measures across teams would allow for more credible observations (but not direct comparisons) of outcomes across projects. The construct of common measures has been a part of cooperative research since the 1960s, when the First Grade Studies (Bond & Dykstra, 1967) required each of its 22 separately funded and enacted projects to use the same measures for both outcomes and key covariates. It was also a feature of the Follow-Through Studies (Stebbins et al., 1977). A core of common measures, with project-specific options, seems both wise and easy to implement. Even better would be initiative-tailored common measures, the very goal intended within the RfU initiative.

**Recommendation 9: Issues of affect and conation should be at the forefront of reading comprehension research.** Since the onset of the cognitive revolution in the 1970s, pedagogical research about reading comprehension has been dominated by cognitive strategies and skills. The same could be said of curricular and pedagogical practices devoted to reading comprehension in our schools. The teams in the RfU initiative sought a path to reading achievement marked by innovative curriculum and dedicated teacher professional development in teaching comprehension. While this innovative disposition resulted in many attempts to bring affective or conative factors into the work, as we documented in Chapter 5 (see the Metacognition section on p. 236 and the Engagement section on p. 240), the teams maintained a strong emphasis on cognitive strategy and skill.

While necessary for reading success, cognitive strategies and skills cannot do the job on their own. Readers must be motivated and engaged, and they must possess the self-efficacy that helps power them through challenging texts and tasks. This is especially so for students who struggle in our schools. These are often students whose affective and conative dispositions are ill fitted with the school version of successful reading. We have a critical mass of research that examines students' affect and conation in relation to reading development and reading achievement. Going forward, innovative research should propose the productive marriage of cognition, affect, and conation. All three are essential to development and achievement, yet it is rare to encounter reading comprehension research or instruction based on this acknowledgment.

## THE GRAND QUESTION

We close by providing our answer to the grand question: What rewards did we reap from the RfU initiative? There are a few different ways to answer the question.

In the core chapters, we have answered that grand question finding by finding, study by study, issue by issue, theme by theme, and insight by insight across the

portfolio of work produced by these six teams across the past decade—the 5 years of funding plus the extensions and the lingering trail of publications.

We can answer the question with numbers. We could tell you that more than 100 scholars worked with hundreds of teachers and thousands of students in scores of schools in the majority of states across the duration of the RfU initiative. They produced more than 300 research publications, the vast majority of which appeared in top-tier refereed journals. They improved the lives of the students reached by their research by achieving average effect sizes for interventions in the very respectable range of .20 to .80—a sizable and reliable advantage over the control groups. We could even tell you how many downloads and citations their research reports and curricular materials have garnered across the years.

Or we could tell you a completely different story to mark their achievement. We could tell you what a singular qualitative achievement it was to persuade all of those teachers (and all of those students) in all of those schools to work hard to acquire new knowledge, new routines, and new expectations for what students can and should do, and what an achievement it was for the teachers to go on to deliver suitable instruction to all of those students—to get the students to do things that were out of their comfort zone.

We could assert that it is hard, but not impossible, to move that stubborn, sticky reading achievement needle. But it takes a lot of effort and stamina to do so. Teachers have to overcome the temptation to pick the low-hanging pedagogical and curricular fruit and search instead for the higher-hanging and more rewarding fruit—close reading, critique, and a search for evidence to support explanations and arguments. Teachers have to ignore, or work around, the barriers of the required curriculum and misguided accountability schemes with all of the test prep.

Even so, we know from the RfU work that the higher-hanging fruit can be reached, and that when the practices hanging up there are implemented with integrity, they can mediate student learning. This suggests key elements of the success of the effort:

- For most of the projects, there were strong and supportive professional learning communities that maintained high standards and offered sustained support in the form of coaching and careful monitoring.
- Those communities allowed teachers to implement and even sustain engaging but challenging practices.
- Those practices promoted wide and deep student engagement in collaborative discussions about interesting and edgy texts.
- Those conversations were on the pathway not just to comprehension but to applying what students learned to explanations and arguments about important ideas.

What this means is that the job of comprehension is not complete until one uses the resulting understanding to *do* something—tell a story, explain a situation, argue with an author or a classmate, or maybe even plan to change the world. In short, one reading of the RfU is that it has given us a glimpse of what an alternative culture of comprehension pedagogy might look like. The RfU initiative has led us part of the way down that path. And the legacy they left us—both in terms of what we learned and still need to learn—is surely an important road map for taking the next steps in unpacking this important pedagogical puzzle around comprehension and learning in the presence of texts.

## REFERENCES

- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M. A., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44*, 68–82.
- Ajayi, L. (2015). Critical multimodal literacy: How Nigerian female students critique texts and reconstruct unequal social structures. *Journal of Literacy Research, 47*(2), 216–244.
- Barnes, J. A., Ginther, D. W., & Cochran, S. (1989). Schema and purpose in reading comprehension and learning vocabulary from context. *Reading Research and Instruction, 28*(2), 16–28.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.
- Begoray, D., Higgins, J. W., Harrison, J., & Collins-Emery, A. (2013). Adolescent reading/viewing of advertisements: Understandings from transactional and positioning theory. *Journal of Adolescent & Adult Literacy, 57*(2), 121–130.
- Bell, Y. R., & Clark, T. R. (1998). Culturally relevant reading material as related to comprehension and recall in African American children. *Journal of Black Psychology, 24*, 455–475.
- Biancarosa, G., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.
- Bloom, H. S., Hill, C. J., Black, A. B., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328.
- Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2*(4), 5–142. doi: 10.2307/746948.
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*, 893–900.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Castek, J., Coiro, J., Guzniczak, L., & Bradshaw, C. (2012). Examining peer collaboration in online inquiry. *The Educational Forum, 76*(4), 479–496.
- Catts, H. W. (2018). The Simple View of Reading: Advancements and false impressions. *Remedial and Special Education, 39*(5), 317–323.
- Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing, 29*(4), 761–779.
- Cho, B. Y., & Afflerbach, P. (2015). Reading on the Internet: Realizing and constructing potential texts. *Journal of Adolescent & Adult Literacy, 58*(6), 504–517.
- Cho, B. Y., Woodward, L., & Li, D. (2017). Epistemic processing when adolescents read online: A verbal protocol analysis of more and less successful online readers. *Reading Research Quarterly*. doi: 10.1002/rrq.190.
- Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12.
- Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria seventh graders use to evaluate the quality of online information. *Journal of Adolescent and Adult Literacy, 58*(7), 546–550.
- Coiro, J., Sekeres, D. C., Castek, J., & Guzniczak, L. (2014). Comparing the quality of third, fourth, and fifth graders' social interactions and cognitive strategy use during structured online inquiry. *Journal of Education, 194*(2), 1–15.
- Collins, J. L., Lee, J., Fox, J. D., & Madigan, T. P. (2017). Bringing together reading and writing: An experimental study of writing intensive reading comprehension in low-performing urban elementary schools. *Reading Research Quarterly, 52*(3), 311–332.
- Connor, C. M., Phillips, B. M., Kaschak, M., Apel, K., Kim, Y. S., Al Otaiba, S., Crowe, E. C., Thomas-Tate, S., Johnson, L. C., & Lonigan, C. J. (2014). Comprehension tools for teachers: Reading for understanding from prekindergarten through fourth grade. *Educational Psychology Review, 26*(3), 379–401. doi: 10.1007/s10648-014-9267-1.

- Connor, C. M., Phillips, B. M., Kim, Y. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & 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.
- Cromley, J., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325.
- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176.
- Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. doi: 10.1016/j.edurev.2018.09.003.
- Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*, 1–44.
- Fitzgerald, M., Higgs, J., & Palincsar, A. S. (2020). *New media, new literacies: Implications for reading for understanding*. Washington, DC: National Academy of Education.
- Foorman, B., Dombek, J., & Smith, K. (2016). Seven elements important to successful implementation of early literacy intervention. *New Directions for Child and Adolescent Development, 154*, 49–65.
- Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the Simple View of Reading to account for variance within and across readers: The Complete View of Reading (CVRI). *Remedial and Special Education, 39*, 274–288.
- Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Tracy, W. Z., Patton III, S., Davis-Perkins, N., Kim, W., Elleman, A. M., & Peng, P. (2018). Evaluating the efficacy of a multidimensional reading comprehension program for at-risk students and reconsidering the lowly reputation of tests of near transfer. *Learning Disabilities Research and Practice, 33*(1), 11–23. doi: 10.1111/ldrp.12162.
- Gaultney, J. F. (1995). The effect of prior knowledge and metacognition on the acquisition of a reading comprehension strategy. *Journal of Experimental Child Psychology, 59*(1), 142–163.
- Gersten, R. (2016). Commentary: The tyranny of time and the reality principle. *New Directions for Child and Adolescent Development, 154*, 49–65.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.
- Goldman, S. R. (2018). Discourse of learning and the learning of discourse. *Discourse Processes, 55*(5–6), 434–453.
- Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, & Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly, 47*(4), 356–381.
- 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. doi: 10.3102/0002831219831041.
- Goldman, S. R., Snow, C., & Vaughn, S. (2016). Common themes in teaching reading for understanding: Lessons from three projects. *Journal of Adolescent & Adult Literacy, 60*(3), 255–264. doi: 10.1002/jaal.586.
- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a Simple View of Reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710–744.

- 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*, 172–177.
- Hoover, W. A., & Tunmer, W. E. (2018). The Simple View of Reading: Three assessments of its adequacy. *Remedial and Special Education, 39*(5), 304–312.
- Hull, G., & Schultz, K. (2001). Literacy and learning out of school: A review of theory and research. *Review of Educational Research, 71*(4), 575–611.
- Hutchinson, A. C., Woodward, L., & Colwell, J. (2016). What are preadolescent readers doing online? An examination of upper elementary students' reading, writing, and communication in digital spaces? *Reading Research Quarterly, 51*(4), 435–454.
- IES (Institute of Education Sciences). (2009). *Request for applications: Reading for understanding research initiative*. Retrieved from [https://ies.ed.gov/funding/pdf/2010\\_84305F.pdf](https://ies.ed.gov/funding/pdf/2010_84305F.pdf).
- Ito, M., Arum, R., Conley, D., Gutiérrez, K., Kirshner, B., Livingstone, S., Michalchik, V., Penuel, W., Pepler, K., Pinkard, N., Rhodes, J., Tekinba, K. S., Schor, J., Sefton-Green, J., & Watkins, S. C. (2020). *The Connected Learning Research Network: Reflections on a decade of engaged scholarship*. Irvine, CA: Connected Learning Alliance.
- Jimenez, L. M., & Meyer, C. K. (2016). First impressions matter: Navigating graphic novels utilizing linguistic, visual, and spatial resources. *Journal of Literacy Research, 48*(4), 423–447.
- Jones, S. M., LaRusso, M., Kim, J., Kim, H. Y., Selman, R., Uccelli, P., Barnes, S. P., Donovan, S., & Snow, C. (2019). Experimental effects of Word Generation on vocabulary, academic language, perspective taking, and reading comprehension in high poverty schools. *Journal of Research on Educational Effectiveness, 12*, 448–483.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447.
- Kaefer, T., Neuman, S. B., & Pinkham, A. M. (2015). Pre-existing background knowledge influences socioeconomic differences in preschoolers' word learning and comprehension. *Reading Psychology, 36*(3), 203–231. <https://doi.org/10.1080/02702711.2013.843064>.
- Kelley, H. M., Siwatu, K. O., Tost, J. R., & Martinez, J. (2015). Culturally familiar tasks on reading performance and self-efficacy of culturally and linguistically diverse students. *Educational Psychology in Practice, 31*(3), 293–313. <https://doi.org/10.1080/02667363.2015.1033616>.
- Kennedy, C., Rhoads, C., & Leu, D. J. (2016). Online research and learning in science: A one-to-one laptop comparison in two states using performance based assessments. *Computers and Education, 100*, 1–21.
- Kesler, T., Gibson, Jr., L., & Turansky, C. (2016). Bringing the book to life: Responding to historical fiction using digital storytelling. *Journal of Literacy Research, 48*(1), 39–79.
- 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.
- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: a.k.a. the remix. *Harvard Educational Review, 84*(1), 74–84.
- LARRC (Language and Reading Research Consortium). (2015a). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*, 151–169.
- LARRC. (2015b). The dimensionality of language ability in young children. *Child Development, 86*, 1948–1965.
- LARRC. (2015c). The dimensionality of Spanish in young Spanish-English dual-language learners. *Journal of Speech, Language, and Hearing Research, 58*, 754–766.
- LARRC. (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*, 1273–1284.
- LARRC & Chiu, Y. D. (2018). The Simple View of Reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education, 39*(5), 289–303.
- 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, Johanson, M., & Arthur, A. M. (2016). Improving the language skills of pre-kindergarten students: Preliminary impacts of the Let's Know! experimental curriculum. *Child & Youth Care Forum, 45*(3), 367–392.
- LaRusso, M. D., Donovan, S., & Snow, C. (2016). Implementation challenges for tier one and tier two school-based programs for early adolescents. *New Directions for Child and Adolescent Development, 154*, 11–30.

- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., Donovan, S. & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research in Educational Effectiveness*, 9, 201–222.
- Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, 52(4), 750–786. doi: 10.3102/0002831215579485.
- Lawrence, J. F., Rolland, R. G., Branum-Martin, L., & Snow, C. E. (2014). Generating vocabulary knowledge for at-risk middle school readers: Contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk*, 19(2), 76–97.
- Lee, C. D., 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.
- Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. Final Report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy. New York: Carnegie Corporation.
- LeMahieu, P. G. (2011, October 11). What we need in education is more integrity (and less fidelity) of implementation [Blog post]. Retrieved from <http://commons.carnegiefoundation.org/what-we-are-learning/2011/what-we-need-in-education-is-more-integrity-and-less-fidelity-of-implementation>.
- LeMahieu, P. G., Grunow, A., Baker, L., Nordstrum, L. E., & Gomez, L. M. (2017). Networked improvement communities: The discipline of improvement science meets the power of networks. *Quality Assurance in Education*, 25(1), 5–25.
- Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, 45(2), 196–228.
- Leslie, L., & Caldwell, J. (2006). *Qualitative Reading Inventory, 4th edition*. New York: Pearson.
- Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly*, 50(1), 37–59.
- Lonigan, C. J. (2016, May). The FCRR Reading for Understanding Network: Challenges and Unanswered Questions in Pre-K to Grade 3. Paper presented at the National Symposium on Reading for Understanding. Alexandria, VA.
- Lonigan, C. J., & Burgess, S. R. (2017). Dimensionality of reading skills with elementary-school-age children. *Scientific Studies of Reading*, 21, 239–253.
- Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the Simple View of Reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, 39, 260–273.
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Research*, 48(3), 158–166. doi: 10.3102/0013189X19832850.
- McCullough, R. G. (2013). The relationship between reader response and prior knowledge on African American students' reading comprehension performance using multicultural literature. *Reading Psychology*, 34(5), 397–435.
- McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165–193.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31(2), 132–141.
- Moran, J., Ferdig, R. E., Pearson, P. D., Wardrop, J., & Blomeyer, R. L. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research*, 40(1), 6–58.
- Neuman, S. B., & Celano, D. (2006). The knowledge gap: Implications of leveling the playing field for low income and middle-income children. *Reading Research Quarterly*, 41(2), 176–201.
- Norris, S. P., Phillips, L. M., Smith, M. L., Guilbert, S. M., Stange, D. M., Baker, J. J., & Weber, A. C. (2008). Learning to read scientific text: Do elementary school commercial reading programs help. *Science Education*, 92(5), 765–798.

- NRC (National Research Council). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- NRC. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st Century*. Committee on Defining Deeper Learning and 21st Century Skills. Washington, DC: The National Academies Press.
- Palincsar, A. S., & Duke, N. K. (2004). The role of text and text-reader interactions in young children's reading development and achievement. *The Elementary School Journal*, 105(2), 183–197.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97.
- Park, J. Y. (2016). "He didn't add more evidence": Using historical graphic novels to develop language learners' disciplinary literacy. *Journal of Adolescent & Adult Literacy*, 60(1), 35–43.
- Pearson, P. D., & Cervetti, G. N. (2017). The roots of reading comprehension instruction. In S. E. Israel (Ed.), *Handbook of research on reading comprehension, 2nd edition* (pp. 12–56). New York: Guilford Press.
- Pearson, P. D., Valencia, S. W., & Wixson, K. (2014). Complicating the world of reading assessment: Toward better assessments for better teaching. *Theory into Practice*, 53(3), 236–246. doi: 10.1080/00405841.2014.916958.
- Phillips, B. M., Kim, Y.-S., Lonigan, C. J., & Connor, C. M. (2015, July). *Language for understanding: Two large-scale studies of small-group language interventions in prekindergarten and kindergarten*. Paper presented at the annual meeting of the Society for the Scientific Study of Reading, Kona, HI.
- Piasta, S., Language and Reading Research Consortium, & Jiang, H. (2016, July). *Targeting lower- and higher-level language skills to support comprehension: Initial results for Let's Know*. Paper presented at the annual conference of the Society for the Scientific Study of Reading, Porto, Portugal.
- Pinkard, N., Erete, S., Martin, C., & McKinney de Royston, M. (2017). Digital Youth Divas: Exploring narrative-driven curriculum to trigger middle school girls' interest in computational activities. *Journal of the Learning Sciences*, 26(3), 477–516.
- Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25(4), 273–295.
- Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 54(3), 469–523.
- 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.
- Sabatini, J. P., & O'Reilly, T. (2013). Rationale for a new generation of reading comprehension assessments. In B. Miller, L. E. Cutting, & P. McCardle (Eds.), *Unraveling reading comprehension: Behavioral, neurobiological and genetic components* (pp. 100–111). Baltimore: Paul H Brookes.
- Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, 48(4), 369–390. doi: 10.1177/0022219413504995.
- Shanahan, C., Bolz, M., Cribb, G., Goldman, S., Heppeler, J. & Manderino, M. (2016). Deepening what it means to read (and write) like a historian: Progressions of instruction across a school year in an eleventh grade U.S. history class. *The History Teacher*, 49(2), 1–30.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40–59.
- Snow, C. E. (2018). Simple and not-so-simple views of reading. *Remedial and Special Education*, 39, 313–316. doi: 10.1177/0741932518770288.
- Snow, C. E., & Moje, E. B. (2010). Why is everyone talking about adolescent literacy? *Phi Delta Kappan*, 91(6), 66–69.
- Sommer, S., Hinojosa, L., Traut, H., Polman, J., & Weidler-Lewis, J. (2017). *Integrating eye-tracking activities into a learning environment to promote collaborative meta-semiotic reflection and discourse*. Philadelphia: International Society of the Learning Sciences.
- Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model* (Vol. IV-A). Cambridge, MA: Abt Associates.



- Sun, J., Anderson, R. C., Lin, T., Morris, J. A., Biller, B. W., Ma, S., Nguyen, K. T., & Scott, T. (2020). Children's engagement during collaborative learning and direct instruction. Manuscript under review.
- Thorndike, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology, 8*(6), 323–332. doi: 10.1037/h0075325.
- Valencia, S., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *Elementary School Journal, 115*(2), 270–289.
- 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*, 31–50.
- Wagner, R. K., Herrera, S. K., Spencer, M., & Quinn, J. M. (2015). Reconsidering the Simple View of Reading in an intriguing case of equivalent models. *Journal of Learning Disabilities, 48*, 115–119.
- Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology, 111*(3), 387–401.
- Wanzek, J., Swanson, E., Vaughn, S., Roberts, G., & Fall, A.-M. (2016). English learner and non-English learner students with disabilities: Content acquisition and comprehension. *Exceptional Children, 82*, 428–442.
- Wanzek, J., & Vaughn, S. (2016). Implementation of a text-based content intervention in secondary social studies classes. *New Directions for Child and Adolescent Development, 154*, 31–48.
- Wilkinson, I. & Son, E. (2011). A dialogic turn in research on learning and teaching to comprehend. In M. L. Kamil, P. D. Pearson, E. Moje, P. Afflerbach (Eds.), *Handbook of reading research (volume IV)* (pp. 359–387). New York: Routledge.
- Wood, D. R., & Lieberman, A. (2000). Teachers as authors: The National Writing Project's approach to professional development. *International Journal of Leadership in Education, 3*(3), 255–273.

## Biographical Sketches of Steering Committee Members and Authors

**Peter Afflerbach** is a professor of education at the University of Maryland at College Park. His research interests are individual differences in reading, reading comprehension strategies for print and digital reading, reading assessment, conation and affect in reading, and the verbal reporting methodology. Afflerbach has served on National Academy of Education and National Academy of Sciences committees related to literacy. He is a member of the National Assessment of Educational Progress (NAEP) 2025 Reading Framework Development Committee, and has served on the NAEP Standing Reading Committee and prior Reading Framework Committees. Afflerbach is Chair of the Literacy Assessment Task Force of the International Literacy Association. He was elected to the International Literacy Association's Reading Hall of Fame in 2009. Afflerbach is the editor of the *Handbook of Individual Differences in Reading: Reader, Text, and Context* (2016), and the co-editor of the *Handbook of Reading Research, 4th Edition* (2010) and *5th Edition* (in press). He is a founding editor of the journal *Metacognition and Learning*, and he has published in numerous theoretical and practical journals, including *Reading Research Quarterly*, *Cognition and Instruction*, *Elementary School Journal*, *Journal of Adolescent and Adult Literacy*, *Language Arts*, *Theory into Practice*, and *The Reading Teacher*. Prior to his work with universities, Afflerbach had broad classroom experience teaching Chapter 1 remedial reading, reading and writing in middle school, and high school English.

**Amy I. Berman** is the deputy director of the National Academy of Education, where she works to advance the strategic and research initiatives for the more than 200-member organization. She also is an adjunct professor of school and the law at The George Washington University Graduate School of Education and Human Development (GSEHD). Prior to that she was an education civil rights lawyer as an Enforcement Director at the U.S. Department of Education's Office for Civil Rights and Section Chief at the U.S. Department of Justice's Civil Rights Division's Educational Opportunities

Section. In those positions, she worked to ensure equal access to education through the vigorous enforcement of civil rights laws, including in the areas of race, national origin, sex, religion, disability, and language. In addition to enforcement, she worked on key guidance documents addressing the use of race in schools, harassment in schools, education of English learner students, and the requirement to educate all students, regardless of immigration status. She also held a fellowship addressing prisoners' civil rights issues, worked as an associate at Sullivan & Cromwell, and clerked for the late Honorable Robert J. Ward of the U.S. District Court for the Southern District of New York. She has been an adjunct professor at The George Washington University School of Law and the American University Washington College of Law. She is a doctoral candidate in education policy at GSEHD, and she holds a J.D. from Harvard Law School and a B.S. from Cornell University.

**Gina Biancarosa** is the Ann Swindells Chair in Education and an associate professor in the Department of Special Education and Clinical Sciences in the College of Education at the University of Oregon (UO). Biancarosa also maintains an affiliation with the Department of Educational Methodology, Policy, and Leadership and serves as the director of its Ph.D. program in quantitative research methods. She conducts her research at the Center on Teaching and Learning. Biancarosa's research centers on questions related to more precise and informative measurement of reading, especially of reading comprehension, and modeling of growth in reading over time. To date, she has led the development of MOCCA, a diagnostic measure of cognitive processes involved in reading comprehension for grades 3–5, and Dynamic Indicators of Basic Early Literacy Skills, 8th edition, which now covers kindergarten through grade 8 and offers better precision and more information for teachers on all learners than previous editions. Biancarosa's research has been published in top peer-reviewed journals, such as the *Journal of Educational Measurement*, *Educational and Psychological Measurement*, *Assessment for Effective Intervention*, *Journal of School Psychology*, *Reading and Writing*, *Learning and Individual Differences*, and *Elementary School Journal*. She serves as an associate editor of the *British Journal of Educational Psychology*, and on the editorial boards of *Educational Assessment*, *Elementary School Journal*, *Journal of Educational Psychology*, and *Reading Research Quarterly*. She has served as a reading, adolescent literacy, and applied measurement expert on several review panels, including panels for the Institute of Education Sciences (IES), National Assessment of Educational Progress (NAEP), Programme for International Student Assessment (PISA), and the Oregon Department of Education. Before coming to UO, she was a reading specialist and an after-school teacher in Massachusetts and completed her master's and doctoral studies in language and literacy at the Harvard Graduate School of Education. She also completed a 3-year IES-sponsored postdoctoral fellowship at Stanford University. She has received more than \$2 million in external funding for her research from IES, the Spencer Foundation, and the Carnegie Corporation of New York, as well as \$100,000 in internal funding for curricular endeavors at UO. Biancarosa regularly teaches courses on measurement and assessment and applied statistics, as well as occasional courses on reading theory and methods.

**Gina N. Cervetti** is an associate professor of literacy, language, and culture at the University of Michigan, where she teaches and studies literacy development and

instruction. Her work involves three central concerns: the potential benefits of content-area instruction for literacy development, the role of world knowledge in literacy development, and the nature of vocabulary/language instruction that supports reading comprehension. She has been particularly interested in science as a context for elementary students' reading, writing, and language development. She has examined how the collaborative, experiential, and knowledge-enhancing qualities of inquiry-based science instruction can fuel students' engagement and growth in literacy. She has served as a principal investigator on several grants investigating integrated science-literacy instruction with a focus on how science might serve as an especially rich pedagogical context for English language learners. Her work has appeared in many books and journals. Cervetti serves on the board of the Literacy Research Association, on the editorial boards of several literacy journals, and on several national panels, including the panel that is currently developing the new framework for the National Assessment of Educational Progress (NAEP). Following her doctoral work in educational psychology at Michigan State University, Cervetti worked for several years as a postdoctoral scholar and researcher at the University of California, Berkeley, on the Seeds of Science/Roots of Reading program. Cervetti joined the University of Michigan following 3 years as an assistant professor at the University of Colorado Boulder.

**Donald Compton** is a professor of psychology at Florida State University/Florida Center for Reading Research. He was formerly a professor and the Chair of Special Education and a John F. Kennedy Center Investigator at Peabody College at Vanderbilt University. He earned a Ph.D. from the Northwestern University School of Communication Sciences and Disorders, with a specialization in learning disabilities. While working on his Ph.D., and for several years after its completion, he was employed as a learning disabilities resource teacher in Skokie, Illinois. Compton then worked for 4 years as an assistant professor in the Department of Curriculum and Instruction at the University of Arkansas, Fayetteville. He then accepted a National Institute of Child Health and Human Development (NICHD) postdoctoral research fellowship at the Institute for Behavior Genetics, University of Colorado. From there he accepted a position at Vanderbilt University that he held until spring 2015. He then accepted his current position with the Florida Center for Reading Research at Florida State University. Compton is experienced in designing, managing, analyzing, and disseminating data from cross-sectional and longitudinal studies as well as randomized controlled studies. His research involves modeling individual differences in the development of children's reading skills and the identification and treatment of children with reading disabilities. Compton has served as principal investigator (PI) and co-PI on multiple NICHD and Institute of Education Sciences studies using randomized controlled trials to evaluate academic interventions for children with learning difficulties. He has more than 75 peer-reviewed publications and is on the editorial boards of the *Journal of Educational Psychology*, *Journal of Learning Disabilities*, *Scientific Studies of Reading*, and *Exceptional Children*. Compton is the past president of the Society for the Scientific Study of Reading and currently serves as an associate editor of *Scientific Study of Reading*.

**Miranda S. Fitzgerald** is an assistant professor and a teacher educator at the University of North Carolina at Charlotte. Fitzgerald's research interest is in instructional

contexts that support literacy development and disciplinary knowledge building in the intermediate grades, particularly those that foster equitable opportunities to learn for diverse and historically underserved students. Using qualitative methods and socio-cultural theories of learning, she investigates meaningful uses of multimodal literacy in intermediate-grade classrooms. Fitzgerald's current research focuses on the integration of language literacy and science instruction, especially in the context of project-based learning, and strategies to support pre- and in-service teachers to enact ambitious and equitable instructional practices in literacy. In 2017, she was awarded the International Literacy Association's Steven A. Stahl Research Grant and an American Educational Research Association (AERA) Division C Graduate Student Research grant to partially fund her dissertation research. She designed and conducted her dissertation study, *Texts and Tasks in Elementary Project-Based Science*, in the context of the Multiple Literacies in Project-based Learning project. For this research, she was the recipient of the 2019 AERA Research in Reading and Literacy Special Interest Group's (SIG's) Graduate Student Award for Literacy Excellence, the 2019 University of Michigan School of Education's Stanley E. and Ruth B. Dimond Best Dissertation Award, and was a finalist for the 2019 International Literacy Association (ILA) Timothy & Cynthia Shanahan Outstanding Dissertation Award. Her work has been published in *Review of Research in Education*, *Reading & Writing Quarterly*, and *International Journal of Educational Research*. Fitzgerald completed her doctorate at the University of Michigan.

**Susan R. Goldman** is the Distinguished Professor of Liberal Arts & Sciences, Psychology, and Education and co-directs the Learning Sciences Research Institute at the University of Illinois at Chicago. Goldman's research focuses on understanding and improving the learning processes of educators as well as those they educate. Her work emphasizes the implicit and explicit literacies inherent in the inquiry practices of a variety of disciplines in the physical and biological sciences, social sciences, mathematics, and language arts and literature. Goldman's work has included curriculum development along with teaching and technological innovations that foster improved learning for children and adults. Most recently, her research has focused on instruction and classroom discussion that promote critical inquiry, deep reading, and reasoning on the part of students in science, mathematics, history, and literature/literary reading. Central to this research is creating opportunities for teachers to learn what they need to know and to successfully support students in developing these 21st-century competencies in a range of academic disciplines. She routinely collaborates with other researchers, teachers, and other educational practitioners to bridge research and practice. Goldman was the principal investigator on Project Reading, Evidence, and Argumentation in Disciplinary Instruction (READI), a collaboration of five institutions and multiple schools and school districts. Current projects are building on that work to engage in Research Practice Partnerships.

**Kenji Hakuta** is the Lee L. Jacks Professor (emeritus) at the Stanford University Graduate School of Education. He received his Ph.D. in experimental psychology from Harvard University in 1979, has held faculty positions at Yale University and the University of California, Santa Cruz, and served as the founding dean of the School of Social Sciences, Humanities, and Arts at the University of California, Merced. He is a fellow of

the National Academy of Education, the American Association for the Advancement of Science, the American Academy of Arts & Sciences, and the American Educational Research Association. His publications are in the areas of psycholinguistics, bilingualism, language shift, the acquisition of English in immigrant students, and education policy. He has served on the boards of the Spencer Foundation and the Educational Testing Service, and currently serves on the board of the William T. Grant Foundation. For 8 years, he chaired the National Educational Research Policy and Priorities Board (predecessor of the Institute of Education Sciences [IES] Board) of the U.S. Department of Education.

**Jennifer Higgs** is an assistant professor of learning and mind sciences and language, literacy, and culture at the University of California, Davis (UC Davis). Using mixed methods and sociocultural theories of learning, Higgs investigates practices around digital tools as well as improvement of digital tool practices. She documents how digital technologies are used in and across school and nonschool settings and works alongside teachers and young people to design interventions that support equitable learning and are attuned to local needs. Currently, as a co-principal investigator on the UC Davis Teachers as Learners project, funded by the McDonnell Foundation, Higgs is helping to design and implement augmented reality scenarios for new teachers learning to facilitate text-based discussions in culturally and linguistically diverse classrooms. What fundamentally drives her work is a desire to learn how the possibilities afforded by new cultural tools can support robust and socially just learning. Prior to pursuing her Ph.D. at UC Berkeley, Higgs was a middle and high school English teacher in Virginia and Illinois public schools and an English teacher in Japan. She was a 2016 National Academy of Education/Spencer Dissertation Fellow, and her dissertation received the 2017 American Educational Research Association Division K Outstanding Dissertation Award and the Peter Lyman Award from the Berkeley Center for New Media.

**Glynda A. Hull** is a professor of education at the Graduate School of Education at the University of California, Berkeley (UC Berkeley), and holds the Elizabeth H. and Eugene A. Shurtleff Chair in Undergraduate Education. Most broadly her research has focused on writing as a sociocultural process. More specifically, over the past 15 years, she has studied meaning making in the context of, and in concert with, digital media and networked spaces, like social media and online courses. How are reading and writing changing in a digital age? What are the impacts on communication of dissolving time and space? How can teachers imagine networked spaces for learning that are ethically alert as well as attuned to the affordances of the digital? These are the questions that animate her research, teaching, and service. Hull began her career in education by teaching middle and high school English in two small towns in Mississippi. A recipient of UC Berkeley's Distinguished Teaching Award, she offers undergraduate and graduate courses on literacy, digital media, and teaching in higher education.

**Matthew Hurt** is a Ph.D. student in the Department of Teaching and Learning, Policy and Leadership at the University of Maryland. He began his career in education as an elementary school teacher after graduating with a B.A. from Virginia Tech. He later attended Vanderbilt University, where he received an M.P.P. in education policy from

the Peabody College of Education and Human Development. During that time, Hurt worked for the Tennessee Department of Education and aided in managing Tennessee's summer reading camp program. At the University of Maryland, Hurt has taught classes on literacy assessment in elementary grades and also on writing development for young students. Hurt's research focuses on quality literacy instruction and on analyzing district- and state-level policies to better understand how quality instructional practices can be effectively scaled up. His work has been presented at both the Literacy Research Association and the American Education Research Association's annual conferences.

**Laura M. Justice** is the Education and Human Ecology (EHE) Distinguished Professor and University Distinguished Scholar (2017) at The Ohio State University's College of EHE. She also is the executive director of the Crane Center for Early Childhood Research and Policy, which seeks to accelerate action-oriented research that improves practices and policies directed toward early childhood.

**Panayiota (Pani) Kendeou** is a professor in the Department of Educational Psychology and Guy Bond Chair in Reading at the University of Minnesota. In her research program, Kendeou examines the cognitive processes of reading comprehension in an effort to advance our theoretical understanding of reading and transform reading instruction in K–12 schools. She develops theoretical models that explain how students acquire and revise knowledge during reading and uses those models to design and test innovative, educational technologies that transform reading instruction and improve reading achievement. The significance of her approach is in the development of theoretically oriented work and its translation into school-based practical applications. Kendeou has published extensively in the area of reading comprehension and her work has received numerous awards, including the Early Career Impact Award from the Federation of Associations in Behavioral and Brain Sciences (FABBS) Foundation in 2015, the Tom Trabasso Young Investigator Award from the Society for Text & Discourse in 2012, and the Research in Literacy Award from the UK Literacy Association in 2009. Kendeou is currently the editor of the *Journal of Educational Psychology* and serves on the editorial boards of *Scientific Studies of Reading*, *Contemporary Educational Psychology*, *Learning and Instruction*, *Educational Psychologist*, *Educational Psychology Review*, and *Reading Research Quarterly*. She is a fellow of the American Psychological Association and a member of the American Educational Research Association, the Society for Text & Discourse, the Society for the Scientific Study of Reading, and the Psychonomics Society.

**Christopher J. Lonigan** is a distinguished research professor of psychology and an associate director of the Florida Center for Reading Research at Florida State University. He received his Ph.D. in clinical psychology from the State University of New York at Stony Brook. After completing his Ph.D., Lonigan was awarded a 2-year National Institutes of Health (NIH) postdoctoral fellowship at the Kennedy Center at Vanderbilt University to study developmental disabilities. In 2016, he was elected as a fellow of the American Psychological Society. For more than 20 years, Lonigan has served as a leading expert in the area of early literacy. He has published more than 160 scholarly articles and book chapters. His current research projects include early identification of reading disability, effectiveness of response-to-instruction models, development of

early academic skills of Spanish-speaking dual-language learners, the development of reading comprehension, and the co-development of self-regulation and academic skills. His research is or has been funded by the National Institute of Child Health and Human Development, the Institute of Education Sciences (IES), the National Science Foundation, and the Administration for Children, Youth, and Families (ACYF). In addition to his research, Lonigan has served in various roles in the education community in the United States. He has chaired or been a member of grant-review panels for IES, NIH, ACYF, and the U.S. Department of Education. He has been a member of various advisory committees for federal and state research and demonstration projects.

**Annemarie Sullivan Palincsar** is the Jean and Charles Walgreen Jr. Chair of Reading and Literacy, Arthur F. Thurnau Professor, and a teacher educator at the University of Michigan. Palincsar's primary research interest is in sensemaking and knowledge building with multiple media, especially in the context of project-based scientific inquiry. She has served on a number of task forces and National Research Council and National Academies of Sciences, Engineering, and Medicine panels, including Preventing Reading Difficulties in Young Children, Preparing Future Teachers, the RAND Reading Study Group, and How People Learn: Volume II. She has been active in professional organizations, serving as a member of the Literacy Research Panel for the International Reading Association (IRA), an executive board member for the National Academy of Education, and a member-at-large for the American Educational Research Association (AERA). She is a member of the National Academy of Education and is an AERA fellow. She was the co-editor of *Cognition and Instruction*. Her research has been published in the *Journal of the Learning Sciences*, *Educational Researcher*, *Harvard Education Review*, *Linguistics and Education*, *Teaching and Teacher Education*, and *Reading and Writing Quarterly*, among others. Palincsar completed her doctorate at the University of Illinois.

**P. David Pearson** is the Evelyn Lois Corey Emeritus Professor of Instructional Science in the Graduate School of Education at University of California, Berkeley (UC Berkeley). Officially retired since 2017, he continues to conduct historical and policy-oriented research on literacy (especially reading) theory, research, and practice and contributes to conversations in key professional organizations that nurture educational and literacy scholarship. Pearson pursues scholarship on reading instruction and assessment practices, with special attention to disciplinary literacy and comprehension instruction. He has many recognitions for his now more than 50 years of research and service to reading research, most notably the William S. Gray Citation of Merit from the International Reading (now Literacy) Association, the Oscar Causey Award from the Literacy Research Association, the Distinguished Contributions to Research in Education Award from the American Educational Research Association, and the Distinguished Alumni Award from the University of Minnesota. In 2010, the Literacy Research Association established the P. David Pearson Scholarly Influence Award to honor research that impacts practice and policy. He was elected to the Reading Hall of Fame in 1990 and the National Academy of Education in 2003. Pearson received his B.A. in history from UC Berkeley, after which he taught elementary school in California for several years, and went on to complete his Ph.D. in reading education at the University of Minnesota. He completed postdoctoral study at The University of Texas at Austin and Stanford University.



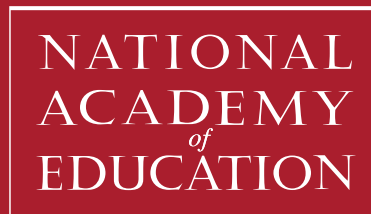
**John Sabatini** is a Distinguished Research Professor in the Institute for Intelligent Systems and the Department of Psychology at the University of Memphis. He was formerly a principal research scientist in the Center for Global Research, Research & Development Division at the Educational Testing Service. His research interests and expertise are interdisciplinary, extending across reading literacy development and disabilities, assessment, cognitive psychology, the learning sciences, and educational technology. In the course of his career, he has had opportunities to conduct research on reading, writing, and critical thinking; develop assessments, instructional programs, and educational technology tools; and participate in evaluation and implementation studies. While his primary focus has been on adolescent and adult learners, he has participated in research inclusive of learners from pre-kindergarten (pre-K) through adulthood, language learners, and neurodiverse subpopulations. He has been the principal investigator of an Institute of Education Sciences–funded grant to develop pre-K through grade 12 comprehension assessments as part of the Reading for Understanding initiative, and to adapt those assessments for use in adult education programs. Other major research efforts include a National Institute of Child Health and Human Development (NICHD)-funded Learning Disabilities Research Center project studying subtypes of reading disabilities in adolescents and an interagency federal grant to investigate the relative effectiveness of reading programs for adults. He provides technical and research advice to national and international surveys including the Programme for the International Assessment of Adult Competencies (PIAAC), the Progress in International Reading Literacy Study (PIRLS), the National Assessment of Educational Progress (NAEP), and the National Assessments of Adult Literacy (NAAL).

**Catherine E. Snow** is the Patricia Albjerg Graham Professor at the Harvard Graduate School of Education (HGSE). After receiving her Ph.D. in developmental psychology from McGill University, Snow worked for several years in the Linguistics Department of the University of Amsterdam before taking up a position at HGSE. She conducts research on language and literacy development, in early childhood and in adolescence, and in monolingual and bilingual learners, topics on which she has published more than 250 articles and books. She chaired the National Academy of Sciences committee that produced *Preventing Reading Difficulties in Young Children* (1998) and the RAND Reading Study Group that produced *Reading for Understanding: Toward an R&D Program for Reading Comprehension* (2000). She was the co-principal investigator of the Reading for Understanding project focused on middle grades (grades 4–8), *Catalyzing Comprehension through Discussion and Debate (CCDD)*; outcomes of that project include curricular materials demonstrated to improve students' academic language and reading comprehension skills (Word Generation) and a program shown to be effective in improving the reading outcomes of struggling grades 6–9 readers (Strategic Adolescent Reading Initiative [STARI]). Much of her recent work has been carried out together with collaborators brought together by the Strategic Education Research Partnership (SERP) in Research-Practice Partnerships.

**Sharon Vaughn**, the Manuel J. Justiz Endowed Chair in Education, is the executive director of The Meadows Center for Preventing Educational Risk, an organized research unit at The University of Texas at Austin. She is the recipient of numerous awards including

the Council for Exceptional Children (CEC) research award, the American Educational Research Association Special Interest Group distinguished researcher award, The University of Texas Distinguished Faculty Award and Outstanding Researcher Award, and the Jeannette E. Fleischner Award for Outstanding Contributions in the Field of LD from CEC. She is the author of more than 35 books and 250 research articles. Several of these research articles have won awards: one the A.J. Harris International Reading Association award for best article published, and another the School Psychology award for best article. She is currently principal investigator on several Institute for Education Sciences, National Institute for Child Health and Human Development, and U.S. Department of Education research grants. She works as a senior adviser to the National Center on Intensive Interventions and has more than six articles that have met the What Works Clearing House Criteria for their intervention reports. She has conducted technical assistance in literacy to more than 30 state departments of education and has worked as a literacy consultant to more than 50 technical assistance projects over a 30-year career.





The **National Academy of Education** (NAEd) advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars.

[www.naeducation.org](http://www.naeducation.org)