



Research Report
ETS RR-13-30

**Preliminary Reading Literacy
Assessment Framework:
Foundation and Rationale for
Assessment and System Design**

John Sabatini

Tenaha O'Reilly

Paul Deane

December 2013

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Managing Research Scientist

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Gary Ockey
Research Scientist

Donald Powers
Managing Principal Research Scientist

Gautam Puhan
Senior Psychometrician

John Sabatini
Managing Principal Research Scientist

Matthias von Davier
Director, Research

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ruth Greenwood
Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

**Preliminary Reading Literacy Assessment Framework:
Foundation and Rationale for Assessment and System Design**

John Sabatini, Tenaha O'Reilly, and Paul Deane
Educational Testing Service, Princeton, New Jersey

December 2013

Find other ETS-published reports by searching the ETS ReSEARCHER
database at <http://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit
<http://www.ets.org/research/contact.html>

Action Editor: Beata Beigman Klebanov

Reviewers: Yeonsuk Cho and Philip Everson

Copyright © 2013 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and LISTENING. LEARNING. LEADING. are
registered trademarks of Educational Testing Service (ETS). CBAL is a
trademark of ETS.



Abstract

This report describes the foundation and rationale for a framework designed to measure reading literacy. The aim of the effort is to build an assessment system that reflects current theoretical conceptions of reading and is developmentally sensitive across a prekindergarten to 12th grade student range. The assessment framework is intended to document the aims of the assessment program, define the target constructs to be assessed, and describe the assessment designs that are aligned with the aims and constructs. This framework report is preliminary in that we are engaged in an iterative process of writing and revising the framework, based on what we learn from efforts to instantiate the ideas in new assessment designs and the results we garner from piloting novel designs. We also anticipate drafting further sections to address issues such as the scoring models and analytic plans once assessments have been designed and piloted.

Key words: reading comprehension assessment, scenario-based assessment, reading components, GISA, reading for understanding

Acknowledgments

This report was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100005 to Educational Testing Service as part of the Reading for Understanding (RfU) research initiative. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education or Educational Testing Service. We are extremely grateful to the Institute of Education Sciences and Educational Testing Service for sponsoring and supporting this research. We would like to also thank Gary Feng, Joanna Gorin, Laura Halderman, Michael Kane, Irwin Kirsch, Robert Mislevy, and Don Powers for their intellectual insights and thoughtful comments; Jennifer Lentini and Kim Fryer for their editorial assistance; and Barbara Foorman, Rick Wagner, Yaacov Petscher, Keith Millis, Katja Wiemer, Karen Douglas, and the researchers of the RfU network who read and commented on earlier versions of this report. We also wish to express our gratitude to all of our RfU and Cognitively Based Assessment *of, for, and as* Learning (CBAL™) colleagues, who have been and are contributing to this ongoing enterprise.

Table of Contents

	Page
Acknowledgments.....	ii
Table of Contents.....	iii
Background and Broad Aims of the Assessment System.....	2
Working Definition for Reading Literacy	5
Designing Assessments for Reading Development.....	10
Dimensions of Reading Literacy	12
Writing (or Print) System	15
Language (or Verbal) System.....	17
Text and Discourse	21
Conceptual Modeling and Reasoning.....	23
Social Modeling and Reasoning	25
Summary.....	27
Deep Comprehension: Going Beyond the Text.....	27
What Is Measured: Components of Reading and Global Comprehension	28
Conclusion	30
References.....	32
Notes	49

In a recently published chapter, Sabatini and O'Reilly (2013) presented a rationale for creating a new generation of reading comprehension assessments that address some of the frequent criticisms of traditional tests of reading comprehension. These criticisms include the failure to incorporate theoretical models of reading into the assessment design, the use of artificial and narrow purposes for reading, an over-reliance on multiple-choice format, the omission of digital texts and multimedia, weak links to instruction and the lack of diagnostic information, too narrow a focus on the product of comprehension rather than the process of how it unfolds over time, and the failure to control for individual differences such as student motivation and background knowledge (August, Francis, Hsu, & Snow, 2006; Bransford, Brown, & Cocking, 2000; Braun, Kirsch, & Yamamoto, 2011; Chou, Tsai, & Chan, 2007; Coiro, 2009; Cutting & Scarborough, 2006; Francis et al., 2006; Glaser & Silver, 1994; Guterman, 2002; Hannon & Daneman, 2001; S. Katz & Lautenschlager, 2001; Magliano, Millis, Ozuru, & McNamara, 2007; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008; Pellegrino, Chudowsky, & Glaser, 2001; Rupp, Ferne, & Choi, 2006; Sabatini, Albro, & O'Reilly, 2012; Sabatini, O'Reilly, & Albro, 2012; Shapiro, 2004).

The chapter (Sabatini & O'Reilly, 2013) presented a conceptual framework based on six principles, grounded in the research literature, to guide assessment designs aligned with 21st century literacy constructs. Briefly, these principles can be summarized as follows:

- Principle 1 states that print skills and linguistic comprehension are each necessary components of reading proficiency, though neither individually is sufficient to ensure proficiency (Adlof, Catts, & Little, 2006; Gough & Tunmer, 1986; Hoover & Gough, 1990; Vellutino, Tunmer, & Jaccard, 2007).
- Principle 2 states that both breadth and depth of vocabulary knowledge are essential for understanding (Anderson & Freebody, 1981; Deane, 2012; Nagy & Scott, 2000; Ouellet, 2006).
- Principle 3 states that readers construct mental models of text meaning at multiple levels, from literal to gist to complex situation models (Kintsch, 1988, 1998; McNamara & Kintsch, 1996).
- Principle 4 states that reading is a purposeful activity, aimed at attaining a coherent understanding of a text that is sufficient for the reader's goal (McCrudden, Magliano, & Schraw, 2011; McCrudden & Schraw, 2007; Rouet, 2006).

- Principle 5 states that skilled reading includes proficiency in evaluating and synthesizing information across multiple texts and that this requirement is driven by the increasing prevalence of digital literacy activities (Britt & Sommer, 2004; Goldman, 2004; Rouet & Britt, 2011).
- Principle 6 states that growth in reading proficiency consists of incremental expansion of knowledge and skills for the understanding of increasingly complex texts and task demands (Snow, 2002).

Collectively, these principles represent a synthesis of a wide range of research in areas that are often investigated in relatively isolated research communities. The Sabatini and O'Reilly (2013) chapter was pitched at a coarse grain of generalization and noted that further work was needed to specify what a new system of assessment would look like. While the chapter concluded by briefly describing a new assessment system, that work is continued here.

In this document, we elaborate on the assessment system alluded to in Sabatini and O'Reilly (2013). In the next section, we provide some background on the broad aims that are guiding the development process for the assessment system. Next, we provide a working definition of reading literacy and some of the conceptual underpinnings that the definition implies. This is followed by a discussion of issues surrounding the design of developmentally sensitive assessments, a key aim of the assessment system approach we are describing. We then elaborate on the broader construct of reading literacy, reviewing theoretical and empirical literature in support of this conception of the construct. We conclude by sketching out two key types of assessments that exemplify the approach we describe: global, integrated, scenario-based assessments (GISA) and component skill assessments.

Background and Broad Aims of the Assessment System

Assessments are designed for specific purposes, uses, and target populations (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Kane, 2006; Mislavy & Haertel, 2006; Mislavy & Sabatini, 2012). For instance, the National Assessment of Educational Progress (NAEP) reading assessment is designed to measure the reading performance of fourth, eighth, and 12th grade students (National Assessment Governing Board [NAGB], 2010). The assessment design allows policy makers to determine whether reading scores are improving or declining in subsequent

cohorts of U.S. students. While the assessment is useful for tracking national and some group trends, it is not useful for tracking growth in reading comprehension for any given student (NAGB, 2010). In other words, the assessment is good at measuring what it was designed to measure, but it is not appropriate for making inferences for which it was not designed. Understanding the purpose of any given assessment is critical for ensuring its appropriate design, use, and interpretation (Baker, 2013; Graesser & Hu, 2012; Mislavy & Sabatini, 2012).

The purpose that guides the assessment framework and system described here was broadly defined in a Request for Applications for the Reading for Understanding initiative from the Institute of Education Sciences (IES, 2009). An assessment team was charged with “developing and validating a series of age-graded summative reading comprehension assessments” (IES, 2009, p. 14). This assessment purpose is very broad and open to interpretation. Subsequent in-person discussions with IES leadership clarified the role of the assessment somewhat by elaborating that the assessment should be “useful” in the broader context of education, but the assessment team was otherwise given broad latitude in achieving the aim. It is also important to note that the emergence of the Common Core State Standards (CCSS; National Governors Association Center for Best Practices and Council of Chief State School Officers [NGACBP & CCSSO], 2010) and funding to support consortia to build new national assessments aligned with the CCSS occurred roughly between the submitting of proposals for the Reading for Understanding initiative (IES, 2009, 2010) and the funding of those proposals.

We interpret the term *useful* to imply that the assessment design should provide added value over the existing summative assessments used in pre-K–12 settings. While existing K-12 summative assessments are generally considered useful for certain purposes, such as rank-ordering students on a unidimensional scale of reading ability or providing comparative information across groups and in relation to national norms, they have been repeatedly criticized on a number of other counts. These criticisms include, but are not limited to, a narrow or outdated view of the construct, the lack of diagnostic information or inactionable data (e.g., scores given at the end of year when there is no time to adjust instruction), ineffective score reports, weak links to instruction, scant measurement of partial knowledge at the lower end of the distribution, and inadequate tailoring of the assessment to the individual needs of the student (August et al., 2006; Bennett & Gitomer, 2009; Bransford et al., 2000; Chou et al., 2007;

Cromley & Azevedo, 2007; Cutting & Scarborough, 2006, 2012; Francis et al., 2006; Glaser & Silver, 1994; Graesser & Hu, 2012; Guterman, 2002; Hannon & Daneman, 2001; Katz & Lautenschlager, 2001; Keenan, 2012; Magliano et al., 2007; Moran et al., 2008; O'Reilly & Sheehan, 2009; Partnership for 21st Century Skills, 2004; Pellegrino et al., 2001; Rupp et al., 2006; Sheehan & O'Reilly, 2011).

While these criticisms span a wide range of areas, many of them are motivated by the desire to achieve a variety of purposes within a single assessment. Existing K-12 assessments are numerous and span a wide range of valued assessment purposes and uses, including screening, placement, diagnosis, progress monitoring, evaluating intervention efficacy, and skill mastery. Addressing these multiple purposes simultaneously is virtually impossible with a single assessment, yet addressing a broader range of purposes is at the heart of the critical commentary. Consequently, we conclude that the development of an *assessment system* that can be tailored for use in a wider range of contexts would provide added value and utility by expanding the construct of reading comprehension, expanding the potential uses of the assessment, and improving the instructional value of the assessment.

To achieve this aim, we envision developing a wider design space consisting of a task and form bank that spans a broad range of ability levels, reading subconstructs, and use cases (Behrens, Mislevy, DiCerbo, & Levy, in press; Frezzo, Behrens, & Mislevy, 2009; Gorin & Svetina, 2012; Mislevy & Sabatini, 2012; Mislevy, Steinberg, & Almond, 2003; Wilson & Moore, 2012). The form bank we envision is a collection of test forms and further associated task sets that are designed for different purposes. Each test form and task set in the bank has an accompanying rationale and justification for the particular purpose of the assessment, the constructs measured, and the justification for the task and item sequences. Depending on the desired uses, distinct test forms would be chosen from the larger form bank to serve particular assessment functions. For instance, if the purpose of the assessment is to screen students, forms would be chosen that maximize efficiency (quick to administer, quick to score) and group discrimination (identify at-risk students). On the other hand, if the assessment function is more instructional and diagnostic, forms would be chosen to maximize depth and precision (e.g., sequences of items to measure summarization skill and its components) and instructional value (items are sequenced to break down complex processes into more manageable, interpretable units).

This is not to say that we are inventing an all-in-one assessment system that serves all assessment functions. However, we believe it would be a significant advance over current single-purpose test designs for us to demonstrate the more modest (though nontrivial) goal of how added value and utility can be achieved by applying this broader approach in a finite set of distinct uses. As noted, we hope to achieve this by broadening the design space of tasks. This seems a sensible solution to a set of concerns raised by researchers and educators, premised on the value in engaging students in complex literacy activities that are not easily and simply reducible to a sequence of random items and passages. Thus, this differs from simple adaptive item banks, which are currently based primarily on decision rules for selecting items that maximize the information value of individual items on a unidimensional proficiency construct. Those systems have their value and purpose, but they do not address the set of criticisms that we seek to redress.

Working Definition for Reading Literacy

In defining reading literacy for an assessment framework, there are always trade-offs between specificity and exclusion: a definition needs to be broad enough to encompass the full construct but specific enough to make it meaningful (see discussion of defining literacy in Venezky, 1990; Venezky, Wagner, & Ciliberti, 1990). In formulating a definition, we consulted a range of current assessment frameworks, including the Progress in International Reading Literacy Study (PIRLS; Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009), the Programme for International Student Assessment (PISA; Organisation for Economic Co-operation and Development [OECD], 2009b), Program for the International Assessment of Adult Competencies (PIAAC; OECD, 2009a), the NAEP (NAGB, 2008, 2010), the National Assessments of Adult Literacy (NAAL; Kutner et al., 2007), and their predecessors. One strength of these large-scale frameworks is the brevity and comprehensiveness of their definitions. As an example, the PIAAC Literacy Framework (OECD, 2009a) definition is as follows:

Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to

develop their knowledge and potential, and to participate fully in their community and wider society. (p. 6)

Words and phrases in a definition are often elaborated in a framework, to further explain the framer's intentions or to clarify how the definition influences the assessment designs. For example, the PIAAC Framework notes that while *communication* is included in the definition to signal that writing is part of the literacy framework, the current instantiation of the PIAAC assessment does not measure writing. Similarly, the PIAAC framework elaborates that print and written materials include digital literacy forms such as those found on the Internet. Definitions from large-scale national and international programs also highlight the importance of literacy skills to personal growth and to non-academic, workforce skills and societal participation.

We also reviewed the CCSS College and Career Readiness Standards for Reading, Writing and Communication (NGACBP & CCSSO, 2010), the reports of the Partnership for 21st Century Skills (2004, 2008), and the report of the RAND Reading Study Group (Snow, 2002). Of these, the CCSS focus on college and career readiness is especially relevant to our charge. The CCSS are designed to prepare students developmentally to possess the literacy skills to learn and thrive beyond the K-12 context. In a similar vein, documents prepared by the Partnership for 21st Century Skills also underscore the importance of preparing students to engage in skills that are critical for success in the 21st century economy (Partnership for 21st Century Skills, 2004, 2008). The RAND Reading Study Group report (Snow, 2002) is often cited for its framing of literacy skills as the interaction of texts, tasks, and individual characteristics in a social context. While not a definition per se, this observation is a foundational insight when contemplating how reading skill may change or develop across the school years and into adulthood.

One conclusion we draw from these sources is that a broad, comprehensive definition and construct of reading literacy should include not only academic reading, but also literacy contexts that are represented in the workplace and society in general. This implies that the construct includes non-academic documents, such as understanding the implications of a rental agreement or an insurance policy (OECD, 2009a) and digital literacy, such as navigating and evaluating websites (Coiro, 2009, 2011; Graesser et al., 2007; Katz, 2007), in addition to the academic literacy typically assessed in standard K-12 assessments.

A related implication is that reading and print are not isolated from the richer literacy practices, purposes, and contexts in which they are embedded (Lave & Wenger, 1991; Snow,

2002). Both the CCSS and the Partnership for 21st Century Skills documents present a more elaborate view of reading and English language arts (ELA) than has traditionally been implemented in traditional reading comprehension assessments. Both approaches discuss how literacy environments may call upon the integrating proficiencies of speaking, listening, writing, collaboration, communication, critical thinking, problem solving, social and cross-cultural skills (e.g., perspective taking), and so forth.

While the heart of any reading literacy assessment continues to be students' ability to understand printed text, in many cases, we see the value in creating assessments that encompass these nontraditional, allied skill sets as well. Simulating communication via e-mail exchanges or comparing printed documents to graphic or video representations are techniques that can be used to gain a richer sense of student ability to apply literacy skills in diverse contexts. While writing skill is not a primary target of this assessment framework, we would expect that students can express their understanding in written constructed responses at higher levels of skill development (see Deane, Sabatini, & O'Reilly, 2012). Thoughtful constructed responses to text have routinely been included in reading comprehension constructs and specifications (e.g., NAEP Reading Framework, NAGB, 2008). Similarly, there is a tradition of expecting critical thinking processes in a reading assessment, with tasks requiring students to analyze, synthesize, evaluate, and interpret information (Bloom, 1956). The challenge has always been, and remains, to articulate the appropriate balance of reasoning that can, and should be, expected of an educated, literate student population.

With this preface in mind, the working definition of reading literacy we propose is as follows:

Reading literacy involves the deployment of a constellation of cognitive, language, and social reasoning skills, knowledge, strategies, and dispositions, directed towards achieving specific reading purposes. Readers seek to acquire high levels of skill to enable them to achieve their personal goals and to participate in a print-rich, literate society.

This definition, while broad, makes some targeted claims about reading literacy. First, reading literacy requires the recruitment of various abilities that are not functionally restricted to text processing. The *constellation* of skills necessary to achieve *specific reading purposes* varies with the types of text sources, task demands, and evidentiary requirements as related to one's

reading goals. While the constellations of skills are not exclusive to reading, when applied in a literacy activity, they operate in synchrony.

Second, reading comprehension is built around and upon one's *language* and linguistic abilities (Catts, Adlof, & Weismer, 2006) and is contextualized in both static (e.g., reading a historical novel) and dynamic settings (e.g., online communication). It is worth noting that in the field of language learning and assessment, reading is considered one of the four modalities or components of language (writing, listening, and speaking comprising the other three). We see value in recognizing this overlap of reading and language skills as a safeguard against attempting to isolate reading as a totally unique set of skills. We would expect that at the highest levels of proficiency, print reading and language skills are largely indistinguishable.

Third, reading literacy is composed of a set of practices created and communicated by and for people, and thus reading skill requires at every moment a sophisticated form of *social reasoning* and modeling. Written texts capture all life experiences encompassing academic, work-related, and social contexts. Within narrative and literary texts, in which goals, motives, and themes are largely derived and dependent on understanding human experience, the centrality of social knowledge, reasoning, and modeling is apparent. However, this social component is not limited to the literary academic domain (OECD, 2009a, 2009b; Snow, 2002), but applies to nonfiction as well. Increasingly, socially mediated contexts (e.g., accessing web sources that present different points of view, discussing and communicating interpretations in person or via e-mail) are a common part of the text understanding and interpretive process experienced by students and adults. An understanding of these social literacy contexts is often necessary and required to demonstrate skilled reading of texts.

Fourth, in a *print-rich, literate society*, print or text sources come in many forms, including materials (e.g., paper, silkscreen, signage, neon) and digital (e.g., e-mail, web, e-books, pdf) sources. The extension of the definition is meant to include all instances in which the alphanumeric writing system is used to convey meaning, as well as any surrounding context that is part of that meaning (e.g., illustrations, graphics, charts, videos). Embedding or making accessible a nontext element in or near a text source may make it part of the context that must be interpreted to achieve certain reading purposes.

Fifth, individuals acquire *knowledge, skills, strategies, and dispositions*. By knowledge and skill here, we refer mostly to knowledge of and skills in print literacy forms and processing

(decoding knowledge and skills; orthographic, grammatical, and typographic conventions; metalinguistic terminology and skills for identifying and processing different text genres, etc.). By strategies, we have in mind “a cognitive or behavioral action that is enacted under particular contextual conditions, with the goal of improving some aspect of comprehension” (Graesser, 2007, p. 6). Researched strategies include question asking (e.g., King, 2007), self-explanation (e.g., McNamara, O’Reilly, Rowe, Boonthum, & Levinstein, 2007), summarization (e.g., Yu, 2003), graphic organizers and tools for making text structure explicit (e.g., Meyer & Wijekumar, 2007). By dispositions, we refer to the motivational and self-regulatory processes that provide the engagement and effort to participate, persist, and learn from print materials (Eccles & Wigfield, 2002; Guthrie & Davis, 2003; van den Broek, Risden, & Husebye-Hartman, 1995; Zimmerman & Schunk, 2001).

Sixth, individuals adjust their reading behaviors to *achieve specific reading purposes*. For the most part, reading literacy is about understanding written text, developing and interpreting meaning, and using that meaning as appropriate (NAGB, 2008). However, for any given print source in a literacy context, there are multiple ways of understanding, interpreting, or using the print to achieve specific purposes. These multiple purposes draw upon different levels of attention, knowledge, skills, and strategies (McCrudden & Schraw, 2007; Rouet, 2006; Sheehan & O’Reilly, 2012). For example, skilled reading behavior includes scanning for a specific place or name in an article when researching, and knowing when to do so, even if the individual never attempts to understand a continuous paragraph of prose for that text. In order to better assess students’ proficiency in text understanding, we would be well served to provide direction with respect to purposes for reading, to better understand how well students can deploy an appropriate constellation and exercise control of reading skills towards achieving that purpose.

Finally, reading literacy continues to be a valued proficiency because of its social value to the individual and society (NGACBP & CCSSO, 2010; OECD, 2009a, 2009b; Venezky et al., 1990). Twelve plus years of compulsory schooling are evidence of the national social commitment to this investment. Including this aspect in the definition is, in part, a safeguard against becoming so enamored of the cognitive and psychological science evidence that one disregards the instrumental human utility that reading literacy is meant to serve. If reading literacy were to stop being of social value, it is unlikely that we would invest or care so much about the nature and quality of proficiency and assessments. While the social value of reading

may evolve over time, it may also change over the course of development. The next section discusses the issue of reading development and how it may progress over time.

Designing Assessments for Reading Development

As mentioned earlier in this discussion, one of the key aims of the reading for understanding research agenda is to develop a set of preK-12 reading assessments that are developmentally sensitive. Assessments must be capable of measuring not only the skills students should be able to master by the end of formal schooling, but also the set of allied skills that enable progression towards these larger aims. While designing for development is a critical aspect of the work, it is not without significant challenges. For example, currently there is no unified, empirically based model of reading development that adequately describes the complex trajectory of individuals towards reading proficiency. Although there is some empirical data that can shed light on aspects of reading development (e.g., Vellutino et al., 2007), the literature is far from complete. In many ways, the closest that one comes to a model of reading development are the various sets of K-12 state reading standards. The CCSS documents are a useful tool in this regard, as they synthesize and streamline the variations represented in the individual state standards. The authors applied their best thinking to synthesizing existing curriculum standards, identifying benchmark reasoning skills according to a Bloom-like taxonomy (describing, explaining, analyzing, evaluating), and spread these across the age/grade span. What changes to expect within versus across grades in specific knowledge and skill sets is yet to be evaluated. While we will draw upon the CCSS documents, we will also use other data-driven techniques to validate them, or incorporate new patterns as they emerge.

In order to understand levels or trajectories of reading development, one must understand the endpoint of development. In this regard, the CCSS provides a useful target: college and career readiness. That is, the purpose of K-12 education is to prepare students for the higher-level reading and thinking that is necessary to succeed in college or in the workforce and society in general. We view this developmental journey as more akin to a continuous trajectory rather than a sequence of discrete stages. That is, mastery of particular skills, processes, or knowledge states can occur at any point in time, and one may never reach the highest levels of development, or conversely, one may reach college- and career-ready proficiency before grade 12. This endpoint in itself is still ill-defined and most likely spans a range of proficiencies. However, it is

a target amenable to empirical investigation that, hopefully, will continue to be clarified over time.

Foundational to our notions of development are some fundamental cognitive principles that have been identified and supported by empirical research. For example, research and theory demonstrate that some cognitive processes grow increasingly automatized with increasing skill, development, and expertise—that is, they are applied accurately and rapidly with minimal conscious awareness or expenditure of cognitive effort (LaBerge & Samuels, 1974). Decoding, word recognition, and oral reading fluency abilities follow this type of trajectory.

Other skills become increasingly strategic—that is, they are called upon to handle the many varied comprehension demands of complex text-task interactions required for learning and participating in a print-rich society. Knowing how to skim, scan, read for gist understanding, study, or memorize a text, and in what situations to choose one approach over another, are examples of strategic processes that are deployed flexibly and adaptively as driven by reader purpose or goals (Carver, 1997).

The distinction between automatic and strategic skills, however, is not necessarily categorical (Kintsch, 2004; Scarborough, 2001; van den Broek, Rapp, & Kendeou, 2005). For the developing reader on a trajectory towards skilled, proficient reading, it may become increasingly difficult over time to distinguish automated versus strategic processing, as even activities that, on the surface, would seem to require thoughtful, conscious, effortful processing, with practice, experience, and familiarity are consolidated as knowledge and skills applied with minimal processing or conscious effort. In particular, the many inferential processes that must occur to support building coherent mental models of texts may vary by individual with respect to whether they were generated automatically, as a consequence of a consciously applied strategy, or initiated via some other external question or prompt (Graesser, 1981; Graesser & Clark, 1985; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998, 2012; Underwood & Batt, 1996).

The importance for a developmental model of reading is that proficient readers must manage a significant amount of complexity of text and tasks to achieve complex purposes, that is, reading literacy and learning. This complexity increases exponentially from prekindergarten through 12th grade academic and social reading expectations. Students are asked to read greater quantities of increasingly linguistically sophisticated texts across many disciplinary domains and to perform a wider and wider variety of thinking and learning tasks. Individuals are only able to

do so because many reading subprocesses have become habitual, routine, and can be orchestrated in such a way as to permit challenging new texts and tasks to be undertaken at each new period.

That being said, we can make a few generalizations about what will vary across this trajectory of reading development, and, consequently, how best to map empirically the terrain. First, the expectation concerning the sophistication of the social (e.g., communication and listening skills of the Common Core ELA standards) and critical reasoning skills (e.g., Metzger, 2007) of most students needs be considered. Second, features of texts affect their developmental appropriateness. Texts vary in terms of linguistic complexity (McNamara, Graesser, & Louwse, 2012; Sheehan, Kostin, Futagi, & Michael, 2010), for example, syntactic and semantic complexity, and, as students develop, they are expected to handle more complex texts.¹ Also, the maturity of the content and themes that the reader might be expected to understand are conditioned on societal norms and expectations. Third, the sophistication of the tasks that the student is expected to perform can vary (see, Embretson & Wetzel, 1987; Mosenthal, 1996; Ozuru, Rowe, O'Reilly, & McNamara, 2008). More advanced readers might be expected to do more advanced tasks with text sources. This may involve producing a written response, comparing information across texts, or detecting bias in an argument.

A fourth way to infer developmental or sophistication levels is to vary the use of support for a given task (see, Bennett, 2011; Bennett & Gitomer, 2009; Deane et al., 2012; O'Reilly & Sheehan, 2009; Sheehan & O'Reilly, 2011, 2012; Vygotsky, 1978). Students at lower levels of development can be given textual aids (audio, video, vocabulary definitions) to help them perform a task targeting a specific proficiency. They may also be given a structured support that reduces difficulty by breaking down a complex task into smaller, more manageable units (e.g., asking students to find main ideas and details before writing a summary). Task difficulty can also be reduced by requiring less-skilled readers to complete only portions of a larger task (e.g., fill in cells of a given graphic organizer vs. complete a graphic organizer from scratch). In summary, assumptions and norms of individual maturity and cognitive development, the demands of the task, the difficulty of the text, and the level of support can all be varied to infer the developmental or sophistication level of the student.

Dimensions of Reading Literacy

The theoretical foundations of the assessments we are designing are meant to be consistent with the more general features of a number of cognitive models and theories of

comprehension. At the same time, we would not want the underlying constructs to be restricted or dependent on any particular theory. Instead, general principles derived from those theories will guide the model, the design of assessments, and the claims about student abilities.

Theoretical models and empirical results inform the technical language used to discuss the framework, especially when that language is used broadly in the research literature. For example, below we refer to distinctions between the text-based versus situation models, terms which are derived from Kintsch's construction-integration (CI) model (Kintsch, 1988, 1998), but we are not proposing that the constructs or assessments designed are a direct instantiation or test of CI theory.

In order to encompass the universe of reading literacy from prereaders to college and career capability, the assessment framework must be, of necessity, broad. As mentioned, we reviewed a host of reading theories and seminal reading frameworks to inform our thinking and development. Much of that thinking has been synthesized and refined over time, and it continues to be elaborated.² A key challenge of this work is to find the right balance between construct specificity and coverage on the one hand, and instructional significance on the other. The assessments need to address student abilities that are meaningful and actionable in light of learning and instruction (Bennett, 2010; Bennett & Gitomer, 2009; Bransford et al., 2000; Connor, Morrison, Fishman, & Schatschneider, 2012; Gordon Commission, 2013; Pellegrino et al., 2001).

Given this aim, we distinguish among five dimensions of reading literacy, employed as an analytic frame to guide the design of these reading literacy assessments (see also Deane, Sabatini, & O'Reilly, 2012). Each dimension is considered a bounded constellation of knowledge and skill processes. However, in typical reading literacy activities, the knowledge and skills for any given dimension may be inextricably overlapping or co-occurring with skills of other dimensions (Cain & Oakhill, 2012; Seo, Chiu, & Roussos, 2008; Steinberg, Sabatini, O'Reilly, & Bruce, 2012). This is particularly the case when dealing with complex, multisource tasks (Wiley et al., 2009; Wiley & Voss, 1999) and their potential reliance on a constellation of subskills represented in other dimensions. Therefore, while this approach is useful for designing tasks or analyzing cognitive task demands, we are *not claiming the dimensions would be identifiable as distinct constructs, factors, or latent variables that could be statistically modeled based on task scores*. The types of assessment tasks we are building often require the blending

and integration of skills across dimensions. An analogy might be how one combines food groups into a well-balanced diet. Any given meal may not include ingredients from each food group. Furthermore, each ingredient may be blended or transformed during preparation of a dish, so as to make it harmonious with other ingredients. However, in planning a balanced diet of reading literacy assessments across development, all five dimensions need be addressed.

The five dimensions we distinguish are the writing (or print) system, language (or verbal) system, text and discourse, conceptual modeling/reasoning, and social modeling/reasoning. The dimensions serve as tools for analyzing assessment activities with respect to the kinds of models, representations, and operations that a reader must bring to bear to successfully accomplish literacy tasks. Stated as claims, reading proficiency is composed of the knowledge, skills, strategies, and dispositions that enable readers:

- to learn and process the visual and typographical elements and conventions of printed texts,
- to learn and process the verbal elements of the English language including grammatical structures and word meanings,
- to learn and process the discourse structures, forms, and genres of print,
- to model and reason about conceptual content, and
- to model and reason about social content.

There is an intended directionality to the dimensions that moves generally from concrete and foundational to more abstract and global. However, the directionality is not always coincident with difficulty, and it is not typically hierarchical. Reading comprehension is a relative term; what it means to understand varies as a function of the demands of the reading situation (Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001). Some reading situations simply require the reader to locate a detail (e.g., what time does the bus arrive?), while other situations require deeper processing (e.g., synthesize two perspectives on the American Revolution). In effect, the demands of the reading situation help dictate what dimensions are required to successfully complete the task at hand (McCrudden et al., 2011). There are texts and tasks that put minimal demands on specific dimensions, and those demands are not always arranged in a strict hierarchy. In the parlance of measurement, tasks load more heavily on some dimensions than others. Demands may be high with respect to print and verbal dimensions, yet account for little variance in scores of proficient readers when an

assessment is meant to discriminate them based on conceptual or social reasoning.

Understanding and controlling the relationship between text, task, and response is fundamental to making inferences about the nature of student ability (Mislevy & Haertel, 2006; Mislevy & Sabatini, 2012).

Most naturally occurring literacy activities are likely to require knowledge or skills across all dimensions, but with varying degrees of emphasis. Thus, multiple dimensions may be necessary, if not sufficient, in successfully completing even a simple task. For example, recognizing the words on the page may be prerequisite to completing a task that primarily requires complex, conceptual reasoning. Conversely, knowing and reasoning about human emotion may be necessary to know why the little boy cried when he dropped his ice cream cone. Thus, we can say that dimensions are engaged simultaneously across cycles of processing as the reader seeks to understand text sources (van den Broek, 2012). For our measurement purposes, we design tasks that span or sample from these dimensions, recognizing that the proficient reader utilizes proficiency at all dimensions to perform simple and complex reading tasks.

The descriptions of the five dimensions of reading literacy that follow also provide some discussion of developmental milestones that we might want to observe or collect evidence of via the assessment tasks. We would like to identify sets or constellations of skills that may be key indicators of progress along the learning trajectory that leads to reading proficiency, that is, sophisticated reading performance.

Writing (or Print) System

Description of the dimension. Of the five dimensions of reading literacy, only the writing or print system is relatively exclusive to the printed or written form of the language. It is composed of the visual symbols and patterns of reading literacy. Print includes all operations, from perceptual to cognitive, that require processing of typographic symbols and organizational structures. The print dimension encompasses letters, numbers, words, spaces, orthography, spelling, punctuation marks, paragraph and tab indents, symbols, font styles and effects (e.g., italics), titles, headers, outline conventions, glosses, indices, and pronunciation keys. In fact, any key on a keyboard or function of a word processor is likely in the universe of print knowledge. One needs to also consider typographical conventions of document literacy such as tables, graphs, charts, maps, forms, and so forth (OECD, 2009a, 2009b). Although referenced as part of the writing system dimension, these proficiencies can become quite advanced, and entangled

with conceptual, domain-specific knowledge, as is the case with scientific notation and representations.

The list of print conventions has recently undergone further elaboration and complexity with the emergence of computer software, the Internet, websites, and other digital technologies (Coiro, 2009). Added to the list are, for example, all the conventions of navigation and search (web links, drop-down menus, point and click, drag and drop, scrolling, tabs), emoticons, visual icons, and so forth.

The proficient reader comes to recognize the rhetorical meaning cued by the entire panorama of typographic markings and has the knowledge and skill needed to interpret each appropriately, given the rhetorical or genre context (e.g., italics signifies emphasis, but in narratives may indicate change of speaker or voice in one's head; quotation marks signify a character speaking in a narrative, or exact quote of an external source in an expository report; a table of contents is in outline format with page numbers; an index is in alphabetical order by topic or author; presence of a scroll bar on a digital text signifies there is more text to follow not visible on the screen; etc.).

Thus, gaining expertise in the arcana of visual and typographical conventions in a world of literacy, as represented in traditional sources, as well as in digital technologies, is itself a complex set of knowledge and skills to master. Even so, much of this knowledge becomes automatized or implicit with experience and practice. Skilled readers are often able to transfer at least some of their knowledge and skills from one social literacy practice environment to another (e.g., book to website); they are often able to anticipate and learn novel conventions as they emerge (e.g., initialisms in texting). This is most readily observed in the ability of literate individuals trained in a mostly nondigital world to adapt to digital forms of traditionally print conventions, such as navigating a linked web table of contents, as in Wikipedia, draws upon one's knowledge of navigating a book or manual table of contents. In fact, this adaptability and flexibility in transferring knowledge and skills to variant forms of conventions is an indicator of expertise and proficiency, that is, it is part of the inferential ability required of the skilled reader. Therefore, it may be an assessable skill set.

Very few conventions of print are learned absent their correlates to verbal/linguistic representations (discussed next).³ Learning how the writing system maps on to the language system is paramount to reading skill development, especially for novice readers of any age

(Perfetti, 2003; Sabatini & Bruce, 2009). An example is learning the visual graphemes or letters of the alphabet in association with their oral labels. Another example is learning the set of sounds or phonemes that specific graphemes typically represent, sounds which are not always synonymous with letter names—and never so in the case of *w* or *h*. Learning to phonologically decode the English alphabetic writing system and, through that skill, acquiring, across one's lifetime, adaptive and flexible visual word recognition and fluency skills, and the associated lexical memory of word forms, is also fundamental to proficient reading skill (Adams, 1990; Daane, Campbell, Grigg, Goodman, & Oranje, 2005; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Perfetti, 1985, 2003; Perfetti & Adlof, 2012; Torgesen, Wagner, & Rashotte, 1999; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). Finally, it is worth noting that the morphological structure of words is often cued somewhat differently in the written form of the language than in the oral form. Visual processing of roots and affixes as they appear in printed written words helps build the learner's knowledge and skills of English at both dimensions we label as print and verbal (Carlisle & Stone, 2003; Kieffer & Lesaux, 2007; Venezky, 1999).

Relevance to assessment. Because the dimension for writing (or print) system has commonalities with the dimension for language (or verbal) system, the relevance for assessment is discussed for both dimensions in the next section.

Language (or Verbal) System

Description of the dimension. The language or verbal dimension is meant to coincide with language or linguistic knowledge and skills as they relate to reading literacy ability. This would encompass in traditional linguistics categories such as syntax, and, to some extent, semantics.⁴ With respect to semantics, this dimension includes specific word knowledge and the semantic network of associations among lexical elements, that is, vocabulary and morphology.

Word knowledge and lexical associations are not acquired solely through reading, nor are they restricted to use when reading during early development; much of the learner's lexical knowledge is acquired via the spoken channel, or at least in a combination of spoken and written, such as, a parent or teacher speaks and explains the meaning of novel words that appear first to the child in a book or printed form (Aouad & Savage, 2009; Catts et al., 2006; Catts, Fey, Tomblin, & Zhang, 2002; Catts, Fey, Zhang, & Tomblin, 1999; Catts, Gillespie, Leonard, Kail, & Miller, 2002; Catts, Hogan, & Fey, 2003; Cutting & Scarborough, 2006, 2012; Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005; Gough & Tunmer, 1986; Wise, Sevick, Morris,

Lovett, & Wolf, 2007). At some point, the proportion of word learning that originates via spoken versus written language is thought to shift towards the print channel in normally developing readers (Nagy & Scott, 2000). That is, one is more likely to encounter novel words in print and read them more frequently in print, than to hear them in oral discourse. However, in academic settings, teachers continue to introduce new words and concepts in classrooms via lectures and discussion, so a learner may still encounter many academic terms in the oral channel, having them reinforced in text sources (Deane, 2012). Furthermore, given the ubiquitous presence of media (video and audio media on web, TV, film, radio, etc.), it may be safest to conclude that the acquisition of lexical knowledge and skill needs to be fluent across the print and verbal perceptual channels. For our purposes, it is important that word meanings can be learned or inferred regardless of modality, but especially important that, once learned, they are accessed and applied when reading text sources.

In addition, a significant body of research points to the importance of morphological awareness and skills as they are related to reading comprehension (e.g., Carlisle, 2000; Carlisle & Stone, 2003; Deane, 2012; Hogan, Bridges, Justice, & Cain, 2011; Kuo & Anderson, 2006; Perfetti & Adlof, 2012; Tong, Deacon, Kirby, Cain, & Parrila, 2011). Morphemes are the basic building blocks of meaning in English. Inflectional morphemes mark syntactic shifts, such as a plural or a verb tense. Derivational morphemes change the meaning or part of speech of a word and include prefixes and suffixes. Anglin (1993) and Nagy and Anderson (1984) estimate that over half of English words are morphologically complex—that is, they are made up of one or more morphemes. As students progress through school, they encounter texts of greater and greater morphological complexity (e.g., a text on government may contain words such as politics, political, apolitical, politician, and politically).

Syntax is traditionally defined as the study of how words combine to form meaningful sentences (Wheeldon, 2011). In written texts, sentences—for the most part—are visually demarcated by punctuation. However, the sentence itself is a collection of idea units—propositions, phrases, clauses, and so forth. These are often unmarked and must be inferred or deduced from linguistic knowledge and skill. One of the advantages of printed text in comparison to live speech is that words, sentences, and other dimensions of language in English are cued by use of spaces, punctuation, and other typographical symbols. These print conventions partially compensate for the absence of tone, inflection, stress, prosody, and other

subtle oral cues to syntactic structure and meaning. Also, print can be visually reexamined, whereas live speech must be processed in real time.⁵

It would seem appropriate to assess whether young learners are acquiring knowledge and skill for mapping these print conventions to their linguistic syntactic knowledge and skills. In the early grades of school, for most children, we expect that the majority of the language forms that learners encounter in printed text in their classrooms are language forms that they have already mastered in their receptive oral language processing ability. That is, if they heard the printed text read to them aloud, the syntactic structure would not be the primary barrier to their understanding of the meaning. This is a testable hypothesis, and an assessable construct. Thus, it would seem appropriate to assess learners' awareness of phrase, clause, and other syntactic boundaries and markers when reading printed text, especially when they are not explicitly marked by typographic symbols like punctuation.

The sentence, whether discussing written or spoken words, is a somewhat arbitrary concept. A sentence can be one word (Stop!) or it can go on for the length of a novel.⁶ One of the advantages of writing (versus live speaking) is that one can organize a complex sequence of ideas into a stable, relatively condensed form of language. The writer does this by reflecting and iteratively editing those ideas into a text. One consequence in the evolution of written text is that complex syntactical structures emerge that cue all variants of subtle relationships among ideas (e.g., cause-effect, sequence, narrative, figurative expressions, etc.). The consequence for the reader, however, is that sentences get longer and more complex, and the ideas embedded in them can be quite abstract, academic, metaphoric, or figurative (McNamara, Crossley, & McCarthy, 2010; Nagy & Townsend, 2012). Through our education system, we have distributed this complexity somewhat across the school years, such that the more complex syntactic structures become more frequent in secondary and postsecondary academic writing. Spoken language in everyday social settings, and even in most classrooms, does not approach the syntactic complexity of academic or literary prose. Thus, over time, there is a transition from mapping print to the learner's existing verbal knowledge in the elementary years towards gaining familiarity and command over understanding how meaning is conveyed via the more complex syntactic structures that appear primarily in print in higher-level academic writing (Nagy & Townsend, 2012). Individuals with limitations in working memory capacity are potentially at a disadvantage in encoding the meaning in complex sentences absent learning and practice

(Cromley & Azevado, 2007; Montgomery, Magimairaj, & O'Malley, 2008). It would be useful to observe how learners encode and integrate the central ideas and meaning in sentences as texts increase in syntactic complexity across the school years.

Relevance to assessment. Collectively, the first two dimensions constitute a local comprehension model, that is, the ability to locate, retrieve, comprehend, or remember small and isolated pieces of a larger text. Items that assess local comprehension require the reader to identify facts, details, places, names, phrases, dates/times, and so forth. As such, local comprehension is only a part of what a complete text intends to convey, but a necessary element of understanding when reading for specific purposes. The print and verbal dimension combine to encompass the constructs *locate/recall* in NAEP, *access and retrieve* in PISA, *access and identify information in the text* in PIAAC, and *retrieving explicitly stated information* in PIRLS.⁷

Note that locate/recall texts may load most heavily on print and verbal dimensions, but that does not mean that they do not require some reliance on other dimensions to accomplish any given task. A knowledge-based social or conceptual inference may be the initiating response necessary to decide to adopt a skim or scan strategy for achieving the task purpose (e.g., finding a specific term in a passage, table, or chart). The load in such a task, however, is relatively lighter on building a complex discourse or situation meaning model of the document or text.

The print and verbal dimension would typically constitute the microstructure or the linguistic parts that would be used to *construct* what Kintsch (1998, 2012) calls the *textbase*. A textbase model represents the propositional structure of what was explicitly stated or very narrowly implied by a passage.⁸ While a textbase model can be rich and complex depending on the reader who forms it and the level of text sophistication, the model is generally restricted to what was stated in the text. For the verbal and print dimensions in this framework, understanding is localized at the phrase or sentence level; we are rarely probing intersentential inference and the global understanding that is characteristic of a complete textbase or situation model. These operations are more characteristic of skills one needs to apply in tasks that load on the text and discourse dimension in the framework (discussed next).

A richer and more embellished representation of a text that goes beyond what is stated is referred to as the *situation model* by Kintsch (1998, 2012). The situation model describes a deeper form of comprehension than the textbase; the situation model augments the text with reader prior knowledge and other sources of information. In the current framework, the situation

model is broken out across three dimensions. Together these dimensions comprise a deeper form of comprehension, driven by the need to engage in conceptual and social reasoning about texts, discussed later.

It is important to note that knowledge and skill at the writing and language system dimensions are acquired across the pre-K-12th grade years and can be considered quite sophisticated and complex domains at the higher levels. The language dimension encompasses all lexical knowledge and associations; thus, it encompasses the idiomatic, polysemic, literary, and metaphoric nature of English language usage at the word and phrase level. A parallel case can be made for understanding and interpreting complex, nuanced single sentences. Nonetheless, knowledge and skills for these dimensions are largely of a distinct and different kind from the dimensions we discuss next.

Text and Discourse

Description of the dimension. Text and discourse in reading is concerned with the ability to read a text source and build a global and coherent mental model of its content and structure. The text and discourse dimension encompasses the processing required to form a global and coherent mental representation of the printed text that reflects the basic meaning. As such, it includes knowledge of rhetorical text structures and genres. It comprises the ability to model the macrostructure of propositional content within a rhetorical frame, whether from an author's or reader's perspective. A skilled reader would be expected to have acquired mental models for a wide variety of genres, each embodying specific strategies for modeling the meaning of the text sources to achieve reading goals.

Most reading situations require more than just understanding a string of isolated facts. To understand the *gist* or global representation of a passage, skilled readers need to integrate segments of text into a larger, more coherent whole. This more global level of comprehension is similar to Kintsch's (1998) notion of the situation model and includes the ability to make inferences across proximal and distal portions of text, as well as the ability to assemble core ideas into a meaningful and coherent model of the passage. However, in considering the boundaries of the text and discourse dimension, we do not wish to encompass all variants of situation models that a reader might construct, but rather a relatively specific type of mental representation.

We refer to a *canonical situation model* as a coherent mental representation of the text meaning that would be most consistently agreed upon by the reading audience to which the text was directed. For this representation, the expectation is that a minimal amount of interpretation is made beyond what the text presents. The formation of a coherent mental model does require that *text-based*, *bridging*, and *knowledge-based* inferences are made as necessary. Bridging inferences connect referential, propositional, and causal elements within the text, aligned with the text structure or genre. Required knowledge-based inferences must be made relative to the expectations the author has of the common knowledge shared by the audience for which he or she is writing. True, the reader is not always privy to all the background knowledge that the writer may have expected of his or her audience (an issue taken up later in discussion of the social and conceptual dimensions). Nonetheless, it is an expectation of the text and discourse dimension that, should the reader possess the requisite background knowledge, he or she connects and applies it toward constructing a coherent model of the text.

Readers who lack key background knowledge should still be able to use rhetorical discourse cues in the text, or knowledge of the genre, to infer gaps in their own knowledge. This is a version of a knowledge-based inference that results in a provisional or tentative understanding of the text. Such models are often confirmed or refuted based on information that appears later in the text or by seeking external knowledge or information from an outside source. Therefore, a skillful reader lacking some key background knowledge should equivalently be able to construct a coherent *provisional* model, with inferences in place to connect text elements sensibly.

To reiterate, to form a coherent mental model as described here, the reader must build a Canonical Situation Model,⁹ that is, a situation model that is coherent and preserves the basic meaning of the text. This model would contain no information that would not be inferred automatically by a competent reader who is reading with the goal of creating a basic, but coherent, representation of the text (Gerrig & McKoon, 1998; Kintsch, 2004; McKoon & Ratcliff, 1992; Millis & Magliano, 2012; Sanford & Garrod, 2005). While this description might sound somewhat minimalistic, building a coherent situation model will draw upon social awareness, background knowledge, and inferencing skills and can go well beyond information directly available in the text, but is unlikely to require interpretive or creative leaps (unless the author is particularly inconsiderate or incoherent). It may call upon more active, effortful forms

of text interpretation, but only as required to achieve a coherent representation of the text. Issues that require extensive interpretation, transformation, and extension of the basic ideas in the texts transcend text and discourse level, and therefore reside in the conceptual and social levels of the model (described next). A good example of one such extension is symbolism, which requires going beyond the basic or literal meaning by making connections to metaphorical or other abstract analogical referents.

Relevance to assessment. In sum, tasks that measure the text and discourse dimension of comprehension assess a test taker's ability to draw bridging inferences, identify key ideas, distinguish redundant and tangential information from key ideas, and assemble the key ideas into a coherent whole. The text and discourse dimension is somewhat similar to the constructs of *integrate/interpret* in NAEP, *integrate and interpret* in PISA, *integrate and interpret (relate parts of text to each other)* in PIAAC, and *make straightforward inferences* and some dimensions of *interpret and integrate ideas and information* in PIRLS. In this framework, execution of discourse comprehension differs slightly from the above instantiations in that we assess not only the ability to make isolated inferences or the ability to identify core ideas, but also the ability to identify and assemble the key ideas into a coherent whole. Tasks that assess the ability to assemble key ideas into a whole include completing graphic organizers (applying text structure knowledge; Meyer et al., 2010) and generating summaries (Gil, Bråten, Vidal-Abarca, & Strømsø, 2010).

Conceptual Modeling and Reasoning

Description of the dimension. Conceptual modeling and reasoning in reading literacy addresses the knowledge and skills that support categorization, logical inference, interpretation, evaluation, integration, causal reasoning, and strategic understanding to achieve more complex goals that extend beyond simple text meaning (Goldman, 2004; Graesser et al., 2007; Metzger, 2007). It is best suited to building, presenting, and evaluating general, decontextualized models of text sources and knowledge as conditioned by external goals or queries. Becoming skilled at conceptual reasoning entails mastery of a range of strategies for exploring, elaborating, analyzing, synthesizing, and evaluating ideas.

A long-standing dilemma and debate in the reading literature concerns where to draw the boundary between reading comprehension, reasoning, and thinking. Most modern reading comprehension frameworks and constructs continue to use some variation of the classic Bloom's

taxonomy categories of analyze, synthesize, interpret, evaluate, and so forth (Bloom, 1956) in generating or categorizing items and task demands. Understanding or inferring cause and effect, comparisons and contrasts, and logical sequences of thought are central characteristics of what we mean to comprehend texts (e.g., Meyer et al., 2010; Meyer & Ray, 2011; National Institute of Child Health and Human Development, 2000).

However, these examples are not in any way exclusive to textual or verbal reasoning; hence, the dilemma: are we confounding reading with reasoning? This framework does not resolve the dilemma, but it does help partition reasoning operations as somewhat separate from those skills used primarily in building a basic coherent model of text meaning. The conceptual (and social) dimensions also help to isolate these cognitive acts within specific purposes for reading. While many texts invite reasoning or critical thinking operations, most do not require them as a prerequisite to basic understanding.¹⁰ Reasoning is most often applied as demanded by the purpose and situation (Narvaez et al., 1999; van den Broek et al., 2001). Thus, it fits within the class of situation models in which one applies specific reasoning and background information to achieve specific purposes. Within content domains, those conceptual operations may be driven by the overarching purpose of disciplinary learning and modes of thinking (Ford & Forman, 2006; Goldman, 2012; Lee & Spratley, 2010). So, for example, understanding scientific theory and knowledge may require inductive and deductive reasoning that are standard tools of scientific inquiry (Rothchild, 2006). In history, understanding and reasoning through the sequence and causation of events is primary to reaching historical understanding and so forth (Britt, Rouet, & Perfetti, 1996; Goldman, 2012; Rouet & Vidal-Abarca, 2002). In short, the conceptual dimension is similar to the notion of critical thinking with text sources.

Relevance to assessment. The conceptual dimension is somewhat similar to the constructs of critique/evaluate in NAEP; reflect and evaluate in PISA; evaluate and reflect in PIAAC; and examine and evaluate content, language, and textual elements and some elements of interpret and integrate ideas and information in PIRLS. However, the current approach to the conceptual dimension differs from the instantiation of the above assessments. The approach taken here is to develop purpose-driven assessments, and, as such, test takers need to not only evaluate individual texts, but also evaluate the relevancy of the sources to achieve the purpose or goals that are provided during the assessment (Lawless, Goldman, Gomez, Manning, & Braasch, 2012). In addition to determining the relevancy of the sources, the assessment also sets

constraints on how the sources should be integrated to solve a problem or to make a decision. These added constraints and purpose of the assessment are one differentiation characteristic to the assessment tasks in relation to other assessments.

Social Modeling and Reasoning

Description of the dimension. One of the key characteristics of this assessment framework is the partitioning of the knowledge and skills required for social representation and reasoning from the other dimensions. Social modeling is, in part, based on the mental representation of other minds and social interactions among agents (Dijkstra, Zwaan, Graesser, & Magliano, 1995; Graesser et al., 1994). It comprises a class of abilities that people exercise when they understand social situations, interpret narratives, and formulate communicative goals. Social elements such as these are underscored in many reading models, such as the Landscape (Linderholm, Virtue, Tzeng, & van den Broek, 2004; van den Broek, Young, Tzeng, & Linderholm, 1999) and Constructionist approaches (Graesser et al., 1994).

The process of reading comprehension does not occur in a vacuum; it is situated in a larger social context that helps define meaning both within texts and in the reason that texts are written in the first place. Texts are written for specific purposes, and authors (and characters) have emotions, intentions, points of view, and perspectives, all of which help to contextualize the meaning of a text. The social knowledge of how people think, behave, and feel is critical for understanding the world and the written expressions of it. In other words, comprehension is more than the ability to understand what texts say, but also the ability to understand the social context surrounding the creation of texts, the social situation described in texts, and the social interactions or implications for people after reading texts.

The social model makes explicit the fact that texts are written by human agents for the purpose of communication. Texts are written by people, for people, predominantly to serve specific functions. Understanding these functions is directly linked to understanding that social model. The social model raises the awareness of the educator that students are often required to bring their understanding of the social world to their understanding of the text. Students need to be capable of interpreting the emotions, actions, goals, or intentions of people in the world, in order to understand such ideas represented in texts. Irony and sarcasm, though prevalent in adolescent youth speech, are not innate and universal in their distribution. Inferring irony in print requires understanding what irony is.

Further, social modeling transcends the dyadic interactions of characters to encompass broader social spheres—human roles as family members, work colleagues, friends, and the social settings in which they participate; diverse cultural group patterns and dynamics; and how societal mores and institutions condition and interact with everyday life events. Thus, the social model captures the social and ethical reasoning that schools help students to develop, as part of their education as literate participants in society. The social dimension (perhaps in interaction with the conceptual) allows students to reason about symbolism and complex themes (e.g., coming-of-age themes) in works of literature (Lee, 2011).

The social dimension of comprehension also serves as a reminder to educators and readers that informational, nonfiction texts, not only narrative fiction, have human authors, often flawed, representing a point of view on the world. Textbook chapters, newspaper articles, editorials, and informational websites are all authored. Sometimes those authors are accurate and reliable; often they are not, whether through intentional or accidental misrepresentation. Sometimes authors are clear and unambiguous in their meaning; often they are not. Often there is rhetorical cuing in the structure and style of a text to help a reader to infer social models. Humor, satire, sarcasm, and irony are among the familiar elements that may be cued in a text, but that also require social modeling and reasoning to interpret.

Regardless, all authors have a sense, implicit or explicit, of the intended audiences for their writing, that is, what knowledge and perspectives they expect the reader brings to the text, though authors vary in their prescience in anticipating their audiences' knowledge, as well as their skill in communicating their own ideas. This empowers questioning and clarification of text meaning as a basic activity of skilled reading, as well as the basic act of evaluating the reliability and truth value of every text (Beck, McKeown, Hamilton, & Kucan, 1997; Cain & Oakhill, 2012; Graesser, Ozuru, & Sullins, 2009; Palincsar & Brown, 1984). It also closes the gap between the student as developing writer, learning to express his or her own understandings and interpretations to an audience, and to apply that social reasoning to the invisible, sometimes anonymous authors of the texts he or she reads (Deane, Sabatini, & O'Reilly, 2012).

Relevance to assessment. Collectively, most large-scale reading assessments measure elements of the social context in which a text was written and the social context described in the text. These tasks, however, are not separately defined as a dimension, but rather are integrated elsewhere, most often in the skills we described previously in the text and discourse and

conceptual dimensions. However, few measure the ability to understand the social context that occurs after people read and react to or interact with texts. Text comprehension is not static, but a dynamic interaction between a text and the reactions, perspectives, and points of views that the reader, other related texts, and other people may have about the texts (Eco, 1979; Leu, 2000). In today's digital world, more than ever, people interact with each other using text in the form of blogs, e-mail, instant messaging, text messaging, social networking sites, and so forth. Understanding these new forms of interaction and communication also requires an understanding of the social context and the communication skills of the participants involved. The importance of communication and collaboration skills in reading comprehension are now being recognized in the CCSS (under speaking and listening skills) and in the Partnership for 21st Century Skills literacy framework.

Summary

In summary, we distinguish a social dimension of reading literacy that encompasses social modeling and reasoning, thus distinguishing the types of representations and operations that have as referent the social world, in contrast to the physical, technical, or logical realms of knowledge. We observe that learners may acquire and develop the knowledge and skills to perform these mental operations in literacy activities, as well as more generally in acting in the social world. For the purposes of this framework, we are concerned with the student's ability to form these representations and reason through them in the context of texts and documents.

Deep Comprehension: Going Beyond the Text

In both research and educational contexts, there is often a distinction between the *deep* and the *superficial* aspects of comprehension (Graesser, McNamara, & VanLehn, 2005). While there is some disagreement on how to precisely define *deep comprehension*, many experts would agree that it involves the recruitment of higher-level processing skills that enable students to go beyond the text. Despite the lack of an agreed upon definition, we can define *deep comprehension* as it relates to this framework as the set of abilities that enable one to go beyond the text by evaluating, reasoning, and thinking about what has been said, as well as the ability to relate what has been stated to existing knowledge and other sources (thus, primarily encompassing the conceptual and social dimensions). Deep comprehension allows readers to extrapolate from the text and apply it to new situations. It also enables readers to *step back* from

the text and evaluate the claims and arguments made, as well as effectiveness of the style and approach for communicating the intended message. It includes the ability to evaluate the trustworthiness of source, the credibility of author, and the ability to identify the use of persuasive techniques and logical fallacies (Graesser et al., 2007; Lawless et al., 2012; Metzger, 2007; Wiley et al., 2009).

Deep comprehension also includes the set of abilities that enable readers to understand multiple sources (Bråten, Britt, Strømsø, & Rouet, 2010; Goldman, 2004; Rouet, 2006; Wiley & Voss, 1999). Multiple sources require readers to integrate, relate, compare and contrast, and synthesize information in different sources (either texts or other media or informational channels). This *multiple-source comprehension* is challenging because readers have to integrate information that was written by different authors, often for different purposes or in different genres. Thus, when reading multiple sources, there is often little cohesion across texts and, as a result, test takers need to generate significant knowledge-based inferences to make the sources cohere. This is an effortful and demanding process because it requires the reader to engage in all the processes that are required to comprehend a single text in isolation, as well as the ability to link texts for specific purposes. In short, deep comprehension in our model is primarily occurring at the conceptual and social levels of the model rather than the discourse, verbal, or print levels.

What Is Measured: Components of Reading and Global Comprehension

This section provides an overview of the initial design approaches used to measure reading literacy in this assessment framework. We intend to measure the various dimensions of reading literacy in global, as well as isolated component assessments of reading skill. In a subsequent report (O'Reilly & Sabatini, 2013), we also describe performance moderators, which, while not typically measured aspects of the reading construct in and of themselves, are potentially critical to interpreting reading score results. Performance moderators may be measured, manipulated, or controlled to provide a more accurate or nuanced picture of a student's reading ability.

In this framework, reading literacy is measured separately in two assessment types. One assessment type measures the components of reading. Components are subskills of reading that can be isolated and assessed independently from measures of global reading comprehension (O'Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012; Sabatini & Bruce, 2009; Sabatini, Bruce, & Steinberg, 2013). Components include constructs associated with print such as word

recognition, decoding, reading fluency, morphology, and vocabulary. We believe there is added value in measuring the components of reading separately from global comprehension due to their diagnostic potential or moderating influences. A low score on a measure of global reading literacy proficiency may be a result of low reading comprehension, or it may signal weakness in core component reading skills. Inefficient component skills can present barriers that hamper the processes underlying general proficiency (Cain & Oakhill, 2012; Cain, Oakhill, & Bryant, 2004; Daneman & Merikle, 1996; LaBerge & Samuels, 1974; Perfetti, 1985; Sabatini, Bruce, & Steinberg, 2013). By measuring both the components of reading and global reading proficiency, the assessment design is potentially more instructionally relevant. Weaknesses in components may signal different intervention strategies than weakness in global reading comprehension.

The second type of assessment is intended to measure global¹¹ reading literacy ability (see Bennett, 2011; Bennett & Gitomer, 2009; Deane et al., 2012; O'Reilly et al., 2012; O'Reilly & Sheehan, 2009; Sabatini, Bennett, & Deane, 2011; Sheehan & O'Reilly, 2011, 2012). More specifically, we refer to this assessment as the Global Integrated Scenario-Based Assessment of reading, or GISA for short. GISA is designed to measure various levels of reading comprehension in the context of simulated reading situations. Test takers are provided with a specific purpose for reading (e.g., studying for a history, language arts, or science test; preparing for a class presentation) and a set of materials (e.g., websites, blogs, newspaper articles, op ed, authoritative texts, unreliable sources). Test takers progress through the set of materials in a structured way that enables them to: demonstrate different dimensions of comprehension (e.g., conceptual and social); learn, remember, and organize what they read; manage their learning through strategy use; and apply, synthesize, and extrapolate what they have learned to solve the problem or to make a decision that satisfies their original purpose for reading.

We believe there is added value in measuring global reading literacy in this manner. First, unlike traditional assessments of reading that present students with a set of unrelated texts and no purpose for reading them, the scenario-based approach is more representative of the contexts for reading that occur in college, the workforce, and leisure; people rarely read without a purpose in these situations (McCrudden et al., 2011; Rupp et al., 2006). By making the reading tasks resemble the types of situations in which students are expected to read, the assessment better reflects cognitive processes we hope they will engage in when confronted with authentic literacy activities. Second, the richness of the scenario-based approach potentially increases student

motivation because the assessment is more purposed, more meaningful, and more engaging (see Guthrie & Davis, 2003). When students are more engaged, they are more likely to expend their best effort and persist; as a result, the test may be more reflective of their true ability than an assessment that is not as engaging (Braun et al., 2011).

Third, the structured sequencing of the tasks allows for maximum flexibility when testing the abilities at different levels of skill development. Tasks can be made easier or more difficult by varying the level of support provided to the test taker. At lower levels of development, the sequence of tasks can be broken down into smaller and more manageable units, hints can be provided, and parts of the task can be completed for the test taker. At higher levels of development, the sequence in the scenario can provide less support, as more is expected of the test taker. In this way, the different levels of student development can be tested and accommodated by varying the sequencing and support in a given scenario. Finally, the use of scenarios also affords opportunities to integrate more diversity into the assessment. Scenarios can include a range of text genres, text types, multimedia, task types, cultural perspectives, and peer interactions. Scenarios can be structured to encourage test takers to interact with virtual peers who may have similar or dissimilar views than their own.

Conclusion

In sum, we have described the overall purpose and goals of a reading literacy assessment framework. We have argued for conceptualizing the project as building an assessment system that permits greater flexibility for innovating in the design space and assembling targeted assessments fitted to specific uses, purposes, and goals. We have consulted other frameworks and seminal documents in order to draft an initial definition of reading literacy that is sufficiently broad, given the goal of producing pre-K through 12th grade level assessments. We have discussed some of the complexities of assessing reading across a wide developmental span, and some possible approaches to address these complexities. We have reviewed an analytic division of reading literacy as composed of five dimensions that interact during reading literacy activities, with theoretical and empirical literature aligned with the dimensions. Finally, we have described the basic structural approach to componential assessment constructs and a global, integrated scenario-based assessment design.

We are cognizant of the dynamic nature of reading and its historical development over time. As society changes, so do the expectations for its citizens and students; what is deemed

unimportant in one time period can become valued in another era. In a similar vein, the design, task models, claims, and interpretative arguments that are implied in this document are expected to change to accommodate accumulated empirical evidence, as well as to meet the evolving demands of the societal definitions and expectations for literacy. However, we hope this framework captures key elements of the emerging theoretical understanding of reading literacy in the early 21st century, thus, serving as a step forward in the evolution of reading assessment design and use.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- 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.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching* (pp. 77–117). Newark, DE: International Reading Association.
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58* (serial 238).
- Aouad, J., & Savage, R. (2009). The component structure of preliteracy skills: Further evidence for the simple view of reading. *Canadian Journal of School Psychology, 24*, 183–200.
- August, D., Francis, D. J., Hsu, H.-Y. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *Elementary School Journal, 107*, 221–238.
- Baker, E. L. (2013). The chimera of validity. *Teachers College Record, 115*(9), 1–26.
- Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text*. Newark, DE: International Reading Association.
- Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (in press). Evidence-centered design for learning and assessment in the digital world. In D. Robinson, J. Clarke-Midura, & M. Mayrath (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. New York, NY: Springer-Verlag.
- Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspectives, 8*, 70–91.
- Bennett, R. E. (2011). *CBAL: Results from piloting innovative K–12 assessments* (Research Report No. RR-11-23). Princeton, NJ: Educational Testing Service.

- Bennett, R. E., & Gitomer, D. H. (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). New York, NY: Springer.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: Handbook I. The cognitive domain*. New York, NY: David McKay.
- Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. (2010). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*, 48–70.
- Braun, H., Kirsch, I., & Yamamoto, K. (2011). An empirical study to examine whether monetary incentives improve 12th grade NAEP reading performance. *Teachers College Record, 113*, 2309–2344.
- Britt, M. A., Rouet, J.-F., & Perfetti, C. A. (1996). Using hypertext to study and reason about historical evidence. In J.-F. Rouet, J. J. Levonen, A. P. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 43–72). Mahwah, NJ: Lawrence Erlbaum Associates.
- Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology, 25*, 313–339.
- Cain, K., & Oakhill, J. (2012). Reading comprehension development from seven to fourteen years: Implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 59–76). Lanham, MD: Rowman & Littlefield Education.
- 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*, 31–42.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal, 12*, 169–190.

- Carlisle, J. F., & Stone, C. A. (2003). The effects of morphological structure on children's reading of derived words. In E. Assink & D. Santa (Eds.), *Reading complex words: Cross-language studies* (pp. 27–52). New York, NY: Kluwer Academic.
- Carver, R. P. (1997). Reading for one second, one minute, or one year from the perspective of Rauding theory. *Scientific Studies of Reading, 1*, 3–43.
- Catts, H., Adlof, S., & Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language & Hearing Research, 49*, 278–293.
- Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, & Hearing Research, 45*, 1142–1157.
- Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*, 331–361.
- Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35*, 510–525.
- 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.
- Chou, C., Tsai, C., & Chan, P. (2007). Developing a web-based two-tier test for Internet literacy. *British Journal of Educational Technology, 38*, 369–372.
- 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.
- Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*, 352–392.
- Connor, C. M., Morrison, F. M., Fishman, B., & Schatschneider, C. (2012). Assessment and instruction connections: The impact of teachers' access and use of assessment-to-instruction software. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an*

- understanding: Innovations in how we view reading assessment* (pp. 81–99). Lanham, MD: Rowman & Littlefield Education.
- Cromley, J., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325.
- Cutting, L., & Scarborough, H. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277–299.
- Cutting, L., & Scarborough, H. (2012). Multiple bases for comprehension difficulties: The potential of cognitive and neurobiological profiling for validation of subtypes and development of assessments. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 101–116). Lanham, MD: Rowman & Littlefield Education.
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES Report No. 2006-469). Washington, DC: National Center for Education Statistics.
- Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.
- Deane, P. (2012). NLP methods for supporting vocabulary analysis. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 117–144). Lanham, MD: Rowman & Littlefield Education.
- Deane, P., Sabatini, J., & O'Reilly, T. (2012). *English language arts literacy framework*. Princeton, NJ: Educational Testing Service. Retrieved from <http://elalp.cbalwiki.ets.org/Table+of+Contents>
- Diakidoy, I. A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of texts at increasing grade levels. *Reading Psychology, 26*, 55–80.
- Dijkstra, K., Zwaan, R. A., Graesser, A. C., & Magliano, J. P. (1995). Character and reader emotions in literary texts. *Poetics, 23*, 139–157.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.
- Eco, U. (1979). *The role of the reader*. Bloomington, IN: Indiana University Press.

- Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension. *Applied Psychological Measurement, 11*, 175–193.
- Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education, 30*, 1–32.
- Francis, D., Snow, C., August, D., Carlson, C., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the Diagnostic Assessment of Reading Comprehension. *Scientific Studies of Reading, 10*, 301–322.
- Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2009). Activity theory and assessment theory in the design and understanding of the packet tracer ecosystem. *The International Journal of Learning and Media, 1*(2). doi:10.1162/ijlm.2009.0015
- Gerrig, R., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes, 26*, 67–86.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*, 30–68.
- Glaser, R., & Silver, E. (1994). Testing and instruction: Retrospect and prospect. In D. H. Linda (Ed.), *Review of research in education* (Vol. 20, pp. 393–419). Washington, DC: American Educational Research Association.
- Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris & D. M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–351). Greenwich, CT: Information Age Publishing.
- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *Future of Children, 22*(2), 89–116.
- 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
- Gorin, J. S., & Svetina, D. (2012). Cognitive psychometric models as a tool for reading assessment engineering. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 169–183). Lanham, MD: Rowman & Littlefield Education.

- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.
- Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York, NY: Springer-Verlag.
- Graesser, A. C. (2007). Theories of text comprehension: The importance of reading strategies to theoretical foundations of reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3–26). Mahwah, NJ: Lawrence Erlbaum Associates.
- Graesser, A. C., & Clark, L. C. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- Graesser, A. C., & Hu, X. (2012). Moving forward on reading assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 153–157). Lanham, MD: Rowman & Littlefield Education.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225–234.
- Graesser, A. C., Ozuru, Y., & Sullins, J. (2009). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (pp. 112–141). New York, NY: Guilford.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.
- Graesser, A. C., Wiley, J., Goldman, S., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*, 89–105.
- Guterman, E. (2002). Toward dynamic assessment of reading: Applying metacognitive awareness guidance to reading assessment tasks. *Journal of Research in Reading, 25*, 283–298.
- Guthrie, J., & Davis, M. (2003). Motivating struggling readers in middle school through an engagement model of classroom performance. *Reading and Writing Quarterly, 19*, 59–85.

- Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103–128.
- Hogan, T. P., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children, 44*, 1–20.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.
- Institute of Education Sciences. (2009). *Request for applications: Reading for Understanding research initiative CFDA Number: 84.305F*. Washington, DC: U.S. Department of Education. Retrieved from http://ies.ed.gov/funding/pdf/2010_84305F.pdf
- Institute of Education Sciences. (2010). *Reading for Understanding initiative*. Washington, DC: U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncer/projects/program.asp?ProgID=62>
- Jenkins, J., Fuchs, L., van den Broek, P., Espin, C., & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719–729.
- Kane, M. (2006). Validation. In R. J. Brennan (Ed.), *Educational measurement* (4th ed., pp. 18–64). Lanham, MD: Rowman & Littlefield Education.
- Katz, I. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology and Libraries, 26*, 3–12.
- Katz, S., & Lautenschlager, G. (2001). The contribution of passage and no-passage factors to item performance on the SAT reading task. *Educational Assessment, 7*, 165–176.
- Keenan, J. (2012). Measure for measure: Challenges in assessing reading comprehension. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 77–87). Lanham, MD: Rowman & Littlefield Education.
- Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *Reading Teacher, 61*, 134–144.
- King, A. (2007). Beyond literal comprehension: A strategy to promote deep understanding of text. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 267–290). Mahwah, NJ: Lawrence Erlbaum Associates.

- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1328). Newark, DE: International Reading Association.
- Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 21–38). Lanham, MD: Rowman & Littlefield Education.
- Kuo, L., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, *41*, 161–180.
- Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., Dunleavy, E., & White, S. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy* (NCES Report No. 2007-480). Washington, DC: National Center for Education Statistics.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.
- Lave, J., & Wenger, E. (1991). *Studied learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lawless, K. A., Goldman, S. R., Gomez, K., Manning, F., & Braasch, J. (2012). Assessing multiple source comprehension through evidence-centered design. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 3–17). Lanham, MD: Rowman & Littlefield Education.
- Lee, C. (2011). Complexities of teaching and implications for equity. *Education Canada*, *51*, 40–43.
- Lee, C., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy. Final report from Carnegie Corporation of New York's Council on Advancing Adult Literacy*. New York, NY: Carnegie Corporation.

- Leu, D. J. (2000). Literacy and technology: Deictic consequences for literacy education in an information age. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 743–770). Mahwah, NJ: Lawrence Erlbaum Associates.
- Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, *37*, 165–186.
- Magliano, J. P., Millis, K. K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 107–136). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*, 57–86.
- McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89–116). Lanham, MD: Rowman & Littlefield Education.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 397–420). Mahwah, NJ: Lawrence Erlbaum Associates.

- Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58, 2078–2091.
- Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, 4, 127–152.
- Meyer, B., & Wijekumar, K. (2007). A web based tutoring system for the structure strategy: theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 347–374). Mahwah, NJ: Lawrence Erlbaum Associates.
- Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P.-W., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly*, 45, 62–92.
- Millis, K., & Magliano, J. (2012). Assessing comprehension processes during reading. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 35–53). Lanham, MD: Rowman & Littlefield Education.
- Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice*, 25, 6–20.
- 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.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–67.
- Montgomery, J. W., Magimairaj, B. M., & O'Malley, M. H. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, 37, 331–354.

- Moran, J., Ferdig, R., Pearson, D., Wardrop, J., & Blomeyer, R. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research, 40*, 6–58.
- Mosenthal, P. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology, 88*, 314–332.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Boston, MA: TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.edu/pirls2011/downloads/PIRLS2011_Framework.pdf
- Nagy, W., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304-330.
- Nagy, W., & Scott, J. (2000). Vocabulary process. In M. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 269–284). Mahwah, NJ: Lawrence Erlbaum Associates.
- Nagy, W., & Townhend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*(1), 91–108.
- Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). Reading purpose, type of text and their influence on think-alouds and comprehension measures. *Journal of Educational Psychology, 91*, 488–496.
- National Assessment Governing Board. (2008). *Reading framework for the 2009 National Assessment of Educational Progress (NAEP)*. Washington, DC: Author. Retrieved from <http://www.nagb.org/publications/frameworks/reading09.pdf>
- National Assessment Governing Board. (2010). *Reading framework for the 2011 National Assessment of Educational Progress*. Washington, DC: Author. Retrieved from <http://www.nagb.org/publications/frameworks/reading-2011-framework.pdf>
- National Governors Association Center for Best Practices, & 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>
- 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. Government Printing Office.
- Organisation for Economic Co-operation and Development. (2009a). *PIAAC literacy: A conceptual framework*. Paris, France: Author. Retrieved from <http://www.oecd-ilibrary.org/content/workingpaper/220348414075>
- Organisation for Economic Co-operation and Development. (2009b). *PISA 2009 assessment framework—Key competencies in reading, mathematics and science*. Paris, France: Author. Retrieved from http://www.oecd.org/document/44/0,3746,en_2649_35845621_44455276_1_1_1_1,00.html
- O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (ETS Research Report No. RR-13-31). Princeton, NJ: Educational Testing Service.
- O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under an RTI framework. *Reading Psychology, 33*, 162–189.
- O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively based assessment of, for, and as learning: A 21st century approach for assessing reading competency* (Research Memorandum No. RM-09-04). Princeton, NJ: Educational Testing Service.
- Ouellet, 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.
- Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40*, 1001–1015.
- Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175.
- Partnership for 21st Century Skills. (2004). *Learning for the 21st century: A report and mile guide for 21st century skills*. Washington, DC: Author. Retrieved from http://www.p21.org/storage/documents/P21_Report.pdf

- 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
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.
- Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading*, 7, 3–24.
- Perfetti, C. A., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 3–20). Lanham, MD: Rowman & Littlefield Education.
- Rothchild, I. (2006). *Induction, deduction, and the scientific method: An eclectic overview of the practice of science*. Madison, WI: Society for the Study of Reproduction, Inc. Retrieved from: <http://www.ssr.org/Documents/2006-01-04Induction2.pdf>
- Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Rouet, J.-F., & Vidal-Abarca, E. (2002). Mining for meaning: A cognitive examination of inserted questions in learning from scientific text. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 417–436). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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. P., Albro, E. R., & O'Reilly, T. (Eds.). (2012). *Measuring up: Advances in how to assess reading ability*. Lanham, MD: Rowman & Littlefield.
- Sabatini, J., Bennett, R., & Deane, P. (2011). *Four years of Cognitively Based Assessment of, for, and as Learning (CBAL): Learning about through-course assessment (TCA)*. Paper

- presented at the Educational Testing Service Center for K–12 Assessment & Performance Management Invitational Research Symposium on Through-Course Summative Assessments, Princeton, NJ. Retrieved from:
http://www.k12center.org/rsc/pdf/TCSA_Symposium_Final_Paper_Sabatini.pdf
- Sabatini, J., & Bruce, K. (2009). *PIAAC reading component: A conceptual framework* (OECD Education Working Papers, No. 33). Paris, France: OECD Publishing.
- Sabatini, J., Bruce, K., & Steinberg (2013). *SARA reading components tests, RISE form* (Research Report No. RR-13-08). Princeton, NJ: Educational Testing Service.
- Sabatini, J., & O'Reilly, T. (2013). Rationale for a new generation of reading comprehension assessments. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 100–111). Baltimore, MD: Brookes.
- Sabatini, J. P., O'Reilly, T., & Albro, E. R. (Eds.). (2012). *Reaching an understanding: Innovations in how we view reading assessment*. Lanham, MD: Rowman & Littlefield.
- Sanford, A. J., & Garrod, S. C. (2005). Memory-based processing and beyond. *Discourse Processes*, 39, 205–224.
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook for research in early literacy* (pp. 97–110). New York, NY: Guilford Press.
- Seo, M., Chiu, C., & Roussos, L. (2008, September). *Evaluating the dimensionality of the 2001 PIRLS reading assessment: An application of DIMTEST with DESM and CFA*. Paper presented at the 3rd IEA International Research Conference, Taipei, Chinese Taipei.
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41, 159–189.
- Sheehan, K., Kostin, I., Futagi, Y., & Michael, F. (2010). *Generating automated text complexity classifications that are aligned with targeted text complexity standards* (Research Report No. RR-10-28). Princeton, NJ: Educational Testing Service.
- Sheehan, K., & O'Reilly, T. (2011). *The CBAL reading assessment: An approach for balancing measurement and learning goals* (Research Report No. RR-11-21). Princeton, NJ: Educational Testing Service.

- Sheehan, K., & O'Reilly, T. (2012). The case for scenario-based assessments of reading competency. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 19–33). Lanham, MD: Rowman & Littlefield Education.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Steinberg, J., Sabatini, J., O'Reilly, T., & Bruce, K. (2012, July). *Dimensionality analysis of a computerized battery of reading measures*. Paper presented at the annual meeting of the Society of Scientific Studies of Reading, Montreal, QC.
- Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103*, 523–534.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.
- Underwood, G., & Batt, V. (1996). *Reading and understanding*. Oxford, UK: Blackwells.
- van den Broek, P. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how to assess reading ability* (pp. 39–58). Lanham, MD: Rowman and Littlefield Education.
- van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*, 1081–1087.
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*, 299–316.
- van den Broek, P., Risdien, K., & Husebye-Hartman, E. (1995). The role of the reader's standards of coherence in the generation of inference during reading. In R. F. Lorch, Jr., & E. J. O'Brian (Eds.), *Sources of coherence in text comprehension* (pp. 353–373). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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 representation during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates.
- Vellutino, F., Tunmer, W., & Jaccard, J. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*, 3–32.
- Venezky, R. L. (1990). Definitions of literacy. In R. L. Venezky, D. A. Wagner, & B. S. Ciliberti (Eds.), *Toward defining literacy* (pp. 2–16). Newark, DE: International Reading Association.
- Venezky, R. L. (1999). *The American way of spelling: The structure and origins of American English orthography*. New York, NY: Guilford.
- Venezky, R. L., Wagner, D. A., & Ciliberti, B. S. (Eds.). (1990). *Toward defining literacy*. Newark, DE: International Reading Association.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41*, 85–120.
- Wheeldon, L. R. (2011). Generating spoken sentences: The relationship between words and syntax. *Language and Linguistics Compass, 5*, 310–321.
- Wiley, J., Goldman, S., Graesser, A., Sanchez, C., Ash, I., & Hemmerich, J. (2009) Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46*, 1060–1106.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 1–11.
- Wilson, M., & Moore, S. (2012). An explanatory modeling approach to measurement of reading comprehension. In J. P. Sabatini, T. O'Reilly, & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 147–168). Lanham, MD: Rowman & Littlefield Education.
- Wise, J. C., Sevick, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills,

- word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language & Hearing Research*, 50, 1093–1109.
- Yu, G. (2003). Reading for summarization as reading comprehension test method: Promises and problems. *Language Testing Update*, 32, 44–47.
- Zimmerman, B. J., & Schunk, D. H. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Notes

- ¹ Linguistic complexity is used here to refer to the linguistic demands such as text cohesion, syntactic complexity, and the level of vocabulary sophistication that may impact a reader's ability to form a coherent model of the text. Linguistic complexity may contribute to item difficulty independent of the task and content demands.
- ² The interested reader is referred to the related, ongoing research program sponsored by ETS called *Cognitively Based Assessment of, for, and as Learning (CBAL™)* (<http://www.ets.org/research/topics/cbal>), where one can find proficiency models encompassing the English language arts (Deane, Sabatini, & O'Reilly, 2012), as well as models for other content areas including mathematics and science.
- ³ We note that much of the acquisition of knowledge and skills in reading literacy occurs through the building of networks of associations and processing skills across dimensions. Skilled word recognition skills can be characterized as a function of the mapping of print-orthographic to phonemic-linguistic codes; memory-based, discourse processing skill and knowledge is built on a foundation of lexical, semantic organization of language; text organization knowledge is built from repeated applications of strategic processes while engaged in conceptual and social reasoning.
- ⁴ The subfield of pragmatics in this framework is more closely aligned with the text and discourse and social dimensions.
- ⁵ There are exceptions, of course. The running banner of print under a live TV broadcast must be processed in real time. Any audio or video recording of speech can be replayed. Still, in the majority of cases we are concerned with in this framework, the distinction will be valid.
- ⁶ *Autumn of the Patriarch* by Gabriel García Márquez is essentially a two-sentence novel.
- ⁷ To the extent that items in these other frameworks draw upon knowledge of text structures and genres, then they would also encompass some of what we label text and discourse as well.
- ⁸ Which is not to say that constructing an accurate textbase representation would not require a great many referential or causal inferences.
- ⁹ Scholars of Kintsch may argue that the canonical situation model as we have described it, in many situations, is merely a text-based model. For any given reader, to the extent a text is

highly explicit, with clear discourse cues and connectors, this may be true. For the purposes of the conceptual foundation of the discourse dimension, we want to emphasize the interaction of individual prior knowledge with building a coherent mental model, producing the canonical situation model. We can then separately query whether an individual has the requisite prior knowledge necessary to a text. We can also separate it from the more complex why questions that might be characteristic of probes into conceptual and social reasoning or multiple text models.

¹⁰ However, a basic understanding of the text (text and discourse dimension) is often necessary or a prerequisite for conceptual reasoning to occur.

¹¹ By using the term *global*, we are referring to a set of performances on assemblies of tasks that, by their nature and their intent, reflect valued reading activities in the real world, and call upon the kinds of proficiency we want students to develop. The term is meant to convey a contrast with traditional comprehension test formats of unconnected passages and isolated skills questions.