Rationale for a New Generation of Reading Comprehension Assessments

John Sabatini

Tenaha O'Reilly

Educational Testing Service

Princeton, NJ 08541

609-734-5842

jsabatini@ets.org

This manuscript is an early draft of a paper published in *Unraveling Reading Comprehension:* and thus may differ slightly from the final published version. Please see below for the official paper:

Sabatini, J., & O'Reilly, T. (2013). Rationale for a New Generation of Reading Comprehension Assessments. In Miller, B., Cutting, L., & P. McCardle (Eds), *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components,* (pp. 100-111). Baltimore, MD: Brookes Publishing.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100005 to the Educational Testing Service as part of the Reading for Understanding Research (RFU) Initiative. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Existing reading assessments have increasingly been criticized by researchers, educators, and policy makers, especially regarding their coverage, utility, and authenticity (e.g., Magliano, Millis, Ozuru, & McNamara, 2007; Pellegrino, Chudowsky, & Glaser, 2001; Rupp, Ferne, & Choi, 2006). Specifically, there is concern that current assessments often: are poorly aligned with contemporary theoretical constructs and empirical findings pertaining to reading processes and development; are insufficiently sensitive for detecting changes in the kinds of skills that are targeted by interventions; use mainly multiple-choice formats that emphasize strategic reasoning rather than understanding of the text; provide little diagnostic information for guiding instruction; and measure comprehension using tasks and texts that do not represent the full range of the purposeful literacy activities of 21st Century reading.

In view of the foregoing criticisms, it is clear that new reading assessments are needed and desired by educators and researchers. These new assessments should draw upon the lessons that have been learned from prior experiences in both classroom and laboratory, over the past several decades. Our aim is to integrate and extend strengths of past approaches in an innovative way, yet adhere to rigorous design principles that will ensure feasible implementation, good construct coverage, and strong psychometric properties. The instruments' validity and educational utility will also be enhanced, if based on contemporary theory and research on reading, learning, and instruction.

A Conceptual Framework for Assessment of Reading for Understanding (RFU)

The conceptual framework for the design of reading comprehension assessments we propose is a distillation of ideas and evidence, drawn from and integrated over several longstanding theories (Sabatini, Albro, & O'Reilly, 2012; Sabatini, O'Reilly, &Albro 2012). At its foundation are six ***principles***, which are consistent with most contemporary models of

reading, acquisition, and reading disabilities. Associated with each principle are key design implications for assessments. These serve as a guide for the development of 21st Century reading assessments and an accompanying research agenda that can test the validity and utility of particular constructs, design approaches to operationalize them, and procedures for implementing, scoring, and communicating results.

There is no extant theory for which empirical support has been collected across the lifespan to test a unified model that integrates all facets of reading. Collectively, however, research and theories have been proposed and tested on a smaller scale in numerous studies. Admittedly, the empirical work is uneven developmentally (i.e., stronger on componential theories of reading in the K-3 range, stronger on mental representation models of understanding at the upper grades) and there remain healthy debates about which specific theories and models best fit the data. Nonetheless, there is sufficient agreement among theories to identify some basic principles in common, and there is ample empirical support for those principles. The framework represents our best synthesis of reading research as it relates to assessment constructs and design.

Principle 1: Print skills and linguistic comprehension are each necessary components of reading proficiency, though neither individually is sufficient to ensure proficiency. This principle captures the essence of the simple view (Gough & Tunmer, 1986; Hoover & Gough, 1990), in which reading comprehension is viewed as the product of two necessary but nonsufficient abilities: word recognition and linguistic comprehension. In the 20 years since its introduction, this model has been well supported in numerous studies (e.g., most recently, Adlof, Catts, & Little, 2006; Johnston & Kirby, 2006; Vellutino, Tunmer, & Jaccard, 2007). There is also extensive evidence for the validity and utility of word reading and text comprehension measures for younger students and for older readers who lack mastery of basic print and

language skills (e.g., Deno, Fuchs, Marston, & Shin, 2001; Perfetti, Landi, & Oakhill, 2005; Sabatini, 2002; Sabatini, Sawaki, Shore, & Scarborough, 2010).

Inefficient print processing (at both the word and text levels) can also detract from reading proficiency, particularly as texts and tasks become longer and more complex. Across development, print processing skills gradually become more efficient, fluent, and automatized, allowing most cognitive resources to be applied to higher order processes that are necessary for comprehension. With regard to understanding, when basic processes (e.g., decoding, word recognition) are not automatized, they require conscious effort, and draw the reader's attention away from higher level comprehension processes (Cain, Oakhill, & Bryant, 2004; LaBerge & Samuels, 1974; Perfetti, 1985). Indeed, it is well demonstrated that reading proficiency correlates well with the fluency of word- and text-reading (Daane, Campbell, Grigg, Goodman, & Oranje, 2005; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

Furthermore, because unsuccessful reading comprehension can arise from word recognition limitations in young students and in struggling older readers, assessing comprehension proficiency solely in the print modality may underestimate the competence of these students. As shown in differential boost studies (e.g., Cahalan-Laitusis, Cook, Cline, King, & Sabatini, 2008; Fletcher, Denton, & Francis, 2005), these students can often demonstrate stronger comprehension skills when provided accommodations for their weak print skills.

Compelling evidence is not available to identify the set of subskills that are necessary and useful to measure for summative purposes. For instance, there are many reliable and valid instruments for assessing phonological decoding, word recognition, word-reading efficiency, and text-reading fluency. However, it is not clear which of these sub-constructs (at what grade or proficiency levels) adds value in assessment. Furthermore, the 'linguistic comprehension'

construct is rather vaguely defined in the literature and has been operationalized in various ways in research (e.g., Hagtvet, 2003; Keenan, Betjemann, & Olson, 2008), but consistently demonstrates strong predictive power at most ages (Aouad & Savage, 2009; Catts, Adlof, & Weismer, 2006; Catts, Hogan, & Fey, 2003).

Design Implications: Reading comprehension difficulties can arise from weaknesses in either print processing or linguistic comprehension. In the lower grades, both skills should be measured directly. For nonproficient readers in higher grades, continuing the direct assessment of components can add value to summative assessment claims and interpretations.

Principle 2: Both breadth and depth of vocabulary knowledge are essential for understanding. As readers mature, their vocabulary knowledge typically keeps pace with increases in world knowledge, and increases largely by the reader inferring the meanings of unfamiliar words from context. Vocabulary knowledge is important for reading comprehension, and strong correlations (r = .6 to .7) between them are typically observed (Anderson & Freebody, 1981; Daneman, 1988; Hirsch, 2003). This association is evident from the start of schooling and strengthens thereafter (Stanovich, West, & Harrison, 1995). Understanding of a written text is likely to be thwarted if the readers inaccurately recognize the meanings of just 5 to 10% or more of the words in that text (Nagy & Scott, 2000). The reader's depth as well as breadth of word knowledge is important because many words have multiple meanings, or idiomatic usages (Ouellet, 2006).

Given its strong link with comprehension, vocabulary is often measured in reading assessments. Many tests employ multiple choice items to evaluate knowledge of synonyms and definitions, and vocabulary items can also be embedded in continuous text passages to examine contextual effects (e.g., Sheehan, Kostin, & Persky, 2006). Oral language measures of receptive

and expressive vocabulary may include picture naming, picture matching, definitions, and synonym production, among others. Although there is evidence linking reading skills to the use of morphological structure to infer word meaning (Carlisle & Stone, 2003; Kieffer & Lesaux, 2007), assessment of morphology is rarely integrated into reading assessment.

Design Implications: It is important to determine what drives weaknesses in vocabulary: lack of supporting knowledge and vocabulary in specific domains, inability to make context-based inferences about new words, or limited general vocabulary. Hence, comprehension assessments must measure lexical knowledge both within and out of context.

Principle 3: Readers construct mental models of text meaning at multiple levels, from literal to gist to complex situation models. These models reflect and depend on the reader's aims and prior knowledge. Contemporary views of reading comprehension emphasize the importance of differing levels or depths of understanding. To exemplify this principle, we rely on the construction integration ("CI") model (Kintsch, 1988, 1998), which posits three levels of understanding that vary in their stability and depth. The *surface* level is a verbatim representation of the literal words, phrases, and structures of the text. It is typically retained only briefly, during which time the reader analyzes semantic and syntactic relationships to build the *textbase* representation or "gist" of the text. At this level, verbatim retention is lost, but critical meaning is represented abstractly as the key propositions and relationships that can be inferred among them.

The intended meaning of a text cannot always be understood from the textbase, because important pertinent information may be left out (Beck, McKeown, & Gromoll, 1989). The reader must infer meaning based on prior knowledge, resulting in a deeper and more robust level of representation, the situation model (McNamara & Kintsch, 1996). Construction of a *situation*

model allows a reader to learn new information from text by integrating unfamiliar terms and ideas with familiar schemas and knowledge.

A useful assessment, therefore, should indicate how well a student can construct different levels of understanding. It is also valuable to assess the extent to which doing so is constrained by limited background knowledge, different reading goals, tasks and instructions.

Design Implications: Because depth of comprehension varies between and within individuals, assessments should consider the knowledge of the reader and distinguish the literal surface code, textbase, and complex mental schemas (situation models).

Principle 4: Reading is ordinarily a purposeful activity, aimed at attaining a coherent understanding of a text that is sufficient for the reader's goal. Successful readers monitor and self-regulate their comprehension, enabling the repair of mental models as needed. Strategies provide a vehicle for driving deeper levels of processing. One's purpose for reading can influence what is attended to, how it is analyzed, what "standard of coherence" (desired level of comprehension) is adopted, and thus how deeply text is comprehended (van den Broek, Risden, & Husebye-Hartman, 1995). When a low standard of coherence is chosen, gaps in understanding are tolerable to the reader whereas readers who adopt a high standard of coherence must expend additional effort to deepen and embellish understanding, to ensure that information is integrated into a coherent situation model, and to monitor and repair breaks in understanding.

When a higher standard of coherence or deeper level of processing is *demanded*, then the reader may call upon reading strategies as a vehicle for how to construct and organize more robust models of the text. A *reading comprehension strategy* is "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), such as 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). Although a reader's goals are often self-selected, they can be influenced by the nature of the task and text. By varying the texts and instructions during assessment, the reader should adjust their standard of coherence to match the task demands, and performance differences can be evaluated, providing key information for understanding the nature of comprehension difficulties for guiding instruction.

Design Implications: More valid inferences about reading will be obtained by specifying goals for reading activities during assessment, because able readers will evaluate the adequacy of their mental models in relation to those goals, and reconstruct the models accordingly.

Principle 5: Skilled reading includes proficiency in evaluating and synthesizing information across multiple texts. This requirement is driven by the increasing prevalence of digital literacy activities. Reading skills will increasingly be deployed in evolving digital environments, and a hybrid of print and digital skills will be essential to proficiency. Even in elementary school, students are expected to consult multiple sources when engaging in literacy activities involving the internet. However, search engines retrieve enormous numbers of documents on many topics. If a report on "the rainforest" is assigned, hits will likely include a Wikipedia webpage, numerous blogs, the Rainforest Café site, government policy statements, and so forth that vary in content, media (pictures, videos), genre (narrative, argument), and intention (entertain, persuade).

To acquire a deep understanding of multiple documents, the reader must also construct a *situations* model (Perfetti, Rouet, & Britt, 1999) that integrates the information from multiple *document nodes.* A situations model requires a skilled reader to (a) encode source information,

(b) evaluate relevance and trustworthiness, and (c) determine which propositional content from the documents should be emphasized in the situations model or product (e.g., report, poster, etc.).

The relatively recent advent and widespread societal use of the internet has expanded the skill set for literacy; 21st Century readers must be facile in navigating and utilizing e-print environments (Partnership for 21st Century Skills, 2008). Today's proficient reader must be able to deploy skills of searching, retrieving, understanding, locating, evaluating, interpreting, and integrating documents in digital contexts, as well as in print. Compared to print, digital environments are likely to provide novel affordances for deploying literacy skills, including: email, blogging, text messaging, using search engines, navigating websites and so forth (Coiro, 2009). These activities do not have exact parallels in the print world and the easily-accessed information is largely unfiltered (for quality and credibility) and imposes a heavier burden on the reader to understand, evaluate, and interpret the information appropriately and wisely.

Design Implications: Assessing the understanding just one text at a time does not cover the full construct. Evaluating situation models constructed from multiple sources can be used to examine students' evaluation, integration, and synthesis of information. New assessments should be designed to be appropriate for evaluating skills in both print and digital environments.

Principle 6: Growth in reading proficiency consists of incremental expansion of knowledge and skills for the understanding of increasingly complex texts and task demands. Growth is driven primarily by the quality and quantity of instruction, experience, and practice, resulting in substantial variability within and between grade levels, schools, and socioeconomic strata. In our discussion of guiding principles, we have noted there are developmental shifts in the relative importance of each principle in accounting for proficiency differences in reading, and consequently the implications for assessment. In our view, these

developmental changes are gradual and incremental. Although differences in reading performance between students in 1st versus 4th grade, or 6th versus 12th grade, can look qualitatively dissimilar, there is no firm evidence for discrete stages; instead, dramatic improvements over the longer term result from relatively small continuous increments -- in the mastery and automaticity of acquired print skills (Principle 1); in the breadth and depth of oral language knowledge and skills (Principles 1, 2); and in the variety and complexity of texts and tasks (Principles 3, 4, 5).

The rate of growth of these aspects of reading for understanding can differ in ways that will guide the design of the proposed assessments. Notably, as grade increases, (a) *less* emphasis will be placed on examining children's acquisition of print skills; and (b) *more* emphasis will be placed on assessing the construction of mental models for differing text types, reading aims, and multiple media sources. For certain older readers, it might also be useful to measure basic skills.

Design Implications: Assessments should include tasks and reading materials along a developmental continuum of increasing proficiency in all aspects of reading that yield valuable information about achievement differences and sources of difficulty in reading comprehension.

Description of a New Assessment System of Reading for Understanding

Building on this conceptual foundation, we have been designing a new theoretically-based, developmentally sensitive assessment system that consists of two main parts: (1) a set of "integrated" comprehension tasks, in which students read for understanding to attain a defined aim; and (2) a set of supplementary tests of component skills, for use with nonproficient readers, to provide information to identify or rule out potential bases for comprehension difficulties.

The logic of this approach is that global, integrated reading texts and task performances afford multiple cues (e.g., inferential, knowledge) that an individual can exploit to bootstrap

performance to compensate for weak individual component subskills (O'Reilly & McNamara, 2007; Walczyk, Marsiglia, Johns, & Bryan, 2004). For proficient readers, the availability of multiple, overlapping sources of information reduces the complexity of processing. For the nonproficient reader, it may actually enable performance at an artificially high level that is not sustainable as the student encounters more complex texts and tasks in subsequent years. Thus, given their utility for predicting potential risk for a decline in achievement, separate tests of component skills are warranted because weaknesses in those areas could be masked in an integrated assessment.

A Global, Integrated Summative Assessment (GISA) of Reading for Understanding

We envision an integrated assessment of reading comprehension that parallels the kinds of activities that students typically engage in. These activities begin with a specific purpose or goal, and proceed with actions that achieve that goal. The actions include searching for relevant information, evaluating its quality and pertinence, synthesizing and integrating it with other information, and producing some product that satisfies the goal. As such, reading for understanding is not usually a passive activity that involves answering comprehension questions on a collection of unrelated passages, but rather a focused and more complex process of constructing meaning from text(s) in order to meet task goals.

We recognize the challenges and pitfalls of previous performance assessments that relied on a small set of complex tasks and that yielded limited information per individual (weakening reliability and discrimination). In contrast, we envision maintaining a large percentage of discrete, objectively scored items, with a smaller mix of constructed response items. We have had success in models stemming from other reading comprehension projects to build upon

(Bennett, 2011; Deane, Sabatini, & O'Reilly, 2011; O'Reilly & Sheehan, 2009; Sheehan & O'Reilly, 2011).

The GISA will examine the student's proficiencies in a) constructing different levels of representations (textbase, situation, and multiple source) (Kintsch, 1998), b) familiarity with text structure and genre differences (Goldman & Rakestraw, 2000), c) deployment of executive/metacognitive processes (Schraw, 2000), and d) application of strategies for attaining a literacy goal (McCrudden & Schraw, 2007). The goal is to broaden the coverage of the assessment, while maintaining standardized testing conditions and adequate measurement.

Component Skills Assessments

For assessments to more appropriately cover the full reading construct range, we hypothesize that measurement of component skills for *nonproficient readers* is justified and will provide more useful information to guide instructional decision-making than relying solely on integrated measures. Separate component skill subsections of a summative assessment targeted at less proficient readers, can solve the problem and provide more specific information of strengths and weaknesses underlying nonproficient reading. However, on their own, component skills do not sum up to reading proficiency. That is, one could hypothetically be proficient in each subskill and still fail to adequately integrate them.

Conclusions & Implications

Reading or reading comprehension always have been and always will be social constructs. Writing systems change, languages evolve, technologies advance, cultures and societies shift, and the social value and meaning of literacy follow along. We cannot define or legislate what reading comprehension is; but we can observe and describe it in the historical moment, use the tools of science to understand and interpret it; and then hopefully help

individuals and groups to better acquire and use this valuable social technology of learning, communication, personal growth, and societal participation (Venezky, 1990). The rationale and assessment system we have described tries to capture key aspects of literacy in this historical moment. It tries to be as explicit as current research permits, in parsing the construct into elements or principles that we anticipate may interact with individual differences among learners.

The complexities of one's language, one's writing system, and the social context of literacy practices and uses will interact with individual differences, resulting in relative advantages or disadvantages for individuals as they learn to read and become literate. Being more explicit about the elements of the construct, will hopefully improve the value and utility of resulting assessment scores in informing how best to help individuals with differences to demonstrate what they can do and what they struggle with, en route to their acquiring proficiency.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100005 to the Educational Testing Service as part of the Reading for Understanding Research Initiative. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or the Educational Testing Service. We wish to give special thanks to Jennifer Lentini for her help in preparing this chapter. We would also like to thank (*ETS Editor, reviewer 1, reviewer 2, Kim Fryer)* for their helpful comments.

References

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.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching* (pp. 77–117). Newark, DE: International Reading Association.

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.

Beck, I. L., McKeown, M. G., & Gromoll, E. W. (1989). Learning from social studies texts. *Cognition & Instruction, 6,* 99-158.

Bennett, R. E. (2011). *CBAL: Results from piloting innovative K-12 assessments* (Research Report 11-23). Princeton, NJ: Educational Testing Service.

Cahalan-Laitusis, C., Cook, L. L., Cline, F., King, T., & Sabatini, J. (2008). *Examining the impact of audio presentation on tests of reading comprehension* (Research Report 08-23). Princeton, NJ: Educational Testing Service.

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., & 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.

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., 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.

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.

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 No. 2006-469). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education.

Daneman, M. (1988). Word knowledge and reading skill. In M. Daneman, G. E. Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 6, pp. 145–175). San Diego, CA: Academic Press.

Deane, P., Sabatini, J., & O'Reilly, T. (2011). *English language arts literacy framework.* Princeton, NJ: Educational Testing Service.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Fletcher, J. M., Denton, C., & Francis, D. J. (2005). Validity of alternative approaches for the identification of learning disabilities: Operationalizing unexpected underachievement. *Journal of Learning Disabilities, 38,* 545-552.

Goldman, S. R., & Rakestraw, J. A. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. II, pp. 311-335). Mahwah, NJ: Erlbaum.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6-10.

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: Erlbaum.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading & Writing, 16,* 505-539.

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator, 27,* 10–31.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19,* 339-361.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300.

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: Erlbaum.

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.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

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: Erlbaum.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19,* 113-139.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: 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: Erlbaum.

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: Erlbaum.

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: Erlbaum.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high stakes" measures of high school students' science achievement. *American Educational Research Journal, 44,* 161-196.

O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively Based Assessment of, for and as Learning: A framework for assessing reading competency* (Research Report 09-26). 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.

Partnership for 21st Century Skills. (2008). *21st century skills and English map.* 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., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford, UK: Blackwell.

Perfetti, C. A., Rouet, J., & Britt, M. A. (1999). Toward a theory of document representation. In

H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Erlbaum.

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. (2002). Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading, 6,* 267-298.

Sabatini, J., Albro, E., & O'Reilly, T. (2012). *Measuring up: Advances in how we assess reading ability.* Lanham, MD: Rowman & Littlefield Education.

Sabatini, J., O'Reilly, T., & Albro, E. (2012). *Reaching an understanding: Innovations in how we view reading assessment.* Lanham, MD: Rowman & Littlefield Education.

Sabatini, J., Sawaki, Y., Shore, J., & Scarborough, H. (2010). Relationships among reading skills of low-literate adults. *Journal of Learning Disabilities, 43,* 122-138.

Schraw, G. (2000). Assessing metacognition: Implications of the Buros symposium. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 297-321). Lincoln, NE: Buros Institute of Mental Measurements.

Sheehan, K. M., Kostin, I., & Persky, H. (2006, April). *Predicting item difficulty as a function of inferential processing requirements: An examination of the reading skills underlying performance on the NAEP grade 8 reading assessment.* Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA.

Sheehan, K. M., & O'Reilly, T. (2011). *The CBAL reading assessment: An approach for balancing measurement and learning goals* (Research Report 11-21). Princeton, NJ: Educational Testing Service.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31,* 811-826.

van den Broek, P., Risden, 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: Erlbaum.

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.

Walczyk, J. J., Marsiglia, C. S., Johns, A. K., & Bryan, K. S. (2004). Children's compensations for poorly automated reading skills. *Discourse Processes, 37,* 47-66.

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.

Yu, G. (2003). Reading for summarization as reading comprehension test method: Promises and problems. *Language Testing Update, 32,* 44-47.