ETS Research Spotlight

Innovation in Reading Assessment

Foreword

Digital copies of this report can be downloaded free of charge from: **www.ets.org/research/spotlight7**

To comment on any article in this issue, write to: **RDWeb@ets.org**

In 2009 and 2011, the results of the National Assessment of Educational Progress (NAEP) reading section pointed to a situation many policymakers found alarming and unacceptable.

The results suggested that, in grades 4 and 8, too few U.S. students were reading at levels many people consider sufficient to support the aims of a nation trying to remain competitive in a global and technologically advanced economy: In 2009, just 33 percent of fourth graders and 32 percent of eighth graders were reading at or above the "proficient" level on this group survey, often known as "The Nation's Report Card." In 2011, 34 percent at each grade level were scoring at or above the proficient level. Meanwhile, in 2011, 33 percent of fourth graders and 24 percent of eighth graders were below the basic level.[1]

As *Education Week* reported in March 2010, "Some officials saw the NAEP results as a call to arms."[2]

At ETS, we also saw the NAEP results as affirming the importance of expanding our research into reading proficiency. This research will help develop the kinds of assessment innovations needed to meet challenges such as those seen in the NAEP results. This edition of *ETS Research Spotlight* summarizes three recent publications that exemplify this research.

The first publication, covered in this issue's Featured Research Synopsis, is an article published earlier in 2012 in the journal *Reading Psychology*. The article's authors include four members of ETS's Research & Development division — Tenaha O'Reilly, John Sabatini, Kelly Bruce, and Srinivasa Pillarisetti — and Carina McCormick of the University of Nebraska. In their article, "Middle School Reading Assessment: Measuring What Matters Under a RTI Framework," the authors propose an approach to reading assessment that aims to address reading problems before students fall behind their peers.

The other publications comprise a two-volume book series covering innovations in the theory and practice of measuring reading comprehension. The summary of these books, *Measuring Up: Advances in How We Assess Reading Ability* (Volume I) and *Reaching an Understanding: Innovations in How We View Reading Assessment* (Volume II), starts on page 7 of this issue.

The book editors are Sabatini, an ETS Principal Research Scientist; O'Reilly, an ETS Research Scientist; and Elizabeth R. Albro, Associate Commissioner for Teaching and Learning at the Institute of Education Sciences (IES). The books are related to ETS's work on the federally funded Reading for Understanding research initiative.

To learn more about the Reading for Understanding initiative, visit **www.ets.org/research/topics/reading_for_understanding**. If you'd like to learn more about the research we conduct to support and improve assessments, visit us on the web at **www.ets.org/research**.

Ida Lawrence
Senior Vice President
Research & Development

---

[1] For more information, see **http://nces.ed.gov/nationsreportcard**.
[2] *Education Week*, March 24, 2010 edition, see **http://www.edweek.com** (subscription required).

FEATURED RESEARCH SYNOPSIS

Middle School Reading Assessment: Measuring What Matters Under a RTI Framework

1) Introduction

The article featured in this issue of *Research Spotlight* proposes a framework for reading assessment of middle school children. A large proportion of students in the United States have a hard time understanding what they read; one in three fourth graders and one in four eighth graders score below basic in reading, according to the results of the 2009 National Assessment of Educational Progress (NAEP). The authors see this as a national issue for the United States, a problem they want to address by focusing on one aspect of it: the nature of some commonly used assessments.

Today's off-the-shelf reading assessments have been criticized for various reasons, and so has the process of designing them. Many tests, the authors note, are measurement-driven and geared towards rank-ordering, but fail to provide clear and specific information that can be used to help students early enough to avoid the need for special assistance. The authors seek an alternative approach rooted in cognitive science and learning theory and built on the principles of evidence-centered design (ECD).

> *The authors created a computer-based screening battery with six subtests (also called the Screener in the article) that could be administered and scored quickly at all middle schools in a school district.*

2) Approach and Methods

Instead of the traditional "blueprint specification" approach to creating reading comprehension assessments, which creates items based on lists of content and skills, the authors suggest a cognitive competency approach that starts with claims about what students ought to know and be able to do. The design also takes into account how the test result should be used and reported. The authors integrate this approach with a framework known as Response to Intervention (RTI) that is intended to prevent students

from being labeled as learning or reading disabled by addressing their difficulties before they fall too far behind their peers. RTI provides early screening for reading problems and monitors students' progress in order to deliver intervention and instruction adapted to the student's needs.

The article introduces a prototype competency model for reading assessment of middle school students. The model — originally created as part of an ETS research initiative known as the Cognitively Based Assessment *of*, *for*, and *as* Learning[3], or *CBAL*™ initiative — aims to guide assessment development, leading to a test design and content that targets three "branches" of competencies or components necessary for successful reading comprehension:

I. *The skills branch:* Decoding texts, recognizing words and fluency in reading.

II. *The strategy branch:* Constructing meaning from decoded text or spoken language.

III. *The text analytics branch:* Ability to make sense out of texts they encounter as readers.

As part of their study, the researchers designed a screener to spot reading problems early, before students are at risk for being classified as having a reading disability. The screener's six subtests covered these areas:

1. Word Recognition and Decoding
2. Vocabulary
3. Morphology — Word Formation
4. Sentence Processing
5. Efficiency of Reading for Basic Meaning
6. Reading Comprehension

The authors proceed to explain how the model can be integrated with RTI, and introduce a competency based *assessment framework* with these three key elements:

a) a universal screener

b) a set of formative assessments

c) an integrated summative assessment

Taken together, these three elements form an assessment framework that the authors believe will yield more information about the students' reading comprehension and at the same time support their learning and skill development. The universal screener is designed to spot reading problems early so that remedial action can be taken sooner for students who experience reading difficulties. The goal of the universal screener is to provide decision makers with unique information on critical reading subskills, but the results at the top and bottom of the scale are often intercorrelated as strong readers tend to do well overall and poor readers tend to do poorly. Between them is a middle layer of struggling readers where it is possible to distinguish how well they are doing for different subskills. For the model to assess these readers in a

---

[3] For more information on the CBAL Initiative, see **www.ets.org/research/topics/cbal**.

meaningful way, the screener must capture sufficient variation in subskills. The authors therefore created a computer-based screening battery with six subtests (also called "the Screener" in the article) that could be administered and scored quickly at all middle schools in a school district.

To examine the Screener, the authors conducted a study that involved 3,372 students in grades six through eight at nine middle schools in a small urban school district. The goal was to evaluate the efficacy of their competency-based RTI framework. To identify struggling readers, the authors compared the results of the Screener with its six subtests to scale scores from the Language Arts Subtest in the state where the school district was located. (The raw scores for each of the Screeners' subtests were first converted to scale scores in a range that made it possible to compare them with those of the state test.)

3) Analysis and Results

The authors split the student population of the test into two groups based on their results on the state test. One group included students who were rated as less than proficient based on their state test results, while the other group consisted of those who had been classified as proficient based on state test results. The latter group scored high on all of the Screener's subtests, which was expected. The result indicates that the Screener test battery is sensitive to the state tests' skill level. The next step was to determine whether the Screener would be useful in identifying specific subskills in which below-proficient readers were having difficulty. This was done using a statistical method known as multiple-regression analysis, which the authors applied sequentially: Each new round added one subtest scale score to the analysis until all six subtests were included. The researchers performed this analysis separately both for the below-proficient group and for the proficient group.

As they expected, they found that the Screener seemed to discriminate among below-proficient readers' individual strengths and weaknesses, but was generally not sensitive to proficient readers' strengths and weaknesses in the six areas assessed. These results were consistent with the authors' goal of creating an assessment tool that generates useful information about the skills of struggling readers: Whereas proficient readers are generally strong across all of the subskills tested, below-proficient readers may be in greater need of remediation in some areas compared with others — a fact

In a concluding discussion, the authors suggest that their computer-based Screener can help teachers and administrators to identify and make decisions about the need for intervention.

that could make the results of the Screener actionable and therefore useful in instruction. Finally, the researchers used a common statistical procedure to check the reliability of each of the groups they studied. For more information on the technical details of this procedure, see the original article referenced in the introduction of this summary.

4) Discussion

In a concluding discussion, the authors suggest that their computer-based Screener can help teachers and administrators to identify and make decisions about the need for intervention. The proposed assessment model can be integrated with RTI and support monitoring students' progress, evaluate interventions, and assess whether students can use reading strategies, and if such strategies have an impact on reading comprehension.

The authors caution that there are limitations to their study as it was built on a relatively small urban sample in one state and was not customized for students with limited English-language skills. The Screener does not cover the full range of skills and activities in which a student might be expected to engage. It also focused on one aspect of the model, while leaving out formative and summative assessments.

But the authors are optimistic overall and believe that this study is a first step towards developing new reading assessments in support of learning and evidence-based decision making.

Recent Books Related to the Reading for Understanding Initiative

ETS researchers are the editors of a recently released two-volume series that covers the science and practice of measuring reading comprehension and emerging innovations in this area of work. The books are related to ETS's work on the Reading for Understanding (RFU) initiative, funded by the U.S. Department of Education's Institute for Education Sciences.

Both books, *Measuring Up: Advances in How We Assess Reading Ability* (Volume I) and *Reaching an Understanding: Innovations in How We View Reading Assessment* (Volume II), were published in 2012 by Rowman & Littlefield. The editors are John P. Sabatini, Tenaha O'Reilly, and Elizabeth R. Albro. Sabatini and O'Reilly are research scientists in ETS's Research & Development division. Albro is Associate Commissioner for Teaching and Learning at the U.S. Institute of Education Sciences (IES).

The books aim to offer a fundamental rethinking of traditional "off-the-shelf" reading comprehension tests, which typically require test takers to read a passage, answer a set of questions about it, then repeat this pattern with additional, disconnected passages.

For the past two years, Sabatini and O'Reilly have led ETS's research consortium team, which is charged with building innovative reading assessments as part of the RFU initiative. The books aim to offer a fundamental rethinking of traditional "off-the-shelf" reading comprehension tests, which typically require test takers to read a passage, answer a set of questions about it, then repeat this pattern with additional, disconnected passages.

In each volume, the chapter authors include leading educational measurement researchers, reading researchers and practitioners in the field. Exploring new and several overlapping concepts, Volume I takes a theoretical approach to defining the need to advance the science and practice of reading assessment and proposes solutions for doing so. The table on page 7 summarizes the contents of Volume I.

Volume II focuses on the development of new assessments and measurement models and identifying innovative ways to better assess reading comprehension in the 21st century. The table on page 8 summarizes the contents of Volume II.

All volume chapters are related to papers that the expert authors initially presented during an invitational conference on Assessing Reading in the 21st Century in Philadelphia, Pa., organized by editors Sabanti, Albro, and O'Reilly in 2008.

Summary: *Measuring Up: Advances in How We Assess Reading Ability*		
Chapter	**Chapter Authors**	**Contents**
1	**Charles Perfetti** **Suzanne M. Adlof**	The authors analyze basic skill components necessary for comprehension, focusing on a number of key elements including word identification, inferences, strategies, vocabulary, and lexical knowledge.
2	**Walter Kintsch**	The author emphasizes the role cognitive factors such as prior knowledge and motivation play in reading comprehension and their implications for the design of the next generation of reading assessments.
3	**Paul van den Broek**	This chapter offers insights on the strategic processes that lead to deep comprehension. The author examines the importance of capturing and connecting key meanings among the central ideas in a text.
4	**Kate Cain** **Jane Oakhill**	The authors examine components that contribute to comprehension. They describe findings from a longitudinal study that explored whether inference and integration, comprehension monitoring, and the use of narrative structure predict growth in reading comprehension. They use their findings to argue for the need to assess these skills while children begin to learn to read in order to effectively shape instruction.
5	**Janice M. Keenan**	What do current measures of reading comprehension measure? Do measures used with different-aged readers function in the same way? The author examines these questions in the context of considering how reading comprehension tests should vary in the skills they assess.
6	**Danielle S. McNamara** **Arthur Graesser** **Max Louwerse**	The authors examine the importance of text cohesion and factors that impede it. They present an automated tool that they devised to identify potential sources of text difficulty.
7	**Robert J. Mislevy** **John P. Sabatini**	The authors look at the critical role ECD plays in integrating cognitive theory into the design of an assessment before the test has been developed. They advocate the use of ECD to make assessments more meaningful and informative for instructional purposes.

8	**André A. Rupp**	In this chapter, the author reviews existing diagnostic classification models, highlighting the theoretical potential and pitfalls of such models to illustrate current core processes and strategies involved in reading processes.
Conclusion	**Arthur Graesser** **Xiangen Hu**	In the concluding chapter, the authors summarize the reading research and provide recommendations for improving the next generation of reading assessments (for example, providing broader diversity of texts and integrating of text with multimedia).

Summary: *Reaching an Understanding: Innovations in How We View Reading Assessment*		
Chapter	**Chapter Authors**	**Contents**
1	**Kimberly A. Lawless** **Susan R. Goldman** **Kimberly Gomez** **Flori Manning** **Jason Braasch**	The authors present empirical data indicating key differences between skilled and less-skilled readers. They argue that users of traditional measures of reading comprehension may overlook these differences. They also argue that the next generation of assessments should include tasks and activities that require test takers to integrate and synthesize information from multiple sources.
2	**Kathleen M. Sheehan** **Tenaha O'Reilly**	The authors emphasize the importance of incorporating authentic purposes for reading directly into the design of an assessment and measuring and capturing key data while a student reads. The authors argue that their approach has promising implications for instruction.
3	**Keith Millis** **Joseph Magliano**	The authors present a system called Reading Strategies Assessment Tool (RSAT). The automated measure is designed to capture both online and offline reading comprehension, processing, and reading strategies. They contend that their tool will be useful not only for understanding how students form a mental representation of text, but also may reveal particular subskill weaknesses, misconceptions, gaps or fundamental errors in understanding.
4	**Donald L. Compton** **Amy M. Elleman** **Hugh W. Catts**	The authors examine instruction techniques targeted to address particular weaknesses early on. They explore subskills at both the word reading and comprehension levels, including subword orthographic-phonological connections. They contend that diagnostic assessments currently used with young readers do not measure many of these subskills.

5	Carol McDonald Connor Frederick J. Morrison Barry Fishman Christopher Schatschneider	A challenge most teachers confront is how to use the information gleaned from current assessments. In this chapter, the authors demonstrate how they believe assessments can be used to inform instruction. They present early findings from their research on a software system they built that groups students according to their instructional needs and aims to match teachers' current curricular materials to those needs.
6	Laurie E. Cutting Hollis S. Scarborough	The authors build further upon the need for multiple measures of reading comprehension skill. They describe how various reading comprehension assessments provide different classifications of readers (for example, one assessment may identify an individual as having poor comprehension skills, while another assessment may identify the same person as having good comprehension skills). The authors also contend that listening comprehension, which encompasses several factors such as speech perception, word recognition, and sentence processing, also deserves close examination.
7	Paul Deane	The author advocates for a more extensive measure of vocabulary as an integral part of the process, arguing that vocabulary acquisition is not an "all or none" process of learning words and their meanings in piecemeal fashion. Instead, the author contends, words are learned over time and to varying degrees relating to their contexts.
8	Mark Wilson Stephen Moore	The authors use psychometric models to show how they believe an individual's personal characteristics can interact with characteristics of a test item to predict the person's performance on the test.
Conclusion	Joanna Gorin Dubravka Svetina	In the concluding chapter of the volume, the authors explore an approach to test construction known as cognitive psychometric models (CPM). In this process, test designers write items that aim to elicit evidence of why students perform poorly on assessments. This is another new technique that the authors believe will allow reading comprehension assessments to be more diagnostically sensitive.

About ETS

At ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English language learning, and elementary, secondary and post-secondary education, as well as conducting education research, analysis and policy studies. Founded as a nonprofit in 1947, ETS develops, administers and scores more than 50 million tests annually — including the *TOEFL®* and *TOEIC®* tests, the *GRE®* tests and *The Praxis Series™* assessments — in more than 180 countries, at over 9,000 locations worldwide.

Scan this code to read the publication
on your mobile device.

ETS

Listening. Learning. Leading.®

www.ets.org