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.
SARA Reading Components Tests, RISE Form:
Test Design and Technical Adequacy

John Sabatini, Kelly Bruce, and Jonathan Steinberg
ETS, Princeton, New Jersey

April 2013
Action Editor: Marna Golub-Smith

Reviewers: Laura Halderman and Rick Morgan

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

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).
Abstract

This paper describes the design, development, and technical adequacy of the Reading Inventory and Student Evaluation (RISE) form of the Study Aid and Reading Assessment (SARA) reading components computer-delivered assessment. The RISE form, designed for middle-school students, began through a joint project between the Strategic Education Research Partnership (SERP), ETS, and a large urban school district, where middle-school literacy had been identified as an area needing improvement. Educators were interested in understanding more about the component skills of their middle-school students, particularly their struggling readers, using an efficient and reliable assessment. To date, results from our piloting of the RISE form with middle-school students have established its technical adequacy. In the future, we plan to expand the RISE to create parallel forms for 6th-to-8th graders as well as to explore development of new test forms at other points in the grade continuum.

Key words: reading components, reading assessments, computer-delivered assessment, middle grades reading skills, state test score comparison, correlation analysis
Acknowledgments

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305G04065 and Grant R305F100005 to ETS as part of the Reading for Understanding Research Initiative. We would also like to acknowledge collaboration and support from the Strategic Educational Research Partnership (SERP) Institute. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or ETS. We would like to like to thank Suzanne Donovan and Catherine Snow for their partnership and support in this endeavor; Rick Morgan for technical support; Ruth Greenwood, Jennifer Lentini, and Kim Fryer for editorial assistance; and Gary Feng, Laura Halderman, Tenaha O’Reilly, and Marna Golub-Smith for helpful comments.
Background

What Are SARA and the RISE Form?

The SARA (Study Aid and Reading Assistant) is, collectively, a system of componential reading assessments that are typically administered electronically. It is also the research program surrounding its development and its application in educational settings. The system is built upon a base of research studies from the fields of cognition, linguistics, and neuroscience and is being developed to target populations of developing readers across the lifespan who may be learning to read in a variety of settings including (but not limited to) elementary, middle, and high schools; community colleges; and adult literacy programs. SARA batteries are computer-delivered and modularized, giving the system the flexibility to respond to the varying purposes and uses of assessment information for decision-making by stakeholders (e.g., students, teachers, administrators, researchers). Different forms are adapted to target different purposes and uses, such as screening, diagnostic, placement, formative, benchmark, outcome, or research evaluation. These different uses may target information at the individual, classroom, school, or other aggregate units.

The RISE (Reading Inventory and Student Evaluation) form is a 45-minute, computer-administered reading-components assessment with six subtests:

- Subtest 1: Word Recognition & Decoding
- Subtest 2: Vocabulary
- Subtest 3: Morphological Awareness
- Subtest 4: Sentence Processing
- Subtest 5: Efficiency of Basic Reading Comprehension
- Subtest 6: Reading Comprehension

The RISE form is a specific subset of SARA test forms that targets sixth to eighth grade students and is designed to provide scores that can inform educational decision-making at the district, school, and teacher levels.

An Overview of the RISE Form

Background. In 2006, the Strategic Education Research Partnership Institute (SERP) and a large, urban school district in the Northeastern United States undertook a collaboration to
identify and to address pressing issues within the district’s schools. One result of this collaboration was the selection of middle-school literacy as an area of need. Simply put, many middle-school students were arriving in classrooms unable to read and understand the material in their textbooks. Furthermore, although students were subjected to numerous tests throughout the school year, the results were not considered particularly useful in terms of providing information that would allow schools to understand the extent of reading component weaknesses within the building or of providing teachers with an understanding of specific component weaknesses of their students.

To illustrate this point, consider a typical scenario: Students take a paper-and-pencil reading comprehension test—for example, an end-of-year state test—which yields a reading score and a descriptive category along the lines of *Advanced, Proficient, Needs Improvement,* or *Warning.* For most students who fall into the top two categories (*Advanced* and *Proficient*), that single score provides all the information a teacher needs. The students are skilled readers who, for the most part, will be able to handle the subjects and materials of their middle-school curricula.

But for the readers who do not perform well on the end-of-year test—those who fall into the bottom categories (*Needs Improvement* and *Warning*)—the single score provides little information on why they failed to do well. Teachers may ask: Are they having trouble with word recognition or decoding? With vocabulary? With understanding how words are formed? Do they understand the variety of syntactic structures present in sentences? Are they struggling to understand passages at a basic, literal level? Are they inefficient readers? Concomitantly, principals may ask: How many students in my school are struggling with basic decoding, vocabulary, fluency, and so forth? Do we have enough, or any, interventions that address their needs?

**Goals.** As a result of the discussions between the district and SERP regarding the state of assessment and the district’s needs, the goals for the RISE form of the SARA were to:

- Provide information on the individual subskills (components) that contribute to efficient reading
- Use technology (school computer labs) wherever possible
- Be administered in one class period
• Return results in a timely manner

**Development.** The development of the RISE was informed by a rich base of research studies from the fields of reading, cognition, and linguistics that support the measurement of the individual components of reading. (See the Subtest and Item Type Design section, in this report.) The content of the RISE subtests was modeled on the kinds of materials (words, sentences, and passages) that middle-school students will encounter in their school curricula, as determined by a review of formal and informal curricular materials targeted for this population. Additionally, a series of pilot projects conducted in multiple middle schools with over 5,000 students between 2006 and 2009 helped to guide the development of the RISE form. These pilots allowed us to select content for the RISE form that spanned a range of difficulty appropriate for middle-school students. This range was particularly important given that a single form was administered to sixth through eighth graders.

**Conceptual Framework**

**The Simple View of Reading**

The initial framework we drew upon in the design of the RISE was the *Simple View of Reading*, which provides a way to conceptualize the reading process. As described by Hoover and Tunmer (1993), “The simple view makes two claims: first, that reading consists of word recognition and linguistic comprehension; and second, that each of these components is necessary for reading, neither being sufficient in itself” (p. 3). Strucker, Yamamoto, and Kirsch (2003) used a similar framework when they described *print components* (e.g., decoding accuracy and fluency) and *meaning components* (e.g., oral vocabulary).

Important to note is that the components do not develop hierarchically; that is, a student does not need to acquire a fully developed set of recognition and decoding skills before any meaning can be constructed from text. While having some recognition skill is foundational to reading, students can handle the meanings of words and deal with sentences and passages even as they are learning the orthography, morphology, and syntactical structures of the language. Thus, during the reading acquisition process, measuring the components separately can yield valuable information to inform instruction.
The Roles of Efficiency and Fluency

The Simple View continues to inform and challenge reading researchers (Hogan, Bridges, Justice, & Cain, 2011; Vellutino, Tunmer, Jaccard, & Chen, 2007). The relationship of efficiency of processing and fluency to word recognition and comprehension has become of key interest to researchers and, therefore, a chief concern of recent work is how and whether to integrate measures of word and text reading speed and fluency into the Simple View (Carver & David, 2001; Kame’enui & Simmons, 2001; National Institute of Child Health and Human Development [NICHD], 2000; Sabatini, 2002; Sabatini & Bruce, 2009).

That skilled reading is associated with fast, accurate, and relatively effortless recognition of words and text is well documented (Adams, 1990; LaBerge & Samuels, 1974; Perfetti, 1985; Share & Stanovich, 1995; Verhoeven & Perfetti, 2011). This finding has been known and studied dating back well over 100 years by notable figures such as James Cattell and Edmond Huey (see Samuels, 2006, for a brief history). In the 1970s, LaBerge and Samuels (1974) built this empirical observation into a theory of automatic information processing, or *automaticity* for short. With respect to early reading, when a learner first starts to learn a skill like decoding or recognizing words, he or she needs to learn the basics of the task such as how to identify and distinguish the letters and sounds of the alphabet, common sound-letter correspondences, and how to use this knowledge to sound out words; that is, how to decode. With some practice, students’ responses may become accurate, but still be slow and require considerable cognitive effort. With more practice, responses become fast, accurate, and require considerably less cognitive effort. We say their performance is becoming more automatic (Samuels, 2006; Stanovich, 1991).

For the purposes of the RISE form, efficiency/fluency has been operationalized in the time limits implemented for each subtest or section within a subtest. Time limits were determined through early pilot studies of the RISE and are based on the amount of time needed for 90% of the participants to finish all items within a subtest. Items that are not reached are considered incorrect. While there is a practical dimension to the implementation of the time limits (i.e., a goal was to design a components battery that could be completed in one class period), efficiency is also central to each of the underlying constructs of the RISE.

Table 1 presents the number of items in each subtest and the corresponding time limits.
Table 1

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>Number of items</th>
<th>Subtest time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding</td>
<td>50</td>
<td>6 minutes</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>38</td>
<td>6 minutes</td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>32</td>
<td>7 minutes</td>
</tr>
<tr>
<td>Sentence Processing</td>
<td>26</td>
<td>7 minutes</td>
</tr>
<tr>
<td>Efficiency of Basic Reading Comprehension</td>
<td>36</td>
<td>9 minutes</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>22</td>
<td>20 minutes</td>
</tr>
</tbody>
</table>

Subtest and Item Type Design

Subtest 1: Word Recognition & Decoding

Most models of reading development recognize the centrality of rapid, automatic visual word recognition to reading ability (Abadzi, 2003; Adams, 1990; Ehri, 2005; Perfetti, 1985; Verhoeven & Perfetti, 2011). Without going into great detail on the mechanisms of word recognition (which are still under study in the psychological sciences), there are two basic behavioral skills that are indicative of proficiency in word recognition. The first is the accumulation of sight-word knowledge of real words in the language, which is facilitated by repeated exposure to text. The second, more fundamental, skill is decoding, which enables the generation of plausible pronunciations of printed words and conversely, plausible phonetic spellings of heard words. Decoding has been described as the fundamental word learning mechanism in alphabetic languages (Share, 1997), and therefore an essential component to measure directly. When students employ their word recognition and decoding skills, they exploit several aspects of word knowledge: the memory trace of the visual representation of words, knowledge of spelling patterns, and an understanding of sight-to-sound correspondences.

Historically, the measurement of word recognition and decoding has been at the core of reading research and the reading test industry. Experimental studies have used lexical decision tasks to investigate the underlying processes of word recognition, while well-respected, standardized tests such as the Woodcock-Johnson III use separate subtests to measure the
participant’s ability to identify letters/real words and to produce plausible pronunciations of nonwords.

Lexical decision tasks typically present an item and require the participant to decide whether that item is a real word or not; the participant is not required to pronounce the word out loud. To select real words, task designers generally use morphological complexity and word frequency as indices for difficulty. They include a range of real words, from those that are highly frequent and easy (e.g., house and natural) to those that are infrequent, morphologically complex, and difficult (e.g., furrow and impenetrable). The nonwords included in the task sets may include a mix of nonplausible spelling patterns (e.g., vext) and plausible spelling patterns (e.g., braff).

Word recognition and decoding tasks like those in the Woodcock-Johnson III, on the other hand, present an item and require the participant to say the word/nonword out loud. Real words in the tasks generally represent a range of word frequencies. Nonwords represent plausible (i.e., pronounceable) spelling patterns.

In the Word Recognition & Decoding subtest of the RISE, we take a somewhat innovative approach to task design in order to measure both a student’s ability to recognize sight words and to decode nonwords. The task set contains three item types:

1. Real words, selected to cover a wide frequency range with a bias toward including the kinds of content area words that middle-school students will encounter in their school curricula. Examples of RISE real words are: elect, mineral, and symbolic.

2. Nonwords, selected to cover a range of spelling and morphological patterns. Examples of RISE nonwords are: clort, plign, and phadintry.

3. Pseudohomophones, nonwords that nonetheless when pronounced sound exactly like real English words. Examples of RISE pseudohomophones are: whissle, brane, and rooler.

Students are presented with one of the item types on the screen at a time and are asked to decide if what they see:

1. Is a real word.
2. Is not a real word.
3. Sounds exactly like a real word.
Because of the combination of item types within the task, students are given practice to understand the decision process they are asked to make. This practice is delivered both outside of the testing situation as part of the student practice guide and is also embedded in a short RISE tutorial just prior to the presentation of the subtest.

**Subtest 2: Vocabulary**

Very simply, a barrier to understanding what one reads is not knowing the meaning of the printed words. One can infer meanings of unknown words from context (while reading or listening), but this typically produces provisional, uncertain, and incomplete word meanings—the understanding of which must be separately verified (e.g., checking definitions in a dictionary). It has been estimated that adequate reading comprehension depends on a person already knowing between 90 and 95% of the words in a text (Nagy & Scott, 2000). Students who are—or will become—skilled readers tend to have larger vocabularies than struggling readers even early in school, and the gap tends to grow dramatically over time (Hart & Risley, 1995).

In middle school, students begin to encounter general purpose academic words—or what might be called Tier 2 words as described by Beck, McKeown, and Kucan (2002)—as well as more specialized content area words—their Tier 3. Comprehension of textbooks and other curricular materials will be affected by the amount and quality of a student’s vocabulary knowledge of Tier 2 words (e.g., *analyze, factor*) and Tier 3 words (e.g., *parliament, hemoglobin*) as well as knowledge of the topical connections among these words.

As the National Reading Panel put it, “The measurement of vocabulary is fraught with difficulties” (NICHD, 2000, p. 4-15). As support, the panel cited the different kinds of vocabularies that can be measured (receptive vs. productive and oral vs. printed) and the fact that only a limited number of words can be tested in any single test. In the design of the RISE Vocabulary subtest, we chose to test written, receptive vocabulary (that is, the student must read the target word and select an answer from a set of choices). In doing so, we acknowledge that word recognition and decoding will play a role in student performance, as both are prerequisite to print vocabulary skill, though they are not the same as knowing word meanings. The primary construct remains vocabulary, with decoding and word recognition as necessary moderators of performance.

The RISE Vocabulary subtest target item set includes both Tier 2 and Tier 3 words. The response sets were designed such that the correct answer was either a synonym of the target or a
meaning associate. The three choices in the response set were selected using several factors, including part of speech and printed word frequency values from a well-known corpus (Zeno, Ivens, Millard, & Duvvuri, 1995). The design goal was to avoid cases where one response (whether correct or incorrect) would stand out because of very low or very high frequency with respect to the other choices.

Correct answers are underlined and placed in the first position in the following examples for this and for all subsequent subtest examples:

Example of a *synonym* item: data (information, schedule, star).

Example of a *meaning associate* item: thermal (heat, bridge, evil).

Because of the combination of synonym and meaning associate item types, students are given practice and examples to understand how to complete the task successfully. This practice is delivered both outside of the testing situation as part of the student practice guide and is also embedded in a short RISE tutorial just prior to the presentation of the subtest.

**Subtest 3: Morphological Awareness**

Morphemes are the basic building blocks of meaning in language. In English, a morpheme may be a single letter, as in the -s in *boys* or *runs*; a prefix or suffix, as in the un- in *unhappy* or the -ity in *acidity*; or an entire word, as in *screen*. When morphemes mark a plural or verb tenses, as they do in *boys* and *runs*, they are called *inflectional morphemes*. When they change the meaning or part of speech of a word, as prefixes and suffixes do, they are called *derivational morphemes*. Anglin (1993) and Nagy and Anderson (1984) estimated that over half of English words are morphologically complex—that is, they are made up of one or more morphemes. As students enter middle school, they encounter texts of greater and greater morphological complexity (e.g., a social studies unit on government may contain words such as *politics, political, apolitical, politician, politically*, and so forth).

Morphological awareness is the extent to which students recognize the role that morphemes play in words—both in a semantic and syntactic sense. A growing body of research suggests that morphological awareness is related to reading comprehension as well as the subskills that underlie reading (e.g., Carlisle, 2000; Carlisle & Stone, 2003; Fowler & Liberman, 1995; Hogan et al., 2011; Kuo & Anderson, 2006; Tong, Deacon, Kirby, Cain, & Parrila, 2011).
Mahony, Singson, and Mann (2000), for instance, found independent contributions of morphological awareness to decoding in elementary school children. Nagy, Berninger, and Abbott (2003) found that morphology contributed to reading comprehension for second grade students and was correlated to word recognition. Using similar techniques, they also found that morphology made unique contributions to reading comprehension, reading vocabulary, and spelling for fourth through ninth graders and to some measures of decoding accuracy and rate in fourth and fifth graders and in eighth and ninth graders (Nagy, Berninger, & Abbott, 2006).

In designing the RISE Morphological Awareness subtest, we chose to focus on derivational morphology—those words that have prefixes and/or suffixes attached to a root word as in: preview = pre [prefix] + view [root]. This choice was driven primarily by the recognition that a great deal of middle-school academic vocabulary includes these kinds of derived words. We also chose to use the cloze (fill-in-the-blank) item type for this subtest. The student sees a fill-in-the-blank sentence with three choices, all of which are derived forms of the same root word, and picks out the one word that makes sense in the sentence. The design of the task was inspired by Mahony (1994) and Mann and Singson (2003).

The sentences we designed featuring straightforward syntactic structures and relatively easy ancillary vocabulary so that the students concentrate on the derived words. We used word frequency lists to assist in the selection of a range of easy, medium, and difficult derived forms. See the examples below.

- The target derived form is *high* frequency:
  
  For many people, birthdays can be times of great _________.
  
  (happiness, unhappy, happily)

- The target derived form is *medium* frequency:
  
  She is good at many sports, but her _________ is basketball.
  
  (specialty, specialize, specialist)

- The target derived form is *low* frequency:
  
  That man treats everyone with respect and _________.
  
  (civility, civilization, civilian)
Practice items for this subtest were delivered both outside of the testing situation as part of the student practice guide and were also embedded in a short RISE tutorial just prior to the presentation of the subtest.

**Subtest 4: Sentence Processing**

A variety of research studies have shown that the sentence is a natural breakpoint in the reading of continuous text (e.g., Kintsch, 1998). A skilled reader will generally pause at the end of each sentence in order to encode the propositions of the sentence, make anaphoric inferences, relate meaning units to background knowledge and to previous memory of the passage as it unfolds, and decide which meaning elements to hold in working memory. Thus, every sentence requires some syntactic and semantic processing.

In middle school, students encounter texts that contain sentences of a variety of lengths and syntactic structures. For example, consider the case of a student who is assigned to read poetry in English language arts, the Declaration of Independence in social studies, a word problem with several embedded steps in math, and a textbook chapter on photosynthesis in science.

Many current measures of sentence processing focus on a single aspect of the construct: grammaticality, working memory, sentence combining, or the disambiguation of meaning. In the RISE Sentence Processing subtest, we chose to focus on the student’s ability to construct basic meaning from print at the sentence level. The cloze items in the subtest require the student to process all parts of the sentence in order to select the correct answer among three choices. In other words, more than one choice may fit the sentence at the phrase level, but only one choice will make sense in terms of the entire sentence. In designing the sentence items, we controlled for vocabulary so that sentence length and syntactic complexity reflected the difficulty in each item. See the examples below.

The dog that chased the cat around the yard spent all night ________.

(barking, meowing, writing)

Shouting in a voice louder than her friend Cindy’s, Tonya asked Joe to unlock the door, but ________ didn’t respond.

(he, she, they)
Practice items for this subtest were delivered both outside of the testing situation as part of the student practice guide and are also embedded in a short RISE tutorial just prior to the presentation of the subtest.

**Subtest 5: Efficiency of Basic Reading Comprehension**

Skilled reading is rapid, efficient, and fluent (silent or aloud). In recent research, a silent reading assessment task design—known as the *maze technique*—has gained empirical support as an indicator of basic reading efficiency and comprehension (Fuchs & Fuchs, 1992; Shin, Deno, & Espin, 2000). The design uses a forced-choice cloze paradigm—that is, in each sentence within a passage, one of the words has been replaced with three choices, only one of which makes sense in the sentence. The incorrect choices may be grammatically or semantically wrong, but in either case, they are designed to be obviously wrong to a reader with basic comprehension skills.

The RISE Efficiency of Basic Reading Comprehension subtest is composed of three expository passages that are modeled on middle-school science and social studies curricular materials. Students have 3 minutes to complete each passage. See below for an excerpt of a social studies passage:

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

Practice items for this subtest were delivered both outside of the testing situation as part of the student practice guide and are also embedded in a short RISE tutorial just prior to the presentation of the subtest.

**Subtest 6: Reading Comprehension**

When a reader engages with a text, multiple processes occur that facilitate the creation of a mental representation of what is being read. A reader decodes words, accesses meanings from the mental lexicon, parses syntactic structures, activates background knowledge (if any), and makes inferences across sentences and ideas. In Kintsch’s Construction Integration model
(Kintsch, 1998), these activities yield three levels of understanding: the surface level (a verbatim understanding of the words and phrases), the textbase (the gist understanding of what is being read), and the situation model (McNamara & Kintsch, 1996), the deepest level of understanding.

In the RISE Reading Comprehension subtest, the task focuses on the first two levels of understanding. The questions are designed to measure the extent to which students can locate information within the text successfully, understand paraphrases of the information, and make low-level inferences across sentences. The passages used in the subtest are the three the student saw previously in Subtest 5 (Efficiency of Basic Reading Comprehension). The passage remains available for the student (on the screen), while he or she answers six to eight multiple choice questions. An excerpt from the Permanent Housing passage and two related questions are presented below:

To build their houses, the people of this Age often stacked mud bricks together to make rectangular or round buildings. At first, these houses had one big room. Gradually, they changed to include several rooms that could be used for different purposes. People dug pits for cooking inside the houses, and they may have filled the pits with water and dropped in hot stones to boil it. You can think of these as the first kitchens.

The emergence of permanent shelters had a dramatic effect on humans. They gave people more protection from the weather and from wild animals. Along with the crops that provided more food than hunting and gathering, permanent housing allowed people to live together in larger communities.

Example Question 1 (Locate/Paraphrase): What did people use to heat water in Neolithic houses? (hot rocks, burning sticks, the sun, mud)

Example Question 2 (Low-level Inference): In the sentence “They gave people more protection from the weather and from wild animals,” the word "they" refers to: (permanent shelters, caves, herds, agriculture)

Practice items for this subtest were delivered outside of the testing situation as part of the student practice guide and are also embedded in a short RISE tutorial just prior to the presentation of the subtest.
Sample Characteristics and Technical Data

Sample Characteristics

In this section, we describe the characteristics of a large-scale pilot conducted in three school districts (two urban, one suburban) in the Northeastern United States in fall 2009. For this pilot, the RISE was administered in the computer labs of the participating schools, with supervision from trained school personnel. Results were transferred to a secure server and then cleaned and analyzed by ETS personnel.

**Number of schools and grades.** The sample for the large-scale pilot project was 4,104 sixth through eighth grade students from 11 schools.

**Participant characteristics.** See Tables 2 and 3 for participant characteristics.

### Table 2
**Participant Characteristics: By Grade and Gender**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total students</th>
<th>% female</th>
<th>% male</th>
<th>% not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1,554</td>
<td>44.0</td>
<td>51.0</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>1,236</td>
<td>46.8</td>
<td>48.5</td>
<td>4.7</td>
</tr>
<tr>
<td>8</td>
<td>1,314</td>
<td>46.2</td>
<td>48.7</td>
<td>5.1</td>
</tr>
</tbody>
</table>

### Table 3
**Participant Characteristics: By Grade and Race/Ethnicity**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total students</th>
<th>% Asian</th>
<th>% Black/African-American</th>
<th>% Hispanic/Latino</th>
<th>% White</th>
<th>% other/not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1,554</td>
<td>2.1</td>
<td>30.1</td>
<td>9.1</td>
<td>21.9</td>
<td>36.8</td>
</tr>
<tr>
<td>7</td>
<td>1,236</td>
<td>1.5</td>
<td>41.5</td>
<td>10.9</td>
<td>27.1</td>
<td>19.0</td>
</tr>
<tr>
<td>8</td>
<td>1,314</td>
<td>2.1</td>
<td>39.7</td>
<td>14.0</td>
<td>25.0</td>
<td>19.2</td>
</tr>
</tbody>
</table>

**Reliability**

Table 4 presents reliability estimates for each RISE subtest for each grade separately. Coefficient alpha (Cronbach, 1951) was used to estimate the reliability. Please note that these reliability statistics have been consistent across all pilot studies we have conducted.
Table 4

Coefficient Alpha Estimates for RISE Subtests

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>Number of items</th>
<th>Raw score range</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding</td>
<td>50</td>
<td>0–50</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>38</td>
<td>0–38</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>32</td>
<td>0–32</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Sentence Processing</td>
<td>26</td>
<td>0–26</td>
<td>0.81</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>Efficiency of Basic Reading</td>
<td>36</td>
<td>0–36</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Comprehension</td>
<td>22</td>
<td>0–22</td>
<td>0.76</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Subtest Means

Table 5 presents raw score means, standard deviations, and standard errors of measurement for each RISE subtest for each grade. The purpose of this information is to demonstrate how the relative difficulty and variability of the subtest distributions vary across grades.

Table 6 presents raw score means for each RISE subtest by grade and state test proficiency category. The purpose of this information is to demonstrate how the relative difficulty and variability of the subtest distributions vary across the proficiency categories and grades. It is evident from the table that there is more within-grade variability in means between proficiency levels than there is between grades in a comparable proficiency group.
<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>Grade 6</th>
<th></th>
<th></th>
<th></th>
<th>Grade 7</th>
<th></th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>SEM</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>SEM</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>SEM</td>
</tr>
<tr>
<td>Word Recognition &amp; Decoding (50 items)</td>
<td>1,554</td>
<td>36.9</td>
<td>9.1</td>
<td>2.8</td>
<td>1,236</td>
<td>37.8</td>
<td>8.9</td>
<td>2.7</td>
<td>1,314</td>
<td>39.2</td>
<td>8.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Vocabulary (38 items)</td>
<td>1,554</td>
<td>25.2</td>
<td>6.7</td>
<td>2.5</td>
<td>1,236</td>
<td>26.6</td>
<td>6.8</td>
<td>2.5</td>
<td>1,314</td>
<td>28.5</td>
<td>6.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Morphological Awareness (32 items)</td>
<td>1,553</td>
<td>23.2</td>
<td>6.7</td>
<td>2.1</td>
<td>1,236</td>
<td>24.0</td>
<td>6.8</td>
<td>2.1</td>
<td>1,314</td>
<td>25.6</td>
<td>6.3</td>
<td>1.9</td>
</tr>
<tr>
<td>Sentence Processing (26 items)</td>
<td>1,547</td>
<td>19.9</td>
<td>4.3</td>
<td>1.9</td>
<td>1,235</td>
<td>20.0</td>
<td>4.3</td>
<td>1.9</td>
<td>1,313</td>
<td>20.9</td>
<td>4.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Efficiency of Basic Reading Comprehension (36 items)</td>
<td>1,513</td>
<td>27.4</td>
<td>7.1</td>
<td>2.2</td>
<td>1,223</td>
<td>28.0</td>
<td>7.0</td>
<td>2.1</td>
<td>1,311</td>
<td>29.9</td>
<td>6.3</td>
<td>1.9</td>
</tr>
<tr>
<td>Reading Comprehension (22 items)</td>
<td>1,473</td>
<td>10.7</td>
<td>4.4</td>
<td>2.1</td>
<td>1,213</td>
<td>11.4</td>
<td>4.5</td>
<td>2.1</td>
<td>1,308</td>
<td>12.3</td>
<td>4.5</td>
<td>2.1</td>
</tr>
</tbody>
</table>
Table 6
RISE Subtest Raw Score Means and Standard Deviations by Grade and State Test Proficiency Category

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>Grade 6</th>
<th></th>
<th></th>
<th>Grade 7</th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding (50 items)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.5 (5.0)</td>
<td>22.5 (5.1)</td>
<td>16.9 (4.5)</td>
<td>30.8 (4.4)</td>
<td>24.6 (5.3)</td>
<td>18.2 (5.9)</td>
<td>32.2 (4.4)</td>
<td>25.7 (5.6)</td>
<td>18.6 (6.4)</td>
</tr>
<tr>
<td>Vocabulary (38 items)</td>
<td>27.5 (4.2)</td>
<td>20.7 (5.8)</td>
<td>14.2 (4.7)</td>
<td>27.9 (3.8)</td>
<td>22.4 (5.6)</td>
<td>18.6 (4.9)</td>
<td>29.0 (3.0)</td>
<td>23.5 (5.6)</td>
<td>16.5 (6.4)</td>
</tr>
<tr>
<td>Morphological Awareness (32 items)</td>
<td>22.2 (2.8)</td>
<td>18.9 (3.7)</td>
<td>14.8 (4.7)</td>
<td>22.1 (2.8)</td>
<td>19.4 (3.5)</td>
<td>15.2 (4.4)</td>
<td>22.7 (2.3)</td>
<td>20.0 (3.6)</td>
<td>16.2 (4.2)</td>
</tr>
<tr>
<td>Sentence Processing (26 items)</td>
<td>31.7 (4.0)</td>
<td>25.2 (6.4)</td>
<td>17.8 (6.0)</td>
<td>31.9 (3.8)</td>
<td>26.8 (6.2)</td>
<td>19.3 (6.8)</td>
<td>32.9 (3.2)</td>
<td>28.3 (5.8)</td>
<td>21.2 (6.8)</td>
</tr>
<tr>
<td>Efficiency of Basic Reading Comprehension (36 items)</td>
<td>13.6 (3.8)</td>
<td>8.8 (3.2)</td>
<td>6.3 (2.1)</td>
<td>14.1 (3.7)</td>
<td>9.8 (3.6)</td>
<td>6.5 (2.5)</td>
<td>14.7 (3.5)</td>
<td>10.2 (3.7)</td>
<td>7.1 (2.4)</td>
</tr>
<tr>
<td>Reading Comprehension (22 items)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. A/P = Advanced/Proficient; NI = Needs improvement; W = Warning. Standard deviations are presented next to means in parentheses.
Correlations

Previous research has suggested that we might expect moderate to strong correlations among component reading subskills such as those measured in the RISE (e.g., McGrew & Woodcock, 2001; Vellutino et al., 2007). Tables 7 to 9 show the correlations between RISE subtest raw scores. These correlations range from .581 to .831. The strongest relationship for each grade level is between vocabulary and morphological awareness, which makes sense, as both subtests measure different aspects of lexical language knowledge. The weakest relationship for each grade level is between sentence processing and reading comprehension, followed by word recognition and decoding. Both correlations are of about the same moderate magnitude in relationship to reading comprehension. In general, the pattern of results is consistent with previous research on component reading skills using published and experimental measures.

While it is encouraging that the tests show moderate correlations with each other, it would not be ideal if they were so highly correlated as to be indistinguishable psychometrically. Disattenuated correlations estimate the correlation between measures after accounting for measurement error. The correlations after correcting for attenuation range from .712 to .928, with a median of .808. There are only two above 0.90—the correlations between Vocabulary and Morphological Awareness in Grades 6 and 8. With these two exceptions, this range of correlations indicates the subtests measured related, but distinguishable, skills.

Table 7

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.745**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>.768**</td>
<td>.804**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Processing</td>
<td>.655**</td>
<td>.643**</td>
<td>.734**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency of Basic Reading</td>
<td>.709**</td>
<td>.712**</td>
<td>.785**</td>
<td>.744**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.597**</td>
<td>.670**</td>
<td>.653**</td>
<td>.581**</td>
<td>.672**</td>
<td>1</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
Table 8

*Grade 7: Pearson Correlations of RISE Raw Subtest Scores*

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.764**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>.772**</td>
<td>.795**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Processing</td>
<td>.649**</td>
<td>.649**</td>
<td>.723**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency of Basic Reading</td>
<td>.730**</td>
<td>.722**</td>
<td>.792**</td>
<td>.743**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.597**</td>
<td>.664**</td>
<td>.665**</td>
<td>.583**</td>
<td>.671**</td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

Table 9

*Grade 8: Pearson Correlations of RISE Raw Subtest Scores*

<table>
<thead>
<tr>
<th>RISE subtest</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recognition &amp; Decoding</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.778**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>.814**</td>
<td>.831**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Processing</td>
<td>.672**</td>
<td>.663**</td>
<td>.743**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency of Basic Reading</td>
<td>.734**</td>
<td>.714**</td>
<td>.787**</td>
<td>.763**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.619**</td>
<td>.677**</td>
<td>.662**</td>
<td>.605**</td>
<td>.658**</td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
Summary, Conclusions, and Next Steps

The RISE form of the SARA was designed to address a practical educational need by applying a theory-based approach to assessment development. The need was for better assessment information of struggling middle grades students—those students who typically score below proficient on state English language arts tests. The theoretical and empirical literature suggested that overall reading comprehension skills are composed of componential reading skills such as decoding, word recognition, vocabulary, and sentence processing. Weaknesses in one or more of these skills could underlay poor reading comprehension performance. Such componential score information is not derivable from traditional reading comprehension tests. We designed subtests targeting six components.

Further design considerations were imposed to meet practicality and feasibility constraints. Specifically, the need for efficient administration (e.g., a 45-minute limit) and rapid, inexpensive turnaround of scores argued for electronic delivery and scoring.

The report describes adequate reliability and other psychometric properties for each of the subtests in the RISE form for each of the grade levels. The sample included students in the middle grades (sixth through eighth) in three school districts comprising a mixture of racial/ethnic groups. Evidence of validity of scores includes strong, but not statistically indistinguishable, intercorrelations among the subtests (see also Mislevy & Sabatini, 2012; O’Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012; Sabatini, 2009; Sabatini, Bruce, & Pillarisetti, 2010; Sabatini, Bruce, Pillarisetti, & McCormick, 2010; Sabatini, Bruce, & Sinharay, 2009). The subtest means demonstrate how the relative difficulty and variability of the subtest distributions vary across state test proficiency categories and grades.

The adequacy of the measurement properties of the RISE provide the basis for school administrators and teachers to interpret test scores as part of the evidence available for making educational decisions. For example, school administrators might use prevalence estimates of how many students are scoring at low levels on subtests of decoding/word recognition to determine how to plan and allocate resources for interventions targeting those basic subskill deficiencies (which are usually implemented as supplements to subject-area classes). Classroom teachers can look at evidence of relative strengths and weaknesses across a range of their students to make adjustments to their instructional emphasis in teaching vocabulary, morphological patterns, or assigning reading practice to enhance reading fluency and efficiency. We have been working
with pilot schools to develop professional development packages to assist teachers in using RISE score evidence in making sound decisions aligned with their instructional knowledge and practices (see Mislevy & Sabatini, 2012; O’Reilly et al., 2012; and Sabatini, 2009, for other applications).

Our next steps, now underway, are elaborating on the SARA system in several directions. First, we are designing parallel forms and collecting pilot data in the middle grades. These forms are designed to expand the use of the tests for benchmarking and summative purposes and for tracking student progress within and across school years. Second, we are building and piloting forms for use in elementary, secondary, and adult literacy settings. Third, we are evaluating the properties of the tests with special populations such as English language learners. Fourth, we are expanding and elaborating on the item types within each of the componential constructs. Fifth, we are exploring ways to develop scale scores for each RISE subtest using item response theory (IRT) so that proficiency levels for each subtest can be developed. Sixth, we are expanding our research on providing interpretative guidance for using results to inform decision-making at the teacher and school levels, for which the development of proficiency levels and profiles will be useful. Finally, we are working on versions of the SARA that can be used in more formative contexts for students and teachers.

In conclusion, the RISE form of the SARA fills an important gap in assessment of reading difficulties in the middle grades. The RISE form is a proof of concept that theory-based instruments can be designed to be practically implemented, scored, and interpreted in middle grades contexts. We are hopeful that the information provided by the RISE is of practical utility to educators above and beyond scores obtained on state exams and traditional reading comprehension tests. Our ongoing research agenda is designed to collect evidence to enhance and improve the SARA system utility and validity in a wide range of contexts.
References


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. Albro, & T. O’Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 119–134). Lanham, MD: Rowman & Littlefield.


