Examining the Measures of and the Relationship between
Vocabulary Knowledge, Oral Reading Fluency, and Reading Comprehension

Doctoral Candidate

University of Oregon
Abstract

Educators disagree on the definition of vocabulary knowledge, the importance of oral reading fluency, and their corresponding roles in reading comprehension. This lack of consensus in construct definition results in the use of multiple assessment measures and difficulty in interpreting results. Authentic assessment is based on instruments that reliably and validly measure the defined construct. This thesis will provide a framework to define vocabulary knowledge, oral reading fluency, and reading comprehension, and the subsequent correlational study will examine the relationships between these constructs in 3rd grade students and explore the relationship between an appropriate vocabulary assessment, oral reading fluency, and reading comprehension.
Examining the Measures of and the Relationship between Vocabulary Knowledge, Oral Reading Fluency, and Reading Comprehension

The National Reading Panel (2000) and the Institute for the Development of Educational Achievement (IDEA) (Kame'enui et al., 2002) identified five dimensions or big ideas of reading in an alphabetic writing system: (a) phonemic awareness, (b) phonics, (c) vocabulary knowledge, (d) reading fluency, and (e) reading comprehension (p. 2). Of these dimensions, vocabulary knowledge is linked with both oral reading fluency and reading comprehension (Baker, 1995; Nagy, 1988; Nelson-Herber, 1986) Additionally, vocabulary knowledge contributes significantly to achievement in the subjects of the school curriculum, as well as in formal and informal speaking and writing (Smith & ERIC Clearinghouse on Reading English and Communication., 1997). Word and concept knowledge are incorporated as essential aspects in models of cognition, intelligence, and verbal reasoning (Wiig & Secord, 1992). This knowledge is considered essential for academic achievement and communication (p. 2). Together, the elements of vocabulary knowledge, oral reading fluency, and reading comprehension are the building blocks of reading (p. 2).

This thesis will examine the relationship between vocabulary knowledge, oral reading fluency, and reading comprehension in 3rd grade students. The following questions will explore the problem:

1. How is vocabulary knowledge defined and what are the most appropriate vocabulary knowledge assessment measures?

2. How is oral reading fluency defined and what are the most appropriate oral reading fluency assessment measures?
3. How is reading comprehension defined and what are the most appropriate reading comprehension assessment measures?

4. What is the relationship between vocabulary knowledge, oral reading fluency, and reading comprehension as defined and assessed by questions one, two, and three?

LITERATURE REVIEW

Educators disagree on the definition of vocabulary knowledge, the importance of oral reading fluency, and their role in reading comprehension (Anderson & Freebody, 1985; Baker, 1995; Irvin, 2001; Jongsma, 2000; Mckeown et al., 1983; Qian, 1999a, 2002; Schum, 1992; Stahl et al., 1991). The National Reading Panel (Panel, 2000b) states, “Reading comprehension is a cognitive process that integrates complex skills and cannot be understood without examining the critical role of vocabulary learning and instruction in its development” (p. 4-1). The panel adds that reading vocabulary is crucial to the comprehension process of a skilled reader (p. 4-3).

Related research and current educational practice suggests a correlation between students’ vocabulary, that is, words they know, and students’ comprehension of what they read (Gersten & Geva, 2003). While there is adequate research to support this conclusion, most research centers more on effective instruction than on appropriate measurement. For example, Gersten & Geva conclude that vocabulary knowledge is a key to reading fluency and accuracy, based on their two-year study of 34 first grade classrooms (p. 11). The authors assessed students on letter naming and phonemic awareness at the beginning of the year, and comprehension, fluency, and accuracy at the end of the year. The study identified reading growth but the authors based their conclusions more on the level and quality of vocabulary instruction, rather than the quality or appropriateness of the vocabulary assessment (p. 7).
Similarly, Stahl (2003) says that the relationship between vocabulary and reading comprehension is a “robust” one and that vocabulary knowledge has consistently been the “foremost predictor of a text’s difficulty” (p. 241). Stahl adds that vocabulary knowledge and reading comprehension are strongly correlated, based on measurement of word difficulty and sentence difficulty (pp. 241-242). Stahl acknowledges that measuring vocabulary and comprehension are “problematic” and concludes that high-quality vocabulary instruction aids reading comprehension (p. 1). Stahl, just as Gersten and Geva focuses more on the importance of instruction and less on the appropriateness of assessment.

Mezynski (1983) discusses the importance of instruction and the appropriateness of assessment by examining eight studies that focused on the relationship between vocabulary knowledge and reading comprehension. Although all eight studies reported increases in students’ vocabulary knowledge through direct and indirect vocabulary instruction, half of the studies reported no gains in reading and half reported mild to moderate gains in reading due to direct vocabulary instruction (p. 259). Mezynski notes that the variation in results is not surprising, given the differences in instructional design and measurement found across the studies (p. 258). Mezynski cites a study by Beck et al. (1980) in which the assessment measures included three components of semantic processing: (a) accuracy of word knowledge, (b) fluency of lexical access, and (c) richness of semantic networks (p. 11). Accuracy of word knowledge was assessed through a multiple-choice test, while a semantic decision task measured the speed of lexical access. Comprehension was measured through a probed or guided recall and written retell of stories. (p. 13) The authors concluded that students who received vocabulary instruction scored significantly better on the test of word knowledge and significantly faster and more accurately on the response-timed tasks. However, the results of the story comprehension task were equivocal,
that is, there was no significant difference between students who received direct vocabulary instruction and those who did not (p. 13). Consequently, the authors concluded that evidence of a causal connection between vocabulary knowledge and reading comprehension was weak, despite strong correlative evidence (p. 253).

McKeown et al. (1983) recreated Beck et al.’s (1980) experimental-control study in an effort to further explore and explain the lack of improvement in comprehension. McKeown et al. state that problems with the stories and the comprehension measure itself suppressed the relationship between instruction and comprehension and, consequently, any significant increase in reading comprehension. Beck et al. specifically identified three problems in the methodology of the story comprehension task. First, some stories had a more complex plot structure than others did. Second, the story plots seemed “over-contrived” due to their construction around a large set of taught words. Third, the use of probed or guided recall and retell as an assessment may have forced children to use the probed structure rather than their own structure to generate recall and retell (p. 5). Probed or guided recall allows students to demonstrate comprehension using story maps or comprehension questions, unlike free recall where students demonstrate comprehension through an unprompted, unaided retelling of the passage or story.

In the replication of Beck et al.’s (1980) study, McKeown et al. (1983) (a) revised the stories to ensure similar plot structures and shortened the number of words taught; (b) replaced probed recall with free recall and retell; and (c) added a set of 25 multiple-choice questions to further assess comprehension (p. 6). The authors studied fourth-grade students from two low SES, neighborhood urban schools, with one class from each school selected as the experimental group and the three remaining classes designated as the control group. Both control and experimental group students in this study could be matched pair wise within three points of their
combined scores on the Reading and Vocabulary subtests of the Iowa Test of Basic Skills (ITBS), which were used as pretests (p. 6). The results addressed two issues: first, whether vocabulary instruction was successful in producing accurate knowledge of the instructed words; and second, whether text comprehension was affected. With regard to accuracy of word knowledge, the authors concluded that the instruction was “successful in enhancing the accuracy of the knowledge of the instructed words” since the mean of correct responses for the experimental group posttest (.80) was higher than that of the experimental group (.32) (p. 10). Similarly, on the text comprehension posttest, the mean correct recall and multiple choice responses for the experimental group (.33 and .66 respectively) was higher than that of the control group (.16 and .42 respectively) (p. 13). Consequently, the authors concluded that since instruction was successful in enhancing the accuracy of the knowledge and the comprehension and recall of passages, the replication of Beck et al.’s study was successful (p. 10).

Qian (1999a; 1999b; 2002) further examined this relationship, choosing to assess the “breadth and depth” of vocabulary knowledge in reading comprehension through his study of 80 ESL university students with a vocabulary size of 3000 words or better (Qian, 1999a). Vocabulary size was measured using the Vocabulary Levels Test (Nation, 1990) which required students to scan groups of six words and three definitions, choose 3 of the words, and match them with their correct definitions. Reading comprehension was assessed using the Test of English as a Foreign Language (TOEFL) (Read, 2000; Service, 1987), and required students to read 4 passages and answer 20 multiple choice comprehension questions. Depth of vocabulary was measured using the Word Associates Format (Read, 2000) which assessed students’ ability to find relationships between a targeted word and word groups. Qian concluded, “Scores on vocabulary size, depth of vocabulary knowledge, and reading comprehension are highly, and
positively correlated; and scores on depth of vocabulary knowledge can make a unique contribution to the prediction of reading comprehension levels” (p. 280). Qian’s conclusions support the need for an appropriate vocabulary assessment that measures more than word recognition.

**Opposing Views**

While *Teaching Children to Read* from the National Reading Panel (2000) has been referenced in this thesis, this document has its critics. Although much of the criticism stems from the report and conclusions on phonics instruction (Camilli & Wolfe, 2004; Coles, 2003; Garan, 2001; Krashen, 2004), Garan (2001) also cites problems with the report’s methodology, data collection, and data reporting. Garan reports that the panel published two different versions of the report, without distinguishing between the editions, and that the original, did not include a complete table of contents or the minority report from the panel (p. 505). Garan further reports that the panel incorrectly listed the studies used in the meta-analysis, making it difficult to “check the veracity of any claim” (p. 506). Additionally, the panel conducted a meta-analysis of only thirty-eight studies, invalidating its claim of “comprehensive, scientific review of reading research on phonics” and “making the reliability of its conclusions questionable from the outset” (p. 507). In conclusion, Garan says, (a) the research base was inappropriate for a meta-analysis, (b) the findings were not generalizable, based on the studies' student populations and sampling sizes, (c) the results were not reliable, and (d) the NRP's research was not valid and did not accurately represented the data in its conclusions (pp. 521-523).

Gersten and Baker (1999) presented a research synthesis of studies conducted between 1979 and 1999 involving learning disabled students, concluding that “Successful reading comprehension is correlated with oral reading fluency and vocabulary knowledge. However,
interventions that focus on improving fluency or vocabulary do not necessarily increase reading comprehension, especially of long passages” (p. 1). Further examination of this statement is offered later in this paper, in the oral reading fluency section.

The above mentioned criticisms notwithstanding, the panel’s reports and conclusions regarding the elements of reading, and the importance of fluency and vocabulary knowledge in comprehension assessment mirror those of other researchers and educators. Further, the importance of vocabulary knowledge in assessing comprehension is a consistent theme throughout the literature.

**Vocabulary Knowledge**

While controversy surrounding vocabulary and its relationship to comprehension continues, educators agree that vocabulary is an important component of reading (Anderson & Freebody, 1985; Irvin, 2001; Jongsma, 2000; Senechal & LeFevre, 2002; Stahl, 2003). There is consensus on the importance of vocabulary knowledge for children and concurrence that this knowledge includes several dimensions of word knowledge (Stahl, 2003). However, researchers often interchange the terms vocabulary knowledge and word knowledge when describing the same construct. In fact, throughout the literature and within individual works, the terms vocabulary knowledge and word knowledge are used interchangeably, leading to possible confusion when defining the construct and examining assessment measures. This thesis acknowledges that vocabulary knowledge and word knowledge are synonymous terms and seeks to clearly define the construct.

Formerly, the measure of word knowledge was the ability to define the word (Graves et al., 2001). Graves et al. add that children learn meanings of words partially, then progress to more refined understandings (p. 81). Graves et al. outlined six stages of vocabulary knowledge
that include learning to read a known word, learning new meanings of known words, learning new words that represent known concepts, clarifying and enriching meanings of known words, and moving words from receptive (listening and reading) to expressive (speaking and writing) vocabulary (p. 81).

Baker et al. (1995) underscore the importance of vocabulary knowledge stating, “Although vocabulary development pervades every subject from reading to mathematics to physical education, it is difficult to isolate for instructional purposes” (p. 2). Further, successful vocabulary knowledge measures: (a) increase word learning, (b) reduce the gap between a poor and rich vocabulary, and (c) expand the depth of knowledge of different contexts in which words are used (Baker et al., 1995; Qian, 1999a).

Wiig & Secord (1992) state, “Word and concept knowledge are essential aspects in models of cognition, intelligence, and verbal reasoning. Word and concept knowledge is essential for academic achievement and that level of word knowledge has been identified as the best predictor of reading comprehension” (p. 2). Other researchers, (Baumann & Kameemui, 1991; Beck et al., 2002; Gersten & Geva, 2003) have agreed on the terms ‘minimal, partial, and full concept knowledge’ to categorize and describe “breadth and depth” (Qian, 1999a) of acquired vocabulary. Using these criteria, students with a minimum level of vocabulary concept knowledge link new words with a specific definition or single context while students with full concept knowledge understand and use words in different contexts, know the varied meanings of multiple-meaning words, and can ascertain a word’s meaning from the meanings of similar words (p. 8). Students with partial concept knowledge fall somewhere in between.

Students learn vocabulary more effectively, i.e. full concept knowledge, when they are directly involved in constructing meaning rather than in memorizing definitions or synonyms.
Vocabulary, Fluency, and Comprehension (Smith & ERIC Clearinghouse on Reading English and Communication., 1997). Effective vocabulary knowledge “includes a component of both oral and written language that encompass the body of words students must know if they are to read increasingly demanding text with fluency and comprehension” (Kame'enui et al., 2002).

Nagy & Scott (2000) write, "A person who knows a word can recognize it, and use it, in novel contexts, and uses knowledge of the word, in combination with other types of knowledge, to construct meaning for a text. Knowing a word means knowing how to do things with it, not write its definition!” (p. 3). Further, good readers know several meanings of many words, thus leading to increased understanding or comprehension of these words, both in and out of context (Brabbham & Lynch-Brown, 2002; Paul & O'Rourke, 1988; Qian, 1999a). Read (2000), however, writes, “Readers do not have to understand every word in order to extract meaning from a text satisfactorily. Some words can be ignored, while the meaning of others can be guessed by using contextual clues, background knowledge of the subject matter, and so on.” (p. 4). Read suggests that prior knowledge and context can compensate for lack of knowledge or contextually irrelevant knowledge when reading text. In fact, there would be many who would support the importance of prior knowledge in comprehension (Adams, 1990; Alderson, 2000; Anderson & Freebody, 1979; Baker et al., 1995; Beck et al., 2002; Hoyt, 1999; Irvin & National Education Association of the United States., 1990; Johnson & Pearson, 1978; Jongsma, 2000; Read, 2000; Schum, 1992; Stahl, 2003). Similarly, Alderson states, “Background knowledge should be recognized as influencing all comprehension (p. 121). If prior knowledge does affect comprehension then appropriate vocabulary and comprehension measurement tools are crucial to authentic assessment.
Vocabulary, Fluency, and Comprehension 12

**Vocabulary Assessment**

“The measurement of vocabulary is fraught with difficulties” (Panel, 2000b), partly because researchers distinguish between different vocabularies, such as receptive or expressive and oral or reading, etc. (p. 4-15). Read (2000) examined the history of vocabulary assessment and says the early assessments were initially objective in nature with material divided into small units, assessed by a single correct answer (p. 75). Vocabulary assessments then incorporated multiple-choice items that could discriminate among learners according to their level of ability (p. 76). Read lists 6 limitations of multiple-choice vocabulary tests:

1. Difficulty in test construction, requiring labor-intensive field-testing and refinement.
2. The learner may know another meaning for the word, but not the one sought.
3. The learner may choose the correct word by process of elimination.
4. Items may test students’ knowledge of distracters rather than their ability to identify the exact meaning of the target word.
5. The learner may miss an item for lack of knowledge of words or lack of understanding of syntax in the distracters.
6. The format permits only a very limited sampling of the learner’s total vocabulary.

(PP. 77-78)

Read (2000) offers three components of exemplary vocabulary assessment measures. Firstly, discrete constructs are measured, e.g., a vocabulary test measuring some aspect of the learners’ knowledge of target language words, or assessing vocabulary knowledge rather than grammatical knowledge or reading comprehension ability. Secondly, selective constructs are measured, i.e., target words are selected and incorporated into test items. Thirdly, the assessment presents context-independent vocabulary, i.e., words in isolation which require students to select
meanings for the words without reference to any linguistic context (p. 10). Read cites The Vocabulary Levels Test (Nation, 1990) as his choice for the test that best meets the above-mentioned criteria and because the tool: (a) is easy to make and easy to mark, (b) provides a low chance of “guessing” correct answers, and (c) tests a large number of words in a short amount of time (p. 261). The Vocabulary Levels Test while popular is one of many widely-used vocabulary assessment tools, measuring different aspects of vocabulary knowledge. As stated before, these multiple measures lead to difficulty in defining the construct and interpreting assessment results (table 1).

Table 1

<table>
<thead>
<tr>
<th>Definition/Measurement</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can read words correctly</td>
<td>Fuchs et al., 2001</td>
</tr>
<tr>
<td>Can use words in novel contexts</td>
<td>Nagy &amp; Scott, 2000</td>
</tr>
<tr>
<td>Learning to read a known word, define a word, learning new meanings of known words, learning new words that represent known concepts, clarifying and enriching meanings of known words, and moving words from receptive (listening and reading) to expressive (speaking and writing) vocabulary.</td>
<td>Graves et al., 2001</td>
</tr>
<tr>
<td>Accuracy of word knowledge, fluency of lexical access, and richness of semantic networks.</td>
<td>Beck et al.</td>
</tr>
<tr>
<td>Linking new words with a specific definition or single context or understand and use words in different contexts; know the varied meanings of multiple-meaning words, and can ascertain a word’s meaning from the meanings of similar words.</td>
<td>Qian, 2002</td>
</tr>
<tr>
<td>Successful measures: increase word learning, reduce the gap between a poor and rich vocabulary, and expand the depth of knowledge of different contexts in which words are used.</td>
<td>Baker, 1995; Qian, 1999</td>
</tr>
</tbody>
</table>
Oral Reading Fluency

In 1992, the National Assessment of Educational Progress (NAEP) conducted a study of the status of fluency achievement in American education (Pinnell, 1995). The study provided a national database for the development of oral reading abilities of fourth grade students and how these abilities relate to overall reading achievement (p. 9). The study also examined oral reading fluency, confirming what teachers already knew: (a) oral reading fluency is one of several descriptors of good reading; and (b) proficient readers not only recognize and read words quickly, but read with a sense of ease and fluidity (p. 13). The researchers also found a “close relationship” between oral reading fluency and reading comprehension (p. 13). However, they acknowledge that, “Despite renewed interest in oral reading fluency and its association with reading development, there is no widely accepted definition of reading fluency” (p. 13). The study offers multiple definitions of oral reading fluency that include:

1.) The ability to recognize words rapidly and accurately;

2.) A demonstrated level of automaticity in word recognition;

3.) The level of expressiveness in students’ oral reading and readers’ use of appropriate phrasing

4.) Focus on meaning construction rather than attending to words on a page (p. 13).

The Institute for the Development of Educational Achievement (IDEA) in the College of Education at the University of Oregon examined and evaluated reading assessment measures and defined fluency as part of their analysis of reading assessment instruments (Kame'enui et al., 2002). IDEA, gathering data from the Reading First subpart of the 2002 No Child Left Behind (NCLB) legislation, defines fluency as “Reading accurately, quickly and with expression” (p. 19). However, fluency is further defined in terms of automaticity (a quick and accurate level of
recognition that occurs with little conscious attention, such as the ability to quickly and accurately associate sounds with letters in order to read words (p. 19).

Fuchs et al. (2001) define oral reading fluency as, “The oral translation of text with speed and accuracy” and hypothesize that oral reading fluency may serve as an indicator of overall reading competence (p. 239). Fuchs et al. studied seventy middle school and junior high school students, each with a reading disability. The Stanford-Binet achievement test was administered, as were four alternative measures (question answering, recall, cloze, oral reading fluency). Criterion validity coefficients for the question answering, the recall, and the cloze measures were .82, .70, and .72, respectively, while coefficient for oral reading fluency was .91 (p. 246). Fuchs et al. concluded that tests for differences between these correlations demonstrated that the correlation for oral reading fluency was significantly higher than the correlation for each of the three direct measures of reading comprehension (p. 246). These study results support the importance of oral reading fluency as an integral component of reading, and despite the variations in definition and measurement, research into oral reading fluency has centered around the ease, rapidity, and accuracy of performance (Pinnell, 1995).

The National Reading Panel (Panel, 2000a) also examined oral reading fluency as a component of literacy. The purpose of this report was, “To review the changing concepts of fluency as an essential aspect of reading and to consider the effectiveness of two major instructional approaches to fluency development and the readiness of these approaches for wide use by the schools” (p. 3-1). The panel concluded that, “Fluent readers can read text with speed, accuracy, and proper expression, and fluency depends upon well developed word recognition skills” (Panel, 2000a). They add, “However, these skills do not inevitably lead to fluency” (p. 3-1). Further examination of this statement reveals that the NRP identified other factors that lead to
fluency, including: (a) practice, or the amount of time spent reading; (b) instructional
interventions such as repeated reading or guided reading; and (c) quality of reading instruction
(pp. 3.5-3.21).

Shinn et al. (1992) define reading fluency in terms of “reading speed and accuracy,”
adding, “The competent (fluent) reader is able to decode automatically without the services of
attention and thus is able to attend to processing meaning” (p. 459). The authors suggest that for
some, reading fluency is synonymous with rapid decoding, acknowledging the increase in
attention to measurement of reading fluency and presenting the construct of reading as a
combination of “decoding and comprehension” (p. 459). The authors include examples of
assessment tools and sub tests within these tools to appropriately define and measure the
constructs of reading, fluency, decoding and comprehension (pp. 459-460). The study includes
research-based evidence that oral reading fluency is a “reliable and valid measure of a student’s
general reading skill, including reading comprehension” and Curriculum-Based Measurement
(CBM) as an effective and valid tool to measure this construct (p. 461).

Oral Reading Fluency Assessment

The National Reading Panel (2000) says, “The ability to obtain meaning from print
depends so strongly on the development of word recognition accuracy and reading fluency, both
the latter should be regularly assessed in the classroom” (p. 7). Further, using oral reading
fluency to develop long-term targets or goals within a dynamic target/goal setting process has
been demonstrated to promote greater student achievement (p. 7). Oral reading fluency can be
assessed in a variety of ways using many measures, for example, counting the number of correct
words while a student reads aloud from text for 1 min, also known as curriculum-based
measurement (CBM) (Shinn, 1989). Research shows that this method for collecting oral reading
fluency data produces a “broad dispersion of scores across individuals of the same age, with rank orderings that correspond well to important external criteria, and that represent an individual’s global level of reading competence” (Fuchs et al., 2001). Teachers can also use these scores to identify discrepancies in performance levels. In addition to generating quantitative scores, CBM can be used to gather qualitative descriptions of performance. As teachers count the number of words read correctly in 1 minute, they can note the types of decoding errors and decoding strategies students use or how miscues reflect students’ “reliance on graphic, semantic, or syntactic language features, or how self-corrections, pacing, and scanning reveal strategic reading processes” (p. 251). IDEA found CBM sufficient to assess outcomes in oral reading fluency in third grade students, as well as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Gray Oral Reading TEST IV (GORT-IV) (p. 98).

Reading Comprehension

Reading can be defined as, “a complex system of deriving meaning from print” (Adams, 1990; Kame'enui et al., 2002). Within this system are a series of identified skills associated with the process of reading and comprehension. The plethora of reading-related skills makes it difficult to discuss reading comprehension without defining reading as a construct. The New York City Board of Education identified thirty-six reading-related skills (Lunzer & Gardner, 1979). Subsequent researchers (Adams, 1990; Alderson, 2000; Anderson & Freebody, 1982; Munby, 1978; Staskowski & Creaghead, 2001; Taylor et al., 2000) narrowed the list to a more manageable set of 10 skills prevalent in good readers, including: recalling word meanings, drawing inferences, and following the structure of a passage.

Alderson (2000) says that reading ability is an “abstract notion” and that reading constructs come from a theory of reading. They are “realized through the texts we select, the
tasks we require readers to perform, the understandings they exhibit, and the inferences we make from those understandings” (p. 117). Consequently, Alderson defines reading through the assessment of a desired construct or constructs, or, as Alderson states, “Constructs of reading are based on a model of reading and the factors that affect reading insofar as these are relevant to the construct” (p. 120). Alderson further defines constructs as “Any variable that has an impact on either the reading process or its product, with regard to test design or validation “ (p. 120). For example, prior knowledge, cultural context, knowledge of language, synthesis and evaluation skills, and purpose are all constructs of reading (pp. 120-122). Alderson concludes that although there are many variables that affect reading and, subsequently, comprehension, the reader’s background knowledge should be recognized as influencing all comprehension (p. 121).

Reading Comprehension Assessment

The Institute for the Development of Educational Achievement (IDEA) in the College of Education at the University of Oregon reviewed and analyzed technical information materials for 29 reading assessment instruments designed for use in kindergarten through third grade. IDEA evaluated these instruments based on the “Reading First” legislation and language from the No Child Left Behind Act of 2001 (NCLB), which established criteria for reading programs and assessment of said programs (Education, 2003). Under NCLB, qualified Reading First programs contained the five essential components of reading: phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. Reading First also identifies three common components of effective reading assessment: screening, diagnostic, and classroom-based instructional reading assessments” (Kame'enui et al., 2002). Kame’enui adds that an effective screening assessment must be valid, reliable, and based on scientifically based reading research and adds, “The instrument must have adequate validity established by independent research as evidenced by any
reported criterion validity (either concurrent or predictive), construct and content validity data (p. 27). The screening assessment is brief and designed as a first step in identifying children who may be risk for delayed development or academic failure, and in need of further diagnosis or additional reading instruction (p. 23). Problems arise when educators and researchers use broad and varied terms to define reading proficiency and comprehension, leading to both disagreement and confusion when attempting to measure these constructs (table 2).

Table 2

Variations in Reading Comprehension Measurement

<table>
<thead>
<tr>
<th>Task</th>
<th>Description/Forms</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Read w/Oral Retell</td>
<td>Probed, Guided, Free</td>
<td>Brown &amp; Cambourne (1990)</td>
</tr>
<tr>
<td>Teacher Read w/ Written Retell</td>
<td>Probed, Guided, Free</td>
<td>Brown &amp; Cambourne (1990)</td>
</tr>
<tr>
<td>Taped Passage w/ Oral Retell</td>
<td>Probed, Guided, Free</td>
<td>Woodcock-Johnson</td>
</tr>
<tr>
<td>Taped Passage w/ Written Retell</td>
<td>Probed, Guided, Free</td>
<td>Woodcock-Johnson</td>
</tr>
<tr>
<td>Multiple Choice--Oral</td>
<td>Multiple test items, 3-5 choices within each test item</td>
<td>Miller (1995)</td>
</tr>
<tr>
<td>Multiple Choice--Written</td>
<td>Multiple test items, 3-5 choices within each test item</td>
<td>Alderson (2000), Miller (1995)</td>
</tr>
<tr>
<td>Miscue Analysis</td>
<td>Running record and/or comprehension checklist during oral retell</td>
<td>Miller (1995), Beaver (1993)</td>
</tr>
</tbody>
</table>

Standardized assessments such as the Iowa Test of Basic Skills (ITBS), the Woodcock-
Johnson Test of Basic Skills (WJIII), and the California Achievement Test CAT use a combination of the assessment measures shown in table 2 within the same test, moving further away from defining a single construct or comprehension skill. Miller (1995) says that any comprehension assessment must include assessment of four sub skills: explicit comprehension (responding to explicit or literal questions), implicit comprehension (inferring or interpreting), critical reading (using judgment or evaluation), and creative comprehension (combining prior knowledge with insights gained after reading) (pp. 75-76). Bachman and Palmer (1996) offer five additional components of reading comprehension assessment: (a) test setting, (b) test rubrics, (c) type of input test: oral, visual, timed, length, etc., (d) type of expected response: oral, visual, timed, length, etc., and (e) relationship between test input and response.

Kame’enui et al. (2002) speak to the importance of a diagnostic reading assessment, stating, “The diagnostic assessment is used to identify a child’s areas of strengths, determine difficulties in learning to read and the potential cause of such difficulties, and help determine possible reading intervention strategies and related special needs” (p. 23). Kame’enui et al. add that classroom-based instructional reading assessment evaluates children’s learning based on, “Systematic observations by teachers of children performing academic tasks that are part of their daily classroom experience, and is used to improve instruction in reading” (p. 23).

IDEA examined 29 reading assessment instruments and rated them either “sufficient” or “not sufficient” based on NCLB and Reading First criteria (Kame'enui et al., 2002). IDEA identified six of the 29 assessments that met Reading First criteria and were sufficient as diagnostic instruments to assess vocabulary and reading comprehension at third grade (table 3), nine that were sufficient to assess vocabulary outcomes at third grade, and 7 that were sufficient to assess comprehension outcomes at third grade (table 3).
Table 3

Diagnosing Vocabulary, Oral Reading Fluency and Reading Comprehension at 3rd Grade

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sufficient for Diagnosing Vocabulary</th>
<th>Sufficient for Diagnosing Oral Reading Fluency</th>
<th>Sufficient for Diagnosing Reading comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Evaluation of Language Fundamentals</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Curriculum Based Measurement</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dynamic Indicators of Basic Literacy Skills</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Early Reading Diagnostic Assessment</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Gray Oral Reading Test IV</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Iowa Test of Basic Skills</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Peabody Picture Vocabulary Test</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Stanford Achievement Test</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TerraNova—CAT</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Test of Language Development</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Test of Word Knowledge</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Texas Primary Inventory</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wechsler Individual Achievement Test</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Woodcock-Johnson Test of Achievement</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Woodcock Reading Mastery Test</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>


IDEA also identified three oral reading fluency measures that were sufficient to assess outcome measures in third grade students (table 4).

Table 4
Assessing Vocabulary and Reading comprehension Outcomes at 3rd Grade

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sufficient for Assessing Vocabulary Outcomes</th>
<th>Sufficient for Assessing Oral Reading Fluency Outcomes</th>
<th>Sufficient for Assessing Comprehension Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Evaluation of Language Fundamentals</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Curriculum Based Measurement</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Dynamic Indicators of Basic Literacy Skills</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Early Reading Diagnostic Assessment</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Gray Oral Reading Test IV</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Iowa Test of Basic Skills</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Peabody Picture Vocabulary Test</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Stanford Achievement Test</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>TerraNova-CAT</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Test of Language Development</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Test of Word Knowledge</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Texas Primary Inventory</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wechsler Individual Achievement Test</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Woodcock-Johnson Test of Achievement</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Woodcock Reading Mastery Test</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>


Retell as an Appropriate Comprehension Measure

Reading comprehension assessment measures differ between classrooms, schools, and school districts. Classroom assessment varies from informal questioning strategies to standardized assessment (Krauss, 1989). Informal assessments have the obvious limitation of
teacher subjectivity in both interpreting answers and measuring fluency (p. 6). Standardized assessment may be inconsistent with established curricula and tend to produce general, non-specific results that are insensitive in diagnosing individual student deficiencies (p. 5). Another form of standardized assessment is cloze where students fill in missing words from a text, using the original words or appropriate substitutions or maze, which is a multiple-choice form of cloze. The construct validity of these tests comes into question when deciding if these tests are measures of comprehension or vocabulary (p. 9). For example, the test may measure the ability to use context to predict omitted words, rather than the ability to understand a passage (p. 9). Also, the nature of cloze test construction (omitting every fifth word of a passage, beginning with the first word, second word, etc.) can produce a different test by starting with the second or third word (Alderson, 2000). Alderson adds, “What an individual cloze test measures will depend on which individual words are deleted,” adding that many reading skills may not be assessed, depending on the deletions (p. 208).

While acknowledging both the strengths and weaknesses of past assessment measures, contemporary authors and researchers (Bintz, 2000; Brown & Cambourne, 1990; Fuchs et al., 2001; Hoyt, 1999; Stahl et al., 1991) offer retelling as an accurate measure of reading comprehension. Bintz (2000) and Stahl et al. (1991) state that traditional forms of reading comprehension assessment, specifically multiple-choice questions on standardized test, do not accurately reflect what we currently know about reading. Bintz also documents recent comprehension reforms and advocates written retelling in the form of free writing as an accurate measure of comprehension (p. 208). Stahl et al.’s study of 10th grade students concluded that vocabulary proficiency enables readers to recall smaller units of text and prior knowledge helps the reader organize and evaluate the importance of these items and, through retell, demonstrate
comprehension and understanding (p. 1). Similarly, Brown & Cambourne (1990) list several benefits of retell:

1. Evidence of incidental learning of text structures, vocabulary, and conventions of written language.
2. Evidence of “linguistic spillover” or reappearance of forms, structures, concepts, and conventions used in retelling sessions.
4. Growth in “reading flexibility” or the ability to “change reading gears” from skimming to intense engagement with the text. (pp. 10-11)

Further evidence of the importance of retell in reading comprehension assessment is evident in many nationally recognized assessments. Measures like the Woodcock-Johnson Test of Achievement (WJ III), the Iowa Test of Basic Skills (ITBS), the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), and the Developmental Reading Assessment (DRA) include a measure of fluency, accuracy and the student’s ability to recall and retell what was read. Additionally, Fuchs et al. (1988) say that oral and written recall measures correspond significantly with students’ performance on standardized reading comprehension tests.

Perhaps Hoyt (1999) offers the most eloquent rationale:

Guided reflection and retelling have the added bonus of ‘teaching’ comprehension, while providing a format for assessing it. When a learner retells the content of a reading selection, the reader takes total responsibility for understanding and then communicating that understanding. The learner is also placed in a situation where he or she must consciously utilize the elements of a story structure to guide the retell while integrating information from the focus text. Through this process, an observing teacher has an
unrivaled opportunity to understand the learner. Retellings offer a window into oral and written language proficiency, depth of understanding of text, ability to connect world knowledge to the reading material, understanding of story structure, and stylistic devices. This increase in ownership, responsibility, and oral language use provide a learning atmosphere that is significantly more powerful than having a child sit passively and respond to teacher-directed questions. (pp. xi-xii)

However, despite the research linking vocabulary with comprehension and comprehension with retell, educators and researchers alike have traditionally resisted both acknowledgement of a direct correlation between vocabulary and reading comprehension and documentation of the reliability and validity of retell in comprehension assessment (Johnston, 1983). Fuchs et al. (2001) reasons that oral reading fluency involves the greatest growth in the primary grades with a “negatively accelerating curve through the intermediate grades” consequently retell ability in older students may be a reflection of reading competence rather than the ability to analyze a passage (pp. 240-241). Another area of concern with retell, as a measure of comprehension, is construct validity (Clark, 1982). Is oral retell a test of comprehension or vocabulary? Is written retell a test of reading comprehension or writing? Are both a measure of comprehension or memory? It follows that retell measures all of the above and I believe that the benefits of this authentic assessment outweigh the threats to construct validity.

Literature Summary

Vocabulary knowledge, oral reading fluency, and reading comprehension have been identified as essential components of reading (Alderson, 2000; Kame'enui et al., 2002; Milller, 1995; Panel, 2000a). Some studies have linked vocabulary knowledge with student achievement (Smith, 1975; Wiig & Secord, 1992), while others have suggested links between vocabulary and
reading comprehension (Block & Pressley, 2002; Gersten & Geva, 2003; Jongsma, 2000; McLaughlin et al., 2000; Qian, 2002; Read, 2000; Stahl, 2003; Staskowski & Creaghead, 2001). Additionally, the research on oral reading fluency suggests a “close relationship” with comprehension (Pinnell, 1995). While the research acknowledges the importance of these constructs in reading, the research has also linked quality of instruction with assessment and produced multiple definitions and applications of these instructional techniques, constructs, and, subsequently, multiple measurement tools and techniques, raising the issue of construct validity in assessment.

Through the methodology outlined below, this study further examines the relationship between vocabulary knowledge, oral reading fluency, and reading comprehension, focusing not on instruction but on specific measurement tools chosen for their ability to adequately and specifically measure the defined constructs and provide adequate data for appropriate analysis.

Methodology

Setting and Participants

Sixty-five third grade students from a 450-student elementary school in the Pacific Northwest will participate in the study. Fifty-four percent of the students in the school are classified as minority (non-white) and forty-five percent are on either free or reduced lunch, based on annual family income. The group of sixty-five third graders contains thirty-two boys and thirty-three girls, thirty-seven of which are white and fifteen of African-American or Hispanic descent. The cultures of the remaining students include Native American, Eskimo, Korean, Samoan, and Russian. Thirty-two of the sixty-five students (23 girls and 9 boys) qualify for free and reduced lunch and fourteen are classified as Limited English Proficient (LEP), with
four of these certified as monolingual. Finally, ten of the sixty-five students are certified in
special education, receiving support services in reading comprehension.

Measures

For this study, measures are selected from the National Reading Panel, IDEA’s list of
sufficient and effective instruments, The Southwest Educational Development Laboratory
(SEDL) Assessment Database, and field-tested reading assessments measures from the
Northwest Regional Educational Laboratory (NWREL). Appropriate measures are defined as: (a)
efficient—taking less time than other measures while providing equally precise information; (b)
sensitive—directly measuring the area of concern and measuring both count and time, e.g. words
per minute; and (c) positively correlated with expert opinion, i.e. teacher judgment.

Consequently, I have selected the Test of Word Knowledge (TOWK) to measure oral
vocabulary and DIBELS to measure oral reading fluency, and the Developmental Reading
Assessment to measure reading comprehension.

Test of Word Knowledge

The Test of Word Knowledge (TOWK) was developed in 1992 by Wiig and Secord as a
measure of receptive and expressive language and students’ ability to understand and use
vocabulary, including figurative language, multiple word meanings, conjunctions, and transition
words (Wiig & Secord, 1992). Further, the TOWK identifies students’ semantics skills and
measures and evaluates semantic knowledge (p. 1). Wiig & Secord state that semantics
represents the “what of communication—the content that is encoded and sent by a speaker and
received and decoded by a listener, as opposed to the how (syntax) and the why (pragmatics) of
the language system” (p. 1). The authors state definitively that students who have difficulty with
semantic development will be “severely hampered in both communication and learning” (p. 1).
The TOWK probes word knowledge at three levels. First, the TOWK assesses students’ ability to match spoken words with referents (receptive language) and name picture referents (expressive vocabulary). Second, the TOWK explores the students’ knowledge of word meanings and shared dimensions of word meanings (opposites and synonyms). Third, higher-level aspects of word knowledge are explored through examination of multiple meanings, multiple contexts, figurative usage, and logical relationships (Wiig & Secord, 1992).

The TOWK was evaluated by IDEA for reliability and the results indicated adequate reliability across measures (Kame'enui et al., 2002).

**DIBELS**

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of standardized, individually administered measures of early literacy development, designed to be one minute fluency measures used to regularly monitor the development of pre-reading and early reading skills (Good & Kaminski, 2002). The measures were developed from the essential early literacy domains discussed in both the National Reading Panel (2000) and National Research Council (1998) reports to assess student development of phonological awareness, alphabetic understanding, and automaticity and fluency with the code (p. 2). Each measure has been thoroughly researched and demonstrated to be reliable and valid indicators of early literacy development and predictive of later reading proficiency to aid in the early identification of students who are not progressing as expected. When used as recommended, the results can be used to evaluate individual student development as well as provide grade-level feedback toward validated instructional objectives (Good & Kaminski, 2002).

DIBELS Oral Reading Fluency (DORF) is a standardized, individually administered test of accuracy and fluency with connected text. DORF is a standardized set of passages and
administration procedures designed to (a) identify children who may need additional instructional support, and (b) monitor progress toward instructional goals. The passages are calibrated for the goal level of reading for each grade level. Student performance is measured by having students read a passage aloud for one minute. The number of correct words per minute from the passage is the oral reading fluency rate (Good & Kaminski, 2002).

*Developmental Reading Assessment*

The Developmental Reading Assessment (DRA) is a criterion-based, performance assessment developed and piloted by Joetta Beaver in collaboration with primary classroom teachers and was field-tested and revised by primary teachers in the Upper Arlington School District in Ohio between 1988 and 1996. This instrument allows teachers to gather information about students' reading interests, use of strategies, comprehension, and attitudes (Williams, 1999). The DRA can: (a) determine a reader's independent assessment reading level, (b) facilitate the effective grouping of students for reading experiences and instruction, (c) provide immediate information for instructional decision making; confirms or redirects ongoing instruction, and (d) document changes over time in reading performance (p. 1). The DRA conference follows a consistent format that varies slightly, depending on the age of the student. Students read aloud a grade appropriate book, selected by the teacher. As the student reads, the teacher takes a running record, recording accuracy and miscues. After completing the story, the students begin a free retell of the story, with the teacher probing or prompting for any missing information. After the session, the tester completes a comprehension rubric based on the recorded answers (Beaver, 2001).

In the spring of 1999 a reliability study was conducted to examine (a) inter-rater agreement of teachers using the assessment, and (b) internal consistency of the Developmental
Reading Assessment (DRA) instrument. This was accomplished through volunteer teachers who submitted three audio taped DRA student conferences, which were then submitted to a second and third teacher for rating. Analyses calculated across 4 facets (raters, students, text levels, and test items) revealed reliability between the originator and second rater was strong, i.e., inter-rater agreement between the first two raters was 0.80, however, inter-rater agreement among all three raters was not as strong (p. 5). The researchers expected the reliability between the second two raters to be higher but were not surprised, given the third rater received the tape eight months after the second rater. The researchers gave no reason for the lapse in time.

Additional data revealed the internal consistency to be strong for the five rating scale items (accuracy, comprehension, stage, phrasing, and reading rate), i.e., item separation reliability (Cronbach’s alpha = 0.98), across all three raters as well as for the DRA assessment texts, i.e., text separation reliability (Cronbach’s alpha = 0.97) (p. 6).

To ensure that the DRA measured what was intended, the validity of the DRA instructional reading level was assessed. To assess its validity, individual scores on the DRA for the second grade population (N=2470) at the end of the 1998-99 school year from a large urban/suburban school district were correlated with the students’ scores from fall of third grade on the Iowa Test of Basic Skills Subscales: Vocabulary, Reading Comprehension, and Total Reading. All correlations were significant at the 0.01 level (2-tailed) using Spearman’s Rho rank order correlation; however, the highest correlation for this assessment was with Total Reading (r = 0.71, p < .01) (Williams, 1999).

Louisiana requires every student to be assessed using the DRA and Arkansas provides state-funded training for teacher in the administration of the DRA (Southwest Educational Development Laboratory, 2004). Finally, the DRA was selected for the study because it is an
authentic performance based assessment in which children are responding to real text through retelling.

**Procedures**

During the fourth academic quarter, 3rd grade teachers will rank their students according to reading proficiency using teacher judgment. Students will then be tested in word recognition and vocabulary using the TOWK, oral reading fluency using the DIBELS Oral Reading Fluency Assessment, and in reading comprehension using the DRA.

**Validity-Reliability**

As part of my administrative training, I received 8 hours of training in administration of the Developmental Reading Assessment and have been administering this assessment for the past six years. Based on my assessment qualifications and background I received pre-qualification from Harcourt Assessment to purchase the TOWK and four hours of on-line training and technical support in test administration. The DIBELS Oral Reading Fluency measure is widely used in my school district and I have been administering this measure for the past six years. I will further address rater reliability through a calibration exercise. A state-certified Special Education Resource Teacher trained in assessment and I have randomly selected and assessed five students each from outside the target group using the measures listed above and compared and calibrated scores. Inter-rater and other test item reliability issues will be addressed as a possible confound in the data analysis section.

**Analysis**

Data will be analyzed during a two-month period immediately following the assessment using bivariate correlational statistics. Scores for performance on the TOWK will be correlated with scores on the DIBELS and the scores on the DRA. A product-moment correlation
coefficient $r$ will be used to determine the magnitude of the relationship between scores since all three variables are expressed through continuous scores and because $r$ has a small standard error. While it is acknowledged that any correlations obtained cannot necessarily establish causality, the multiple sub tests of the TOWK specify the variables equated with vocabulary knowledge proficiency, thereby reducing the likelihood that performance on the criterion variables was caused by any unnamed variable. Test appropriateness and efficiency will be evaluated based on the vocabulary and comprehension criteria established earlier. Test scores would be compared with teacher ratings to further examine assessment validity. I hypothesize that there would be a positive correlation between teacher rating and scores on reading comprehension. I would also hypothesize a positive correlation between scores on the TOWK, scores on the DIBELS, and scores on the DRA.
References


Washington, DC: National Center to Improve the Tools of Educators College of Education University of Oregon;

U.S. Dept. of Education Office of Educational Research and Improvement Educational Resources Information Center.

In J. J. Flood, Jensen, D. Lapp, J. R. Squire (Ed.), *Handbook of research on teaching the English language arts* (pp. 604-632). New York: Macmillan.


New York: International Reading Association ;

Teachers College Press.

