
Chapter 12

Vocabulary Assessment

*Making Do with What We Have
While We Create the Tools We Need*

**P. David Pearson
Elfrieda H. Hiebert
Michael L. Kamil**

After a long absence, vocabulary has returned to a prominent place in discussions of reading; it is alive and well in reading instruction and reading research. This is good news for teachers, teacher educators, and researchers. There is good reason to teach vocabulary more aggressively. After all, as many of the contributors in this volume illustrate, vocabulary is a strong predictor of comprehension, and the gap in the vocabularies of the “haves” and the “have-nots” is substantial on school entry. Without strategic and concerted efforts in schools, this initial gap may never be eliminated and, in fact, can easily be broadened. If effective instruction is to become commonplace, we must first address the vexing question of how we assess vocabulary knowledge and growth.

In this chapter, we argue that vocabulary assessment is grossly underdeveloped, both in its theoretical and practical aspects. On the theoretical side, it has been driven by tradition, convenience, psychometric standards, and a quest for economy of effort rather than a clear conceptualization of its nature and relation to other aspects of reading expertise, most notably, to comprehension. On the practical side, it has provided teachers with scores

that tell them only how students perform in relation to other students (e.g., percentiles or grade norms); it has not given them data that indicate how well students have mastered particular domains of vocabulary, such as all of the key vocabulary associated with a fourth-grade reading level or with the central concepts of a topic such as habitats and ecosystems. We hope that this chapter serves as one small step in stimulating the development of strategic instruction and curriculum needs that students deserve.

We examine the literature—research, common practices, and theoretical analyses—on vocabulary assessment to answer three questions:

1. What do vocabulary assessments (both past and current) measure?
2. What could vocabulary assessments measure, as illustrated by conceptual frameworks, and what development and validation efforts are needed to make such assessments a reality?
3. What vocabulary assessments should teachers use, create, or modify while we wait for the research that is needed for wide-scale change?

VOCABULARY ASSESSMENTS: PAST AND PRESENT

A working definition of a domain is useful, prior to embarking on either the design or the review of the assessment of that domain. In the case of vocabulary, definitions that come from dictionaries are typically vague, such as “the words of a language” or “the stock of words used by or known to a particular people or group of persons” (Flexner, 2003). Because such definitions lack the specificity required for instruction or assessment, educators (e.g., National Institute of Child Health and Human Development, 2000) have typically turned to a two-by-two matrix where the use of vocabulary is on one axis (productive, receptive) and the mode of communication (written, oral) is on the other. This matrix implies four types of vocabulary: listening, speaking, writing, and reading.

The assessment of vocabulary, as it pertains to reading comprehension, has almost exclusively emphasized the receptive dimension of vocabulary. At least on large-scale reading tests, the mode involves reading, although within the most prominent norm-referenced vocabulary measure—the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007)—the mode is listening. Rarely is the productive aspect of vocabulary examined, especially as it relates to comprehension; for example, when students encounter (or are taught) new words in relation to new texts or topics in subject-matter classes, it would be useful to know whether these words spontaneously emerge in their speaking and writing. Small-scale analyses

have documented students' spontaneous use of new, complex vocabulary in writing as a result of participating in intensive vocabulary instruction (Flinspach, Scott, & Vevea, 2010), but such projects are labor-intensive because students' writing must be collected and analyzed over extended periods of time.

More typically, the approach to assessment of vocabulary follows the patterns that can be traced back to the origins of standardized tests. In early tests of intelligence developed by Binet and Thurstone (see Pearson & Hamm, 2005), students were asked to define or explain words that they were likely encounter in schoolbooks (e.g., the function of a *fork*). With the movement toward mass testing prompted by the need to test recruits for World War I (Resnick & Resnick, 1977) came the need for easily administered and scorable assessments, which led to standardized, multiple-choice tests with items such as those illustrated in the first row of Table 12.1.

That sort of item dominated formal vocabulary assessment until the 1970s (Read, 2000), when changes in thinking about language and reading motivated more contextualized vocabulary assessments (see the second row of Table 12.1). The press for contextualization increased systematically, at least in the most ambitious context for vocabulary assessment, English as a second language (see Read, 2000; Nation, 2001), resulting in a progression of items as illustrated in the final three rows of Table 12.1.

Despite this history, many major assessments still use fairly isolated approaches. To illustrate the nature of current vocabulary assessments, we have analyzed items on four prominent vocabulary assessments that are among those that a national panel identified as fitting the criteria for use in Reading First (Kame'enui, 2002), a large-scale, federally funded national reading initiative: two individually administered assessments—the PPVT (Dunn & Dunn, 2007) and the Woodcock Reading Mastery Test (WRMT; Woodcock, McGrew, & Mather, 2007)—and two that are group administered—the Iowa Test of Basic Skills (ITBS; Hoover, Dunbar, & Frisbie, 2003) and the Stanford Achievement Test (SAT; Stanford Achievement Test, 2006). Items characteristic of those included in these assessments are presented in Table 12.2, except for the PPVT, which uses illustrations from which students select the word spoken by the test administrator (e.g., the stimulus for the word *surfing* might be a picture set of someone surfing, someone playing water polo, someone swimming, and someone driving a speedboat). In all of the items in Table 12.2 as well as in the PPVT example just described, there are rather simplistic word associations. None can be described as capturing the deep or multifaceted features of words that we know are important to vocabulary learning and, hence, to assessment.

TABLE 12.1. Sample Items of Different Eras

Time period	Sample item(s)		
1915–1920: Decontextualized vocabulary assessment	<p>Pick the word that fits in the blank.</p> <p>A _____ is used to eat with.</p> <p>—plow —fork —hammer —needle</p> <p>Pick the best meaning for the underlined word.</p> <p><u>Foolish</u> —clever —mild —silly —frank</p>		
1970s: Early efforts to contextualize vocabulary	<p>Pick the best meaning for the underlined word.</p> <p>The farmer <u>discovered</u> a tunnel under the barn.</p> <p>—wanted —found —traveled —captured</p>		
1980s: Steps toward contextualization	<p>In a (1) <u>democratic</u> society, individuals are presumed innocent until proven guilty. The (2) <u>establishment</u> of guilt is often a difficult task. One consideration is whether or not there remains a (3) <u>reasonable</u> doubt that the suspected persons committed the act in question. Another consideration is whether the acts were committed (4) <u>deliberately</u>.</p> <p>For each item, select the choice closest in meaning to the underlined word corresponding to the number.</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top;"> (2) 1. attribution 2. business 3. creation 4. absolution </td> <td style="width: 50%; vertical-align: top;"> (4) 1. both 2. noticeably 3. intentionally 4. absolutely </td> </tr> </table> <p>Among a set of comprehension items, you might find the following:</p> <p>In line 2, it says, “Two reasons are usually advanced to account for this <u>tardy</u> development; namely the mental difficulties . . . ”</p> <p>The word <u>tardy</u> in line 2 is closest in meaning to</p> <p>1. historical 2. basic 3. unusual 4. late</p>	(2) 1. attribution 2. business 3. creation 4. absolution	(4) 1. both 2. noticeably 3. intentionally 4. absolutely
(2) 1. attribution 2. business 3. creation 4. absolution	(4) 1. both 2. noticeably 3. intentionally 4. absolutely		
Late 1990s: Computerized format	<p>The Southwest has always been a dry country, where water is scarce, but the Hopi and Zuni were able to bring water from streams to their fields and gardens through irrigation ditches. Because it is so <u>rare</u>, yet so important, water played a major role in their religion.</p> <p>Look at the word <u>rare</u> in the passage. Click on the word in the text that has the <u>same</u> meaning.</p>		

TABLE 12.2. Simulated Items of Vocabulary Tasks on Three Norm-Referenced Tests

Test	Prototypical item(s)
Iowa Test of Basic Skills	To <u>sink</u> in the water play rest wash go down
Stanford Achievement Test	Item type #1: To <u>cut</u> is to— slice bark run save Item type #2: Put the money in the <u>safe</u> . In which sentences does the word <u>safe</u> mean the same thing as in the sentence above? The puppy is <u>safe</u> from harm. I am <u>safe</u> at home. It is <u>safe</u> to go out now. Michael opened the <u>safe</u> . Item type #3: Ron only has one hat, but he has <u>several</u> coats. <u>Several</u> means— funny some hungry large
Woodcock Reading Mastery Tests	Subtest 1: Antonyms (Read this word out loud and then tell me a word that means the opposite.) near (far) dark (light) Subtest 2: Synonyms (Read this word out loud and then tell me another word that means the same thing.) cash (money) small (little) Subtest 3: Analogies (Listen carefully and finish what I say [text is visible but experimenter reads the text].) Dark—light night (day) Rain—shine wet (dry)

VOCABULARY ASSESSMENTS: DIRECTIONS FOR THE FUTURE

Words may seem like simple entities, but they are not. Their surface simplicity belies a deeper complexity. Scholars of vocabulary suggest a much richer view of vocabulary—and its assessment—than captured by current tools. Two of the most influential perspectives come from the work of Nagy and Scott (2000) and Read (2000).

Complexity of word knowledge is evident in Nagy and Scott's (2000) identification of five aspects of word knowledge used in reading. The first of these, *incrementality*, refers to the fact that knowledge about a word/idea becomes a little deeper and a little more precise with each encounter, leading to nuanced understanding and flexible use. As an example, consider the word *éclat*, in a phrase from a novel: "the *éclat* of love at twenty years of age" (Dumas, 1845). The use is ambiguous enough that a 21st-century reader may be a little unclear about its meaning. But, when encountered a second time, in another novel (Austen, 1790), it becomes clearer: "where we spent our little fortune with great *éclat*." On a third encounter—this time perhaps in an online review of an author's presentation at a local bookstore, where the audience's response to an author who has received otherwise modest acclaim is described this way, "They gave him more *éclat* than he really deserved," the reader is likely to feel fairly confident that the word means *enthusiasm* and might even be tempted to use it in a dinner-table conversation.

Multidimensionality, the second aspect, refers to qualitatively different types of word knowledge, including understanding nuances of meaning between words such as *glimpse* and *glance* or typical collocations of words (e.g., we can have a *storm front* but not a *storm back*).

The third aspect is one of the most critical and challenging for learning English—*polysemy*. Many words have multiple meanings, and the more common the word, the more meanings it is likely to have. The common word *set*, for example, has dozens and dozens of meanings, including ones that are part of unique phrases (e.g., *movie set*) and that take on different parts of speech (e.g., *set* the table, a *set* of drawers, a *set* direction). By contrast, an uncommon word such as *poinsettia* is likely to have a single meaning.

Interrelatedness involves learning or knowing a word, which often entails derivation (*commence–commencement*) or association with the meanings of related words, either in a linguistic context (*dogs bark* or *buffaloes roam*) or in one's semantic memory store (dogs are members of the canine category and related to cats because they share the attribute that they can be domesticated).

Heterogeneity refers to a word's meaning, which differs depending on its function and structure (e.g., frequency in written English, and syntactic roles). Contrast, for example, the sentences "I spilled the cocoa. Get a

broom” with “I spilled the cocoa. Get a mop.” Over time, by encountering a word such as *spill* in different contexts, we learn more about the range of its applications and the shades of meaning instantiated by different contexts and applications. In the case of the spillage of cocoa, comprehension depends on knowing that cocoa can be a bean, a powder, and a liquid and that brooms and mops, although both cleaning tools, have different properties.

Although these categories illustrate the complexity of vocabulary, few studies of vocabulary attend to these variables in any systematic fashion, especially when it comes to choosing the words for instructional interventions or assessments (Scott, Lubliner, & Hiebert, 2005). At the present time, these distinctions are unlikely to be highly productive as filters for reviewing assessments that are commonly used in large-scale assessment. These variables do, however, suggest important new directions for vocabulary research in both assessment and instruction.

A second framework comes from Read (2000), who has examined existing assessments and scholarship to understand how different test formats represent different kinds of vocabulary knowledge. Read’s interest lies in the design of appropriate assessments for individuals learning English as a second language, but we believe that they are equally germane to the design of assessments for children learning to read English. Read has identified three continua for designing and evaluating vocabulary assessments: (1) discrete-embedded, (2) selective-comprehensive, and (3) contextualized-decontextualized.

The *discrete-embedded* continuum addresses whether vocabulary is represented by its own separate set of test items (discrete) or as part of the larger construct of text comprehension (embedded). In the vocabulary assessments summarized in Table 12.2, vocabulary is reported separately from comprehension (i.e., discrete). Most norm-referenced silent reading assessments and the National Assessment of Educational Progress (NAEP)—a Congressionally mandated assessment administered to a representative group of students in each state every few years—take an embedded stance, wherein contextualized vocabulary items contribute to an aggregate comprehension score. But the direction of the NAEP is changing. The 2009 NAEP framework (National Assessment Governing Board, 2005; Salinger, Kamil, Kapinus, & Afflerbach, 2005) mandated a discrete measure of vocabulary. This change could be reflected in the assessments that two state consortia—SMARTER Balanced Assessment Consortium (2010) and the Partnership for Assessment of Readiness for College and Careers (2010)—are developing in response to the Common Core State Standards (Common Core State Standards Initiative, 2010).

The *selective-comprehensive* distinction refers to the relationship between the sample of items in a test and the hypothetical population of

vocabulary items that the sample represents. Thus, if one assesses students' grasp of the vocabulary in a story from an anthology or a chapter in a science text, the sample is inherently selective. Most vocabulary assessments, however, fall on the comprehensive side of the continuum, to the point where the domain sampled is a mystery. For example, consider the items in Table 12.2. Because the items on real tests are copyrighted and cannot be shared publicly, we had to create items that paralleled those of assessments. To create these items, we could find no guidelines, theories, or frameworks within test makers' technical reports or materials for users. We ended up choosing our parallel words by matching the word frequency and decodability of the target words in the actual items. Such a lack of clarity on the source of vocabulary is typical of current, large-scale assessments. Most of our current vocabulary assessments have no theoretically defined population of words at all or, if they do, are difficult (if not impossible) to infer from available documents.

We came away from the process concluding that psychometric criteria, not theoretical frameworks, drive the test-development process. Test developers obtain a group of words that is administered to a sample of students at known levels of reading development or expertise. The words are sorted by their difficulty, expressed often as the percentage of students in a particular population who answered the question correctly. Ultimately, scores for individuals on such a test derive their meaning from comparisons of particular students' performance on a particular test with the population of students—not from words, at large—which is why we call them norm-referenced tests. Under such circumstances, all we know is that a given student performed better, or worse, than the average student on the set of words that happened to be on the test. We know nothing about what the scores say about students' vocabulary knowledge of any identifiable domain or corpus of words.

In the next section, we illustrate a framework for a theoretically grounded selection of words for instruction and assessment. Theoretically grounded frameworks have yet to be used in designing assessments, but such frameworks exist and they suggest that it may be possible to move assessments toward the selective end of the continuum. Only when theoretically grounded frameworks drive assessments will it be possible to make claims that "The average student in a given school (or classroom) exhibits basic mastery over $X\%$ of the words in a given corpus (e.g., the words encountered in a given curriculum in a given grade level)."

Contextualized–decontextualized refers to the degree that textual context is required to determine the meaning of a word on an assessment. Table 12.3 includes several examples of the continuum. Item 1 falls firmly on the decontextualized side of the continuum. Even though context is provided for item 2, it is not needed if someone knows the meaning of

TABLE 12.3. Degrees of Contextual Reliance

1. <u>consumed</u> a) Ate or drank b) Prepared c) Bought d) Enjoyed	2. The people <u>consumed</u> their dinner. a) Ate or drank b) Prepared c) Bought d) Enjoyed
3. The people <u>consumed</u> their dinner. a) Ate or drank b) Used up c) Spent wastefully d) Destroyed	4. The citizens <u>consumed</u> their supply of gravel through wanton development. a) Ate or drank b) Used up c) Spent wastefully d) Destroyed

consume as eating or drinking. In item 3, because all four meanings denote one or another meaning of *consume*, context is essential for zeroing in on the meaning as used in the sentence. Item 4 is even trickier than item 3. Unlike item 3, which requires the selection of the most common meaning of *consume*, item 4 requires a student to reject the default (most common) meaning in favor of a more arcane sense of *consume*. Note also that a very fine semantic distinction is required in item 4 to select *spent wastefully* over *used up*.

We know of no vocabulary initiative, be it a research study, a curriculum scheme, or an assessment development effort, with the aim of bringing together the criteria related to word knowledge (Nagy & Scott, 2000) and the criteria related to assessments formats (Read, 2000). There is a degree of overlap between the two schemes. For example, items that fall at the completely contextualized end of the continuum are almost without exception polysemous words because it is nearly impossible to assess vocabulary in context without distractor sets that reflect at least two of the meanings of each assessed word. On the other features, however, the two schemes represent important but unique aspects of the design of vocabulary assessments. However, if educators are to develop a true measure of their students' knowledge of vocabulary, just such an integration of word knowledge features and assessment features is needed.

In the framework that we unpack in the next section, we have not achieved such an integration of all variables in both schemes, but we begin a line of work that might lead to integration by addressing one issue in depth: the selection of words on assessments. This feature of assessments is at the core of understanding what students are learning and what needs to be taught and learned. It falls squarely into Read's selective-comprehensive issue and it also addresses, at least to some degree, some of the features of word knowledge. It is in the design of individual items, however, that issues raised by Nagy and Scott will be played out. Researchers are engaged

in efforts that should provide insight into the manner in which features such as incrementality, multidimensionality, interrelatedness, polysemy, and heterogeneity can be best measured (Scott, Flinspach, Miller, Vevea, & Gage-Serio, 2009; Scott, Vevea, & Flinspach, 2010). With the validation of new item formats from projects such as these, educators can anticipate that richer views of the nature of students' vocabulary knowledge will be possible. For these rich views of students' vocabulary knowledge to ensure a closing of the vocabulary gap, it is imperative that the words on the assessments are *the right words*. By "the right words," we refer to information on the selection of words. It is that topic to which we turn our attention next.

An Illustration of a Theoretically Grounded Framework for Vocabulary Assessment

There are numerous ways of parsing the words of a vocabulary. Written words are, after all, multidimensional, as Nagy and Scott (2000) demonstrate in their analysis; they differ in orthographic forms, syntactic functions, and semantic roles, to name several important distinctions. In this section, we illustrate one way of organizing vocabulary that might increase our capacity to measure vocabulary acquisition in ways that will be more useful to educators across the disciplines of literature, science, social studies, and mathematics. The illustrative framework for vocabulary assessment is one on which one of us (Hiebert, 2011) has been working. We offer this categorization scheme not as the "be-all and end-all" of vocabulary schemes but to illustrate the manner in which assessments can be organized.

The three parts of the scheme, as depicted in Figure 12.1, are (1) *core vocabulary*: the 5,600 most common words (along with words that share the root word but have inflected endings) that constitute approximately 90% of running academic text; (2) *extended vocabulary*: the other 300,000–600,000 (Leech, Rayson, & Wilson, 2001; Oxford Dictionaries, 2010) words in English that, while occurring rarely, give detail to narratives and precision to informational texts; and (3) *word flexibility*: the capacity to negotiate the subtleties of word use (mainly due to polysemy and interrelatedness) in a variety of contexts. We illustrate the focus of each of the three components of the scheme, reminding readers that we provide this scheme primarily as an illustration of ways of specifying the words for vocabulary assessment.

Core Vocabulary

The vocabulary of English is distributed unevenly. A small group of words—approximately 5,600 (Carroll, Davies, & Richman, 1971; Zeno, Ivens,

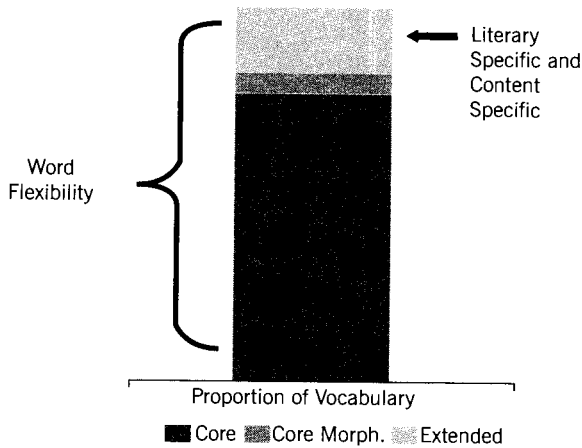


FIGURE 12.1. Framework for vocabulary. From Hiebert (2011). Reprinted with permission from the author.

Millard, & Duvvuri, 1995)—accounts for about 80% of all the words in all English texts, including those of college and the workplace. When inflected words (e.g., *looked* plus *look*; *happier* plus *happy*) are added, this group accounts for about 90% of the total words in English texts.

The other 10% of the total words that we encounter in printed English come from an enormous body of words—approximately 300,000–600,000, depending on how a word is defined—that occur with low frequency in texts. This 10% of the rarest English words is often the source for the items on vocabulary tests. To foreshadow our position, we will provide an alternative stance toward this 10%, what we will call *extended vocabulary*.

Ironically, it is the *core vocabulary* that has often been overlooked in vocabulary assessments. Typically, core vocabulary is viewed as the Dolch words—a small group of mostly abstract function words—which, in actuality, form only a small part of the approximately 5,600 words. But beyond the first 100 or so most frequent words, there are numerous core content words that carry the semantic load of sentences. Typically, these words represent general concepts (e.g., *mysteries*, *property*, *value*, and *interior*) and are versatile—hence, their frequency. By *versatile* we mean that these words often have multiple meanings, change their function as parts of speech (e.g., be a *force*, *force* an issue), share a root word with numerous other words (e.g., *spectacles*, *inspect*, *respect*, *spectrometer*), and often “travel” across disciplines (history, science, mathematics, and literature).

The 10% of the lexicon that occurs rarely provides much of the specificity in texts—words such as *anthill*, *commandant*, *rectory*, and *dirigible*.

But it is unlikely that a reader will be able to understand *dirigible* if he or she is not facile with the core vocabulary, at least in part, because core vocabulary provides the contextual and lexical fabric that allows us to infer the meanings of the rarer extended vocabulary. Consequently, an essential part of vocabulary assessment pertains to students' knowledge of the vocabulary that accounts for the majority of the texts that they read. If students can read these words, it is highly likely that they can figure out the "other 10%"—or at least make a sufficiently plausible inference about the meaning of these rarer words to make sense of the text as a whole. If students aren't facile with the core vocabulary, they will never be successful as readers—no matter how many engaging vocabulary lessons are taught that include challenging words from the 10%.

Despite the versatile nature of the core vocabulary, a basic issue has been viewed as an obstacle to textbook publishers, in designing instructional materials, and to test-makers in creating assessments. The issue is this: Just because two words happen to be in the same "band" of frequency, it does not follow that knowing one means that students are likely to know the other. Consider this set of words that share a predicted low frequency of 19 appearances per million: *mayor*, *measurement*, *mercury*, *microscope*, and *mysterious*. Words that have high imagery values or concreteness, such as *microscope*, *mercury*, and *mayor*, are fairly straightforward and likely tied to students' exposure to these words in everyday life or even through text. Words such as *measurement* and *mysterious* are more abstract and have nuanced meanings. If a student knows the meaning of *microscope*, it does not follow that he or she will know the meaning of *mysterious*.

Although membership in a particular band of frequency provides no indication of semantic relationship, it can be taken as an index of two important phenomena: (1) students' mastery of this important band of words and (2) the amount of exposure students have had to texts. The 5,600 words are predicted to occur at least 10 times or more per million words (with most occurring substantially more). Ten repetitions is the "rule of thumb" that has been associated with gaining working knowledge of a word (McKeown, Beck, Omanson, & Pople, 1985), although, as Landauer, Kireyev, and Panaccione (2011) have shown, concrete words representing specific objects or people (e.g., *microscope*, *mayor*) may require fewer repetitions. If, however, fourth graders have read approximately 500,000 words in school texts (a reasonable expectation, given what we know about curriculum and reading volume; Anderson, Wilson, & Fielding, 1988), they should have had exposure to a substantial portion of the core vocabulary. Thus, an assessment of words from particular bands of the core vocabulary could be taken as evidence that students have read critical benchmark amounts of text.

Extended Vocabulary

Now what about those 300,000–600,000 words in the extended vocabulary? They should be assessed somehow, for surely it is mastery of this vocabulary that sets apart the truly literate from the functionally literate individuals in our society. The extended vocabulary, as we discuss in the next section, differs as a function of text type. And the vocabulary assessments we use or design need to take those differences into account.

The percentage of extended vocabulary appearing in text is highly similar across narrative and informational categories through the elementary period, at which point the proportion of extended vocabulary in informational texts begins to creep ahead of the proportion in literary texts—roughly 15% of the total tokens in the text represent extended vocabulary.

What differs between narrative and informational text is the conceptual complexity and the relationships among the words in a text. Hiebert and Cervetti (Chapter 16, this volume) report that science words are longer, more conceptually complex, and more thematically related to one another and to the topic of the text than are narrative words. These differences suggest the need for unique instructional approaches and types of assessment. In science, for example, where unique words are conceptually complex but interrelated, assessing students' knowledge of relationships among key terms or their ability to use these words in talking or writing about the big ideas in the curriculum makes sense. Because the unique words of narrative texts represent familiar concepts with unfamiliar labels (e.g., *dejected/sad*, *meticulous/careful*), assessments would emphasize either synonyms or perhaps the special nuance of meaning that is gained when a writer uses a rare word, such as *gargantuan*, rather than common synonyms such as *big* or *large*.

Nagy, Anderson, and Herman (1987) operationalized conceptual difficulty as the relationship of a new word's meaning to students' existing knowledge, with the key distinction made between words that are new labels for concepts already known to the learner (e.g., *apologize* for the known concept *saying you're sorry*) and words that require the acquisition of new factual information or a new system of concepts to learn (e.g., *divide* as the boundary between river systems). Nagy et al. (1987) found that conceptual difficulty was the factor that most strongly influenced whether students learned a given word from context while reading a text. Other word features that did not have a significant impact on learning from context in this study were word length, part of speech, morphological transparency, and proportion of students who reported knowing the word before reading. There was no evidence of incidental learning from context for words at the highest level of conceptual difficulty. In their analysis of text-level factors,

Nagy et al. found that the proportion of conceptually difficult words was also a significant negative predictor of word learning from context.

Semantic relatedness among word meanings is another critical part of conceptual difficulty: A word can be conceptually difficult because it is part of a system of related meanings that is new, not simply because its meaning is unknown. When the focus is on knowledge acquisition, these systems of related words become the unit of instruction, rather than individual words, particularly for content-area texts (e.g., all the words required to explain photosynthesis or the legislative process). The contrast is important: The unknown words in a narrative text are, for the most part, unlikely to be either related to one another or encountered again, whereas the new terms of one topic in a content area are (1) likely to be encountered many times in the chapter in which they initially occur, (2) likely to be related to one another, and (3) highly likely to be the conceptual foundation for the new ideas in the next chapter or topic.

Word Flexibility

This third issue of word flexibility involves the stance that learners can take as they encounter new words while reading. We have labeled it *word flexibility*, drawing liberally on Spiro and Jehng's (1990) notion of cognitive flexibility to define it. According to Spiro and Jehng, cognitive flexibility is the ability to adapt cognitive strategies to the particular situation we find ourselves in as learners. This adaptability is just as essential in navigating English vocabulary as it is in any other learning situation. When it is applied to vocabulary, however, it takes on a particular character that distinguishes it from other kinds of learning. To understand how word flexibility works, it is useful to revisit several of Nagy and Scott's (2000) dimensions introduced earlier, specifically, polysemy, interrelatedness, and multidimensionality.

To cope with polysemy, readers need to understand that the same written form of the word on the page can mean something different in different contexts; moreover, they must suppress alternate meanings in order to focus on the meaning at hand. The manner in which the word *conflict* is used, for example, in the sentence *The two groups have been in conflict with each other for years* is different from its use in the sentence *Armed conflict might be unavoidable*. These alternate meanings may seem slight, but they imply very different meanings about the world.

Word flexibility also involves an awareness of interrelatedness, particularly the expectation in complex texts that particular words often appear together. Using the word *conflict* again, the term *conflict of interest* illustrates a unique meaning for which a reader needs to have word flexibility. With respect to multidimensionality, readers need to anticipate that words take on different functions (i.e., parts of speech) and, in some cases, even

different pronunciations. In the case of the word *conflict*, the stressed syllable changes when it moves from a noun (CON flict) to a verb (con FLICT). Another aspect of multidimensionality is the recognition of shared meanings across words. Many words, especially in the academic texts of the disciplines, come from the Romance language historical heritage of English. This means that adding prefixes and suffixes generates new forms of the word. The word *recommend* illustrates derivation of a word. *Recommend* is a member of a family that comes from the base word *commend*, and both *commendation* and *recommendation* represent additional derived forms. When readers possess the knowledge that English permits this sort of generative flexibility in forming derived forms, their capacity to understand and compose increasingly complex texts is greatly enhanced.

These three elements—core vocabulary, extended vocabulary, and word flexibility—are the concepts in our approach to vocabulary assessment—and instruction! We believe that they could be used to transform the way in which we assess vocabulary in schools as well as the decisions we make about what words to teach and how to teach them.

A Research Agenda for the Next Decade

We have raised this three-pronged framework of core vocabulary, extended vocabulary, and word flexibility as one means of identifying the parts of the lexicon for assessments. There is an enormous amount of work that needs to be done before such a framework—or others—can be used on a large-scale to assess students' vocabulary knowledge and growth that, in turn, guide practice and policy. In this section, we raise some of the research issues that need to be addressed before such assessments can be used on a large scale.

Criteria for Selection

We have offered one framework for making determinations along the selective–comprehensive continuum. The usefulness of assessments that capture bands within the core and extended vocabularies in guiding instructional practice is of the essence. Are such assessments any better than those that fall in the extremes of the comprehensive end of the continuum in supporting educators in closing the vocabulary gap?

Additional schemes require validation as well, such as Biemiller's (2009) designation of words worth teaching on the basis of data on students' familiarity with words. Beck, McKeown, and Kucan (2002) advocate focusing on words that are relatively rare synonyms of known concepts—just the sort of words that characterize literary texts. Determining bodies of words that belong to particular semantic clusters (e.g., ways that are used to communicate excitement) might be another way to narrow the sample of words

within an assessment. These and other ways of configuring words need to be studied, as does the relationship of assessments with different bodies of vocabulary. In particular, performances on vocabulary assessments need to be considered in relation to different modes of use: receptive-productive and oral-written. How is vocabulary on a particular measure related to students' use of grade-appropriate vocabulary in writing? How well does a vocabulary assessment relate to students' ability to understand complex vocabulary in a listening or speaking mode? As much as we want to get the assessment of reading vocabulary right, we cannot do so at the cost of rich approaches to assessing it in other language modes.

Word Flexibility

For the large population of extended vocabulary, what we really need to know is how capable students are of determining the meanings of words that they have seldom, if ever, encountered. That presents a formidable challenge. What seems most promising to us is to conceptualize the process of determining word meaning as a problem-solving activity in which sentence contexts (both the one at hand and other "collocations" we could provide to students), morphological connections (it's a member of the *spec* family so it must have something to do with vision), and semantic connections (it's in the same conceptual family as other machines we use to measure eyesight) become the resources that students use to infer the meanings of unknown words they encounter in text. In short, we should move toward the assessment of students' word flexibility levels.

Vocabulary and Text Genre

It is clear that informational text typically carries a heavier vocabulary burden than does literary text. Currently, that difference is a hidden variable in many studies. Research is needed to untangle the relationship between text genre and vocabulary: (1) how words are chosen for instruction, (2) what's new for students—the ideas, the words, or both, and (3) how likely are words to be repeated in the texts in which they first occur. Regardless of the answers, they will have profound implications for vocabulary instruction and transfer. Because vocabulary is generally considered in a holistic fashion, one dividend might be to differentiate methods of vocabulary instruction by text genre.

Transfer of Vocabulary Knowledge

The preceding points all converge on the issue of transfer of vocabulary knowledge to other components of reading. The research alluded to here

would almost certainly offer insights into the difficulties we have raised about issues of transfer and the specific effects of vocabulary instruction on comprehension; for example, which aspects of novelty (new ideas or new words) cause the most havoc with comprehension? More important is the explicit attention to the issues of transfer, both near and far, for the tasks under investigation. In addition, the strength and durability of transfer over time should be a part of this effort, particularly given the relatively short duration of many vocabulary instruction interventions in the literature. It is one thing to teach a word, but if it is forgotten in a few days, it hardly seems worth the effort.

Computerized Assessments of Vocabulary

Finally, we need to seriously address computerized assessments of vocabulary domains. In a better world, we would not be limited to conventional norm-referenced assessments of vocabulary acquisition, where our only benchmark for gauging vocabulary growth is the average performance of other students. We could opt instead for estimates of mastery over particular domains of interest (e.g., all of the words in a given curriculum or a given frequency band) or estimates of control over other characteristics that might prove to be effective indexes of vocabulary learning (e.g., all words with a common morpheme, such as *spec*, or all the words denoting some aspect of *power*). The work recently reported by Landauer et al. (2011) suggests that we may not be far off from providing what they describe as *personalized* assessments of students' vocabulary. For example, they are refining a word maturity algorithm to determine students' knowledge of words in a variety of contexts (i.e., heterogeneity). They can, over time, capture students' learning of a word (incrementality), and they can provide sufficient contexts for capturing at least some levels of interrelatedness. However, at this point, their digital assessment provides only one format that does not do justice to the dimensions of polysemy or multidimensionality. Nevertheless, efforts such as those of Landauer et al., indicate that we may not be far off from adaptive assessments that permit insight into the breadth and depth of students' knowledge of words.

VOCABULARY ASSESSMENT FOR THE PRESENT: WHAT TEACHERS CAN DO RIGHT NOW

So what's a teacher or a principal or a curriculum coordinator to do about vocabulary assessment in his or her situation while we wait for the revolution in research and development that we have just described as a set of goals for the future? Educators still have to make decisions about who possesses

what levels of knowledge about what vocabulary. They still have to place students in programs, decide whether students have met standards, and determine whether their curriculum is doing the job they hoped it would do. So what can and should be done in the interim? Two general strategies, each supported by a set of specific tactics, seem appropriate to the situation educators face: (1) continue to use the norm-referenced tests currently in use but with explicit knowledge of their limitations clearly in mind, and (2) develop local assessments based on the advice offered in this chapter and tailored to a particular situation.

Making Judicious Use of the Current Crop of Commercially Available Assessments

In one sense, our critique of standardized comprehension assessments is culturally unfair. After all, these norm-referenced assessments have been used to make decisions about students' vocabulary knowledge for over a century. And what's wrong with knowing how a student "stands" in relation to other students around the country? Nothing . . . and everything. The problem with norm-referenced assessments is that they don't offer a clue about instruction. Just because you know that Tommy scores at the 32nd or 75th (or any other) percentile tells you nothing about what steps (or what specific words) you might teach to improve his vocabulary. A norm-referenced score thus provides a "screen" or an "early warning system" to let you know that something more ought to be done in this school or that you, as a teacher, need to learn about Tommy's vocabulary repertoire in order to zero in on an instructional plan (low norm-referenced scores = do something, quickly!), but they cannot and will not provide the plan. At best, they lead to prescriptions such as "Read more books, experience more of the world, and learn more words." In short, they tell you whether Tommy knows more or fewer words than the "average bear—nothing more, nothing less." As long as they are used with these limitations in mind, they can be helpful sources of information for evaluating the effectiveness of curriculum and instruction or determining who needs instructional attention. For greater instructional guidance, we will need tests that focus on students' relative mastery of particular collections of words.

Taking Our Advice Seriously

When it comes to vocabulary, what a teacher wants to know, in addition to how to teach it (a topic well addressed by the other chapters in this book), is how much of the corpus that Tommy *should* know does he *actually* know? And can the teacher learn that? It is, at once, the simplest and most complex of tasks. It is simple because it involves identifying the corpus (or

population) of words of interest and then sampling words from that population to get good estimates of the proportion of words Tommy knows. It is complex because everything matters, as the following sections make clear.

Defining the Corpus

Twenty years ago, it would have been difficult to determine how many new words were introduced in a textbook, a unit, or a story. There were just too many, and it was too hard to keep track of them. Not so in the 21st century. Computers have changed all that. It is now possible, with most textbook curricula, to answer the question “How many new words are there?” for just about any unit of curriculum. One can usually get a computerized list, and with some help from your district technology experts, create a computerized portfolio of relevant lists of words. Users need to be aware of their options here: One can define the corpus very broadly (all the words in the K–12 curriculum) or very narrowly (all the new words introduced in this story or all the words that include the *spect-* morpheme).

Dividing the Corpus

Sometimes it is useful to divide the corpus in ways that provide answers to questions that matter about student performance. For example, we believe that it is useful to make separate judgments about students’ mastery of words in different bands of frequency. Recall that a big dividing line for us is the core (the 5,600-plus inflected forms) and extended (the rest of the words in English) vocabularies. We would want to be able to answer questions like this: In this history text, what percentage of the new core words did my students know at the start of the semester (or year, depending), and what percentage have they learned through the semester (or year)? We would want to be able to answer the same question for extended vocabulary. Others might want to know the proportion of “academic” words known or learned that consist of a subset of words, mostly core, that are also on the Coxhead (2000) Academic Word List.

One can also imagine a plan in which at key points throughout a course, a unit, or a text, you and your colleagues assess mastery of vocabulary words that you have already taught versus those that are yet to be taught, to distinguish between intentional and incidental learning of vocabulary.

Keeping Track of Your Corpus

It’s a good idea to create a spreadsheet to keep track of all of the relevant variables involved in classifying words. For each word, classify it by several categories (e.g., zone membership, letter–sound transparency, number

of syllables, number of prefixes, number of suffixes, or inflection status). There are web resources to aid with this task (see Table 12.4 for a list of several, currently available sites).

Drawing Samples

Once a corpus has been defined, the task is to draw a representative sample of words from it, usually with an eye toward determining students' level of mastery over its contents. Drawing random samples of the words in the corpus is really the only defensible and appropriate strategy; the minute we begin to exercise judgment (let's pick this word rather than that word because it is more obscure or interesting) in selecting words, we compromise our ability to answer questions about level of mastery. Once you have committed to the logic of drawing random samples, you create a lot of options for yourself. The beauty of selecting random samples, provided that the corpus is sufficiently large, is that you can create an indefinitely large number of "forms" of the test—where a form equals a particular random sample of size N (# of items) from the corpus. And if your school has a large (and nimble) computer assessment capacity, a simple program could generate 30 different samples of words of a given size for the 30 students in your classroom. The advantage of this sort of system is that you can assess mastery of the corpus frequently (say, weekly or monthly) throughout the unit,

TABLE 12.4. Sources for Online Databases and Text Analyzers

Resource	Contents	Website
Web Vocabulary Profiler	This program tells you how many word types the text contains from the following four frequency levels: (1) the list of the most frequent 1,000 word families, (2) the list of the second 1,000 word families, (3) the Academic Word List, and (4) the words that do not appear in any of the preceding lists.	<i>www.er.uqam.ca/nobel/r21270/textools/web_vp.html</i>
Online Readability Index	On this University of Nebraska–Lincoln site you can input text to have it analyzed by Hiebert's word zones. You can also create custom zones using your own lists of content-specific words.	<i>cehs07.unl.edu/reading/zone</i>

semester, or year without giving the students the “same” form of the test—a real benefit for monitoring progress. Experience with curriculum-based assessment (Deno, 1984) tells us that occasionally one will draw an “odd” sample, but over time, if assessments are given frequently, the “bumps” due to this sort of fluctuation will smooth out, giving us accurate and useful assessments of the domain of words at hand.

Determining Sample Size

Determining how large a sample of words to select for an assessment to be representative of a corpus of words is a complex task, and we have not made enough progress on this approach to assessment to offer unassailable advice. But there are a few rules of thumb, taken from the test development literature. For example, it is hard to achieve even a modicum of reliability for measuring a particular phenomenon (e.g., objective, standard, or domain) of interest with fewer than 10 items, and that number assumes that the items are developed with enough care to exhibit small errors of measurement. With lower-quality items, more items would be needed. For classroom assessment purposes, especially when teachers are gathering lots of evidence about mastery of a standard over several days, the 10-item rule of thumb can be relaxed a bit. And when there are fewer than 10 words in a particular domain, as there might be for a particular domain like “all of the new words in this selection,” then the whole population of relevant words can be measured, and sampling is not an issue. Our advice is not to regard this issue lightly for the important reason that making an unreliable judgment about individual performance leads inevitably to poor pedagogical decisions for that student.

Item Format

The advice here is straightforward: Pick a format that matches your purpose for assessment. So, for example:

- If you want to know whether a student can “define” a word, then ask him or her to write or select a definition.
- If you want to know whether a student can determine the meanings of particular words in context, then ask him or her to write or select a definition after observing it in a sentence or paragraph context (even better, after seeing it used in three or four sentence contexts).
- If you want to know where a word “fits” into a knowledge base or a semantic network of related words, then assess student knowledge of words that are closely related to that word.

The Role of the Student

For years, teachers have been bringing students into the assessment process by asking them to rate their knowledge of particular words, either before or after a lesson or unit in which those words are used or taught. Scott et al. (2010) reported on a method of engaging students in estimating their personal knowledge of particular words. They asked students to make judgments about a word in four distinct and progressively discriminating test items:

1. Have you ever seen or heard this word before?
 - a. Yes
 - b. No
2. How confident are you about what you know about the word?
 - a. I've heard it, but I'm not sure what it means.
 - b. I think I know what this word means.

Items 3 and 4 for each word focus respectively on the word's semantic field and choosing a definition that also recognizes the part(s) of speech it exemplifies.

3. The word *flick* most likely has something to do with
 - a. Dolphins
 - b. Looking
 - c. Fruit
 - d. Movement
4. *Flick* means
 - a. The seed of a jungle fruit.
 - b. The way dolphins swim.
 - c. Looking carefully.
 - d. Moving something quickly.

The point is to involve students in the process of determining their own level of knowledge about words as early as they can begin to make those judgments. This can begin in grade 1, but it gets more important and more feasible in grades 3 and 4, and it can become a routine part of the introduction of new texts in middle and high school.

CODA: RECAPPING VOCABULARY ASSESSMENT

We wish we could offer our readers a more positive account of the state of vocabulary assessment, one in which we were able to either (1) express

more confidence about the validity, appropriateness, and utility of the current crop of commercially available vocabulary assessments; or (2) provide a definitive set of guidelines that readers could use to do it on their own. We can't offer that sort of account because our reading of the evidence leaves much to be desired (and much to be done!). We can, however, be more confident about a future in which we adopt a different framework (core and extended vocabulary coupled with word flexibility) for thinking about vocabulary and look to different assessment tools and constructs (the ideas in our suggestions for future research section) to improve the way we assess knowledge of words and their linguistic environments. In the meantime, we can begin to improve the quality, relevance, and utility of the assessments we use in our classrooms, schools, and districts by keeping the limitations of current assessments clearly in mind as we use them, and by building a new generation of local vocabulary assessments that adhere to as many of the principles of a new generation of vocabulary assessments as we can manage. Finally, we hope that whoever writes this chapter a decade from now will have a happier tale to tell. Whether they do unwittingly depends upon what all of us who care about vocabulary learning and its assessment do over that decade.

ACKNOWLEDGMENT

Portions of this chapter, particularly the section titled "Vocabulary Assessments: Past and Present," are based on and adapted from Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282–296. Copyright 2007 by the International Reading Association. Adapted by permission.

REFERENCES

- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285–303.
- Austen, J. (1790). *Love and freindship* [sic]. www.gutenberg.org/files/1212/1212-h/1212-h.htm
- Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.
- Biemiller, A. (2009). *Words worth teaching: Closing the vocabulary gap*. Columbus, OH: SRA.
- Carroll, J., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. Boston: Houghton Mifflin.
- Common Core State Standards Initiative. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and*

- technical subjects*. Washington, DC: CCSSO & National Governors Association.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2) 213–238.
- Deno, S. (1984). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Dumas, A. (1845). *Twenty years after*. www.gutenberg.org/files/1259/1259-h/1259-h.htm
- Dunn, L., & Dunn, P. (2007). *Peabody Picture Vocabulary Test* (4th ed.). San Antonio, TX: Pearson Assessments.
- Flexner, S. B. (Ed.). (2003). *Random House Webster's unabridged dictionary* (2nd ed.). New York: Random House. Retrieved October 15, 2001, from dictionary.infoplease.com/vocabulary
- Flinispach, S. L., Scott, J. A., & Vevea, J. L. (2010). Rare words in students' writing as a measure of vocabulary. In R. T. Jimenez, V. J. Risko, D. W. Rowe, & M. Hundley (Eds.), *59th annual yearbook of the National Reading Conference* (pp. 187–200). Oak Creek, WI: National Reading Conference.
- Hiebert, E. H. (2011, June 7). The 90–10 rule of vocabulary in increasing students' capacity for complex text. Retrieved from textproject.org/frankly-freddy/the-90-10-rule-of-vocabulary-in-increasing-students-capacity-for-complex-text
- Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2003). *Iowa Tests of Basic Skills*. Rolling Meadows, IL: Riverside.
- Kame'enui, E. J. (2002). *An analysis of reading assessment instruments for K–3*. Eugene, OR: Institute for the Development of Educational Achievement, University of Oregon.
- Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for word knowledge. *Scientific Studies of Reading*, 15(1), 92–108.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British National Corpus*. London: Longman.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 20, 522–535.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237–270.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah, NJ: Erlbaum.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- National Assessment Governing Board. (2005). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: American Institutes for Research.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: Author.

- Oxford Dictionaries. (2010). *Oxford dictionary of English* (3rd ed., rev.). New York: Oxford University Press.
- Partnership for Assessment of Readiness for College and Careers. (2010). *The Partnership for Assessment of Readiness for College and Careers (PARCC) application for the Race to the Top Comprehensive Assessment Systems Competition*. Retrieved from www.fl DOE.org/parcc/pdf/apprtcasc.pdf
- Pearson, P. D., & Hamm, D. N. (2005). The history of reading comprehension assessment. In S. G. Paris & S. A. Stahl (Eds.), *Current issues in reading comprehension and assessment* (pp. 13–70). Mahwah NJ: Erlbaum.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Resnick, D. P., & Resnick, L. (1977). The nature of literacy: An historical exploration. *Harvard Educational Review*, 47, 370–385.
- Salinger, T., Kamil, M. L., Kapinus, B., & Afflerbach, P. (2005). Development of a new framework for the NAEP reading assessment. In C. M. Fairbanks, J. Worthy, B. Maloch, J. V. Hoffman, & D. L. Schallert (Eds.), *55th yearbook of the National Reading Conference* (pp. 334–349). Oak Creek, WI: National Reading Conference
- Scott, J. A., Flinspach, S., Miller, T., Vevea, J., & Gage-Serio, O. (2009, December). *Vocabulary growth over time: Results of a multiple level vocabulary assessment based on grade level materials*. Paper presented at the National Reading Conference, Albuquerque, NM.
- Scott, J. A., Lubliner, S., & Hiebert, E. H. (2005). Constructs underlying word selection and assessment tasks in the archival research on vocabulary instruction. In J.V. Hoffman, D.L. Schallert, C.M. Fairbanks, J. Worthy, & B. Maloch (Eds.), *55th Yearbook of the National Reading Conference* (pp. 264–275). Oak Creek, WI: National Reading Conference.
- Scott, J. A., Vevea, J. L., & Flinspach, S. L. (2010, December). *Vocabulary growth in fourth grade classrooms: A quantitative analysis of Year 3 in the VINE project*. Paper presented at the annual meeting of the Literacy Research Association, Fort Worth, TX.
- SMARTER Balanced Assessment Consortium. (2010). *Race to the Top Assessment Program application for new grants: Comprehensive assessment systems CFDA Number: 84.395B*. Retrieved from www.k12.wa.us/SMARTER/RTTTApplication.aspx
- Spiro, R. J., & Jehng, J. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163–205). Hillsdale, NJ: Erlbaum.
- Stanford Achievement Test. (2006). *Stanford Achievement Test Series* (10th ed.). San Antonio, TX: Pearson Assessment.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock–Johnson III*. Rolling Meadows, IL: Riverside.
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. New York: Touchstone Applied Science Associates.