Flinspach, S., Scott, J., Vevea, J. (2010) Rare Words in Students' Writing as a Measure of Vocabulary. In R. Jimenez, V. Risko, M. Hundley, D. Rowe (Eds.) *59th Annual Yearbook of the National Reading Conference* (pp. 187-200). Oak Creek, WI: National Reading Conference.

Susan L. Flinspach
Education Department
University of California – Santa Cruz
1156 High Street
Santa Cruz, CA 95064

Judith A. Scott
Education Department
University of California – Santa Cruz
1156 High Street
Santa Cruz, CA 95064

Jack L. Vevea
School of Social Sciences, Humanities and Arts
University of California – Meced
5200 N. Lake Road
Merced, CA 95343

For more information on the VINE Project, see: http://vineproject.ucsc.edu

Rare Words in Students' Writing
as a Measure of Vocabulary

Susan Leigh Flinspach
Judith A. Scott
University of California, Santa Cruz

Jack L.Vevea
University of California, Merced

> Knowing a word means knowing not only the meaning, but knowing the contexts in which that word is used; it means knowing related words and ideas; it means knowing when and where to use the word. Therefore, to assess word knowledge, we need to consider the behaviors and actions that demonstrate what it means to know a word (Harmon, Hedrick, Soares, & Gress, 2007, p. 138).

Assessing word knowledge through examination of the full scope of behaviors and actions that demonstrate what it means to know a word, as suggested by Harmon et al. (2007), sounds reasonable, but it is ambitious. Most assessments range from tasks asking students to identify a synonym or a definition for a decontextualized word to vocabulary items embedded in tests of reading skills (Pearson, Hiebert, & Kamil, 2007; Read, 2000). Taking the latter approach, for example, the National Assessment of Educational Progress (NAEP) recently developed items that produce reliable and valid vocabulary measures by functioning "…both as a measure of passage comprehension and as a test of readers' specific knowledge of the word's meaning as intended by the passage author" (National Assessment Governing Board, 2008, p. iv). Built on years of research documenting the strong link between vocabulary and reading comprehension (Davis, 1942; Freebody & Anderson, 1983; Just & Carpenter, 1987; National Institute of Child Health and Human Development, 2000; Whipple, 1925), the new NAEP assessment represents a leap forward in the study of what it means to know a word through the vocabulary-reading nexus.

Encouraging as recent progress has been, there are still gaps in the field of vocabulary assessment. Writing, for example, is an action or behavior that can demonstrate what it means to know a word, yet student writing is rarely considered a medium for the study of vocabulary. When scholars examine the words in students' writing, typically their purpose has been to assess the quality of the writing rather than the child's word knowledge. Word choice and word usage are important in the writing process (Culhan, 2003; Fletcher, 1993; Hayes & Ahrens, 1988; Samway, 2006), so not surprisingly, writing assessments draw on a variety of measures of word choice, word sequence, content words, number of words, word diversity, syllable length, word length, and word frequency (Deno, Marston, & Mirkin, 1982; Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Grobe, 1981; Olinghouse & Leaird, 2009). Despite their dependency on lexical attributes, these measures, and assessments based on them, reveal relatively little about students' vocabulary knowledge in the expressive dimension.

This paper makes an argument for studying vocabulary knowledge and growth through students' writing. It reviews the literature on vocabulary assessment both to provide a rationale for venturing into this relatively unmapped territory and to examine possible assessment tools. It then

turns to one vocabulary assessment informed by this perspective—an analysis of the rare words in fourth-graders' personal narratives. A final section highlights the implications of this assessment for further study of elementary students' vocabulary in their written work.

LITERATURE REVIEW

Vocabulary Assessment and Word Learning in the Expressive Dimension

In their recent essay on vocabulary assessment, Pearson et al. (2007) criticize the vague use of the term "vocabulary" across distinct domains of words, across multiple text genres with varying vocabulary loads, and across the different dimensions of word learning needed to listen, speak, read, and write. The authors find that this undifferentiated notion of "vocabulary" leads to undifferentiated testing, which in turn fails to inform the educational community adequately about students' knowledge and skills, and about the relationship of prior knowledge, instruction, and skills such as comprehension to vocabulary learning. They recommend that the research agenda for the next decade pay greater attention to vocabulary distinctions, like the expressive versus receptive dimensions, and align assessments more specifically to reflect those distinctions (Pearson, et al., 2007, p. 294).

Current vocabulary assessments focus on the receptive dimensions of vocabulary (listening and reading), largely ignoring the expressive dimensions (speaking and writing) (Pearson, et al., 2007). Baumann, Kame'enui, and Ash note that "…expressive vocabulary requires a learner to know a word rather well before using it; not knowing a word is likely to result in the learner not using the word at all" (2003, p. 755). Thus testing the expressive dimension of vocabulary has the potential to inform researchers about how students master words by recalling and using them at will. In contrast, the word knowledge needed to complete receptive tasks varies in its demands; students can do some receptive vocabulary tasks without knowing the word at all by figuring out its meaning from the context (Baumann, et al., 2003). Consequently, current vocabulary assessments are limited in what they can say about the full trajectory of word learning.

Vocabulary scholars agree that "knowing a word" is not a binary variable; rather, word learning is a process. Several authors have suggested that word learning occurs along a continuum from "not knowing" to "knowing a word" (Beck, McCaslin, & McKeown, 1980; Carey, 1978; Clark, 1993; Kame'enui, Dixon, & Carnine, 1987; Stahl, 2003; Stahl & Nagy, 2006), and that correct word generation, production, or expression in speech and writing is evidence that a person is familiar with the meaning of the word.

> According to Beck et al. (1980) and Stahl, a word is really known when a child is able to retrieve that word from memory rapidly and use it correctly in an uninstructed context. This standard of knowing word meanings is akin to the standard we discussed earlier for expressive vocabulary (Baumann, et al., 2003, p. 756).

Receptive vocabulary assessments, especially in their current state, do not yield information about word learning near the end of this continuum.

With little information available about word learning and writing, those interested in expressive vocabulary assessment must look to studies of children's early word development as their point of departure. A review of that literature is beyond the scope of this paper, but it is important to include a few key ideas that may influence vocabulary use in writing. Bloom defines knowing the meaning of a word as having "1) a certain mental representation or concept 2) that is associated with a certain form" (2001, p. 17). So word learning can be viewed as a process of mapping a concept onto the right lexical form. The notion of "fast mapping" (Carey, 1978) explains a very efficient and rapid matching process that yields preliminary and tentative links between words and meanings in young children. Studies show that early word mapping is influenced by biological, psychological, and social factors, and that children tend to take new words as exemplars of taxonomies (Mayor & Plunkett, 2010), suggesting the importance of semantic relationships across words. Children's vocabularies grow incrementally; word knowledge tends to build gradually in complexity and completeness through additional exposures to a word and attempts to use it over time (Nagy & Scott, 2000). Word play and experimentation with words can demonstrate, and reinforce, the incremental nature of word learning (Bauman, Ware, & Edwards, 2007; Blachowicz & Fisher, 2004; Graves & Watts-Taffe, 2008; Scott, Skobel, & Wells, 2008).

Most vocabulary tests have been developed without reference to the nature of word learning (Scott, Hoover, Flinspach, & Vevea, 2008), yet Pearson et al. (2007) argue that incrementality should play a larger role in vocabulary assessment (p. 290). Scholars have studied incrementality through fairly discrete and decontextualized multiple-choice assessments that ask explicitly about, and test, levels of word knowledge (Dale, 1965; Paribakht & Wesche, 1997; Scott, et al., 2008) and that explore differing knowledge by varying the degree and type of contextualization across distractors and items on the same word (Stallman, Pearson, Nagy, Anderson, & Garcia, 1995). Although these instruments break new ground by incorporating the incremental aspect of word learning into assessments, they do not extend its study into the expressive dimension.

The literature assessing vocabulary instruction with measures derived from students' writing is piecemeal, but it can be divided into two approaches: 1) instruction explicitly designed to teach target words, and 2) instruction based on word-rich classrooms with word-conscious and word-learning strategies. In the first case, following instruction on target words, educators examine student writing for the frequency of the instructed words. Duin and Graves (1987), for instance, counted the number of target words in student essays both before and after instruction. Similarly, Papadopoulou (2007) compared a treatment group and a control group on the number of instructed words used in a story-writing exercise. Some teachers who teach target vocabulary in content areas like science and geography have examined vocabulary use in students' writing (notebooks, journals, and compositions) to evaluate knowledge and comprehension of the terms (Aschbacher & Alonzo, 2006; Gregg & Sekeres, 2006).

In a study of one word-rich classroom, Baumann and colleagues note that they integrated composition, and some assessment of the vocabulary embedded in writing, into their research plans to honor the teacher's belief in the importance of writing to literacy development (Baumann, Ware, & Edwards, 2007, p. 110). In this formative assessment, the pre- and post-test measures included parent and student questionnaires, a receptive assessment of listening vocabulary, the *Expressive Vocabulary Test* (Williams, 1997) to measure growth in spoken vocabulary, and, because of the

teacher's beliefs about writing, the overall word count plus the number of low-frequency words used in students' writing samples. Their measures documented vocabulary growth using a comprehensive and strategy-driven approach to word learning.

Tools for Assessing Vocabulary Knowledge in Student Writing

Thanks to specialists in reading, speech, and writing, many word measures for use on texts currently exist. In the readability studies of the past century, researchers often counted the number of syllables per 100 words to determine the difficulty of reading passages (Stahl & Nagy, 2006). Many assessments of the quality of writing include the total number of words in the composition (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Malecki & Jewell, 2003), and some employ measures based on the number of different words written (Gajar, 1989; Grobe, 1981; Morris & Crump, 1982; Olinghouse & Leaird, 2009) and on word size (Deno, et al., 1982; Gajar, 1989; Gansle, et al., 2002). As noted earlier, these measures have not been used to learn much about vocabulary or expressive vocabulary development.

The assessment goals of importance here include developing measures of students' growing familiarity with academic vocabulary in their writing and their ability to use such words expressively. Measures of curricular and instructional coverage, morphological complexity, word frequency, semantic networks, and conceptual difficulty would be promising measures for such assessments. Although researchers are working to capture some of these elements of word knowledge, measures of word frequency are readily available. Nagy and Hiebert (in press) caution that word frequency has, in the past, served as a proxy for word familiarity, conceptual difficulty, and other not-necessarily-related constructs because of its availability and ease of use. Hence word frequency may be a starting point for assessing vocabulary through writing, especially given the dearth of scholarly attention to the subject, but the field will be limited until a wider set of language and concept measures can be used.

Under the assumption that academic and literary words occur less frequently than other words in children's writing, word frequency measures can contribute to an understanding of expressive word learning. Word frequency has been measured in numerous ways, and the resources for constructing such measures are expanding rapidly. Early resources, generally lists of higher frequency words, include: Dale and O'Rourke's Living Word Vocabulary (Dale & O'Rourke, 1976); Finn's undistinguished word list (Finn, 1977); the Dale-Chall high-frequency list (Chall & Dale, 1995); and the Harris-Jacobson high-frequency list (Harrison, 1980). Several vocabulary researchers prefer the use of U values, a standardized measure of frequency per million words adjusted for variation in distribution, such as that found in the Educator's Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995). The U values allow researchers to compare the frequency of individual words or to set cut-off values that identify a set of frequent words or of rare words. Google and Wikipedia provide frequency information derived from online sources, and the Frequency Dictionary of American English is based on the prevalence of the words from the Corpus of Contemporary American English in several written genres and in speech (Davies & Gardner, 2010). Researchers who want to measure word frequency have options.

Unfortunately, there are problems associated with the use of word frequency. Stahl and Nagy (2006) note that word frequency measures distinguish neither idiomatic uses of words nor multiple

meanings, so that, for example, "bear" meaning an animal, to carry, or to endure are conflated into one frequency. Often proper nouns are not differentiated from common nouns. Most frequency lists do not provide frequency values to help users understand differences among the words on the lists (Stahl & Nagy, 2006, pp. 19-20), although newer resources generally provide a frequency score for each word. The use of word frequency measures requires attention to these shortcomings.

Children's writing holds clues about their comprehension of words. Teachers can examine errors or experimentation in written work as indications of students' knowledge, and researchers could follow suit if they had the appropriate tools. Identifying and coding students' experimentation with language is not a vocabulary measure; however, the information gleaned from such coding may inform measure development for later assessments and provide insights about expressive word learning. It is included as part of the following rare-words assessment for that reason.

RARE WORDS IN FOURTH-GRADERS' PERSONAL NARRATIVES: A VOCABULARY ASSESSMENT

Based on personal narratives collected from fourth graders, this section presents an assessment of the word knowledge and growth that the students demonstrated in their writing. The assessment used a measure of word frequency that identifies the rare words in the writing samples. The assessment also included an analysis of the students' experimentation with rare words. The results and discussion highlight the need for more studies to advance the field of expressive vocabulary development through writing.

Data Collection and Analysis

The research participants, data, and analyses are part of the VINE (Vocabulary Innovations in Education) Project, a three-year study of the importance of word consciousness in vocabulary development. The study helped teachers create fourth-grade classrooms that celebrated rich language and approached word learning as a generative process. The intervention sought to improve vocabulary instruction by changing teachers' metalinguistic and metacognitive knowledge about word learning and their own engagement with words. The premise of the study was that teachers' heightened word consciousness would then nurture word consciousness in their students, which would lead, in turn, to greater student acquisition and use of academic language.

Instead of teaching specific sets of words, VINE intervention teachers designed their instruction to develop general understandings about how words work in English and to develop student dispositions to pay greater attention to word use. Teachers in VINE intervention classrooms met together as a collaborative learning community throughout the school year and shared ideas for building enthusiasm for word learning in their classrooms. See Miller, Gage-Serio, & Scott (in press, this volume) for a description of one VINE intervention teacher's word-conscious instruction.

The 2007-2008 VINE participants were 16 classroom teachers—8 intervention teachers and 8 control teachers without access to the VINE intervention—and their 381 fourth-grade students. These classrooms were in five school districts located in metropolitan and smaller-city settings of California. Forty-five percent of the fourth-graders spoke just English at home. Ten percent had more than one home language but had always been fluent in English. Thirty-two percent of the

students were learning English, and ten percent had been reclassified as English proficient after having been English learners in school. The students came from backgrounds representing over 20 home languages, but 76% of the English learners were Spanish speakers. For this vocabulary assessment of the students' writing, the sample was delimited to the 300 fourth graders who took the narrative prompt in both the fall and spring and for whom the districts provided demographic and test-score data.

The vocabulary assessment draws on narrative writing collected during the 2007-2008 school year. The writing task asked students to compose a personal narrative in 30 minutes in response to the prompt: "Think about a memory that you would like to write about. It should be an event or experience you remember well and would like other people to read about." The research team administered the prompt in standardized fashion. The instructions suggested that students spend five minutes planning their piece, 20 minutes writing, and five minutes re-reading and revising, and the researchers reinforced these instructions with time cues and supplemental directions during the 30 minutes. Teachers generally remained in the room doing other work during the administration of the prompt, but some teachers helped answer questions, translated directions, or acted as scribes for students with restricting conditions. Spanish-speaking newcomers were allowed to write their narratives in Spanish, but those narratives were not used in the analyses. The prompts were administered in the fall of 2007 and again the following spring.

The rare-words measure described in this paper is a secondary analysis of the personal narrative data. Most of the 300 students wrote about a different memory in their fall and spring narratives. For example, in the fall one student wrote about her family's day at the beach, and in the spring her story was about a friend's birthday party at a pizza restaurant. Hence the academic and literary vocabulary that students used at each time period tended to be unrelated, limiting the possibilities for studying word learning. The VINE research team decided that a word-frequency measure might provide useful information about growth in academic vocabulary and the effectiveness of the study intervention despite the discontinuities created by topic.

The VINE research team developed a word-frequency measure based on the rare words identified by *The Educator's Word Frequency Guide* (Zeno, et al., 1995). The team set the cut-off for rare words at U value equal to 25 words per million in written text; words with U values of 25 or less were considered rare words. The team believed that this cut-off point would be sufficiently low to capture changes in word usage that demonstrated growth in word knowledge.

Teachers with experience deciphering developmental spelling helped clarify student writing and misspellings on the prompts. The prompts were converted into electronic documents and run through the word-frequency program to identify rare words. A researcher checked the computer results against the students' writing to delete unacceptable words (proper names, words with apostrophes, abbreviations, and nonwords). She also distinguished words like *Jade*, the gem, from *Jade*, a girl's name, and determined if *Sharks* referred to the marine animals or a sports team; the latter usages, respectively, were also dropped from the analysis. After these adjustments, the computer identified and summed the number of rare words in each piece of writing. The 300 students received a rare-words count for both their fall and spring narratives.

These data were analyzed using a three-level hierarchical linear model (HLM; Raudenbush, Bryk, Cheong, & Congdon, 2004). Fall to spring growth in rare words was estimated for each student at the lowest level of the HLM model. At the second level, differences in growth were investigated using English proficiency and home language as moderators. The third level recognized the nesting of children within classrooms and compared the intervention and control classrooms in the study.

To provide convergent evidence for the validity of the word-frequency results (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), the rare-word counts for the fall and the spring were correlated with three other measures from the students. The other measures were the students' performance on the fall and spring VINE Vocabulary Tests, their scores on the vocabulary subscale of the English Language Arts section of the California Standards Tests in 2007 (taken at the end of third grade) and in 2008 (taken at the end of fourth-grade), and their scale scores on the English Language Arts tests overall in 2007 (from third grade) and in 2008 (from fourth grade).

In order to maximize what might be learned about vocabulary development from the rare-words measure, a researcher examined and coded the identified rare words in their narrative contexts. She looked for examples of experimentation, defined as a marked or non-standard use of a rare word. Examples were coded as experimentation with semantic relationships, experimentation with lexical or morphological relationships, and/or experimentation with word choices.

RESULTS

On average, the 300 VINE fourth-graders in this analysis included 7.55 rare words in their fall narratives (s = 4.93) and 9.50 rare words in their spring narratives (s = 6.33). Their mean growth in rare words from fall to spring was 1.95 words, t_{598} = 5.429, p < .0001. As might be expected, students at different levels of English proficiency (English speakers, English learners, and fluent bilinguals) used dramatically different numbers of rare words, but there was no evidence of difference in their growth rates from fall to spring (χ^2 = 0.72 on 3 *df*, *p* = .870). Similarly, home language made no difference in rate of growth for these students (χ^2 = 1.44 on 2 *df*, *p* = .487). These results indicate that all fourth-graders' expressive knowledge of rare words grew fairly evenly over the school year.

The HLM analysis also compared students taught by VINE intervention and control teachers. Students in the intervention classrooms grew by 2.35 rare words, and students in the control classrooms grew by 1.57 words, but the difference was not significant (t_{597} = 1.144, *p* = .253). One of the assessment critiques made by Pearson et al. (2007) deals with the lack of alignment between different aspects of vocabulary and vocabulary assessments. Given that the rare-words measure was not particularly aligned with the VINE word-consciousness intervention, this finding, however disappointing, is not so surprising.

If the rare-words measure is indeed a measure of growth in the breadth of vocabulary knowledge of fourth graders, it should overlap with similar measures. Table 1 presents correlations between the students' rare-word counts in the fall and in the spring and their scores on three other vocabulary-related measures. The three other measures are the scores on the VINE Vocabulary Tests, the score on the vocabulary subscale of the English Language Arts test (California Standards Tests),

and the scale score on the English Language Arts test (California Standards Tests). Students wrote the narrative prompt, took the VINE Vocabulary Tests, and took the standardized English Language Arts tests within a span of a month and a half during the spring of 2008, but the spring 2007 standardized tests were taken at the end of the students' third-grade year, several months before the fall 2007 VINE assessments. The time lag may help account for two of the lower correlations in the fall column of the table. For the tests taken at approximately the same time, the correlations range in value from .318 to .444. This is convergent evidence that the rare-words counts validly measure something about vocabulary growth in fourth graders.

Table 1. Convergent Evidence for the Validity of the Rare-Words Measure

Other Vocabulary-Related Measures	Correlation with Rare-Words Measure Fall 2007	Correlation with Rare-Words Measure Spring 2008
VINE Vocabulary Test Fall 2007	$r = .318$ ($N=260, p < .0001$)	
VINE Vocabulary Test Spring 2008		$r = .444$ ($N=260, p < .0001$)
Vocabulary Subscale, English Language Arts, California Standards Tests Spring 2007 (third grade)	$r = .255$ ($N=273. p < .0001$)	
Vocabulary Subscale, English Language Arts, California Standards Tests Spring 2008		$r = .338$ ($N=291. p < .0001$)
English Language Arts Scale Score, California Standards Tests Spring 2007 (third grade)	$r = .332$ ($N=269, p < .0001$)	
English Language Arts Scale Score, California Standards Tests Spring 2008		$r = .350$ ($N=293, p < .0001$)

The analysis of word experimentation in the narrative writing built on the rare-words measure. Only 35 of the 5115 rare words in this assessment were coded as clear examples of student experimentation with words. Each example was identified as a marked or non-standard use of a rare word in the narrative. These varied from "bestest" as in "…the first bestest person…," to "zephyr" in "…speaking of the whistle, Brenda blew the whistle harder than a zephyr of wind,…". Omitted from this analysis were experimentation at the phrase or multiple-word level ("in the crack of the sunlight morning" for "at the crack of dawn"), probable misspellings ("bread" instead of "breed" of hamster), slight orthographic irregularities (repeated references to "dye eggs" rather than "dyed eggs"), inflectional errors, and experimentation with words that were not identified as rare.

The examples were coded as experimentation with semantic relationships, experimentation with lexical or morphological relationships, and/or experimentation with word choice. In general, examples in the semantic category showed that the student both understood and misunderstood something about the meaning of the word—perhaps knowing only a partial meaning or another meaning not appropriate in the narrative context he or she had constructed. After describing a happy vacation disrupted only by a painful bee sting that caused her to faint, a girl concluded, "It turned out to be a very fun, hurtful vacation." In his story about the

flag football season, one student wrote "versed" for "competed against." "I versed the Cheifs [*sic*] and Buccaneers that day. When I versed the Cheifs [*sic*], it was a close game!" Four of the semantic experiments occurred in written dialogue as proposed substitutes for "said." One child wrote that his mother "…pronounced 'well hurry up do you want to get a good parking spot [*sic*]' I pronounced back 'yes!' " Another described his mother's reason for allowing his sister to go on a ride at an amusement park, but not him: " 'She is big and your [*sic*] small,' my mom recommended." Thirteen of the 35 examples of rare-word experimentation were coded as exploring semantic relationships.

Eighteen of the examples reflect experimenting with lexical or morphological creations or mismatches, including the most common one, "humongous." A swimmer consistently used the word "components" for her opponents or competitors at a swim meet. One girl wrote "visibly" for "visibility" in her account of snorkeling in very clear water. Another described something as "near an algae infected river" rather than near an algae-infested one. A boy lunged for his little brother when the brother stole some of his Halloween candy, but he wrote, "I plunged at him with flames in my eyes…".

The third coded category, experimenting with word choice, consists of embellishments that fail to fit seamlessly into the narrative account. Some children, like the girl who wrote about the whistle being blown harder than "…a zephyr of wind," are playing with words. Others, though, seem to be adding modifiers solely to complicate their sentences. One boy wrote, "Finally the enthralling game started" without ever saying more to show that the game was indeed enthralling. Another described a long drive to a campground with his family and added, " 'We're here' I randomly hollered," where "randomly" leaves the reader guessing about the author's meaning. Although few, these examples of rare-words experimentation offer evidence important to the study of word learning.

DISCUSSION

This preliminary assessment of the vocabulary knowledge in fourth graders' narratives has both strengths and weaknesses. Its primary strength rests with its success in demonstrating that student writing is a promising vehicle for learning about vocabulary knowledge and development. The results add to what is known about the acquisition of expressive vocabulary—that both overall and across all subgroups, students' use of rare words increased in their narrative writing over the course of fourth grade. They also indicate that word-frequency measures in general, and the counting of rare words in particular, can be useful and valid tools for assessing the breadth of vocabulary that students employ in their writing. The analysis of rare-word experimentation provides evidence that some processes that characterize earlier word learning may also influence the words students write. The not-quite-right use of rare words suggests that the students' morphological, lexical, and semantic understandings are growing incrementally, with partial knowledge being mapped onto lexical forms. Thus the results from the assessment of rare words tantalizingly reveal the promise of putting greater focus on writing and writing measures in vocabulary assessment.

The convergent evidence for the rare-words measure presented in Table 1 is reasonable confirmation of the measure's validity. Vocabulary learning and vocabulary assessments are both multifaceted, and the results of any one vocabulary test should correlate with the results of another

to the extent that the two measure the same thing. The rare-words measure is an indicator of the upper range of vocabulary breadth in writing. In contrast, two of the comparison tests in Table 1, the VINE Vocabulary Tests and the Vocabulary Subscale of the English Language Arts test (California Standards Tests), are based on receptive word knowledge across a broad continuum of fourth-grade words. The VINE Vocabulary Tests assess students' incremental knowledge of words (Scott, et al., 2008), and the Vocabulary Subscale focuses on students' understanding of concepts through knowledge of roots, affixes, derivations, synonyms, antonyms, and idioms (California Department of Education, 2002). In fourth grade, the English Language Arts scale score (California Standards Tests) includes vocabulary and a writing test, but it also covers an array of other literacy knowledge and skills (California Department of Education, 2002). Thus the overlap between the rare-words measure and any of the other measures in the table is limited. The correlations are sufficiently consistent and strong to provide convergent evidence for the rare-words measure; they are also sufficiently distinct to show that the measure adds new dimensions not tested by the other assessments.

The rare-words assessment also had flaws. Like other word-frequency indicators, the rare-words measure was insensitive to words with multiple meanings, which introduced error into the identification of rare words in the students' narratives. In addition, the rare-words measure relied on the word frequencies assigned by *The Educator's Word Frequency Guide* (Zeno, et al., 1995), and some words that are rare in the *Guide*, such as "pizza" and "mall," are not really rare for fourth graders. The U-value cut-off level distinguishing rare words from others was based on the authors' prediction about rare words in fourth-grade writing; it would be worthwhile to reanalyze the data using different cut-off levels to evaluate the accuracy of that prediction.

A final source of error came from the writing task itself. The assessment was a secondary analysis of VINE data collected for another purpose, and the writing prompt usually elicited completely different personal narratives in the fall and in the spring. Whereas the subject of a story does not necessarily influence the use of rare words, narratives about a horseback riding lesson or a field trip to a science museum tended to include more technical terms flagged as rare words than a story about a day spent shopping. In the future, researchers gathering data specifically for a growth analysis of vocabulary knowledge and use in student writing should consider developing tasks or prompts that encourage students to write about the same subject at both testing periods.

The VINE research team was hoping that the assessment would link the study's intervention to increasing use of rare words, but the greater rate of growth in the intervention classrooms was not significant. The generative word-learning strategies and word-rich activities characteristic of VINE intervention classrooms help students learn more words—both rare and not rare. A complete assessment of VINE vocabulary instruction would require multiple measures of word consciousness and vocabulary learning from multiple sources. In a VINE word-conscious classroom, increasing students' use of rare words is just one thread in the fabric of word learning.

CONCLUSIONS: IMPLICATIONS FOR THE FUTURE OF VOCABULARY ASSESSMENT

The results of the assessment support two conclusions about vocabulary assessment. First, the vocabulary knowledge encoded in student writing has been largely ignored by vocabulary researchers, a reality that has curtailed knowledge of vocabulary development and possibly of effective vocabulary instruction. Reading and writing are taught to most elementary students as complementary, mutually reinforcing skills. Both skills contribute to, and benefit from, word learning. Similarly, both present opportunities for vocabulary assessment that merit attention. The study and assessment of vocabulary in student writing is overdue.

Second, the measurement of written vocabulary is territory staked out and developed by several academic fields (writing, reading, and second-language acquisition to name a few) but not claimed by vocabulary researchers, who tend to focus on reading comprehension instead. Word-frequency measures, like the rare-words count in this paper, capture something about the breadth of students' expressive vocabulary knowledge, but other expressive measures based on words (word diversity, word length, number of syllables, total words, etc.) say remarkably little about an author's word knowledge. To learn more about vocabulary development through writing, scholars should lay aside the indicators of other fields and develop new measures to help them understand the instructional, lexical, morphological, and semantic attributes of the words in children's writing.

Pearson et al. wrote the following characterization of the field of vocabulary assessment today:

> …vocabulary assessment is grossly undernourished, both in its theoretical and practical aspects…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 comprehension (2007, p. 282).

Given this state of affairs, vocabulary researchers interested in assessment have much to do. The assessment presented in this paper suggests a promising new drection for this work.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, D.C.: Author.

Aschbacher, P., & Alonzo, A. (2006). Examining the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment, 11*(3 & 4), 179-203.

Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 752-785). Mahwah, NJ: Lawrence Erlbaum Associates.

Baumann, J. F., Ware, D., & Edwards, E. C. (2007). "Bumping into spicy, tasty words that catch your tongue": A formative experiment on vocabulary instruction. *The Reading Teacher, 61*(2), 108-122.

Beck, I. L., McCaslin, E. S, & McKeown, M. G. (1980). *The rationale and design of a program to teach vocabulary to fourth-grade students* (LRDC Publication 1980/25). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Blachowicz, C. L. Z., & Fisher, P. J. (2004). Keep the "fun" in fundamental: Encouraging word awareness and incidental word learning in the classroom through word play. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 218-237). New York, NY: Guilford Press.

Bloom, P. (2001). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

California Department of Education. (2002). *STAR CST blueprints, English language arts blueprints grades 2 to 11*. Retrieved from http://www.cde.ca.gov/ta/tg/sr/blueprints.asp

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, MA: MIT Press.

Chall, J., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge, UK: Cambridge University Press.

Culhan, R. (2003). *6+1 Traits of writing: The complete guide grades 3 and up*. Northwest Regional Educational Laboratory. New York, NY: Scholastic.

Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English, 42*, 82-88.

Dale, E., & O'Rourke, J. (1976) *The living word vocabulary*. Elgin, IL: Field Enterprises Educational Corporation.

Davis, F. B. (1942). Two new measures of reading ability. *Journal of Educational Psychology, 33*, 365-372.

Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: Word sketches, collocates, and thematic lists*. New York, NY: Routledge.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48*, 368–371.

Duin, A. H., & Graves, M. F. (1987). Intensive vocabulary instruction as a prewriting technique. *Reading Research Quarterly, 22*(3), 311-330.

Finn, P. J. (1977) Computer-aided description of mature word choices in writing. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 69–89). Buffalo, NY: State University of New York at Buffalo.

Fletcher, R. (1993). *What a writer needs*. Portsmouth, NH: Heinemann.

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly, 18*(3), 277-294.

Gajar, A. H. (1989). A computer analysis of written language variables and a comparison of compositions written by university students with and without learning disabilities. *Journal of Learning Disabilities, 22*, 125–130.

Gansle, K. A., Noell, G. H., VanDerHeyden, A., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternative measures for curriculum-based measurement in writing. *School Psychology Review, 31,* 477–497.

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review, 35*(4), 435-450.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*, 170–182.

Graves, M. F., & Watts-Taffe, S. (2008). For the love of words: Fostering word consciousness in young readers. *The Reading Teacher, 62*(3), 185-193.

Gregg, M., Sr. fcJ., & Sekeres, D. C. (2006). My word! Vocabulary and geography learning. *Journal of Geography, 105*(2), 53-58.

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 15*, 75–85.

Harrison, C. (1980). *Readability in the classroom*. Cambridge, MA: Cambridge University Press.

Harmon, J. M., Hedrick, W. B., Soares, L., & Gress, M. (2007). Assessing vocabulary: Examining knowledge about words and about word learning. In J. R. Paratore & R. L. McCorack (Eds.), *Classroom literacy assessment: Making sense of what students know and do* (pp. 135-153). New York, NY: Guilford Press.

Hayes, D. P., & Ahrens, M. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27,* 572-585.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston, MA: Allyn & Bacon.

Kame'enui, E. J., Dixon, D. W., & Carnine, R. C. (1987). Issues in the design of vocabulary instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based writing probes. *Psychology in the Schools, 40*, 379–390.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117*(1), 1-31.

Miller, T. F., Gage-Serio, O., & Scott, J. A. (in press). Word consciousness in practice: Illustrations from a fourth-grade teacher's classroom. *59th Yearbook of the National Reading Conference.* Oak Creek, WI: National Reading Conference.

Morris, N. T., & Crump, W. D. (1982). Syntactic and vocabulary development in the written language of learning disabled and non-learning disabled students at four age levels. *Learning Disability Quarterly, 5,* 163–172.

Nagy, W., & Hiebert, E. H. (in press). Toward a theory of word selection. In M. Kamil, P. D. Pearson, E. Moje, & P. Afflerbach (Eds.), *Handbook of Reading Research* (Vol. 4). New York, NY: Routledge.

Nagy, W., & Scott, J. (2000). Vocabulary processes. In R. Barr, P. Mosenthal, P. D. Pearson, & M. Kamil (Eds.), *Handbook of reading research* (Vol. 3, pp. 269-284). Hillsdale, NJ: Lawrence Erlbaum Associates.

National Assessment Governing Board (2008). *Reading framework for the 2009 National Assessment of Educational Progress.* U.S. Department of Education. Washington, D.C.: U.S. Government Printing Office.

National Institute of Child Health and Human Development, NIH, DHHS. (2000). *Report of the National Reading Panel: Teaching children to read: Reports of the subgroups (00-4754).* Washington, D.C.: U.S. Government Printing Office.

Olinghouse, N. G., & Leaird, J. L. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*(5), 545-565.

Papadopoulou, E. (2007). *The impact of vocabulary instruction on the vocabulary knowledge and writing performance of third grade students.* (Doctoral dissertation, University of Maryland). Retrieved from http://www.lib.umd.edu/drum/handle/1903/7649

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 174-200). Cambridge, UK: Cambridge University Press.

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.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Chicago, IL: Scientific Software International.

Read, J. (2000). *Assessing vocabulary.* Cambridge, England: Cambridge University Press.

Samway, K. D. (2006). *When English language learners write: Connecting research to practice, K-8.* Portsmouth, NH: Heinemann.

Scott, J. A., Hoover, M., Flinspach, S. L., & Vevea, J. L. (2008). A multiple-level vocabulary assessment tool: Measuring word knowledge based on grade-level materials. In Y. Kim, V. J. Risko, D. L. Compton, D. K. Dickinson, M. K. Hundley, R. T. Jiménez, K. M. Leander, & D. W. Rowe (Eds.), *57th Yearbook of the National Reading Conference* (pp. 325-340). Oak Creek, WI: National Reading Conference.

Scott, J. A., Skobel, B. J., & Wells, J. (2008). *The word conscious classroom: Building the vocabulary readers and writers need.* New York, NY: Scholastic.

Stahl, S. (2003). How words are learned incrementally over multiple exposures. *American Educator, 27*(1), 18-19.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings.* Mahwah, NJ: Lawrence Erlbaum Associates.

Stallman, A. C., Pearson, P. D., Nagy, W. E., Anderson, R. C., & Garcia, G. E. (1995). *Alternative approaches to vocabulary assessment* (Report No. 607). Urbana-Champaign, IL: Center for the Study of Reading, University of Illinois.

Whipple, G. (Ed.). (1925). *The 24th Yearbook of the National Society for the Study of Education: Report of the National Committee on Reading.* Bloomington, IL: Public School Publishing.

Williams, K.T. (1997). *Expressive vocabulary test.* Circle Pines, MN: American Guidance Service.

Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science Associates.

AUTHORS' NOTE

Susan Leigh Flinspach, Education Department, University of California, Santa Cruz; Judith A. Scott, Education Department, University of California, Santa Cruz; Jack L. Vevea, School of Social Sciences, Humanities and Arts, University of California, Merced.

Correspondence should be addressed to Susan Leigh Flinspach, Education Department, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064. E-mail: flinspac@ucsc.edu