Comprehension is the overall goal of reading and as such, researchers and practitioners wrestle with how to facilitate students’ routes to understanding what they have read. This is not an easy task given the complexities of reading where students are required to process a symbol system rapidly enough to identify and hold written words in memory as they glean meaning from connected text (Smith, 2004). Rasinski (2003) and Samuels (2002) similarly conveyed that reading comprehension requires fluency, the rate and degree of expression, to overcome an information “bottleneck” caused by the interference of a lack of word recognition and word meaning. Slower reading taxes the physiological limitations of short-term memory and this detracts from comprehension. For years, researchers have documented the correlational relationship of vocabulary knowledge to reading comprehension and to overall academic success (Baumann, Kame’enui, & Ash, 2003; Becker, 1977; Davis, 1942). Early word exposure varies across socioeconomic levels. Hart and Risley (1995) and Snow (1991) found that language development prior to entering school either empowers children with language or sends them off into the world with a word deficit that is hard to overcome.

Vocabulary Learning and Instruction: Contexts and Materials

When they enter this world, infants immediately begin to learn the words of their language through exposure. They initially take in entire streams of sound and later begin to distinguish words and later their individual phonemes (Lyon, 1995). Around 5 months of age, infants begin to communicate through babbling to those around them. Risley and Hart...
ent. According to Pearson, Hiebert, and Kamil (2007) “the multiple contexts in children’s home environments and stark contrast to earlier word learning that arose naturally in Systematic vocabulary instruction of targeted words stands in purpose and the environmental contexts that supported earlier is the next step for word learning. In school, an immediate (Wren, 2006; Beck & McKeown, 1991; Chall, 1996).

This process, that aligns words with their concrete referents, word “milk” through aural, visual, and tactile sensory input. She gains a deep and lasting knowledge of the meaning of the word when caregivers give milk to the infant or young child, he or she gains a deep and lasting knowledge of the meaning of the word “milk” through aural, visual, and tactile sensory input. This process, that aligns words with their concrete referents, is repeated hundreds of times and results in the learning of 2,500 to 6,000 words before children even enter school (Wren, 2006; Beck & McKeown, 1991; Chall, 1996).

Vocabulary instruction in the formal institution of school is the next step for word learning. In school, an immediate purpose and the environmental contexts that supported earlier word learning are diminished or not present (Pinker, 1994). Systematic vocabulary instruction of targeted words stands in stark contrast to earlier word learning that arose naturally in multiple contexts in children’s home environments and was centered on their needs and interests. In Holmes and Canterbury’s (2012) review of core reading teacher’s manuals, they found that the recommended strategies for involving students in multiple contexts almost always were through multiple written contexts. Recommendations to continue the earlier concrete materials-based contexts to teach vocabulary were almost nonexistent. School-age children, who initially learned many of their words through interactions with concrete materials, now use flat two-dimensional materials, written words, and pictures to learn word meanings. We must clarify, however, that our search consisted primarily of scanning the research literature base addressing vocabulary acquisition.

Looking through the lens of cognitive psychology, the proximal and distal relationship between the materials and the students may play a vital role in the quality of the contexts for word learning (Dunn, Beach, & Kontos, 1994). Proximal materials are those in, or closest to, the students and their environment thus enabling them to engage with actual representations of words rather than abstract definitions and remote descriptions. Their authenticity makes it easier for students to move beyond simple word definitions to a deeper level where they can examine, analyze, and categorize attributes of the word’s meaning. Because context is critical to word meaning, the authors sought to explore the use of proximal contexts in K-3 classrooms where students can interact with materials to determine nuances of meaning through guided or discovery exploration of the word’s referent. According to Pearson, Hiebert, and Kamil (2007) “the more contexts in which a word is encountered, the greater the likelihood that its meaning will be acquired” (p. 290).

The need to document the materials teachers use to teach the meaning of words arose when Holmes noted through classroom observations that student teachers rarely used concrete materials to teach new vocabulary. A commonly taught lesson on the stages of the butterfly often contained only oral and written descriptions of words such as pupa and chrysalis. With just a little imagination and effort, these lessons could have included authentic materials, or at least their representations such as models, that would have given students a more in-depth and memorable experience with the words they were learning. To determine whether teachers in K-3 classrooms use authentic materials to teach vocabulary, the authors developed the Vocabulary Observation Instrument (VOI; see Figure 1.). This article reports on the results of a pilot study that confirmed the reliability and validity of VOI. From the data collected through the use of VOI, the researchers are looking to develop a theory of practice to lend specificity to the term multiple contexts by labeling the contexts according to their proximal and distal relationships to students. To test this instrument, 15 observers piloted it during three different vocabulary lessons (45 trials) to determine the ease of use and whether the categories were sufficient for gathering the intended data on material use. Observers were required to list the materials and how they were used in the lesson. Since the pilot study, this instrument has been used to collect data in a descriptive study in more than 500 K-3 classrooms. (See a description of these online lessons and the link to them in the following section.)

**Reliability and Validity of the VOI**

Reliability and validity were assessed in the following manner: (a) training of raters via video analyses of three vocabulary lessons, (b) discussion and resolution of differences in ratings, (c) multidimensional scaling (MDS) of rater scores from Grade K-3 classrooms (n = 30).

**Reliability via Observer Video Training With Discussion**

The pilot lesson. Observer training utilized three video cases of vocabulary instruction to identify events in teaching episodes. The vocabulary lessons were conducted for three different grade levels and are accessible online at http://www.scoe.org/pub/htdocs/archer-videos.html. The three lessons were Kindergarten, 1st grade, and 2nd grade hyperlinks. Two methods were used to increase reliability of ratings: (a) insight-oriented analysis enabling accurate and descriptive entries on methods and (b) coding-oriented analysis to generate categories of materials and types by teachers and students (Castorr et al., 1990).

Collaborative viewing of three vocabulary instruction videos and follow-up discussion enabled the trainees to
become confidently grounded in their assertions gleaned from the evidence. We categorized all of the instruments by video episode (three different grade level lessons) and then checked to see whether the participants recorded the same information from each of the videotaped lessons. In addition to discussing rater responses post observation of the three videos, the authors engaged the raters in discussion of different ratings of the same video episode to reconcile differences. We also further analyzed the ratings of the three video episodes to determine the interrater reliability of the instrument. Many have argued that reliability of rater judgments increases via discussion of rating differences (Mott, Etsler, & Drumgold, 2003) and also functions to increase rater competency for understanding the construct domain being observed.

For the first trial, we noted some discrepancies for two categories of materials: drama and word lists. For the purpose of this study, we defined drama as actions, facial expressions, and variations of voice that conveyed the meaning of the word. Raters were confusing the video episode component when the teacher provided a definition using affective speech characteristics such as exaggerated voice tone, facial expressions, and so on. Thus, the author-researchers provided a rationale, via discussion, for separating out teacher sharing of a word list with authentic dramatic activity. For an accurate documentation on the use of visuals, we requested that participants include a brief description of the word lists on overhead transparencies and charts so we would not miss any pictures and graphic organizers that accompanied the words. We also recommended a brief description of book use as well (certainly books may contain all kinds of visuals but we wanted to know when a book was used versus solely a chart or word list). In conclusion, all participants were in agreement on how to use the observation instrument.

Furthermore, an analysis of the instruments used during 30 pilot lessons showed congruity in the documentation of data.

**VOI Validity**

The validity of an assessment is the degree to which it assesses what it purports to assess. The VOI seeks to enable observers to accurately record material use and context. In addition to the face validity of the instrument established through the genesis of its creation derived from the research literature, both experts in K-adult literacy instruction and research, construct validity was technically evaluated to support the instruments use in this study. According to Messick (1996), construct validity is a key component in the hierarchy of validity and is the degree to which an assessment measures an underlying construct, in this case vocabulary learning. To determine construct validity, MDS of vocabulary instruction observation ratings (n = 45) were applied to visually portray similarities and dissimilarities in the data set.

MDS detects meaningful underlying dimensions allowing the current researchers to explain observed similarities or dissimilarities between observation items. Sireci and Geisinger (1995) summarized the benefits for using MDS in establishing the construct validity of an assessment. The rationale being that an ordination or the tendency of like items to be similar in proximity and unlike items to be further away from each other provides evidence that the instrument is indeed discriminating material-type between defined construct dimensions, in this case “control” and “number of senses.” In other words, the dissimilarity of ratings along categories of “senses” and “control” lead us to believe that when “teacher control” is high and “number of senses” used is low, this indicates that the VOI is yielding important
information about the tendencies occurring in a lesson. In this case, “active” students (student control) utilized more “senses” during the lesson. This merely indicates that the VOI data enable us to distinguish patterns and relationships occurring in the lessons. Consider the alternative: An observation instrument that yields all “similar” data without dissimilarity among observations along predefined categories might not be sensitive to the vocabulary lesson. Please note that MDS requires the researcher to define the dimensions and thus is only an initial step in establishing validity (Sireci & Geisinger, 1995).

Thus, the vocabulary observations might reveal evidence for material-use characteristics along “student control” if those item responses cluster in proximity and “teacher control” if item responses cluster with each other but distant to “student control.” Similarly, “high number of five senses” (auditory, visual, kinesthetic, olfactory, and taste) and “low number of senses” indicating that observations consistently identified material-type along the dimensions. “Book” clustered in “teacher control” and “high number of senses” and “drama” clustered in “student control” and the midpoint between low and high “number of senses.” The dissimilarity between “drama” and the other materials provides evidence that the VOI observer ratings are indeed discriminating material-type and placing that material-type in a context of “number of senses” used and “student control.” “Drama” contained the highest number of senses used and thus the researchers expected to see dissimilarity with “drama” and the other materials (and similarity between “wordlist,” “poster/chart,” and “projector”). “Book” material-type clustered in “high number of senses” but it is important to note that this is due to the fact that this material-type was only identified by a few observers and thus contains an inflated “number of senses” value. It is safe to identify “book” as an outlier. Thus, “book” should not necessarily be viewed as being associated with a high number of senses used given the low number of “book” data points.

Conclusion

To assess the degree to which observers would be able to identify materials teachers use to teach vocabulary, it was necessary to conduct a pilot study of the vocabulary instrument VOI. Results showed that the instrument measured what it was intended to do. In other words, the data similarities and dissimilarities revealed by the MDS analyses provided information about how the materials distinguished themselves along dimensions (the categories defined by the researchers based on MDS coefficients revealed in the scatterplot) generated by observations. Observers reported that the instrument was simple and easy to use for collecting data. Participants in the pilot study were less confident during the first classroom observation, but gained the necessary confidence and expertise to rate the two subsequent classroom observations accurately. The authors found that the VOI provided the appropriate categories and space for the transcription of relevant data. In addition to documenting the materials teachers used, the instrument enabled observers to document the context teachers used with the materials. In the analysis of the data gathered during this study, the authors were able to compare the incidences of word learning in written contexts versus learning through environmental contexts where meaning is supported by visuals, representations, artifacts, and real-world materials.

There is evidence for more holistic incorporation of sensory-rich experiences to promote vocabulary development in early childhood classes (Moyeda, Gomez, & Flores, 2006). Understanding teacher material selection and the number of senses incorporated with specific materials will enable the researchers to develop a theory of vocabulary learning for early childhood educational settings. Accurately documenting the
use of materials in K-3 classrooms will provide the researchers with essential information for understanding effective materials-rich methodologies early childhood educators can use to promote vocabulary learning given the ever-present constraints of time and resources. Through VOI, observers were able to document whether the parents’ and caregivers’ incorporation of sensory-rich experiences into vocabulary lessons continued in school, specifically Grades K-3. By expanding the word context to include concrete materials, the authors have added specificity to contextual learning.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research and/or authorship of this article.

Note
1. Multidimensional scaling (MDS) is a method that represents information similarity (or dissimilarity) among pairs of objects as distances between points of a low-dimensional multidimensional space. The data can be any observed and recorded behavior, measured in any number of ways, of interest to the educator, in this case vocabulary lessons. MDS enables the researcher to quickly convey structure, as a visual portrayal, to colleagues across disciplines and professional backgrounds. MDS is appropriate to use for data mining, in this case, what is going on in the pilot lessons and how are raters responding to the phenomena they are observing? For more on MDS, see Forest Young: http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html.

References
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 on research on teaching the English language arts (2nd ed., pp. 752-785). Mahwah, NJ: Erlbaum.
Beck, I. L., & McKeown, M. G. (1991). Social studies texts are hard to understand. Mediating some of the difficulties. Language Arts, 68, 482-490.
Becker, W. C. (1977). Teaching reading and language to the disadvantaged—What we have learned from field research. Harvard Educational Review, 47, 518-543.
Caldwell, K. A. (2005). Non-participant observation: Using video tapes to collect data in nursing research. Nurse Researcher, 13, 42-54.
Castor, A. H., Thompson, K. O., Ryan, J. W., Phillips, C. Y., Prescott, P. A., & Soeken, K. L. (1990). The process of rater training for observational instruments: Implications for inter-rater reliability. Research in Nursing & Health, 13, 311-318.
Chall, J. S. (1996). Learning to read: The great debate (1967). New York, NY: McGraw-Hill.
Dana, C., & Rodriguez, M. (1992). TOAST: A system to study vocabulary. Reading Research Instruction, 31, 78-84.
Davis, F. B. (1942). Two new measures of reading ability. Journal of Educational Psychology, 33, 365-372.
Dunn, L., Beach, S. A., & Kontos, S. (1994). Quality of the literacy environment in day care and children’s development. Journal of Research in Childhood Education, 9, 24-34. doi: 10.1080/02568549409594950
Hart, B., & Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD: Paul H. Brookes.
Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. American Educator, 22, 4-9.
Holmes, K., & Canterbury, B. (2012). An analysis of materials recommended for instruction in core reading textbooks. Unpublished manuscript.
Lyon, R. (1995). Toward a definition of dyslexia. Annals of Dyslexia, 45, 3-27.
Messick, S. (1996). Validity and washback in language testing. Language Testing, 13, 241-256.
Mott, M. S., Etsler, C., & Drumgold, D. (2003). Applying an analytic writing rubric to children’s hypermedia narratives. Early Childhood Research & Practice, 5(1). Retrieved from http://ecrp.uiuc.edu/v5n1/mott.html
Moyeda, X. G., Gomez, I. C., & Flores, T. P. (2006). Implementing a musical program to promote preschool children’s vocabulary development. Early Childhood Research & Practice, 8(1). Retrieved from http://ecrp.uiuc.edu/v8n1/galicia.html
National Reading Panel. (2000). Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child Health and Human Development.
Newport, E. L., Gleitman, H., & Gleitman, L. A. (1977). Mother, I’d rather do it myself: Some effects and non-effects of maternal speech style. In Snow and Ferguson (Eds.), Talking to children: Language input and acquisition (pp. 109-149). Cambridge, UK: Cambridge University Press.
Pearson, D. P., Hiebert, E. H., & Kamil, M. C. (2007). Vocabulary assessment: What we know and what we need to learn. Reading Research Quarterly, 42, 282-296.
Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? Lingua, 92, 377-410. (Reprinted in L. Gleitman and B. Landau (Eds.), The acquisition of the lexicon. Cambridge, MA: MIT Press)
Rasinski, T. V. (2003). The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension. New York, NY: Scholastic.
Risley, T. R., & Hart, B. (2006). Promoting early language development. In N. F. Watt, C. Ayoub, R. H., Bradley, J. E. Puma, & W. A. LeBoeuf (Eds.), The crisis in youth mental health: Critical issues and effective programs. Volume 4: Early intervention programs and policies (pp. 83-88). Westport, CT: Praeger.
Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 166-183). Newark, DE: International Reading Association.

Sireci, S. G., & Geisinger K. F. (1995). Using subject matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement, 19*, 241-255.

Smith, F. (2004). *Understanding reading: A psycholinguistic analysis of reading and learning to read* (6th ed.). Hillsdale, NJ: Lawrence Erlbaum.

Snow, C. E. (1991). Diverse conversational contexts for the acquisition of various language skills. In J. Miller (Ed.), *Research on child language disorders: A decade of progress* (pp. 105-124). Austin, TX: Pro-Ed.

Wren, S. (2006). *The reading assessment database for Grades K-2*. Southwest Educational Development Laboratory, Austin, TX. Retrieved from http://www.sedl.org/reading/rad

**Bios**

Michael S. Mott, PhD, is assistant professor, Department of Curriculum & Instruction at The University of Mississippi and Faculty, Center for Excellence in Literacy Instruction.

Kerry Holmes, EdD, is associate professor, Department of Curriculum & Instruction at The University of Mississippi.