Article

Journal of Early Childhood Literacy
0(0) 1–23
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1468798413501184
ecl.sagepub.com

Assessing vocabulary learning in early childhood

Jessica L Hoffman
Miami University, USA

William H Teale
University of Illinois at Chicago, USA

Kathleen A Paciga
Columbia College Chicago, USA

Abstract

There is widespread agreement with in the field of early childhood education that vocabulary is important to literacy achievement and that reading aloud can support vocabulary growth. However, there are unexplored and significant problems with the ways we assess young children's vocabulary learning from read-alouds. This paper critically reviews the forms of vocabulary assessment commonly used with young children, examining the benefits and drawbacks of each. This review found: (a) general vocabulary measures are not practical, meaningful measures for vocabulary learning of specific words from books read aloud, (b) researcher-developed measures for specific words from books read aloud that mimic normed general vocabulary measures include serious threats to validity and reliability, and (c) other forms of measurement, such as soliciting definitions from children, are difficult to score reliably. This critical review of existing vocabulary assessments of word learning from read-alouds concludes that researchers and practitioners should carefully consider their needs for assessment data so as to choose, design and balance the uses of assessments to meet their needs for meaningful, reliable data.

Keywords

Early childhood, language development, assessment, storybook reading, young children

Corresponding author:
Jessica L Hoffman, Miami University, 210 E. Spring St., 401 McGuffey Hall, 210 E. Spring St. Oxford, OH 45208, USA.
Email: hoffmajl@miamiOH.edu

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Researchers, practitioners and policymakers agree that vocabulary is a critical factor in literacy development, both during early childhood (National Early Literacy Panel (NELP, 2008) and throughout schooling (Baumann, 2008). Vocabulary is central to oral language development, but it also plays an integral role in developing domain-specific knowledge and in supporting reading comprehension (Beck and McKeown, 2007; Cunningham and Stanovich, 1997; Scarborough, 2001; Snow et al., 1998; Stahl and Nagy, 2006). The issue of vocabulary learning and instruction has become especially crucial during the preschool years. Spurred especially by the research of Hart and Risley (1995, 1999), whose studies showed dramatic differences in vocabulary knowledge between children from affluent and low-income backgrounds by the age of three, early childhood educators and programmes now strongly advocate putting increased emphasis on vocabulary learning in early childhood education. Vocabulary was (a) specified by Early Reading First (ERF) as an area of focus (''oral language''; ERF was a US federal grant programme from 2002 to 2009 that supported ''the development of early childhood centers of excellence'' with a focus on language and literacy development; U.S. Department of Education, 2008, p.4), (b) an area deemed by the What Works Clearinghouse (WWC) as important to examine for improvement outcomes (2012b), and (c) virtually universally discussed as an ''important area of development'' (Bennett-Armistead et al., 2005) in textbooks and other professional development materials focusing on preschool literacy instruction (e.g. the International Reading Association Preschool Literacy Collection (Roskos et al., 2005)). In particular, interactively reading aloud to preschool children enjoys support theoretically, philosophically, anecdotally and empirically among researchers and early childhood educators as an appropriate and effective context for vocabulary instruction with young children (e.g. Mol et al., 2009; National Early Literacy Panel, 2008; Van Kleeck et al., 2003). But there is also evidence that vocabulary instruction through read-alouds alone may not be powerful enough to narrow the gap with at-risk students (Marulis and Neuman, 2010) and that more integrated approaches are potentially more effective (Marulis and Neuman, 2010; Neuman and Dwyer, 2011; Neuman et al., 2011).

So, there is widespread agreement within the field of early childhood education that vocabulary is important and that reading aloud is one significant context for supporting vocabulary growth, especially when integrated with vocabulary instruction in other contexts. However, with an eye toward data-based decision-making, we believe there are several unexplored problems with the ways in which young children's vocabulary learning is assessed.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

This paper begins with a theoretical framework for how we conceptualize vocabulary learning and then overviews our context for enquiry into preschool vocabulary assessment. The remainder of the paper focuses on the critical analysis of existing approaches to vocabulary assessment with preschool children and uses of the data in practice, especially with regard to children's learning from classroom read-aloud experiences. The analysis concludes with recommendations for filling an extremely important need in the field of literacy research: reliable and valid assessment of preschool children's vocabulary learning in relation to classroom instruction.

What do we mean by "vocabulary?" Breadth versus depth of word knowledge

There exists a well-documented vocabulary gap between students at risk of language and literacy difficulties and those at low risk (Biemiller and Slonim, 2001; Hart and Risley, 1995). Studies and discussion of this gap emphasize the difference in the breadth of word knowledge of different groups of students, defined as the estimated number of root word meanings an individual can understand and use in oral language (see Biemiller and Slonim, 2001; Hart and Risley, 1995). In other words, research has been primarily concerned with how many words children know, with an eye on how to expand the size of children's vocabulary. Typically, intervention efforts to expand the breadth of children's vocabulary involve embedded instruction, with brief definitions or explanations of words incorporated into existing oral language interactions or interactions around texts (e.g. during read-aloud discussions; Coyne et al., 2009). Instruction and intervention efforts are effective at growing children's vocabulary, but they also consistently fail to narrow the gap between students at risk and their typically developing peers (Marulis and Neuman, 2010). Although we certainly value vocabulary instruction that expands children's vocabulary, it is also important to question the practical significance of such growth. For example, an intervention study may succeed in teaching children on average two new words a day and demonstrate a statistically significant difference between the intervention and comparison groups. But is that number "enough?" What kinds of words are they learning? What are they learning about the words? That is to say, will that learning result in a meaningful difference in the children's language or literacy development? Conceiving of vocabulary as breadth alone is only one piece of the puzzle.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

In contrast, depth of word knowledge involves the question of what it means to "know" a word. Knowledge of word meaning ranges from completely unknown, to varying degrees of partial knowledge, to complete knowledge, which is typically defined as the ability to define a word clearly or to use it appropriately in context (although even deeper levels of knowledge than definitional ability exist). Interpretations of an individual's depth of word knowledge often rely on articulations of the relationships among words, as well as the understanding that word meanings rely heavily on context and thus are not constants, all of which implies the need for some qualitative analyses of assessment data.

The goal of vocabulary instruction in terms of depth of word knowledge is, typically, to know a word well enough to support comprehension of the language involving the word, e.g. in reading, to understand a word well enough to comprehend the meaning of the text using it (Coyne et al., 2009). Studies of depth of vocabulary knowledge typically employ integrated approaches to vocabulary instruction involving multiple instructional contexts and outcome measures that attempt to measure more than one feature of word knowledge, for example meaning and context (Beck and McKeown, 2007), or full versus partial knowledge (Coyne et al., 2009).

Vocabulary instruction and assessment: General versus specific word knowledge

The long-term goal of vocabulary instruction is to expand and deepen children's general vocabulary, the entire repertoire of words a child understands (receptive vocabulary) and can use (expressive vocabulary). But this goal is achieved through the learning of specific words; in other words, general vocabulary comprises individual words that are learned and assessed. Instruction efforts may focus on developing general vocabulary through implicit and embedded instruction (or even simply increased efforts to engage in more sophisticated oral language interactions; Dickinson and Porche, 2011), or may include direct and explicit instruction in the meanings of specific words (Beck and McKeown, 2007; Neuman et al., 2011). Both approaches are effective and important to developing vocabulary (Coyne et al., 2009; Marulis and Neuman, 2010). Similarly, assessments of vocabulary may estimate general vocabulary or measure knowledge of specific words. The distinction between these two approaches to instruction and assessment is essential to the critical analysis of the affordances and drawbacks of assessment approaches in our remaining discussion.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Our context

We have conducted three ERF (UIC ERF) projects with preschool children over the past seven years (www.education.uic.edu/erf), and vocabulary has been a central instructional emphasis of this work. Teachers in our projects are encouraged to involve children in as much conversation as possible so as to promote general language development during less structured periods of the day such as lunch and transition times. They are also coached to employ a specific curriculum to engage children with activities and materials to teach specific words intentionally in multiple instructional contexts such as read-alouds, small groups and centre instruction focusing on specific cross-curricular themes.

To drive data-based decision-making, we sought assessment options that informed instruction toward all four goals—breadth and depth of vocabulary and general vocabulary development and knowledge of specific words taught. As required for all ERF projects (ERF, 2007), we assessed children's breadth of general vocabulary learning with the Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn, 2007). However, we struggled to identify and even to design a valid and reliable measure of preschool children's breadth or depth of learning of specific words they were taught. After a review of virtually every common vocabulary measurement technique for specific words used in current research and attempts to apply these techniques in our context, we concluded that the complexities involved in measuring preschool children's learning of specific words that results from instruction has gone either unrecognized or unreported in previous research.

Early childhood vocabulary assessment

We extensively reviewed existing research on vocabulary learning in early childhood, in particular learning from classroom read-aloud experiences, and found that assessments used in the field fell into two main categories: measures of general vocabulary or knowledge of specific words, either of which may assess the depth or breadth of word knowledge. In other words, we perceived the existing options as falling somewhere on the two-by-two grid depicted in Table 1.

General vocabulary measures

Normed general vocabulary measures: Measures of breadth. Normed measures of general vocabulary (receptive and expressive) focus on measuring the breadth of

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Table 1. Examples of types and vocabulary measures reviewed.

		Conceptual understanding of words	
		Breadth of knowledge	Depth of knowledge
Intended source of words assessed	General vocabulary measures	• PPVT • EOWPVT • EVT	• Definitional measures of general vocabulary (e.g. Biemiller and Slonim, 2001)
	Specific word measures	• PPVT-like formats • EOWPVT- or EVT-like formats • Yes/no questioning format	• Variations of yes/no questioning format (e.g. Beck and McKeown, 2007; Coyne et al., 2009) • Definitional measures of specific words (e.g. Blewitt et al., 2009; Coyne et al., 2009)

individuals' vocabulary (Table 1). PPVT (Dunn and Dunn, 2007) is the most common measure of receptive vocabulary used in early literacy research today. It is an individually administered assessment in which the examiner shows the child a page with four separate illustrations and asks him/her to point to the one that shows the target word. Almost all studies measuring the effects of reading aloud on general vocabulary use the PPVT as a measure of general vocabulary (e.g. the 19 studies reviewed by Dickinson and Smith, 1994; Hargrave and Sénéchal, 2000; NELP, 2008; Reese and Cox, 1999; Schwanenflugel et al., 2005; Walsh and Blewitt, 2006). We have expressed concerns about the PPVT (Hoffman, 2008), as have others (e.g. Pearson et al., 2007), but overall we find it adequate as a general measure of receptive vocabulary. Related normed measures also assess expressive vocabulary, two common examples being the Expressive One Word Picture Vocabulary Test (EOWPVT; Brownell, 2000) and the Expressive Vocabulary Test (EVT-2; Williams, 2007). In these expressive measures, rather than pointing to a picture portraying a target word, the administrator prompts the child to name the word portrayed in a picture.

Although the PPVT, EOWPVT and EVT are the most commonly used normed measures of general vocabulary in literacy research, many other examples exist. The Woodcock-Johnson III Picture Vocabulary Subtest (Woodcock et al., 2001) is widely used in clinical practice, although less frequently in literacy research. This assessment includes a receptive measure

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

(similar to the PPVT) and an expressive measure (similar to the EOWPVT and EVT). Other more holistic assessments of oral language development appropriate for preschoolers typically include a vocabulary component. For example, the Systematic Analysis of Language Transcripts (SALT, Miller, 2012) measures the number of different root words in a spoken language sample, and the Preschool Language Scale (PLS-4, Zimmerman et al., 2002) includes receptive and expressive vocabulary components. Because assessments like SALT, PLS and other similar assessments (e.g. Clinical Evaluation of Language Fundamentals (CELF); Semel et al., 2003) are more often used by speech-language pathologists for decision-making in clinical practice and have not been commonly used in studies of preschool vocabulary learning from classroom read-alouds, our remaining discussion of general vocabulary assessments focuses on those most relevant to our specific area of analysis, namely the PPVT, EOWPVT and EVT.

Measures like PPVT, EOWPVT and EVT are useful instruments because they are normed and standardized, statistically determined to be valid and reliable over thousands of administrations, easy to train for administration and simple to score. These measures virtually guarantee a normal distribution of scores conducive to inferential statistical analyses and can support the conclusion that gains on the test represent general vocabulary growth by comparing the individual to the general population curve. Thus, PPVT, EOWPVT and EVT are useful tools for indicating the impact of preschool instruction on the growth of children's general vocabulary over the course of their year(s) in school and for estimating the impact of vocabulary on later reading achievement (Dickinson and Porche, 2011; National Institute of Child Health and Human Development Early Child Care Research Network (NICHD ECCRN), 2005; Neuman et al., 2011; Storch and Whitehurst, 2002). The value of standardized measures like PPVT lies in their well established validity and reliability, which enable rigorous quantitative comparisons for programme evaluation and research.

In terms of measuring the breadth of vocabulary learning specifically attributed to instruction, gains on general vocabulary measures like the PPVT are typically small (Marulis and Neuman, 2010; National Early Literacy Panel, 2008). Research using other measures of vocabulary learning as well as anecdotal evidence suggest that reading aloud positively impacts on children's vocabulary, and so it is unreasonable to conclude that the small effect sizes measured by assessments like PPVT indicate that the children in those studies did not learn new vocabulary from instruction (Marulis and Neuman, 2010). Rather, we suspect that what students did learn about new words was not well

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

captured by the normed general vocabulary measures being used in such research. Other sources have raised similar concerns about the relative insensitivity of general vocabulary measures to detect small amounts of growth, in particular the National Reading Panel (NICHD, 2000), which suggested researchers should use both researcher-created and standardized measures to best examine vocabulary gains.

Definitional measures of general vocabulary: Measures of depth. A very different approach to vocabulary assessment involves directly prompting children for their definitions of words. Definitional measures provide information about depth of word knowledge in two ways: (1) they require the child to use their own language to describe word meanings, which provides rich qualitative data for insights into how the child understands a particular word in relation to other words and (2) they are typically scored on a scale (as opposed to correct/incorrect), which allows measurement of levels of word knowledge (Table 1). However, data about the depth of word knowledge inevitably lead to challenges with scoring the responses of young children. As opposed to the straightforward right/wrong nature of picture-based measures, with definitional measures one is forced to confront which features of a definition constitute ''knowing'' a word?

Different researchers have used different approaches to scoring definitional responses. Snow et al. (1989) assessed the ability of older children (grades 2 to 5) to give formal definitions of words, which had to contain an equivalency statement (e.g. *a* ''*dog*'' *is*... or ''*dog*'' *means*...) and superordinate (a broader category to which the word belongs, e.g. *an animal*). However, this may not be a realistic expectation for younger children, even though young children can communicate much about what they do know about a particular word. Alternatively, Biemiller and Slonim's (2001) study of vocabulary of children in preschool through to grade six scored children's informal definitions using a three-point scoring system: ''known'' (1), ''possibly known'' (0.5) or ''not known'' (0). Using this definitional scoring system, Biemiller and Slonim estimated the size of children's general vocabulary based on the percentage of children in the entire sample who knew each word at each grade level (their measure of word difficulty). Because definitional measures rely on indications of the level of word knowledge, they are also useful in assessing the depth of children's word knowledge in ways that are not possible with traditional receptive and expressive measures. However, definitional measures of general vocabulary like these are much less commonly used in research, especially research on the effects of instruction, than PPVT, EOWPVT and EVT, most likely due to the complexity of scoring.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

General vocabulary measures like those we have reviewed can provide valuable insights into the breadth and depth of children's word knowledge. However, what is not so useful about general vocabulary measures for work with teachers and children in preschool classrooms is their inability to provide insights into children's knowledge of specific words. This is problematic because a range of preschool activities, and especially read-alouds, target specific words that children are meant to learn, and in order to drive data-based instructional decisions, teachers need information about what words children are learning and what they are learning about them.

Specific word vocabulary measures

A number of studies have attempted to measure children's learning of specific words taught (e.g. Coyne et al., 2009; Hargrave and Sénéchal, 2000; Justice, 2002; Neuman et al., 2011; Sénéchal and Cornell, 1993; Silverman, 2007; Walsh and Blewitt, 2006; Wasik and Bond, 2001; Whitehurst et al., 1994), the clear advantages being that the researcher can target specific words related to concepts in the curriculum and/or in books to be read aloud, and simply pre- and post-test the words to measure vocabulary growth. In practice, teachers benefit from data like these because they lead to better understanding of what children are learning from instruction, in terms of both breadth and depth of word knowledge.

PPVT-like formats: Measures of breadth. The most commonly developed formats for measuring the breadth of specific word vocabulary reported in the literature are researcher-developed receptive vocabulary measures that mimic the PPVT. The researcher shows four pictures (1 correct and 3 distractors) for each of the target words and prompts the child, "Point to the picture that shows [target word]." Another related type of pictorial assessment is a researcher-developed measure of expressive vocabulary (structured like EOWPVT or EVT). In these expressive measures, the researcher shows a picture of the target word and asks the child to name the picture. Measures like these have been used extensively with preschool students to measure the breadth of word learning from instruction (e.g. Ard and Beverly, 2004; Collins, 2005; Hargrave and Sénéchal, 2000; Justice, 2002; Walsh and Blewitt, 2006; Wasik and Bond, 2001; Whitehurst et al., 1994; Table 1).

Both of these types of tests appear straightforward, but in actuality are problematic in terms of design and use as measures of specific word vocabulary learning with preschoolers. First of all, there is the issue of the three distractors included for each item in the receptive instruments. The PPVT has

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

been through numerous rounds of item analysis and refinement to present items systematically in a progression of increasing difficulty to fit scored responses to a normal curve, thus enabling the instrument's purpose of measuring general vocabulary knowledge and growth compared to the normed sample. Researcher-developed tests, on the other hand, have not monitored or controlled for the difficulty of items in this or any closely related way. To illustrate, in the PPVT, the distractors become more similar to the target as the items increase in difficulty in order to capture the increasingly fine-grained meanings of the target words. However, in researcher-developed tests that are not designed to fit a normal curve, how does the researcher determine the appropriate level of difficulty of each item, or more accurately, the appropriate level of disparity between the target and the distractors? For example, one of the items we wished to assess was lightning. What should the distractors look like for this word? Would illustrations of a lamp, a kitten and a dish serve the purposes of distractors? Or should the illustrations represent other weather forms, like rain, snow and wind, or conform to some other pattern? We searched through numerous studies of early childhood vocabulary learning resulting from the instructional practice of reading aloud and could find no discussion of this issue. The absence of such discussion indicated to us that early literacy researchers typically did not address this potential measurement problem at all, and simply chose distractor words and illustrations that were categorically similar to the target.

A second item construction issue to consider is the actual visual portrayal of the target words. It is interesting to note that the PPVT items administered to three- and four-year olds typically consist of mostly nouns (72% of the first 72 items in PPVT-4). There are relatively few verbs (21%), and almost no adjectives (only 5%) that are tested. Having to represent vocabulary words pictorially creates a bias toward including nouns and simple verbs and adjectives because they are the easiest to portray. This does not pose much of a problem for standardized measures of general vocabulary because it is relatively easy to find items that work well from among the thousands of words in a language, but it creates significant problems when one is seeking to assess specific words from a book or related to a concept or unit that has been taught. To get the greatest payoff in terms of instruction, teachers are encouraged to teach "Tier 2" words (i.e. words that are not frequently heard in everyday conversation but are typical of written text (Beck et al., 2002)) and words that are central to the meaning of the texts being read. Very often such words are not easily represented in picture items (Beck and McKeown, 2007). For example, one word that we wished to assess was *nudge*. Efforts to portray the action *nudge*

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

in a still format resulted in illustrations in which animals appeared to be sniffing more than nudging. Therefore, even though it was a rich, Tier 2 vocabulary word worthy of emphasis in the story, *nudge* posed problems for assessment using this format.

The third item construction issue relates to children's scores on these types of measures. In using a receptive measure with four choices, the child inevitably has a one in four chance of getting the item correct, regardless of her/his knowledge of the word meaning. For a normed measure like PPVT, this issue is largely irrelevant: standardized scores account for probability, and since each child's score is based on dozens of responses, a child's results are not intended to be used as indicators of knowledge of particular words. However, when assessing only a limited number of words from relatively few instructional activities, achieving 25% correct by chance can greatly diminish the practical significance of the scores, rendering teacher understanding of an individual child's knowledge of specific words less than robust.

Fourth, there is a significant problem with interpretations of results from such an assessment format when it is used to examine preschoolers' vocabulary growth as a result of read-aloud experience. As previously stated, studies of the effects of vocabulary instruction in read-alouds using specific word vocabulary measures not surprisingly demonstrate larger effect sizes than those reported for general vocabulary measures (Marulis and Neuman, 2010; NICHD, 2000). Thus, data from specific word measures support the use of reading aloud to promote vocabulary learning of specific words, but they cannot be used to conclude that reading aloud promotes practical, significant general vocabulary growth, as they are sometimes interpreted. This distinction can have far-reaching effects on policy and practice. As one example, the WWC, which serves as "a central and trusted source of scientific evidence for what works in education" (2012a), reported strong evidence of the positive effects for Dialogic Reading as a read-aloud intervention supporting general oral language development despite their own report of a lack of statistically significant findings for the majority of measures used across all of the reviewed studies (2007). Four studies were included in the WWC intervention ratings (met WWC evidence standards; Lonigan et al., 1999; Lonigan and Whitehurst, 1998; Wasik and Bond, 2001; Whitehurst et al., 1994). Of these four studies, only one reported a significant difference on a normed receptive vocabulary measure (Wasik and Bond, 2001, as reported in the original publication—these data were missing from the WWC analysis in the intervention report), and only one reported a significant difference on a normed expressive vocabulary measure (Whitehurst et al., 1994). Effect sizes

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

calculated from data from normed measures reported in these studies averaged only 0.02 for receptive and 0.13 for expressive vocabulary. Effect sizes on researcher-developed measures (Wasik and Bond, 2001; Whitehurst et al., 1994) were much higher, averaging 1.58 for receptive and 1.13 for expressive vocabulary. Through careful analyses of the included studies, it is clear that this recommendation of the WWC is attributed to the strong effect sizes of one study using researcher-developed specific word vocabulary measures (Wasik and Bond, 2001). However, the WWC made no distinction between the types of measures used across different studies, and, in fact, in a related online resource developed for teachers by the U.S. Department of Education, Doing What Works (2008), clearly promoted Dialogic Reading as a research-based practice for enhancing oral language for all children. Our goal here is not to dispute studies of Dialogic Reading that have found positive effects for some children, but rather to highlight how the use of evidence from specific word measures can be and has been inappropriately applied in recommending instructional practices for general vocabulary growth, thus promoting misapplications of research in practice.

Yes/no questioning format. An alternative to representing words pictorially is the use of yes/no questions about the meanings of target words (as used in Beck and McKeown, 2007 and Coyne et al., 2009 with kindergarten and older children). In this form of assessment, each target word has one question for which the correct response is yes and another with the correct response no. To demonstrate knowledge of the target word, both questions must be answered correctly. For example, for the target word *trail*, the child should respond yes to, ''Could you walk along a trail?'' and no to, ''Could you send a trail to your friend?'' in order to be considered as knowing *trail*.

Because these forms of assessment do not require pictorial representation, it is probably easier to assess a wider variety of target words. In addition, variations of this design can indicate children's depth of word knowledge (Table 1). For example, Beck and McKeown (2007) included questions focusing on basic definitional meaning as well as those using the word within a specific context, and Coyne et al. (2009) designed yes/no questions that required full versus partial knowledge of the word in order to respond correctly.

Nevertheless, we believe there are challenges to yes/no assessments that should be carefully considered before using them with preschool children. When we assessed preschool children with this approach, we found that they were easily confused by the structure of yes/no questions. Our assessment experience was similar to teaching experiences asking yes/no questions in large group discussions with young children, where half of the class earnestly

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

answers "yes" while the other half just as earnestly answers "no". This behavior suggests children are responding to what they think the teacher has in mind, rather than comprehending and responding to the meaning of the actual question. In other words, this assessment form confounds comprehension of yes/no questioning formats with word knowledge in a way that is problematic for interpreting an individual child's knowledge of a specific word. In addition, the 50% probability that each question will be correct by chance corresponds to a 25% probability that any two questions for a target word will be correct by chance.

Definitional measures of specific words: Measures of depth. Other studies have drawn on research using definitional measures of depth of general vocabulary to design assessments that elicit definitions of specific words taught (Table 1). As explained in our discussion of general vocabulary assessments, definitional measures inevitably involve the challenge of scoring a child's explanation of word meaning. As with general vocabulary definitional measures, different researchers have used different approaches to scoring children's definitions of specific words taught. Blewitt et al. (2009) analyzed children's definitions for the number of information units specified: superordinate category (e.g. "A cat is a kind of animal"), synonyms (e.g. "A ship is a boat"), perceptual properties (e.g. "A carrot is orange"), functional properties (action, "A ship floats", or use, "You eat carrots") or parts (e.g. "A tree has a branches"). Children earned one point for each information unit for each word and the points were totalled across all words tested. Coyne et al. (2009) used a similar three-point system to that employed by Biemiller and colleagues (Biemiller and Boote, 2006; Biemiller and Slonim, 2001), which simply indicated whether the word was unknown, partially known or completely known.

Neither of these scoring approaches can overcome the complexity of scoring levels of word knowledge in definitions from children of this age. To demonstrate some of the complexities involved, consider the following difficulties we found. After multiple encounters with/discussions of the word *reflection* during read-alouds (a word from the book *Raccoon on His Own* (Arnosky, 2003) in which the raccoon sees his reflection in water) as well as in hands-on guided explorations of reflective surfaces, we prompted children to define the word reflection. The following is a sample of five children's responses:

1. "It means that your reflection is yourself. It means that there is another person that looks just like you."
2. "Means if you see yourself in stuff and you see your reflection."
3. "Is like when you look in something, like water, you can see yourself."

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

4. "It mean your face go in the water."
5. "That means if you the same skin as him, you blend in."

From a scoring perspective, which response evidences the most complete knowledge of word meaning? We contend that responses 1–3 contain evidence that the child understands the basic premise of the word meaning, while 4 and 5 are far more questionable. Still, there is variation within the more accurate responses (1–3). The first response could be interpreted to include certain misinformation, that is, there is not actually another person that looks just like us. Responses 2 and 3 seem similar, except that the second generalizes the word meaning "in stuff", while the third includes the example, "like water". To aid readers' evaluation of the last two responses, we offer this additional insight into the context of the classroom and the reading of the text: response 4 is most likely drawing on details from one of the illustrations in the book, which depicts a raccoon leaning over water and gazing downward at his reflection on the surface, and response 5 is almost certainly a confusion between reflection and camouflage, another concept recently introduced in the classroom by the teacher and brought up by a student during the book reading that preceded this assessment.

The characteristics of these five responses are not isolated examples. They are representative of common challenges we experienced in using a definitional measurement approach to assess knowledge of specific words. First, when assessing vocabulary with children this young, measurement of word knowledge through definitions is inevitably confounded with the ability to convey that knowledge in language. The concept of reflection is a complex one to explain, even for an adult (try to think of how you would formally define this word), but especially so for a four-year-old. In response 1, for example, it is impossible to determine whether the child actually understood a reflection to be another person, or just struggled to convey the idea in words.

Second, because the context in which these responses were obtained focused on assessing word knowledge from a specific book, the issue of the generalization of word meanings is highlighted. Formal definitions are intended to convey the most general meaning of the word while still distinguishing it from similar words. However, when preschool children have encountered a word in one context, they cannot be expected to generalize its meaning to others; that is to say, they do not yet know the boundaries of that word's meaning vis-à-vis other words with a family resemblance (Wittgenstein, 1953). For example, both responses 2 and 3 evidence some knowledge that reflections can occur in something more general than water

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

("in stuff" or "in something"), but both also seem to indicate the limited understanding that a reflection is only an image of oneself, with no acknowledgement of the possibility of seeing a reflection of another object or person.

Third, and related to the issue of the generalization of meanings, is the inclusion of book-specific details in definitions, such as in response 3, "like water". Children had a tendency to include such details in their definitions, which in hindsight is not at all surprising, but which definitely complicates scoring because formal definitions are not "supposed" to include such details. Yet, and especially for children of these ages, their presence is so naturally characteristic of answering a question about a word in a book that was read to you that dismissing them in analysis would be tantamount to saying that the context of the assessment is irrelevant.

In addition, problems with the definitional measurement of word learning arise from the opposite perspective: when the child offers a definition not in line with the word use in the book. A prime example arose with the word *chill*, as used in the book that was read to the children: "A chill ran down his spine." During the assessment, many children proffered a more colloquial definition of the word. A favorite of ours was: "When you want to lay down and watch TV—and eat nachos." Thus, when the measure and the task are designed to value a particular response, how does one evaluate a different, but equally appropriate response? In this "chill" example, the child does not provide any evidence of having learned the word meaning as used in the book, but was demonstrating a level of knowledge of the word. More importantly, the word usage described by the child is arguably relevant for his interpretation of the usage of the word in the text, in that, in one sense, they are both metaphors for the connection between emotional state and body temperature (Lakoff and Johnson, 1980).

With respect to the specific scoring approaches for definitions, we attempted both analyzing information units and the three-point scoring system. When analyzing information units, any one unit can be problematic to analyze. For example, for the target word *swamp*, does the phrase "animals live there" evidence an understanding that a *swamp* is a habitat and therefore represents an emergent use of a superordinate? Or does it constitute a use of a *swamp*? For the word downstream, one child responded, "It means you're going lots of water [sic] that's going really slow." How should the inaccurate additional information, "that's going really slow", be analyzed? Does this count against the score in some way, or does one ignore it? In analyzing the definitions our students provided, we struggled to achieve a reasonable degree of inter-rater reliability and were unable to designate any specific

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

combination of these formal definitional traits (i.e. superordinates, synonyms, properties) that consistently captured the holistic quality of communicating the basic premise of the word meaning. In other words, a child could convey an understanding of a word only using a list of traits (raccoon—"he got a black mask, white eyes, brown hands, brown fur, brown legs") or one superordinate and one trait (raccoon—"it's a kind of animal that's sneaky"). If scored based on the inclusion of information units using Blewitt and colleagues' system (2009), the first response would receive a 5 and the second response a 2; however, we do not believe that first response indicates a holistically deeper understanding than the second, and thus we question the validity of the measure as an indication of the depth of word knowledge.

The only approach to scoring with which we were able to achieve a reasonable degree of inter-rater reliability was on a holistic three-point scale like that used by Coyne et al. (2009) and Biemiller and colleagues (Biemiller and Boote, 2006; Biemiller and Slonim, 2001). Quantitatively, a three-point scale only provides one additional level over the correct/incorrect measures previously described. Qualitatively, however, children's definitions provide a wealth of information about how individual children are thinking about words.

Aside from the complexity of scoring, definitional measures offer much information about children's word knowledge. Because they are completely open-ended, they do not rely on indirect measurement of word knowledge through some instantiation of multiple-choice, nor are they confounded with the issue of pictorial representation. In addition, instead of being restricted to analyzing children's responses as only right or wrong, definitional measures can capture more about how individual children understand and use words. This is useful information for teachers and researchers alike, because the resulting data emphasize the complexity of word learning, as became clear through the discussion of the intricacies of children's knowledge of words in our examples. In other words, children's responses to definitional measures (like those reported in this section) highlight an understanding of words as interrelated and deictic components of language rather than discrete, isolated objects with stable meanings.

Complementary assessment design for early childhood vocabulary learning

Our analysis of the benefits and challenges of each form of assessment led us to conclude that only a careful design of complementary assessment approaches can achieve valid and reliable measures of preschool children's

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

vocabulary learning. And thus, only complementary assessment approaches can provide the data needed to indicate child learning that results (or does not result) from vocabulary instruction. We propose that teachers, administrators, policy-makers and researchers closely examine their purposes and needs for vocabulary assessment in order to choose and design complementary measures that assess: (1) both the breadth and depth of children's vocabulary and (2) children's general vocabulary development as well as learning of specific words. To aid these audiences in this process of assessment design that balances these goals, we propose the following guidelines for appropriate and effective applications of the assessment formats we have reviewed.

Normed general vocabulary measures

For teachers and researchers, normed general vocabulary measures like the PPVT are a useful tool for screening groups of students for the breadth of their general vocabulary development. For policy-makers or researchers such normed measures are useful for programme evaluation and experimental comparison based on the breadth of children's general vocabulary. But precisely because they are normed general vocabulary measures, they do not provide practically significant information to teachers about children's learning of specific words from instruction, nor do they capture well the depth of individual children's knowledge of specific words. In conclusion, measures like the PPVT have a place and a role in assessing the breadth of children's general vocabulary development, but interpretations of data from such measures should not reach beyond these limitations.

Specific word measures

First of all, we have not been able to locate any published evidence to date on the use of yes/no assessments of specific words with preschoolers. Thus, we recommend that they not be used until further evidence suggests they are appropriate and meaningful assessments for preschool children. Second, we propose that researcher- or practitioner-designed PPVT-like assessments of specific words are not useful for school administrators or practitioners to use in planning instruction because the drawbacks in terms of validity and reliability (e.g. ease of pictorial representation, opportunities for an item to be correct by chance) far outweigh the benefits (namely, ease of administration and scoring). Researcher-developed PPVT-like assessments may be useful tools for measuring word learning for groups of preschool students in experimental

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

studies (when an individual child's knowledge of a specific word is less relevant), but only if they are crafted in careful, systematic ways with reported measures of reliability and validity. In addition, research using these measures should specify much more clearly how items were developed and should remind readers that effect sizes reported using such measures are not comparable to effect sizes from normal general vocabulary measures.

Definitional measures

We believe definitional measures of word knowledge are a useful but underutilized assessment tool in early childhood vocabulary assessment. In particular, definitional measures have the potential to provide insights into individual preschool children's depth of word knowledge of specific words taught. For teachers, it is important to note that definitional measures need not be highly structured one-on-one assessments. Teachers can glean much information from children during normal classroom activities through informal prompts for word meanings, while simply jotting down notes to record children's definitions and/or scoring their responses on the three-point scale previously discussed. Thus, this assessment form is conducive to on-the-fly informal assessment opportunities that are developmentally appropriate for preschool classroom instruction, but which also provide the depth of information about individual children's knowledge of specific words that a teacher needs to plan future instruction. Nevertheless, it is important for teachers to recognize that definitional measures inevitably confound children's word knowledge with their ability to convey their knowledge through language.

It must also be noted that there is limited research on the use of definitional measures with young children and their relationship to planning vocabulary instruction. More research into the application of definitional vocabulary measurement in classroom practice is needed, and teachers will likely require professional development in how to analyze reliably and interpret appropriately the qualitative data that definitional measures provide.

In order to build knowledge of the reliability, validity and utility of definitional measures in classroom practice, we recommend that such measures be incorporated more often into research studies of word learning. Definitional measures offer one way to measure gradients of word knowledge of the rich words that teachers actually teach without the confounds of difficulties associated with providing pictorial representation, and so we also recommend putting more research focus on deeper understandings of word learning than simple recognition.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Conclusion

Given the complexities involved with each of the reviewed approaches to early childhood vocabulary assessment, we were surprised to discover in our review that researchers have been virtually silent about the challenges involved in measuring this construct. In this paper, we have attempted to show that each approach to vocabulary measurement has its benefits and challenges, and that a careful and balanced approach to assessment can provide the variety of information teachers and researchers need to make informed decisions. As Einstein was fond of noting, "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted." Given that the majority of vocabulary studies involving young children conducted over the past two decades relied heavily on quantitative measures that may hold relatively little practical significance, it appears that we have succumbed to counting what can be easily counted. When relying solely on such measures, research and instruction that merely focus on recognition are granted precedence over that which teach and assess word meaning in a more robust sense, the degree of word meaning that has payoff for reading comprehension later in children's development (i.e. the kind of multi-faceted vocabulary knowledge discussed by Nagy and Scott, 2000). By choosing, designing and balancing uses of assessments to meet various needs for information, researchers and practitioners can best capitalize on the strengths of each form of assessment to inform instructional design and ultimately improve student learning.

Funding

Parts of the work reported here were supported by Early Reading First grants S359B060073, S359B080021 and S359B090099 from the U.S. Department of Education to the University of Illinois at Chicago. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

References

Ard L and Beverly B (2004) Preschool word learning during joint book reading: effect of adult questions and comments. *Communication Disorders Quarterly* 26(1): 17–28.

Arnosky J (2003) *Raccoon on His Own*. New York: Puffin.

Baumann J (2008) Vocabulary and reading comprehension. In: Israel S and Duffy G (eds) *Handbook of Research on Reading Comprehension*. New York: Routledge, pp. 323–346.

Beck IL and McKeown MG (2007) Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal* 107(3): 251–273.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Beck IL, McKeown MG and Kucan L (2002) *Bringing Words to Life: Robust Vocabulary Instruction*. New York: Guilford Press.

Bennett-Armistead VS, Duke NK and Moses AM (2005) *Literacy and the youngest learner*. New York: Scholastic.

Biemiller A and Boote C (2006) An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology* 98(1): 44–62.

Biemiller A and Slonim N (2001) Estimating root word vocabulary growth in normative and advantaged populations: evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology* 93(3): 498–520.

Blewitt P, Rump KM, Shealy SE et al. (2009) Shared book reading: when and how questions affect young children's word learning. *Journal of Educational Psychology* 101(2): 294–304.

Brownell. (2000) *Expressive One-Word Picture Vocabulary Test*. Novato, CA: ATP.

Collins MF (2005) ESL preschoolers' English vocabulary acquisition from storybook reading. *Reading Research Quarterly* 40(4): 406–408.

Coyne MD, McCoach DB, Loftus S et al. (2009) Direct vocabulary instruction in kindergarten: teaching for breadth versus depth. *Elementary School Journal* 110(1): 1–18.

Cunningham AE and Stanovich K (1997) Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology* 33(6): 934–945.

Dickinson D and Smith M (1994) Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly* 29(2): 104–122.

Dickinson DK and Porche MV (2011) Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development* 82(3): 870–886.

Doing What Works (2012) Use interactive and dialogic reading. Available at: http://dww.ed.gov/Preschool-Language-and-Literacy/Use-Interactive-and-Dialogic-Reading/learn/?T_ID=15andP_ID=31 (accessed 16 January 2012).

Dunn L and Dunn D (2007) *Peabody Picture Vocabulary Test-4*. Minneapolis, MN: Pearson.

Early Reading First (2007) Guidance for the early reading first program. Available at: http://www.ed.gov/programs/earlyreading/legislation.html (Accessed 16 January 2012).

Hargrave A and Sénéchal M (2000) A book reading intervention with preschool children who have limited vocabularies: the benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly* 15(1): 75–90.

Hart B and Risley TR (1995) *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Brookes.

Hart B and Risley TR (1999) *The Social World of Children Learning to Talk*. Baltimore: Brookes.

Hoffman JL (2008) Story comprehension and vocabulary learning in the preschool read aloud. In: *National reading conference*, Orlando, FL.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Justice L (2002) Word exposure conditions and preschoolers' novel word learning during shared storybook reading. *Reading Psychology* 23(2): 87–106.

Lakoff G and Johnson M (1980) *Metaphors We Live By*. Chicago: University of Chicago Press.

Lonigan CJ, Anthony JL, Bloomfield BG et al. (1999) Effects of two shared-reading interventions on emergent literacy skills of at-risk preschoolers. *Journal of Early Intervention* 22(4): 306–322.

Lonigan CJ and Whitehurst GJ (1998) Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly* 13(2): 263–290.

Marulis LM and Neuman SB (2010) The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research* 80(3): 300–335.

Miller JF (2012) *Systematic Analysis of Language Transcripts (SALT)*. Madison, WI: SALT Software, LLC.

Mol SE, Bus AG and de Jong MT (2009) Interactive book reading in early education: a tool to stimulate print knowledge as well as oral language. *Review of Educational Research* 79(2): 979–1007.

Nagy WE and Scott JA (2000) Vocabulary processes. In: Kamil ML, Mosenthal P, Pearson PD and Barr R (eds) *Handbook of Reading Research (Vol. 3)*. Mahwah, NJ: Erlbaum, pp. 269–284.

National Early Literacy Panel. (2008) *Developing Early Literacy: A Scientific Synthesis of Early Literacy Development and Implications for Intervention*. Washington, DC: National Institute for Literacy/National Center for Family Literacy.

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 (NIH Pub. No. 00–4769)*. Washington, DC: U.S. Government Printing Office.

National Institute of Child Health and Human Development Early Child Care Research Network (2005) Pathways to reading: the role of oral language in the transition to reading. *Developmental Psychology* 41(2): 428–442.

Neuman SB and Dwyer J (2011) Developing vocabulary and conceptual knowledge for low-income preschoolers: a design experiment. *Journal of Literacy Research* 43(2): 103–129.

Neuman SB, Newman EH and Dwyer J (2011) Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly* 46(3): 249–272.

Pearson PD, Hiebert E and Kamil M (2007) Vocabulary assessment: what we know and what we need to learn. *Reading Research Quarterly* 42(2): 282–296.

Reese E and Cox A (1999) Quality of adult book reading affects children's emergent literacy. *Developmental Psychology* 35(1): 20–28.

Roskos KA, Tabors PO and Lenhart LA (2005) *Oral Language and Early Literacy in Preschool*. Newark, DE: International Reading Association.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Scarborough H (2001) Connecting early language and literacy to later reading (dis)-abilities: evidence, theory, and practice. In: Neuman SB and Dickinson D (eds) *Handbook of early literacy research*. New York, NY: Guilford, pp. 97–110.

Schwanenflugel PJ, Hamilton CE, Bradley BA et al. (2005) Classroom practices for vocabulary enhancement in prekindergarten: lessons from PAVEd for success. In: Hiebert EH and Kamil ML (eds) *Teaching and Learning Vocabulary: Bringing Research to Practice*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 155–177.

Semel E, Wiig EH and Secord WA (2003) *Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4)*. San Antonio, TX: Pearson.

Sénéchal M and Cornell E (1993) Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly* 28(4): 360–374.

Silverman R (2007) Vocabulary development of English-language and English-only learners in kindergarten. *Elementary School Journal* 107(4): 365–383.

Snow C, Burns MS and Griffin P (1998) *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Snow C, Cancini H, Gonzalez P et al. (1989) Giving formal definitions: An oral language correlate of school literacy. In: Bloome D (ed.) *Classrooms and Literacy*. Norwood, NJ: Ablex, pp. 249–233.

Stahl S and Nagy W (2006) *Teaching Word Meanings*. Mahwah, NJ: Lawrence Erlbaum.

Storch SA and Whitehurst GJ (2002) Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology* 38(6): 934–947.

U.S. Department of Education (2008) Application for new grants under the early reading first program. Available at: http://www.ed.gov/programs/earlyreading/2008-359a.pdf (accessed 19 January 2012).

Van Kleeck A, Stahl S and Bauer EB (eds) (2003) *On Reading Books to Children: Parents and Teachers*. Mahwah, NJ: Erlbaum.

Walsh BA and Blewitt P (2006) The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal* 33(4): 273–278.

Wasik BA and Bond MA (2001) Beyond the pages of a book: interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology* 93(2): 243–250.

What Works Clearinghouse (2007) Intervention: dialogic reading effectiveness. Available at: http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=135 (accessed 29 April 2013).

What Works Clearinghouse (2012a) What works clearinghouse. Available at: http://ies.ed.gov/ncee/wwc/aboutus.aspx (accessed 16 January 2012).

What Works Clearinghouse (2012b) What works clearinghouse. Available at: http://ies.ed.gov/ncee/wwc/Topic.aspx?sid=4 (accessed 16 January 2012).

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013

Whitehurst GJ, Arnold DS, Epstein AL et al. (1994) A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology* 30(5): 679–689.

Williams KT (2007) *Expressive Vocabulary Test, Second Edition* (*EVT*-2). San Antonio, TX: Pearson.

Wittgenstein L (1953) *Philosophical Investigations*. London: Blackwell Publishing.

Woodcock RW, McGrew KS and Mather N (2001) *Woodcock-Johnson Tests of Achievement* (*WJ III*®). Rolling Meadows, IL: Riverside Publishing.

Zimmerman IL, Steiner VG and Pond RE (2002) *Preschool Language Scale, Fourth Edition* (*PLS*-4). San Antonio, TX: Pearson.

Downloaded from ecl.sagepub.com at Univ of Illinois at Chicago Library on September 14, 2013