Journal of Literacy Research
http://jlr.sagepub.com/

Developing Vocabulary and Conceptual Knowledge for Low-Income Preschoolers : A Design Experiment

Susan B. Neuman and Julie Dwyer

Journal of Literacy Research 2011 43: 103 originally published online 14 April 2011
DOI: 10.1177/1086296X11403089

The online version of this article can be found at:
http://jlr.sagepub.com/content/43/2/103

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Literary Research Association

Journal of Literacy Research
43(2) 103–129

DOI: 10.1177/1086296X11403089
http://jlr.sagepub.com

Developing Vocabulary and Conceptual Knowledge for Low-Income Preschoolers: A Design Experiment

Susan B. Neuman[1] and Julie Dwyer[2]

Abstract

The purpose of this design experiment was to research, test, and iteratively derive principles of word learning and word organization that could help to theoretically advance our understanding of vocabulary development for low-income preschoolers. Six Head Start teachers in morning and afternoon programs and their children ($N = 89$) were selected to participate in the World of Words, a 12-min daily supplemental vocabulary intervention; six classes ($N = 89$) served as a comparison group. Our questions addressed whether the difficulty of words influenced the acquisition and retention of words and whether learning words in taxonomies might support vocabulary development and inference generation. We addressed these questions in two design phases for a total intervention period of 16 weeks. Pre- and post-unit assessments measured children's expressive language gains, categorical development, and inference generation. Significant differences were recorded between treatment and comparison groups on word knowledge and category development. Furthermore, children in the treatment group demonstrated the ability to infer beyond what was specifically taught. These results suggest that instructional design features may work to accelerate word learning for low-income children.

Keywords

early literacy, vocabulary development, conceptual development, design experiment, preschoolers

[1]University of Michigan, Ann Arbor, MI, USA
[2]Boston University, Boston, MA, USA

Corresponding Author:
Susan B. Neuman, University of Michigan, 610 E. University, Ann Arbor, MI 48109, USA
Email: sbneuman@umich.edu

The essence of all instruction is helping children learn new concepts and the words that signify them. In particular, word knowledge—oral language vocabulary—plays a critical role in children's reading achievement (Dickinson, McCabe, & Essex, 2006; Kibby, 1995). Extensive research demonstrates that the size of a child's vocabulary in kindergarten is an effective predictor of reading comprehension in the middle elementary years (Biemiller, 2005). Furthermore, orally tested vocabulary at the end of Grade 1 is a significant predictor of reading comprehension in high school (Cunningham & Stanovich, 1997; Stahl & Nagy, 2006).

Children's knowledge of word meanings is cumulative (Chall, Jacobs, & Baldwin, 1990). The more words children know, the easier it is to learn new words. Children with highly elaborated semantic knowledge are likely to have more ready and more fluent access to this information, and it is this rich interconnected knowledge of concepts, not just individual words, that drives comprehension and reading proficiency in later grades (Vellutino et al., 1996).

Consequently, the well-documented gap in vocabulary knowledge between economically disadvantaged children and their middle-class peers prior to entering the elementary school years (Hart & Risley, 1995) becomes of great concern if educators are to improve reading achievement and decrease the knowledge disparities among poor and middle-income children (Farkas & Beron, 2004). Moats (1999), for example, estimates that the difference at entry into first grade may be as large as 15,000 words, with linguistically disadvantaged children knowing about 5,000 words, compared to their advantaged peers, who have 20,000 words. Hart and Risley (2003) argue that the accumulated experiences with words for children who come from poor circumstances compared with children from professional families may constitute a 30-million word catastrophe that is difficult, if not impossible, to close over time.

Therefore, the earlier children can acquire a large and richly structured vocabulary, the greater their reading comprehension is likely to be in the later grades (Hirsch, 2003). Nevertheless, available evidence (Beck & McKeown, 2007) indicates that there is little emphasis on the acquisition of vocabulary in school curriculum. For example, in a recent content analysis of 10 published early literacy programs adopted by Early Reading First recipients (Neuman & Dwyer, 2009), we found little evidence of an instructional regime or a deliberate effort in curriculum materials to teach vocabulary to preschoolers. Unfortunately, current instructional materials appear to offer little guidance to teachers who want to do a better job of teaching vocabulary to young children.

Despite the importance of vocabulary in predicting later achievement (Senechal, LeFevre, Thomas, & Daley, 1998; Storch & Whitehurst, 2002), few intervention studies have made significant gains in closing the gap in word knowledge between middle- and low-income students (Juel, Biancarosa, Coker, & Deffes, 2003). To date, storybook reading (Mol, Bus, & de Jong, 2009) has been regarded as the most potent source for teaching vocabulary in early childhood. Studies, however, suggest that the effects of reading aloud to children may not be powerful enough to enhance low-income children's word knowledge (Elley, 1989; Penno, Wilkinson, & Moore, 2002). In a now-classic study, Elley (1989) demonstrated that 7-year-olds showed an average vocabulary gain of 15% from an oral storybook reading when the words in the text were frequently mentioned, depicted in illustrations, and redundant in the surrounding context. However, less than half of this gain was demonstrated in a second story with different characteristics. Furthermore, in a recent meta-analysis of 31 experiments (Mol et al., 2009), researchers found that the strongest effect sizes appeared in highly controlled settings executed by examiners, not classroom teachers. Teachers seemed to have difficulty fostering growth in young children's language and literacy skills. Together, this evidence suggests more intensive interventions might be needed to narrow the gap for less advantaged children.

In this article, we report on a supplemental multimedia vocabulary curriculum, known as the World of Words (WOW; Neuman, Dwyer, Koh, & Wright, 2007), designed to engage low-income preschoolers in these kinds of learning activities. It is based on a framework that capitalizes on word learning through category formation. Considered a major component of word learning (Bloom, 2000), category membership is one of the first pieces of information a child learns about a word (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998; Whitmore, Shore, & Smith, 2004). Learning to identify a furry, four-legged animal as a dog, for example, involves not just mapping the label *dog* to one's household pet but actually establishing a concept of what is and what is not a dog. Existing evidence suggests that children use categories to gain information about unfamiliar terms (Gelman & O'Reilly, 1988; Kalish & Gelman, 1992), which, therefore, may potentially help bootstrap word learning.

Traditionally, studies of curriculum development, such as ours, and educational research have been considered two distinct enterprises. Goals related to curriculum development have been to produce instructional materials; scientific research, on the other hand, the creation of knowledge. Such distinctions, however, have not served the educational community well and could be a reason that curriculum development has not reliably improved (Clements, 2007). For example, although knowledge is usually created during curriculum development, this knowledge has seldom been explicated, published, or used to better understand the mechanisms that underlie learning.

Consequently, in this study, we approached curriculum from a design perspective, focusing our efforts on learning how our pedagogical design might support vocabulary and conceptual development. From this perspective (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), learning and development are both generative and dynamic; what is known about a learning process is applied to the anticipated curriculum; what is learned from enacting the curriculum is used to revise and better understand the conditions for learning. In this respect, the goal of our design experiment was not merely to empirically fine-tune "what works." Rather, as a design experiment (Reinking & Bradley, 2008), it was equally important to understand how, when, and why the instructional design might work and its implications for developing theory about word learning.

Overview of the Instructional Design

The WOW curriculum (Neuman et al., 2007) is an embedded multimedia program designed to foster children's vocabulary and conceptual knowledge in pre-K. The 12-min supplemental curriculum uses multimedia (video, pictures, books) to augment children's content learning at the same time it is designed to teach critical early literacy skills of vocabulary and conceptual learning. The curriculum consists of two science-based units, living things and healthy habits, each organized across four topics, with each lesson taught during an 8-day sequence. It is designed to supplement, not supplant, the curriculum used in these Head Start classrooms.

As an illustration of the kind of instruction provided, consider the vocabulary instruction from the topic "Insects." The 8-day sequence begins each day with a "tuning-in"—a rhyme, song, or word-play video clip that is shown from a DVD to bring children together.[1] The teacher follows this activity with additional examples, engaging the children in a briskly paced call-and-response set of interactions.

The tuning-in is followed by a "content" video that introduces children to the definition of the category. The first video is designed to act as a prototype of the category, a particularly salient exemplar of the topic (i.e., a katydid). After the video, the teacher engages the children, focusing on "*wh*" questions. She might ask, "Where does a katydid live? What is an insect?" The words are then reinforced using an information book (i.e., in this case, on insects) specially designed to review the words just learned (e.g., examples of Tier 2 words: *antennae*, *segments*, *camouflage*, *familiar*, *wings*, *outside*) and to provide redundant information in a different medium.

On subsequent days, the teacher increasingly supports children's vocabulary learning using additional videos that focus on new words in and outside the category, helping to build children's knowledge of the properties (e.g., insects have six legs and three body segments) that are related to the category. In addition, videos and teacher's questions deepen children's knowledge of the concept by providing information about the topic (e.g., insects live in a habitat that has the food, shelter, and weather they like). Following the video, the teacher uses the information book and picture cards to engage children in sorting tasks, including words that are not clearly in or out of the category (e.g., Is a bat an insect?), challenging children by giving them problems to solve, such as "Time for a challenge." Last, the children review their learning through journal writing activities that involve developmental (phonic) writing.

The 8-day instructional sequence is designed to help teachers scaffold children's learning. In the beginning, for example, the teacher's lesson plan focuses on explicit instruction, helping children to "get set"—providing background information—and "give meaning" to deepen their understanding of the topic. As the instructional sequence progresses, the teacher begins to "build bridges" to what children have already learned and what they will learn (establishing intertextual linkages across media). Here, the teacher begins to release more control to the children. Finally, the teacher is encouraged to "step back," giving children more opportunities for open-ended discussion. At the end of the instructional sequence, children are given a "take-home" book—a printable version of the information book used in the lesson. Throughout the sequence, familiar words are used for helping children talk about a topic and for incorporating the approximately 5 to 7 Tier 2 new words for each topic into more known contexts. All eight topics follow a similar instructional design format.

Overview of the Methodological Design and Research Questions

Design research investigates the process of design. It emerged as a recognizable field in the early 1960s, resulting from a conference on design methods at Imperial College London in 1962 (Cross, 1984). What followed was the founding of the Design Research Society in 1966, with its stated aim, "to promote the study of and research into the process of designing in all its many fields" (Design Research Society, n.d.) Since then, the field has become a coherent discipline of study, with many new journals, books, and conferences devoted to a design perspective (for a review, see Laurel, 2003).

Design experiment in education, a term most closely associated with Brown (1992) and Collins (1992), was originally introduced in education to develop theories of instruction rather than merely empirically identify what works (Shavelson, Phillips, Towne, & Feuer, 2003). Often seen as a test bed of innovation (Cobb et al., 2003), it involves introducing an intervention in a naturalistic setting and then studying how it functions to support learning. The purpose of a design experiment is to better understand the evolution of learning that takes place and, ultimately, the procedures and instructional tools that work in real-world classrooms. There is the expectation that researchers may have to systematically adjust various aspects of a design so that each adjustment serves as a type of experiment allowing researchers to generate and test theories in classrooms (e.g., instead of in a laboratory). (Also see Reinking & Bradley's [2008] discussion of *formative experiment*, a term often used synonymously with *design experiment*.) The designed context is subject to study and revision, and the successive iterations that result play a role similar to that of systematic variation in experiments. As a practical matter, design experiments are usually conducted in a limited number of settings (Clements, 2007).

For us, the design experiment was an ideal methodology to systematically focus on two facets of word learning: word selection and word organization. Specifically, our goal was to implement an intervention-in-development in classrooms (e.g., under Goal 2, Institute of Education Sciences, 2011, p. 8), to study it, and then to adjust it in ways that could help us better understand how these two facets of word learning influenced children's inferences and generalizations. We describe each pedagogical issue in greater depth below.

Word Selection

Surprisingly little research has focused on word selection in vocabulary training (Beck & McKeown, 2007). More often than not, curriculum builders have selected words subjectively on the basis of what might be considered unfamiliar to children or opportunistically on the basis of the existing instructional materials in hand. Recently, however, there have been two approaches proposed for word selection. Beck, McKeown, and Kucan (2002), for example, have argued that words for vocabulary instruction should be selected from the portion of word stock that comprises high-utility sophisticated words (Tier 2) that are characteristic of written language (e.g., *commotion* for *noisy*). Biemiller (2006), on the other hand, has argued for greater breadth of word knowledge, focusing instruction on words children will learn more readily—words that constitute between 40% and 70% of a target student group's knowledge—because the greatest gains can be made on these words. Although both approaches have been independently examined (Beck et al., 2002; Biemiller & Boote, 2006), the relationship between the degree of word difficulty and word retention has not been explored.

To examine this issue, we selected words in each topic on the basis of content areas (e.g., science and health) designated in early learning standards. Of these words, 5 to 7 were considered Tier 2 words, or academically sophisticated words, and 10 partially familiar words. These words were analyzed for difficulty level using the lexical norming sample of the MacArthur-Bates Communicative Developmental Inventories (MCDI; Dale & Fenson, 1996) to identify words considered known (acquired) by normally developing 3-year-olds. This database is a set of parent-report inventories of child language and communication designed to yield information on the course of language development within a population. The MCDI has strong concurrent and predictive associations with other measures of vocabulary, language, and cognitive development.

We also used a set of corpora from the CHILDES database (MacWhinney, 2000), a database consisting of transcriptions of adult-child spoken interactions in different home and lab settings around the world. We selected a combination of English-American corpora focusing on young children younger than 5 years of age from a variety of socioeconomic backgrounds ranging from high-risk families to professional families. The corpora represented a broad range of speech samplings (2,740 sessions) from 489 children.

After analysis of these two instruments, we created a norming database of word difficulty, that is, words reported to be typically acquired by 3 years (Dale & Fenson, 1996). These words had a mean rate of acquisition of 96% in the CHILDES corpora (MacWhinney, 2000); only 5% were words not found on the MCDI. We selected approximately equal proportions of familiar and unfamiliar words (on the basis of the above corpora), with 56% of the primary words considered unfamiliar to preschoolers in Unit 1 and 54% in Unit 2. As part of our iterative design, our goal was to look at the types of words learned immediately following instruction in the first unit and to adjust our instructional design for the second unit if the pattern of learning observed indicated the need for changes.

Word Organization

The organization of knowledge is a central feature of cognitive ability in early and later learning, and conditions of instruction can significantly influence the kinds of knowledge structures that are acquired (Glaser, 1984). Because children's learning of words, even the simplest names for things, such as *dog*, involves mapping a form onto a concept, the final design feature of WOW was to organize words within topics to take advantage of their category membership. For example, we proposed that learning words such as *antennae*, *segments*, and *wings*, all taught as properties of insects, could help children develop both a mental representation and a strategy for efficiently storing new information.

Richly structured categories tend to be taxonomic (groupings of like things, e.g., pets) rather than thematic (groupings of things that interact, e.g., things you do in a grocery store; Markman & Hutchinson, 1984). They have similar properties (i.e., pets: dogs and cats are animals that live with people) and fall into an intermediate level of abstraction. Studies suggest that category formation can be both economic and generative (Deak, 2002; Murphy & Lassaline, 1997): economic because it allows individuals to attend to and store only the key properties that determine membership, and generative because it supports generalizations and inferences (Murphy, 2004; E. E. Smith, 1995).

Research suggests that preschool-age children typically demonstrate a preference for thematic thinking but that this preference becomes taxonomic with the onset of schooling (Scheuner, Bonthoux, Cannard, & Blaye, 2004; Sharps, 1992; Smiley & Brown, 1979). Therefore, we proposed that by teaching words through taxonomies, one might induce an earlier shift in preference, reflecting a change in semantic organization.

To examine this instructional design feature, we semantically grouped words in each topic that helped to explain the properties of a topic; we then also included words that helped teachers talk about the topic. For example, semantically grouped words in exercise included *muscles*, *bones*, *heart*, *lungs*, and *intestines*. Words used to talk about these features included *strengthen*, *movement*, *activity*, and others (see appendix).

Therefore, the purpose of the design experiment was to research, test, and iteratively derive principles of word learning and word organization that could help to engineer a learning process and theoretically advance our understanding of vocabulary development for low-income preschoolers. Specifically, we asked the following questions: (a) How might word difficulty influence knowledge and retention of words? (b) Could learning words in categories support vocabulary and inference generation? and (c) How might treatment children differ from a comparison group that did not receive the intervention? We addressed these questions in two design phases, focusing on two 8-week units of instruction of four topics each, for a total intervention period of 16 weeks.

The Design Experiment: Phase 1

Method

Research sites and participants. This study was conducted in two elementary school Head Start programs. The programs were selected on the basis of proximity and initial enthusiasm for and commitment to the project on the part of each program's supervisors and the target population served, specifically, low-income preschoolers. Three teachers were selected to participate in each site in the treatment; all had morning and afternoon classrooms.

Table 1. Sample and Classroom Characteristics

Characteristic	Treatment (n = 6 classrooms)	Comparison (n = 6 classrooms)
Participating teachers		
Average age	41	34
Number of years teaching	10	8
Education		
BA	100%	100%
Ethnicity		
Caucasian	100%	83%
African American	17%	
Participating children		
Average age (in months)	51	50
Ethnicity		
Caucasian	63%	49%
African American	25%	30%
Middle Eastern	12%	21%
PPVT mean score	87.29	87.28
Classroom environment		
ELLCO mean score (124 possible)	83	70

Note: PPVT = Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997); ELLCO = Early Language and Literacy Observation (M. Smith & Dickinson, 2002).

Design experiments do not usually require comparison or control classrooms as in conventional experiments (Reinking & Watkins, 2000). Rather, researchers often establish a "counterfactual" (Hollister & Hill, 1995), a comparison group that can allow one to obtain some initial evidence for how an intervention or innovation might influence behavior. The goal is to develop a textured picture of what is happening through instruction and its implications for adapting the intervention to meet the pedagogical goal.[2] To enrich our data and understanding of the intervention's potential to enhance vocabulary, six additional classrooms were selected at random from the same sites to serve as a comparison group.

Demographic characteristics of the sample are provided in Table 1. All teachers were female. One was African American; the others, Caucasian. All had a bachelor's degree and considerable teaching experience (ranging from 4 to 20 years). Class size ranged from 15 to 18 children in each session (treatment = 89; comparison = 89), for a total sample size of 178 children. All treatment and comparison classrooms reported to use High/Scope (Hohmann & Weikert, 1995) as their curriculum framework and to follow the Head Start child outcome standards in health and science.

Prior to the start of the study, members of the research team individually administered a WOW expressive vocabulary test (described below) and the Peabody Picture

Vocabulary Test (PPVT) to examine children's receptive vocabulary in treatment and comparison classrooms. Six graduate students in educational psychology, trained and certified prior to their work in the field, conducted the assessments. *T* tests revealed statistically significant differences between groups on the WOW expressive language test, $F(1, 176) = 3.21$, $p < .05$, in favor of the treatment condition. However, there were no significant differences between groups on the PPVT standardized scores, $F(1, 176) = .78$, *ns*.

The Early Language and Literacy Observation (ELLCO; M. Smith & Dickinson, 2002) was conducted to examine the quality of the literacy supports in the environment in classrooms. Two trained research assistants in language and literacy observed classrooms for approximately 1.5 hr prior to the intervention and assigned ELLCO scores. Interrater reliability between these observers was .90. Resulting mean scores are displayed in Table 1. There was no significant difference in classroom quality found between treatment and comparison classrooms as measured by the ELLCO, $F(1, 10) = .51$, *ns*.

Approach to analysis. The first phase of our design experiment was conducted in collaboration with the six teachers in the treatment group, four graduate research assistants, and a project director. Throughout the experiment, we viewed our relationship as a partnership (Reinking & Bradley, 2008). As part of the process of refining the curriculum, we held monthly debriefing sessions in which we shared and interpreted observations with teachers and discussed and planned for making subsequent adjustments.

Our research team visited classrooms twice each week, generating a comprehensive record of activities in progress throughout the experiment. Using the lesson plans as our guide, we documented the evolving issues to be discussed within our research team as supporting or questioning our conjectures and then shared these with teachers. Teachers did the same, keeping detailed diaries of their activities throughout the project. We also provided teachers with video cameras to document examples of children's activities and their evolving conjectures.

In the first phase of the study, we were interested in how WOW worked exclusively for the treatment group. We devised three sets of instruments to determine how the instructional design of WOW might influence word selection and word organization.

WOW expressive vocabulary test. To examine the extent to which children learned instructed words, we developed an expressive vocabulary test. Five words were randomly selected from each of the instructional topics. Children were shown picture cards for each of the 40 words from a single randomized set and were asked to name each picture. Cronbach's alpha indicated acceptable reliability (alpha = .80). Scores represented the number of cards correctly identified out of 40.

To examine children's immediate recall, 10 words, five "easy" (as identified by Biemiller's criteria) and five "harder" words (as identified by Beck's criteria) were randomly selected and assessed after each topic was taught (Biemiller, 2006; Beck & McKeown, 2007). These "end-of-topic" expressive measures were similar in format to the pre- and postassessment and were designed to measure immediate retention for children in the treatment group.

Word organization. The second task was known as "Picky Peter" and was designed to tap (a) growth in conceptual or categorical knowledge and (b) the use of categorical knowledge to bootstrap children's learning of unfamiliar words. Adapted from a puppet task used by Waxman and Gelman (1986), children were shown a puppet, Peter, and were told, "Peter is picky and only likes insects" (a specific category label). Children were then shown 20 items (some of the items were category exemplars; others did not belong to the category) and were asked to help Peter find the items he would like and to justify their choices. Ten of the items were from the curriculum, and 10 items were novel and not taught in the curriculum. Items were depicted on picture cards with some contextual background (e.g., a raccoon in the woods) because children were expected to infer category membership on the basis of cues from the item. Scores were calculated independently for taught items and not-taught items, allowing us to examine what was learned from the curriculum and what could be inferred from the curriculum.

Children were asked, "What is this?" If the child was unable to name the referent, the assessor provided the label for the child. Then, the assessor asked, "Is it a wild animal?" and "Why do you think a . . . is (is not) a wild animal?" Children's justifications were recorded and transcribed. Cronbach's alpha indicated acceptable reliability ($\alpha = .89$).

Tell Me. The "Tell Me" task was an open-ended measure designed to examine children's use of words and properties of categories prior to and after the unit. Four picture cards, depicting a category (such as pets) for each topic, were shown to child. The assessor said, "I'm going to show you a picture. Tell me what you know about it." Children were given 1 min to talk about the picture. These conversations were audiotaped and transcribed verbatim. Children received a point for each correct word label and 1 point for each property in the categories ($\alpha = .85$).

Procedures. We introduced WOW to teachers in the winter term through a daylong workshop that explained the approach and the instructional design behind its development. Materials were provided, including DVD player, DVD with video clips, information books, picture cards, and instructional guides for each of the topics. Teachers agreed to use the supplementary curriculum during the whole-group circle time for the 10- to 12-min instructional period each day.

A primary goal for our design experiment was to improve the initial design by assessing and revising our conjectures on the basis of our analysis of children's gains in word knowledge and concepts. We used an iterative process that included three sources of data: information from teachers' feedback, observations of the enactment of the curriculum, and children's assessment scores.

Prior to the beginning of the study, pretests were individually administered to children. We then began our iterative cycle. Before each topic, pretests assessed children's productive labeling and understanding of the properties of the categories; following each topic (2-week duration), they received an end-of-topic word assessment. After the unit was completed (8-week duration), they received a post–Tell Me assessment and a Picky Peter task. We reviewed observations among the research team and met with teachers biweekly.

Table 2. Percentage of Words Correctly Identified by Treatment Group in Phase 1

Source	Easy	Hard
Pretest average	78	29
End of topic	87	57
Posttest average	83	46
Average growth pretest to end of topic	9	28
Average drop in retention	4	11

During the enactment of the topic, we conducted weekly observations specifically focused on the alignment between the written lesson plan and its enactment. Our efforts were not designed to examine the fidelity to the lesson plan, in particular; rather, we wanted to learn how teachers might use or adapt the lesson to meet the children's needs. Using the lesson plan as our guide, we took notes on the implementation of the lesson, focusing on such issues as student engagement and attention. We reviewed these notes weekly, comparing our observations across settings.

Phase 1 of this design experiment concluded with a series of focus group discussions to review and debrief with teachers regarding their experiences with WOW. We then began to analyze the quantitative evidence, focusing specifically on our theoretical conjectures about word learning and concept development.

Results: Phase 1

Word selection and word retention. Our first analysis was designed to examine which words were most likely learned and retained. Using our norming criteria, we grouped pretest words into two categories reflecting Biemiller's criteria (e.g., partially familiar, or easy) and Beck's (e.g., sophisticated, or hard; Biemiller, 2006; Beck & McKeown, 2007). Shown in Table 2, the pretest average percentage correct for easy words exceeded the upper range of Biemiller's criteria (e.g., 70%), with children demonstrating knowledge of 78% of these words. In contrast, only 29% of the hard words were correctly identified, indicating that the word selection was, in fact, more difficult. Immediately following instruction, greater gains were made in learning hard words. There was a greater percentage of hard words learned (an increase of 28%) compared to easy words (an increase of 9%).

At posttest, however, the trajectory changed: Children retained more of the easy words than the hard words. There was a dropoff of 4% of the easy words compared to 11% of the hard. These results portray both the growth potential and the retention problem for hard words. Put simply, children appeared to learn them and lose them more frequently than easy words.

Word organization. The next analysis was designed to examine the degree to which children used categorical knowledge for word learning and inference generation. Using the Tell Me task, we coded the average number of words used per category to determine whether these words would be incorporated in their open-ended responses

Table 3. Knowledge of Categories and Properties by Treatment Group in Phase 1

Source	Pretest	Posttest
Average number of targeted words used per category	2.59 (1.57)	2.76 (1.63)
Average properties per category	0.23 (0.48)	0.54 (0.64)
No. of correct sorts		Taught words: 8/Not taught: 8

Note: Standard deviations shown in parentheses.

to a contextually based picture. As shown in Table 3, prior to instruction, the average number of words produced per category was minimal; on average, children used fewer than three words related to the category prior to instruction.

Following instruction, posttests scores showed little improvement. In fact, scores were essentially flat. A similar pattern was evident in the average number of properties children produced prior to and following instruction. On average, children could name fewer than one property of a category prior to instruction; following instruction, although some growth was recorded, scores showed little improvement. These results seemed to suggest that the instructional design did not promote the organizational prosthetic that could help children accelerate word learning and inference generation.

At the same time, results from the Picky Peter task seemed to disconfirm this finding or at least call it into question (see number of correct sorts, Table 3). Unlike the Tell Me task, the Picky Peter measure did not require expressive language; rather, children responded by physically sorting or pointing to a particular category. On the basis of this task, the results seemed to indicate that children were able to use categories to make inferences. They correctly sorted 8 of the 10 words taught in the appropriate category, indicating that they had learned what had been taught. But they also sorted 8 of the 10 words that had not been specifically taught. These results appeared to suggest that children used their understanding of the properties of categories to infer category membership for new words. In other words, categories seemed to serve as a strategy for inference generation.

However, sorting activities are prone to guessing. Given these conflicting results, it was evident that our assessment techniques needed further refinement to examine children's use of categories as a bootstrap for word knowledge. Furthermore, teachers' feedback and observations indicated that we had underspecified the properties of categories and the relation to the words children were learning. We decided to address both of these issues of assessment and instructional design in Phase 2 of our design experiment.

Therefore, in preparation for Phase 2, we developed a series of adaptations to the instructional design both to exploit the opportunities provided in the WOW instruction and to meet its challenges. The purpose of the Phase 2 research was to examine the outcomes of these adaptations, extending our understanding of the theoretical premise underlying the instructional design.

The Design Experiment: Phase 2

Adaptations to the Instructional Design

Word selection. Given evidence of the differential retention rates of easy and hard words, several adaptations were made to the instructional design. From our observations and teacher reports, children appeared to enjoy and use many of the hard words taught throughout the lessons. Teachers' feedback suggested that they liked the complexity and the sounds of many of the words, such as *camouflage* and *habitat.* Consequently, we did not believe the differential retention rates were attributable to children's lack of enthusiasm for learning hard words. Rather, we conjectured that hard words, defined by their nature as outside children's existing lexicon, might need additional practice and review. Previous research by Beck and her colleagues (2002) and Nagy, Anderson, and Herman (1987), for example, had shown that the frequency of exposure to words was tied to word growth. Children needed to hear, say, and practice these words more frequently and in different contexts.

Therefore, we made two design changes to lessons in the second phase of our design experiment. The first was to include a review of hard words. Together with teachers, we developed a strategy that would involve the children in quick call-and-response questions designed to elicit each of the difficult words. This strategy could allow for a brief but frequent review of words, keeping the daily lesson within the 10- to 12-min time constraint. For example, in the insect topic, the teacher might ask, "What body part do insects use to feel?" "Where do insects live most of the time?" and "How do moths hide from their enemies?" Teachers would repeat the responses and cue children to do so as well.

The second design change was to include more review and practice across topics. Teachers believed that a process of continual review, essentially building meaning of these words throughout the topics, would be useful. Because topics within units were tied to superordinate categories, the repetition and practice of words would allow teachers to extend children's understanding, build bridges between topics, and provide opportunities for greater frequency of word use as well as greater depth of understanding. We decided jointly that the "time-for-a-challenge" activity, asking children why words belong either in or out of category, could prove to be an optimal time to practice hard words in different contexts. Together, these adaptations would provide teachers with approximately a 20% increase in opportunity to review and practice hard words more frequently.

Word organization. Given the conflicting results about the potential power of taxonomic categories to scaffold word learning and inference generation, we made several adaptations to the instructional design. Furthermore, to better understand children's use of categories in word learning, we made an additional adaptation to our assessment approach.

The first instructional design change was to highlight the properties that were integral to each topic. For example, with insects, we included properties such as these: Insects always have six legs, they have three different parts or segments, they have antennae that they use to smell and feel things, and they most often live outdoors. These properties had been integrated throughout the teacher-child interactions in lessons; however, now we placed them on the front page of the instructional guide in a special box to call attention to their importance.

The second design change was to place a special column on the left-hand side throughout the lesson plan to give teachers the rationale for each phase of the activity and its importance for conceptual development, for example, "This part of the lesson links the category-related words and provides them with additional information and additional vocabulary words that they can use to describe and explain a category." By making these design changes, our goal was to help teachers focus on the "big ideas," or concepts to essentially prioritize certain aspects of the lesson. In other words, we wanted to ensure that teachers recognized that word learning and word organization worked hand in hand and that the words were taxonomically clustered. These changes were discussed and demonstrated in a workshop with teachers prior to the beginning of the topics in Unit 2. Therefore, these changes were designed to help teachers redistribute rather than extend the time devoted to the WOW lesson (e.g., review and practice difficult words).

To examine how these changes might affect children's ability to make inferences, we also made an adjustment to the Picky Peter task. First, we went back to our first set of lessons and qualitatively coded children's justifications for making sorting decisions. To conduct this analysis, justifications were compared with properties of the category. Two trained graduate assistants independently read each transcribed justification for 10 children. For example, children received a point for each justification that included properties common to the category; a total content score was calculated for each child and averaged across classrooms. The assistants compared their ratings and interrater reliability was .95. After reliability was established, research assistants independently coded all transcriptions.

Second, using words not taught, we asked children to justify their responses to words in categories. According to the MCDI index (Dale & Fenson, 1996), each of these words would be outside of children's average working vocabulary. Our goal was to understand how categorical instruction might support children's abilities to talk about a topic. In addition, we wanted to compare the differences in responses for children in the treatment and comparison groups.

Together, these adaptations in word selection and word organization were designed to better our understanding of whether our instructional design could enhance, and potentially accelerate, vocabulary development.

Participants

In Phase 2 of the study, we examined growth in word knowledge and categorical learning for both treatment and comparison classrooms.

Table 4. Percentage of Words Correctly Identified by Treatment Group in Phase 2

Source	Easy	Hard
Pretest average	80	42
End of topic	88	64
Posttest average	92	70
Average growth pretest to end of topic	8	22
Average growth in retention	4	6

Procedures

We revised lessons for the treatment group according to the adaptations described. Materials were distributed to teachers, and similar procedures were followed. Before the intervention, treatment and comparison children were assessed on WOW expressive language assessment, the Tell Me task, and the Picky Peter task. However, only treatment children were assessed on word knowledge following each topic to examine immediate gains in word knowledge as a result of the intervention. In addition, we continued to collect observational data during this second phase.

Results: Phase 2

Word selection. Our first analysis was to examine differences between treatment and comparison groups on word knowledge. Given that words were curriculum specific, it was not surprising to find that there were significant differences between groups. Analysis of covariance, with pretest as covariate, indicated that treatment children scored significantly higher on the expressive language assessment than those children in the comparison group, $F(2, 175) = 16.68$, $p < .001$; furthermore, these gains were educationally meaningful (Cohen's $d = .64$), as demonstrated by the strong effect size.

Next, we analyzed word growth specifically for the treatment group after the curriculum adaptations were put in place. Table 4 describes the differences in children's word knowledge at pretest. As in the first phase, differences in word knowledge for easy words compared to hard words are stark. Children knew almost double the number of easy words compared to hard words. Similar to the first phase, following instruction, children made greater growth for hard words than easy. Average percentage growth from pretest to end of unit for hard words was 22%, compared to easy words at 8%.

Retention of word knowledge, instead of declining as in Phase 1, however, increased for both easy and hard words. Average growth in knowledge actually continued to increase: For easy words, increases were modest, but for hard words, increases were more substantial.

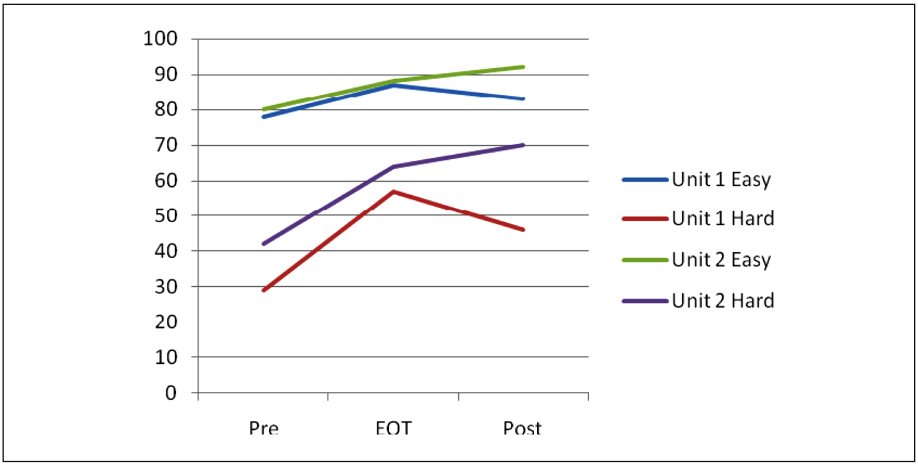

Figure 1. Learning and retention of words in Phase 1 and Phase 2 by word type

Last, we looked at a comparison of gains made in Phase 1 compared to Phase 2. As shown in Figure 1, we found that right after instruction, children seemed to have learned a similar proportion of hard words in both Phases 1 and 2. However, at posttest, children actually continued to gain word knowledge; this was true for easy words as well, albeit less pronounced than with hard words.

These results indicated that initially in both units, children recognized fewer hard words than easy words at pretest. With revision and added frequency of exposure to hard words in Phase 2, children demonstrated increased retention of hard words. Furthermore, it seemed that the cross-topic reviews might have demonstrated greater utility for retention than immediate review. These results suggest that instructional design features of review and practice in different contexts enhanced word knowledge. It also showed that children were capable of learning and retaining hard words.

Word organization. Next, we examined differences between treatment and comparison in children's ability to use words and properties to describe pictures and their ability to identify words in categories. Shown in Table 5, the differences between groups were significant and educationally meaningful, $F(2, 175) = 26.46$, $p < .001$. Children in the treatment group scored significantly higher in expressive language and word properties of categories in their descriptions; furthermore, they were significantly more likely to sort words (both taught and not taught) in appropriate categories, $F(2, 175) = 45.13$, $p < .001$. However, no differences were seen for the uses of word labels on the Tell Me measure at posttesting.

We then conducted a more stringent analysis of children's ability to use categories. We compared treatment and comparison children's justifications, focusing particularly on words that were not taught in the curriculum. For example, shown a picture card of an insect—in this case, *spider*, a word that had not been taught—a child was asked,

Table 5. Treatment and Control Group Means, Standard Deviations, and Differences in Word Knowledge and Categories

Characteristic	Treatment	Comparison	ES
Expressive language (WOW)			
Pretest	16.11* (5.20)	14.11 (4.92)	
Posttest	20.47 (5.68)***	16.98 (5.27)	0.64
Word labels (Tell Me)			
Pretest	17.8 (8.94)	17.0 (10.25)	
Posttest	20.29 (10.09)	18.65 (10.05)	0.16
Word properties			
Pretest	1.32 (2.02)	1.04 (1.49)	
Posttest	3.01 (2.39)***	1.29 (1.64)	0.84
Sorting			
Taught	7.29 (1.12)***	5.90 (1.22)	1.19
Not taught	7.46 (1.11)***	6.34 (1.22)	0.99

Note: Standard deviations shown in parentheses. WOW = World of Words (Neuman, Dwyer, Koh, & Wright, 2007); ES = effect size.
*$p < .05$. **$p < .01$. ***$p < .001$.

"How do you know that it is not an insect?" To receive points, a child would have to provide a justification that included a property of the category, such as "Because it doesn't have six legs."

Our analyses indicated that treatment children were better able to correctly justify their inferences than children in the comparison group. Figure 2 presents the percentage of correct sorting and justifications by group. In short, the treatment group was better able than the comparison group to talk about why they made their choices in ways that demonstrated an understanding of how these words semantically clustered. At the same time, they also used these properties to describe the exclusion of certain words and terms from a particular category, demonstrating cognitive flexibility across tasks and topics.

Table 6 provides several examples of the differences between groups in justifying their choices. In response to the question, "Is a heel a part of the body?" a child who had received instruction reported, "Yes, because it helps you walk," whereas a comparison child not receiving instruction said, "'Cause." Similarly, treatment children were able to apply their categorical information, suggesting that they were using the semantic information about categories to make inferences and generalizations. Categories, therefore, appeared to give children a way to organize words, which became helpful for learning new words. In contrast, children who did not have such information often searched for a rationale that was most immediate to them.

Taken together, these results suggest that the design enhancements appeared to enable children to identify common properties associated with categories and to use

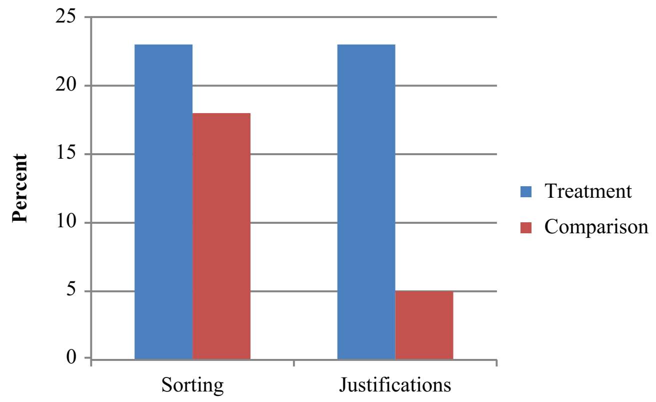

Figure 2. Percentage of correct sorting choices and justifications on categorization: Differences between treatment and comparison groups

this information to make inferences beyond what was taught in the curriculum. By increasing the explicitness of category membership, lessons appeared to better support children's reasoning and knowledge of the concepts words represented. Furthermore, it seemed like taxonomic knowledge acted like a bootstrap for making inferences.

Discussion

Children's vocabularies play an enormously important role in their lives and future possibilities (Beck & McKeown, 2007). A large and rich vocabulary is strongly associated with reading proficiency (National Reading Panel, 2000). Recent cognitive models of reading (Cunningham & Stanovich, 1997; Storch & Whitehurst, 2002) have demonstrated that facility in vocabulary makes a critical contribution to comprehension.

Nevertheless, starting as early as 2 years old (Halle et al., 2009), there are profound differences in vocabulary knowledge among learners from different socioeconomic groups. Particularly disheartening is the finding that when established, differences in vocabulary knowledge remain throughout schooling (Cunningham & Stanovich, 1997). Consequently, there is an emerging consensus that early intervention is critically needed if educators are to substantially improve children's achievement and begin to close the gap on reading performance.

Given that more intensive instruction is needed to increase vocabulary, decisions must be made about the kinds of activities that can potentially accelerate its acquisition. Selecting which words to teach may seem like a primary issue; however, it has received strikingly little attention in curriculum—especially at the preschool level (Neuman & Dwyer, 2009). In fact, Coyne, McCoach, and Kapp (2007) point out that although knowledge about how to teach vocabulary is accumulating, what to teach remains elusive.

Table 6. Examples of Justifications for Children's Category Selections

Word/Picture prompt	Treatment group response	Comparison group response
	Topic: Wild animals	
Key concept/Properties: They live outside/away from people; they live in different habitats (grassland, jungle, or water); they cannot be pets.		
Raccoon (in category)		
How do you know a raccoon is a wild animal?	It lives in the forest.	C: 'Cause.
	'Cause it lives with trees.	C: 'Cause the arrow is pointing to it.
	Because it don't live with people.	
	It lives in the woods.	'Cause it is.
		(silence)
Fish in bowl (not in category)		
How do you know this fish is not a wild animal?	'Cause it lives in a fish bowl.	'Cause.
	It's a pet.	It goes there.
	'Cause it could live with people.	I don't know.
	Topic: Insects	
Key concept/Properties: They are very small creatures/animals; they mostly live outside; they have three body parts called segments; they have six legs; they have special ways to protect themselves from bigger animals; many insects have wings and fly.		
Wasp (in category)		
How do you know a wasp is an insect?	It lives on trees.	'Cause.
	Because it has legs, antennae, wings, mouth, eyes.	I don't know.
	Because it flies.	Because it wants to.
	Three body segments.	It is.
Mouse (not in category)		
How do you know a mouse is not an insect?	It only has one body segments.	'Cause.
	It doesn't have six legs, it just only has four.	Because him want to.
	No, 'cause it don't got this many legs (holds up six fingers).	Because it's not.
		I don't know.

We used the iterative process of a design experiment to extend our theoretical understanding of word selection and word organization. Both Biemiller (2006) and Beck and colleagues (2002) have proposed a heuristic for word selection, though neither has established a normative definition. Using the extant databases of MacArthur-Bates and the CHILDES, we established a normative estimate of each heuristic for words in our curriculum and subjected these to systematic analyses.

Initial evidence suggested that although greater growth occurred for hard words, these words were less likely to be retained. In this respect, it seemed to provide support for Biemiller's position that words should be selected from the portion of word stock that was partially familiar (Biemiller, 2006). However, following our instructional design changes in Phase 2, which involved additional review and practice, growth and retention for hard words substantially improved. Children learned more difficult words and retained them at a higher rate than easy words. These differences were significant when we compared growth in word knowledge between treatment and comparison groups.

Given that instructional time is precious, these results suggest that it may be most facilitative to teach hard words—if sufficient practice and review are provided. These are words that are not only characteristic of written language; they are critical to content learning. Our words, for example, were selected on the basis of content standards regarded by the Fordham Foundation as exemplary (Finn, Petrilli, & Julian, 2006). As Beck and McKeown (2007) have argued, it is precisely these words, comprising sophisticated words of high utility in content areas, that are least likely to be learned outside of school.

The iterative process of the design experiment also allowed us to make conjectures about the role of categorical learning and word knowledge. Studies of early language acquisition (Gopnik & Meltzoff, 1987; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991) have shown a simultaneous growth in the ability to categorize and to acquire new vocabulary. It has been suggested that these two phenomena, the ability to learn new words and knowledge of categories, may be related in a synergistic fashion. Borovsky and Elman (2006), for example, recently tested this assumption through computational simulations, finding in each that improvements in category structure were tightly correlated with subsequent improvements in word learning ability. We proposed in this design experiment that by teaching words in taxonomies along with an articulated set of categorical properties, educators may potentially improve children's word learning and conceptual development. These findings appear to support Gelman and Markman's (1986) research on the generative capacities of categorical learning.

The first test of our theory found equivocal results; the expressive task indicated little improvement on word labeling or identifying properties associated with categories. At the same time, children's ability to sort words in categories, both taught and not taught, seemed to support our thesis. By making the properties of categories more explicit and central to the instructional design, and by seeking children's justifications for sorting, we attempted to learn more about their thinking process in the assessment selection.

Based on these changes, results of the category tasks indicated that treatment children were able to slot familiar words into appropriate categories and to provide a sound rationale for why they were doing so. Furthermore, their knowledge of categories enabled them to better slot words that were not taught into appropriate conceptual groupings. When we examined their justifications, these differences appeared especially striking when compared with other children who had not had such training. These findings suggest that teaching words in categories may represent an important instructional design scaffold for efficiently and economically storing vocabulary. It might also provide greater capacity to attain new information. Schema theorists (Anderson & Pearson, 1984; Rumelhart, 1980), for example, have argued that such frameworks act as a kind organizational prosthetic, serving to diminish information processing load. Given the stark differences in word knowledge between middle-income and low-income children (Hart & Risley, 2003), these results could suggest a significant pathway for accelerating vocabulary development.

Further research is needed, however, to examine whether or in what ways the embedded multimedia might have contributed to our intended goals of word knowledge and word organization. Chambers and her colleagues (Chambers et al., 2008; Chambers, Cheung, Madden, Slavin, & Gifford, 2006), for example, have shown that the use of embedded multimedia can enhance learning, reporting a moderate effect size when compared with instruction without media. Basing their research on the dual coding theory (Paivio, 2008), these colleagues found that the visual and verbal information helped children develop models of knowledge that could be stored and retrieved for subsequent use. Previous studies conducted by the first author (Neuman, 2008) have supported this thesis, suggesting a synergy among high-quality media, which subsequent studies have shown to be particularly effective for word learning (Silverman & Hines, 2009). Others (Goldsen, 1977), however, suggest that multimedia might be distracting for young children in the early years. Given the increasingly complex multimedia environment, we intend to examine this feature in our further research.

The decision to use a design experiment was shaped by the daunting nature of challenges in curriculum development. Too often, research has merely compared the effectiveness of one instructional program against another (Reinking & Bradley, 2008). In such experiments, researchers work to control the influence of design factors rather than to understand them. In contrast, in this research, we worked from a strong theoretical foundation to guide our iterative process toward the goal of improving vocabulary development for low-income children. Design experiments, therefore, may fill an important methodological gap by enriching our understanding of how interventions work and revealing insights that may be useful in future efficacy and effectiveness trials.

Nevertheless, design studies are not stringent experiments. For example, we did not randomly select teachers to participate in our projects; similarly, we did not randomly select treatment and control groups. Rather, following the guidance of previous design research (Reinking & Pickle, 1993), we established a counterfactual to calibrate how our treatment children might react to our stimulus materials compared to others who did not receive them. Consequently, although our findings are promising, we must be cautious in our interpretations until more controlled studies are implemented.

With these limitations in mind, our effort in this design experiment was to expand and deepen our knowledge of instructional practice and how instructional design features can be changed to reach the targeted outcomes. In this respect, we focused on two dimensions of information: the outcomes of new instructional practice and the instructional features that are required to engage in that practice. Knowledge along both dimensions is critical for the creation of evidence-based instructional materials.

Appendix

Words Taught in Curriculum

Unit 1: Living things

Pets	Wild animals	Animals in water	Insects
dog	polar bear	goldfish	ant
puppy	coyote	shark	moth
rabbit	giraffe	whale	bee
cat	leopard	stingray	katydid
kitten	rhinoceros	dolphin	ladybug
bird	elephant	starfish	butterfly
hamster	zebra	seahorse	grasshopper
goldfish	gorilla	octopus	wasp
lizard	deer	crocodile	mosquito
horse	tiger	crab	lizard
snake	seal	gills	snail
pig	monkey	fins	worm
feed	alligator	scales	bat
food	lion	ocean	centipede
water	rooster	lake	spider
play	deer	pond	antennae
tame	habitat		segments
love			wings
			camouflage

Unit 2: Healthy habits

Exercise	Emotions	Healthy foods	Parts of the body
jumping	happiness (happy)	vegetable	face
dancing	cheerful	carrot	checks
jogging	sadness (sad)	broccoli	eyebrows
swimming	lonely (loneliness)	celery	forehead
hopscotch	frustrated (frustration)	lettuce	knees
riding	loving (love)	tomato	elbows

(continued)

Appendix (continued)

Exercise	Emotions	Healthy foods	Parts of the body
playing	angry (anger)	fruit	torso
biking	mad	apple	organs
climbing	afraid	banana	shoulders
stretching	scared	strawberry	abdomen
strengthen	hungry	dairy	heart
healthy	tired	yogurt	lungs
muscles	feelings	protein	brain
movement	alone	grains	attached
heart	bother	energy	senses
bones	company	nutritious	smell
active	comfortable	'junk food'	taste
			hear
			sight

Declaration of Conflicting Interests

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

Funding

This program of research was funded by the Corporation for Public Broadcasting/Public Broadcasting System through the Office of Innovation and Improvement, U.S. Department of Education.

Notes

1. All clips have been specially selected from the archives of *Sesame Street* and *Elmo's World*. Clip length varies from 40 s to 1.5 min.
2. In conventional experiments, researchers try to control the influence of most situational factors; in design experiments, they try to understand them. In this respect, the use of a comparison group allowed us to understand what happens when the intervention is implemented that might be different from traditional instruction (Dickinson & Caswell, 2007).

References

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. Kamil & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255-291). New York, NY: Longman.

Beck, I., & McKeown, M. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, *107*(3), 251-271.

Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life.* New York, NY: Guilford Press.

Biemiller, A. (2005). Addressing developmental patterns in vocabulary: Implications for choosing words for primary grade vocabulary instruction. In E. H. Hiebert & M. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223-242). Hillsdale, NJ: Lawrence Erlbaum.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 41-51). New York, NY: Guilford Press.

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

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, *33*(4), 759-790.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141-178.

Chall, J., Jacobs, V., & Baldwin, L. (1990). *The reading crisis: Why poor children fall behind.* Cambridge, MA: Harvard University Press.

Chambers, B., Abrami, P., Tucker, B., Slavin, R., Madden, N. & Cheung, A. (2008). Computer-assisted tutoring for Success For All: Reading outcomes for first graders. *Journal of Research on Educational Effectiveness*, *1*(2), 120-137.

Chambers, B., Cheung, A., Madden, N., Slavin, R., & Gifford, R. (2006). Achievement effects of embedded multimedia in a Success For All reading program. *Journal of Educational Psychology*, *98*(1), 232-237.

Clements, D. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*(1), 35-70.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9-13.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). New York, NY: Springer-Verlag.

Coyne, M., McCoach, B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction and incidental exposure. *Learning Disabilities Quarterly*, *30*(2), 74-88.

Cross, N. (Ed.). (1984). *Developments in design methodology.* Chichester, UK: Wiley.

Cunningham, A. E., & Stanovich, K. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934-945.

Dale, P., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, and Computers*, *28*, 125-127.

Deak, G. O. (2002). Categorization and concept learning. In J. W. Guthrie (Ed.), *Encyclopedia of education* (2nd ed.). New York, NY: Macmillan.

Design Research Society. (n.d.) Retrieved from http://www.designresearchsociety.org/joomla/index.php

Dickinson, D., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, *22*, 243-260.

Dickinson, D., McCabe, A., & Essex, M. (2006). A window of opportunity we must open to all: The case for preschool with high quality support for language and literacy. In D. Dickinson & S. B. Neuman, *Handbook of early literacy development* (Vol. 2, pp. 11-28). New York, NY: Guilford Press.

Dunn, L., & Dunn, L. (1997). Peabody Picture Vocabulary Test-III. Circle Pines, MN: American Guidance Services.

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174-187.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, *33*(3), 464-497.

Finn, C., Petrilli, M., & Julian, L. (2006). *The state of state standards.* Washington, DC: Fordham Foundation.

Gelman, S., Coley, J., Rosengren, K., Hartman, E., & Pappas, A. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, *63*, 1-194.

Gelman, S., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183-209.

Gelman, S., & O'Reilly, A. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, *59*(4), 876-887.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, *39*(2), 93-104.

Goldsen, R. (1977). *The show and tell machine.* New York, NY: Dial Press.

Gopnik, A. & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, *58*, 1523-1531.

Halle, T., Forry, N., Hair, E., Perper, K. Wander, L., Wessel, J. & Vick, J. (2009). *Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study–Birth Cohort.* Washington, DC: Child Trends.

Hart, B., & Risley, T. (1995). *Meaningful differences.* Baltimore, MD: Brookes.

Hart, B., & Risley, T. (2003). The early catastrophe. *American Educator*, *27*(1), 4, 6-9.

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, *27*(1), 10, 1316-1322,1328-1329, 1348.

Hohmann, M., & Weikart, D. (1995). *Educating young children.* Ypsilanti, MI: High/Scope Press.

Hollister, R., & Hill, J. (1995). Problems in the evaluation of community-wide initiatives. In J. Connell, A. Kubisch, L. Schorr, & C. Weiss (Eds.), *New approaches to evaluating community initiatives* (pp. 127-172). Washington, DC: Aspen Institute.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*, 236-248.

Institute of Education Sciences. (2011). *Request for applications: Education research grants*. Retrievedfrom http://ies.ed.gov/funding/pdf/2012_84305A.pdf

Juel, C., Biancarosa, G., Coker, D., & Deffes, R. (2003). Walking with Rosie: A cautionary tale of early reading instruction. *Educational Leadership*, *52*, 12-18.

Kalish, C., & Gelman, S. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development*, *63*(6), 1536-1557.

Kibby, M. (1995). The organization and teaching of things and the words that signify them. *Journal of Adolescent and Adult Literacy*, *39*(3), 208-223.

Laurel, B. (2003). *Design research: Methods and perspectives.* Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Markman, E., & Hutchinson, J. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1-27.

Moats, L. (1999). *Teaching reading IS rocket science.* Washington, DC: American Federation of Teachers.

Mol, S., Bus, A., & de Jong, M. (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.

Murphy, G. L. (2004). *The big book of concepts.* Cambridge, MA: MIT Press.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 93-131). East Sussex, UK: Psychology Press.

Nagy, W., Anderson, R. & Herman (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237-270.

National Reading Panel. (2000). *Teaching children to read.* Washington, DC: National Institute of Child Health and Development.

Neuman, S. B. (2008). The case for multimedia presentations in learning: A theory of synergy. In A. Bus & S. B. Neuman (Eds.), *Multimedia and literacy development* (pp. 44-56). New York, NY: Routledge.

Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-K. *Reading Teacher*, *62*(5), 384-392.

Neuman, S. B., Dwyer, J., Koh, S., & Wright, T. (2007). *The World of Words: A vocabulary intervention for preschoolers.* Ann Arbor, MI: Ready to Learn Research Program.

Paivio, A. (2008). The dual coding theory. In S. B. Neuman (Ed.), *Educating the other America* (pp. 227-242). Baltimore, MD: Brookes.

Penno, J., Wilkinson, A., & Moore, D. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, *86*(1), 139-153.

Reinking, D., & Bradley, B. (2008). *Formative and design experiments.* New York, NY: Teachers College Press.

Reinking, D., & Pickle, M. (1993). Using formative experiment to study how computers affect reading and writing in classrooms. In D. J. Leu & C. K. Kinzer (Eds.), *Examining central issues in literacy research, theory, and practice* (pp. 263-270). Chicago, IL: National Reading Conference.

Reinking, D., & Watkins, J. (2000). A formative experiment investigating the use of multimedia book reviews to increase elementary students' independent reading. *Reading Research Quarterly*, *35*(3), 384-419.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 34-58). Hillsdale, NJ: Lawrence Erlbaum.

Scheuner, N., Bonthoux, F., Cannard, C., & Blaye, A. (2004). The role of associative strength and conceptual relations in matching tasks in 4- and 6-year-old children. *International Journal of Psychology*, *39*, 290-304.

Senechal, M., LeFevre, J., Thomas, E., & Daley, K. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly*, *33*(1), 96-116.

Sharps, M. (1992). Facilitation of taxonomic recall in preschool children. *Bulletin of the Psychonomic Society*, *30*(2), 137-139.

Shavelson, R., Phillips, D., Towne, L. & Feuer, M. (2003). On the science of education design studies. *Educational Researcher*, *32*(1), 25-28.

Silverman, R., & Hines, S. (2009). The effects of multimedia instruction on the English language learners and non-English language prekindergarten through second grade. *Elementary School Journal*, *101*(2), 305-314.

Smiley, S., & Brown, A. (1979) Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, *28*, 249-257.

Smith, E. E. (1995). Concepts and categorization. In E. Smith & D. Osherson (Eds.), *An invitation to cognitive science* (2nd ed., pp. 3-33). Cambridge, MA: MIT Press.

Smith, M., & Dickinson, D. (2002). *Early language and literacy observation.* Baltimore, MD: Brookes.

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

Storch, S. & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934-947.

Vellutino, F., Scanlon, D., Sipay, E., Small, S., Pratt, A., Chen, R., Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes for specific reading disability. *Journal of Educational Psychology*, *88*(4), 601-638.

Waxman, S. R., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification. *Cognitive Development*, *1*(2), 139-156.

Whitmore, J., Shore, W., & Smith, P. (2004). Partial knowledge of word meanings: Thematic and taxonomic representations. *Journal of Psycholinguistic Research*, *33*, 137-164.

Bios

Susan B. Neuman is a professor in educational studies specializing in early literacy development. Previously, she has served as the U.S. assistant secretary for elementary and secondary education. She has directed the Center for the Improvement of Early Reading Achievement and currently directs the Michigan Research Program on Ready to Learn. Her research and teaching interests include early childhood policy, curriculum, and early reading instruction for children pre-K through Grade 3 who live in poverty.

Julie Dwyer is an assistant professor of education in the Early Childhood Department at Boston University. Her research focuses on early language and literacy learning and teaching, with a particular focus on vocabulary development and the interplay between conceptual knowledge and vocabulary development.