# Educational Effects of a Vocabulary Intervention on Preschoolers' Word Knowledge and Conceptual Development: A Cluster-Randomized Trial

Susan B. Neuman
*University of Michigan, Ann Arbor, USA*

Ellen H. Newman
*IE University, Segovia, Spain*

Julie Dwyer
*Boston University, Massachusetts, USA*

ABSTRACT

The purpose of this study was to examine the hypothesis that helping preschoolers learn words through categorization may enhance their ability to retain words and their conceptual properties, acting as a bootstrap for self-learning. We examined this hypothesis by investigating the effects of the World of Words instructional program, a supplemental intervention for children in preschool designed to teach word knowledge and conceptual development through taxonomic categorization and embedded multimedia. Participants in the study included 3- and 4-year-old children from 28 Head Start classrooms in 12 schools, randomly assigned to treatment and control groups. Children were assessed on word knowledge, expressive language, conceptual knowledge, and categories and properties of concepts in a yearlong intervention. Results indicated that children receiving the WOW treatment consistently outperformed their control counterparts; further, treatment children were able to use categories to identify the meaning of novel words. Gains in word and categorical knowledge were sustained six months later for those children who remained in Head Start. These results suggest that a program targeted to learning words within taxonomic categories may act as a bootstrap for self-learning and inference generation.

Students' knowledge of words and their meanings play an essential role in reading proficiency (Cain, Oakhill, Barnes, & Bryant, 2001; Farkas & Beron, 2004). A large and rich vocabulary is one of the strongest predictors of reading comprehension (Beck & McKeown, 2007). Studies have demonstrated that the size of an individual's word knowledge is related not only to comprehension in elementary grades (Scarborough, 2002; Storch & Whitehurst, 2002) but also to fluency and comprehension in high school (Cunningham & Stanovich, 1997).

Although there is some controversy about the findings and their implications, children from economically disadvantaged circumstances tend to have less extensive vocabularies before they enter school than their middle-class counterparts (Hart & Risley, 1995; Hoff, 2003; for a critique of this research, see, e.g., de Villiers & Johnson, 2007; Miller, Cho, & Bracey, 2005; Stockman, 2010). From this perspective, vocabulary differentials are an important factor contributing to the achievement gap between poor and middle-income students (Farkas & Beron, 2004; Hart & Risley, 2003). By second grade, middle-class students are likely to have acquired around 6,000 root word meanings, whereas students in the lowest quartile on the living word vocabulary list (E. Dale & O'Rourke, 1981) have acquired around 4,000 root words, a gap estimated to equal about two grade levels (Biemiller, 2006).

Compelling as these figures are, they may underestimate the problems associated with vocabulary differentials and school learning (Neuman, 2009). As students get older, they will increasingly need academic vocabularies (Spycher, 2009). These words and their precise meanings are often central to content area understanding and differ from general meanings of even the same terms (Beck, McKeown, & Kucan, 2002). For example, the words *operation* and *sign* have very specific meanings in mathematics. Such academic terms and their specialized meanings may pose the greatest challenge to students who lack a rich network of words and concepts (Stahl & Nagy, 2006).

The question then becomes, How do we effectively intervene with very young students who need more intensive vocabulary instruction? Moreover, how may we potentially accelerate its development? In this study, we attempt to address these questions by evaluating the effectiveness of a vocabulary program designed to promote word learning and conceptual development for preschoolers who lived in urban communities.

## Vocabulary Development and Instruction for Preschoolers

For young children, oral language is the primary source from which they learn new words (Harris, Golinkoff, & Hirsh-Pasek, 2011). Studies have shown that mealtime conversations (Beals, DeTemple, & Dickinson, 1994), daily activities and chores (Tizard & Hughes, 1984), and play (Neuman & Roskos, 1992) provide interactive contexts for word learning. Yet, students are likely to need a wider and more sophisticated vocabulary than what they generally hear in everyday conversations. It is for this reason that book reading, more than any other context, has been the source of study for vocabulary training in the early years. Even simple stories for toddlers like *Over in the Meadow* by Ezra Jack Keats (1999) include complex vocabulary and literary phrases such as "basked in the sun" (n.p.) with a much higher incidence than daily communication (Cunningham & Stanovich, 1997).

Nevertheless, a plethora of studies on the effectiveness of storybook reading have shown equivocal results. Scarborough and Dobrich (1994) and Bus, van Ijzendoorn, and Pellegrini (1995), for example, were among the first to provide a narrative summary and a meta-analysis of the impact of book reading on early literacy skills. Their results provided contrasting views of the power of the effects for shared book reading, with Scarborough and Dobrich calling into question the positive effects often claimed for reading, and Bus and colleagues demonstrating more substantial effects.

Three meta-analyses have subsequently explored the effects of interventions that primarily or entirely focus on shared book reading. Mol, Bus, and de Jong (2009), for example, avoided previous confounds in meta-analytic studies of oral and print-based vocabulary (Elleman, Lindo, Morphy, & Compton, 2009) by separating out the effects on oral language outcomes and print-related skills. Focusing specifically on the impact of interactive storybook reading, they reported a modest effect size for expressive vocabulary (0.28) and a slightly more modest effect size for print knowledge (0.25). However, the largest effect sizes appeared to be present only in experiments that were highly controlled and were executed by the examiners. Teachers appeared to have difficulty fostering the same growth in young students' language skills as researchers did when implementing interventions.

In another recent meta-analysis examining the effects of parent–child storybook readings on oral language development, Mol, Bus, de Jong, and Smeets (2008) found moderate effects for children in the 2–3-year-old age group (0.59) but not for children in the 4–5-year-old age group (0.14). Further, they reported that two groups did not appear to benefit from the intervention: children at risk for language and literacy impairments and kindergarten students. Using a more rigorous set of screening criteria (e.g., studies published only in peer-reviewed journals, randomized controlled trials, quasi-experimental studies), the National Early Literacy Panel (2008) reported moderate effects of storybook reading interventions, with an effect size estimate ranging from 0.35 for composite measures of oral language (e.g., grammar, ability to define vocabulary, listening comprehension) to 0.60 for simple vocabulary.

In short, these meta-analyses appear to support a growing concern voiced by a number of scholars (Beck & McKeown, 2007; Biemiller & Boote, 2006). Although shared book reading represents a fertile ground for vocabulary development, it may not be intensive enough by itself to improve expressive and receptive language development for children at risk. Even in the best of circumstances, Biemiller and Boote found that interventions only yield 20–40% improvement in target word learning, and few read-aloud interventions have shown effects on general knowledge as measured on standardized assessments.

Marulis and Neuman (2010), in the most recent meta-analysis, attempted to address these concerns by examining the full corpus of experimental interventions targeted to enhancing students' oral language development. The researchers examined 67 published and unpublished studies for a total of 216 effect sizes. Their results indicated an overall effect size of 0.88—a gain of nearly one standard deviation on vocabulary measures. However, effect sizes were significantly lower for economically disadvantaged students.

Marulis and Neuman (2010) were able to conduct moderator analyses to try to explain the heterogeneity of variances in effect sizes among studies. These analyses revealed several design features that appeared to be associated with larger effect sizes. For example, providing students with explicit instruction of words in storybooks as well as other materials, discussing words in meaningful contexts, and reviewing words on several occasions was found to be more effective than implicit, embedded instruction alone. Further, training teachers to enact the treatment with fidelity was associated with larger effect sizes. Finally, using assessment measures that were targeted to the specific intervention program showed greater vocabulary gains in studies than did standardized measures. Consequently, these features were among those incorporated into the intervention design.

## Conceptual Development Support for Vocabulary Development

Although students may demonstrate word knowledge through fast mapping (Carey, 1988)—making a connection between an object label and referent within a few instances—recent studies have shown that these mappings can be notoriously fragile over time and with future learning (Gershkoff-Stowe & Hahn, 2007; Wilkinson, Ross, & Diamond, 2003). Students may develop partial knowledge of words from initial exposures, but this knowledge will be far from complete. Rather, depth of processing, which requires meaningful elaboration, appears to support memory for words, stories, and events (Levin, 1988). For example, studies have shown that students who process information more deeply retain information better than those who engage in shallow processing (Stahl & Fairbanks, 1986).

There is an emerging body of evidence indicating that the way in which words are semantically clustered may support word learning (Booth, 2009; Chi & Koeske, 1983; Glaser, 1984). Recent research has shown that when students undergo a vocabulary spurt (McMurray, 2007), a point in development in which the pace of word learning increases rapidly, they also begin to display the ability to categorize. The co-occurrence of these abilities has led researchers to speculate that there is a synergistic relationship between them. Borovsky and Elman (2006), for example, in three computational simulations, manipulated the amount of language input, sentential complexity, and the frequency distribution of words within categories. In each of these simulations, the researchers found that improvements in category structure were tightly correlated with subsequent improvements in word learning ability. The results were consistent with previous research by Gopnik and Meltzoff (1987), who have argued for the bidirectional interaction of categorization as a tool for learning language.

In a recent study, Nelson, O'Neil, and Asher (2008) found that 3- and 4-year-old children learned the labels (i.e., words) for novel artifacts more readily when paired with additional conceptual supporting information about each artifact's function than when paired with supporting information about the artifact's shape or incidental information about the object (e.g., "my sister gave it to me"). Booth (2009) replicated this finding, reporting that 3-year-olds demonstrated greater retention of words when given their conceptual property descriptions as compared with those with nonconceptual properties. Each of these studies suggests that supplementing new word labels with supporting conceptual information may improve vocabulary learning for preschool-age students.

As a facet of conceptual information, category membership has been shown to have a unique potential to bootstrap word learning by linking word labels to existing knowledge through inductive processes (Gelman & O'Reilly, 1988; Medin, Lynch, & Solomon, 2000). That is, once a category has been established, a student may use information about the category to generalize to new instances and make inferences (Rehder & Hastie, 2004). For example, when told that the novel word *katydid* refers to an insect, an individual can infer properties about a katydid based on his or her knowledge of other insects. Children as young as 2 years of age have been shown to use category membership to make novel extensions and inferences (Gelman et al., 1998). Invoking category membership as part of word learning, therefore, may provide a rich background of conceptual and semantic scaffolding for new words. Further, if meaningful elaboration allows for better memory, there is a potential for word learning to be facilitated through concept development. This design feature is examined in this study.

## Vocabulary and Conceptual Development Instructional Design Features

Despite the centrality of word knowledge for developing comprehension and reading proficiency, evidence indicates that there is a lack of attention to its instruction in schools (Beck & McKeown, 2007). Numerous studies have reported the paucity of vocabulary instruction in school curricula (Beck & McKeown, 2007; Biemiller, 2006; Juel, Biancarosa, Coker, & Deffes, 2003). Summarizing much of this research, Biemiller and Boote (2006) found that limited time is spent on increasing vocabulary over the course of students' schooling. For example, Scott, Jamieson-Noel, and Asselin (2003), in their study of 23 upper elementary classrooms, found that teachers did much mentioning and assigning but little actual teaching of new vocabulary.

Unfortunately, available evidence suggests a similar pattern in preschool literacy instruction as well. Wright and Neuman (2010), in a study of 55 kindergarten classrooms, reported virtually no incidences of explicit vocabulary instruction. Further, a recent content analysis of published early literacy programs found little evidence of a deliberate effort to teach vocabulary to preschoolers (Neuman & Dwyer, 2009). The authors reported a mismatch between explicitly stated goals in the scope and sequence, a general pattern of acknowledging the importance of vocabulary but sporadic attention to addressing the skill intentionally, little attention to developing background knowledge, and limited to no opportunities to practice, review, and monitor students' progress.

In short, current instructional materials appear to offer little guidance to teachers who want to do a better job of teaching vocabulary to young students. This means that until instructional materials are developed that emphasize vocabulary and conceptual development early on, less advantaged students may continue to lag behind their middle-income peers even if they master reading the written words.

Given these shortcomings in the research, our goal was to develop an intervention designed to promote vocabulary and conceptual development for preschoolers. Three guiding principles of vocabulary instruction anchored the approach that we developed and evaluated in this study. First, given the limitations of instructional time in preschools, there is an increasing consensus that word selection in vocabulary instruction must be more intentional. Beck and McKeown (2007), 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. These words are domain general and likely to relate to more refined labels for concepts that may enhance students' verbal functioning. Studies of text talk, a strategy used by Beck and her colleagues (2002) to engage students in rich language instruction, have shown impressive results, with kindergarten and first-grade students demonstrating vocabulary gains about twice as large as those in read-aloud studies (Beck et al., 2002). Teaching young students high-utility sophisticated words (Tier 2) even earlier, in the preschool years, therefore, may have great potential to generate vocabulary growth.

Second, words need to be semantically clustered to support conceptual development. Students have been shown to use a variety of different types of category relationships to organize information. Thematic categorization involves the grouping of objects together using relational criteria; *ball* and *bat*, for example, both belong to the schema for *baseball* and are thematically related. Taxonomic categories involve the grouping of objects based on shared properties; for example, *bird* is a taxonomic category consisting of sparrows, robins, and so forth. The shared properties that define taxonomic categories include not just perceptual similarity (e.g., looking the same) but also a shared essence, based on principles of class inclusion between lower and higher level categories (Gelman, 2003). Clustering words within taxonomic categories, therefore, might facilitate inference generation and making inferences and extend word learning.

Third, recent studies have shown that the use of embedded multimedia, strategies in which animations and other video are woven into teachers' lessons, may enhance vocabulary development (Chambers, Cheung, Madden, Slavin, & Gifford, 2006). The use of embedded multimedia is based on two related theoretical premises. One premise is that multimedia can support word learning and concept development through a synergistic relationship (Neuman, 1995). Combining verbal and visual content (i.e., words, pictures) gives learners multiple pathways to retention and comprehension. Kozma (1991) demonstrated that students learned significantly more from multimedia instructional presentations than when materials were presented in one medium alone (see Kozma, 1991, for a review). Further support comes from Mayer and his colleagues (Mayer, 2001; Mayer & Moreno, 2002), who have demonstrated in a series of studies that the addition of moving images, diagrams, and pictures allows for better retention than information held in only one memory system.

The second premise comes from Paivio's (1986) dual coding theory, which posits that visual and verbal information are processed differently, creating separate representations for information processed in each channel. Chambers and her colleagues (2006, 2008), for example, have shown that the use of embedded multimedia can enhance learning, reporting a moderate effect size when compared with instruction without media. Silverman and Hines (2009) also found a positive effect for English learners in prekindergarten through second grade as a result of multimedia-enhanced vocabulary instruction.

Taken together, these design principles form the basis of the World of Words (WOW), a supplemental vocabulary program for preschoolers. The intervention focuses on teaching carefully selected words through richly structured taxonomic categories that are designed to help organize students' understanding of these words and enhance their ability to store ideas efficiently in memory. Lessons are highly interactive, using embedded multimedia—video, audio, and picture enhancements—to support instructional, relational, and conceptual words to extend students' uses of new vocabulary to describe things, solve problems, and draw generalizations and inferences. Together, these features represent the active ingredients of an intervention designed to accelerate word learning and improve conceptual development with the intention of promoting students' long-term achievement.

### *Present Study*

The primary goal of the present study was to examine the effects of WOW, based on the instructional principles delineated previously, for use in preschools with high numbers of economically disadvantaged learners to bolster their vocabulary and conceptual development. The study was designed as a cluster-randomized experiment to meet the criteria for establishing causal predictions, that is, to use random assignment to examine the effects of a vocabulary intervention compared with a control condition among classrooms representing a similar demographic constituency. The study addressed four specific research questions:

1. What is the impact of the vocabulary intervention on word knowledge for preschoolers who come from an economically disadvantaged urban area?
2. Does the intervention enhance students' ability to develop conceptual and categorical development associated with these words?
3. Do potential gains in conceptual development improve students' ability to make inferences and generalizations about novel words and their meanings, providing some initial evidence of cognitive bootstrapping?
4. Are potential gains in word and conceptual development sustained beyond the immediate treatment period?

## Method

### *Study Design and Research Participants*

The study was designed as a prospective cluster-randomized trial between Head Start classrooms. Head Start is a federally funded preschool program targeted to low-income students and designed to promote school readiness through the provision of educational, health, nutritional, social, and other services to enrolled students and families. Having selected the countywide area based on the match between the purpose of the intervention and the instructional goals and outcomes of the Head Start program, the Head Start executive director, directors in schools, and site coordinators agreed to participate in the study. Schools were located in a severely economically depressed urban area in the Rust Belt region of the United States, reporting over 15% unemployment. The Head Start program offered morning and evening classes and full-day programs for students ages 3 and 4 for four days a week, eight months a year. All classrooms served mixed-age groups. Class size was limited to 18 students.

Together with the Head Start management team, 12 schools from five delegate agencies were identified throughout the county to participate; two of the schools were selected from each of four agencies and four of the schools from the fifth agency. Six schools were randomly assigned to treatment and six to the control group. Within schools, classrooms were stratified according to half-day and full-day programs. For each group in five of the schools, one full-day and one half-day classroom were randomly selected, and in the sixth school, two full-day and two half-day classes. In total, 28 classrooms were included in the full-year experiment: 14 classrooms (7 full-day, 7 half-day) in the treatment group and 14 classrooms (7 full-day, 7 half-day) in the control group.

Study participants included 604 3- and 4-year-old students and their head teachers and aides. Table 1 reports the demographic characteristics of our sample. There was comparability across the sample with the

**Table 1. Demographic Characteristics of Treatment and Control Students**

| | Treatment group ($N$ = 294) | Control group ($N$ = 310) |
|---|---|---|
| Average age | 47 months | 47 months |
| Woodcock-Johnson pretest (scaled scores) | 98.4 | 97.5 |
| Female | 55% | 51% |
| Minority | 74% | 75% |
| White | 26% | 25% |
| Black* | 53% | 46% |
| Hispanic | 1% | 2% |
| Asian | 10% | 8% |
| Middle Eastern | 3% | 7% |
| Multiracial | 7% | 12% |
| English as primary language | 96% | 96% |
| Parents' education | | |
| • High school or less | 92% | 90% |
| • Associate's degree | 7% | 8% |
| • Bachelor's degree | 1% | 2% |
| Free or reduced lunch | 100% | 100% |

*$p < .05$.

exception of ethnic status; the treatment group included significantly more African American students than the control group. Most students spoke English as their primary language (96%). Reflecting the eligibility criteria for the Head Start program, all of the students received free or reduced lunch. The majority of their parents had a high school diploma or had dropped out of school.

Twenty-eight teachers participated in the study: 14 teachers (7 full-time, 7 part-time) in the treatment group and 14 teachers (7 full-time, 7 part-time) in the control group. Participating teachers varied in respect to ethnicity, education, and teaching experience. More than a third of the teachers in the treatment group were African American or of African descent (five), two were Middle Eastern, and the remaining 50% were Caucasian; in the control group, 50% of the teachers were African American, and 50% were Caucasian. The control group teachers had significantly more teaching experience (11 years or more; $p = .05$) than the treatment group (5–10 years). Teachers in the control group also had more formal education; they were likely to have an associate's degree ($p = .05$), compared with the treatment group, who were likely to have a child development associate's degree (three courses in child development). The average age of the teachers did not vary significantly between groups; their ages ranged from 41 to 47.

All of the Head Start programs used the HighScope curriculum as their core program (Hohmann & Weikart, 1995). Classrooms assigned to the treatment group received the supplemental WOW intervention (described later) for 12–15 minutes daily in addition to their core program; classrooms assigned to the control group received supplemental activities for a similar time period each day. These control classrooms used materials from the Growing Readers Early Literacy Curriculum (DeBruin-Parecki & Hohmann, 2006), a supplemental program with storybooks and activities in vocabulary, print knowledge, and phonological awareness skills. From activity cards, teachers selected specific strategies or game-like activities to use along with storybook reading. For example, an activity card might include vocabulary words to be identified prior to reading the story and open-ended questions following the reading; the cards can be used flexibly to meet the needs of the students. Consequently, although the focus varied, the treatment and control groups received roughly equivalent amounts of instruction in early literacy–related activity. All programs adhered to the early learning outcome standards approved by the national office of Head Start.

## The WOW Intervention Program

The WOW curriculum (Neuman, Dwyer, Koh, & Wright, 2007) is a supplemental intervention to support vocabulary instruction and conceptual development for pre-K students. Structurally, the curriculum is organized by topics that represent animate taxonomies (e.g., insects) with properties identified for each taxonomic topic (e.g., insects have three segments and six legs). Topics represent content standards in health, science, and mathematics in states that received the highest quality ratings from the Thomas B. Fordham Foundation (Finn, Julian, & Petrilli, 2006). In each state, for example, early learning standards require an emphasis on life sciences through plants and living things, and words that describe the physical characteristics which differentiate plants from animal life.

Within the curriculum, words are selected that represent labels within the category structure (e.g., shoulder, eyebrows). Recognizing that words are conveyors of knowledge, these words, and their meanings, are likely to be encountered repeatedly later on and represent an essential foundation for content learning. We used two databases of children's early language development to calibrate the level of difficulty of words in the curriculum: the MacArthur-Bates Communicative Developmental Inventories (MCDI; P. Dale & Fenson, 1996) and a collection of recordings of child–adult/parent interactions from the CHILDES data set. The MCDI 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 (P. Dale & Fenson, 1996).

We also used a set of corpora from the CHILDES database (MacWhinney, 2000), which consists of transcriptions of adult–child spoken interactions in different home and laboratory settings around the world. We selected a combination of English-language corpora focusing on young children under 5 years of age from a variety of socioeconomic backgrounds ranging from high-risk families to professional families. From this source, we created three norming databases—one for typically developing children, one for bilingual children, and one for high-poverty children—to examine word frequency within and across databases.

In the first set of topics in WOW, we selected approximately equal proportions of familiar and unfamiliar words, based on the previously discussed corpora, with 56% of the primary words considered unfamiliar to preschoolers. In the second set, we nearly doubled the number of words but kept the difficulty level fixed. In the third set, we both added the number of words and increased the difficulty level. In addition, words that challenge children to think about the category structure were also included (e.g. *hair*, *tears*), along with words to support children's conversations about the taxonomies and their properties. Table 2 provides the target words and their difficulty level, as well as the supporting and challenging words in the units of instruction.

**Table 2. Sample From the World of Words Curriculum Matrix**

Unit 1: Healthy Habits—Sample topics
Total number of words: 50
Percentage acquired by age 3 (MCDI): 44; percentage not acquired by age 3: 56
Ratio of frequency of target words to total lexicon: 1:13

| Topic | Phonological awareness skill | Main concepts | Vocabulary[3] |
|---|---|---|---|
| 1. Emotions | • Rhyming | • Your emotions are your feelings.<br>• Emotions are things you feel inside.<br>• Other people can know about your emotions if you tell them how you are feeling.<br>• Your family, friends, and teachers can help you feel happy.<br>• Sometimes, other people can know about your emotions when they look at your face or your body position. People look different when they feel different ways. | ***happy, happiness, cheerful, sad, sadness, lonely, loneliness, frustrated, frustration, loving, love, angry, anger, mad, afraid, scared, tall, short, curly hair, straight hair,*** *hungry, tired, feelings, feel, smile, laugh, fun, cry, bad, better, nobody around, alone, company, bother, interrupted, hug, hit, push, safe, comfortable* |
| 2. Healthy foods | • Rhyming | • Healthy foods are foods that are good for your body.<br>• Healthy foods give your body energy. Energy keeps your body active.<br>• Some healthy foods help make your bones and muscles strong.<br>• Healthy foods come in many different colors.<br>• There are different types of healthy foods. You should have each type of food every day.<br>• Healthy foods taste delicious! | ***vegetable, carrot, broccoli, celery, lettuce, tomato, fruit, apple, banana, strawberry, dairy, milk, yogurt, cheese, protein, meat, chicken, fish, eggs, grains, bread, rice, pasta, cookie, candy, ice cream, French fries,*** *pizza, cereal, energy, good for you, edible, diet, colors, green, orange, delicious, sweet, nutritious, red, yellow, bones, muscles, snack, sugar, oily, greasy, junk food, balanced* |

Unit 2: Living Things—Sample topics
Total number of words: 80
Percentage acquired by age 3 (MCDI): 46; percentage not acquired by age 3: 54
Ratio of frequency of target words to total lexicon: 1:20

| Topic | Phonological awareness skill | Main concepts | Vocabulary |
|---|---|---|---|
| 1. Pets | • Rhyming | • Pets are animals, and all animals are living creatures.<br>• Pets are animals that live with people. They are tame.<br>• We take care of pets by giving them food and water, loving them, and taking care of them when they are hurt or sick.<br>• Pets eat special food, and good pet food makes them healthy.<br>• There are some ways in which pets are the same, and some ways in which they are different. | ***dog, puppy, rabbit, cat, kitten, bird, hamster, goldfish, lizard, elephant, giraffe, tiger, bear,*** *horse, snake, pig, feed, food, water, play, love, take care of, tame, petting, exercise* |
| 2. Wild animals | • Rhyming | • Wild animals are animals that live outside and away from people.<br>• Wild animals take care of themselves.<br>• Many wild animals are ferocious.<br>• Wild animals have their own habitats. | ***polar bear, coyote, giraffe, leopard, rhinoceros, elephant, zebra, gorilla, deer, tiger, seal, monkey, alligator, lion, cat, cow, hamster, rooster,*** *bird, horse, snake, zoo animals, takes care of itself, finds food, outside, ice, Arctic, fish/fishing, hunt/hunting, desert, ferocious, tame, grasslands, big, plants, survive, carnivore, herbivore, jungle, woods, habitat, river* |

*(continued)*

**Table 2. Sample From the World of Words Curriculum Matrix (continued)**

Unit 3: Math—Sample topics
Total number of words: 50
Percentage acquired by age 3 (MCDI): 40; percentage not acquired by age 3: 60
Ratio of frequency of target words to total lexicon: 1:15

| Topic | Phonological awareness skill | Main concepts | Vocabulary |
|---|---|---|---|
| 1. Geometric shapes | • Word families | • A shape describes how something looks. A geometric shape is a special kind of shape. Geometric shapes have special names.<br>• Each geometric shape has a different number of sides.<br>• Some geometric shapes have corners and some do not.<br>• Things in our world come in many different geometric shapes.<br>• Geometric shapes come in a variety of colors and sizes, but they are still the same shape because of the number of sides and corners. | ***triangle***, ***rectangle***, ***circle***, ***square***, ***pentagon***, ***hexagon***, ***octagon***, ***semicircle***, ***cone***, ***sphere***, ***ice cream cone***, ***house***, *squiggle, cloud, three, sides, corners, points, lines, connected, sail, four, door, ruler, narrow, wide, curved, round, wheel, equal, pizza box, stop sign, solid, party hat, ball* |
| 2. Numbers | • Alphabet<br>• Alliteration | • We use numbers to count things.<br>• When we count, we say one number for each thing that we are counting.<br>• You can use numbers to count big things, and you can use numbers to count small things.<br>• Numbers always go in the same order (e.g., 1 is always next to 2, and 2 is always next to 3).<br>• We can add numbers, and we can take away numbers.<br>• You can use numbers to count and see if something is more, less, or the same amount as something else.<br>• Zero is a special number that we use when there is nothing there. | *Number words:* ***one***, ***two***, ***three***, ***four***, ***five***, ***six***, ***seven***, ***eight***, ***nine***, ***ten***; ***group***, ***patterns***, ***multiply***, ***subtract***, ***addition***, ***measure***, ***calculate***, ***guess***, *calendar, clock, count, forward, backward, before, after, add, more, take away, less, more than, less than, nothing, none* |

*Note*. MCDI = MacArthur-Bates Communicative Developmental Inventories. Bolded words are target vocabulary words; the underlined ones are supporting words.

Structurally, the curriculum is organized across three units: healthy habits, living things, and mathematical concepts. There are four topics in each unit, and each topic is taught over an eight-day period. For example, consider the topic of insects. Each day begins with a tuning in—a rhyme, song, or wordplay video clip shown from a DVD[1] to bring students together to the circle and engage them in playing with language for approximately one to two minutes. Although only a modest amount of time was devoted to this type of language play, we recognized that phonological awareness and vocabulary development have a reciprocal relationship (Dickinson, McCabe, & Essex, 2006), reflecting an interaction between multiple aspects of early language processing (Scarborough, 2002).

As more and more words enter students' lexicon, their underlying representations are thought to become more phonologically detailed to differentiate newly learned words from existing words (Metsala & Walley, 1998). Therefore, we used the tuning in to help students come together as a group and engage in language activity. The tuning in is followed by a content video introducing students to the definition of the category. The first video is designed to act as a prototype of the category, a particularly salient example of the topic (e.g., a katydid). Research in concept development has shown that a prototype may act as a mental model for developing key properties of a category (Gelman & Kalish, 2006).

After the video, the teacher engages the students, focusing on *wh*– questions. She might ask, "Where does a katydid live? What is an insect?" Words are then reinforced using an information book (in this case, on insects) specially designed to review the words just learned (e.g., *antennae*, *segments*, *camouflage*, *familiar*, *wings*, *outside*) and provide redundant information in a different medium. Here, the teacher would read about the topic in a different, meaningful context. Based on research in multimedia, working memory can be increased by using a dual modality rather than a single one (Mayer, 2001). That is, it is more effective to target both the visual and auditory processors of working memory.

On subsequent days, the teacher provides increasing supports to develop these words and uses additional videos that focus on new words within and outside the category, helping build students' knowledge of the properties (e.g., insects have six legs and three body segments) that are related to the category. New words and properties are introduced, and previous ones are reviewed. In addition, videos and teacher questions deepen students' knowledge of the concept by providing information about the topic (e.g., insects live in a habitat that has the food, shelter, and the weather they like).

Following the video, the teacher reads a section from a specially designed information book in which the target words are presented in a new context. Picture cards are then used as a strategy for reviewing information and to engage students in sorting tasks. Students are presented with "time for a challenge" items that require them to problem-solve about the category (e.g., is a bat an insect?). These challenge items are designed to encourage students to apply the concepts that they have acquired to think critically about what may or may not constitute category membership (Wellman & Gelman, 1998).

Last, the students review their learning through journal writing activities that involve developmental writing. In this review, the students engage in expressing their ideas through pictures and print, providing an opportunity to extend what they have learned about the topic (cf. Dyson, 1993). In this respect, the WOW intervention took advantage of multiple instructional components that have been shown to highly support vocabulary development (Neuman & Roskos, 2007): singing, interacting and playing with words, shared book reading of informational text, and writing. Each of these components enabled teachers to support explicit and implicit interactions with words and concepts in a multimedia format.

The eight-day instructional sequence is designed to help teachers scaffold students' learning of words and concepts. In the beginning, for example, the teacher's lesson plan focuses on explicit instruction, helping students get set—providing background information—and give meaning to deepen their understanding of the topic. For example, the teacher would introduce the category of insects by explaining that they have antennae, wings, and segments, and that one type of insect is known as a katydid. As the instructional sequence progresses, the teacher begins to build bridges to what students have already learned and what they will learn by establishing intertextual linkages across media. She might ask the students to compare what they saw on the video about katydids with what they just read in the information book. Here, the teacher begins to release more control to the students during the teacher–student language interactions. Questions engage students in more open-ended responses.

Finally, the teacher is encouraged to step back and give students more opportunities for open-ended discussion. At this point, the teacher might help students focus on what they had learned throughout the topic, and their interests in pursuing more information. At the end of the instructional sequence, students 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 students talk about a topic and incorporating the approximately 10–12 content-specific words for each topic into more known contexts. All topics follow a similar instructional design format. A description of the curriculum sequence is provided in Figure 1.

## *Procedures*

The study began in September 2007 and continued throughout the year, ending in May 2008, with a follow-up, delayed posttest in November 2008. Before classes began in the fall of 2007, Head Start treatment and control teachers received two full days of professional development training. Teachers who were assigned to the WOW condition participated in training on the curriculum. At the same time, control group teachers attended workshops on Head Start early learning outcomes as well as the supplemental curriculum that the teachers used. Both groups attended a four-hour refresher workshop in early winter and received ongoing supervision by site directors once per month during the academic year. For both trainings, district supervisors emphasized the alignment of the curriculum and the Head Start standards. Posttests were completed in May 2008. In November of the following school year, the treatment and control students who remained in Head Start were once again assessed on word knowledge and conceptual development.

## *Student Assessments*

As detailed in Table 3, we administered a battery of standardized and author-created assessments throughout the study. Our purpose was to understand how the curriculum might influence word knowledge and conceptual development. By conducting frequent assessments, we could examine how students were responding and whether the author-created measures that we developed might adequately tap students' understanding. They received pretests for each unit of instruction, followed by eight weeks of instruction and ending with the appropriate posttests.

Prior to the start of the study, a standardized measure was given as a pretest; in the middle of the year, a

**Figure 1. Curriculum Sequence of the World of Words Instructional Program**

| | Component 1: Phonological awareness | Component 2: Content video | Component 3: Book reading | Component 4: In-category picture cards | Component 5: Out-of-category picture cards | Component 6: Challenge words | Component 7: Journal (Writing activity) | Component 8: Review |
|---|---|---|---|---|---|---|---|---|
| Day 1 (Lesson 1A) | X | X | X | | | | | |
| Day 2 (Lesson 1B) | X | X | X | X | | | | |
| Day 3 (Lesson 2A) | X | X | X | X | | | | |
| Day 4 (Lesson 2B) | X | X | X | X | | | | |
| Day 5 (Lesson 3A) | X | X | X | X | X | | | |
| Day 6 (Lesson 3B) | X | X | X | X | X | | | |
| Day 7 (Lesson 4A) | X | X | X | X | X | X | | |
| Day 8 (Lesson 4B) | X | X | X | X | X | X | X | X |

different form of the measure was administered, and a posttest was given at the conclusion of the study.

Students were assessed by trained assessors in a single session prior to and after each unit. Each session lasted 15–30 minutes. Graduate students in educational psychology were trained and certified prior to conducting the assessments. No formal breaks were built into the testing protocol; however, if the tested student showed signs of tiring or inattention, we stopped or picked up at a later time. The order of task presentation was held constant. Students were individually assessed in a quiet area within the school. Throughout the data collection period, the site coordinator monitored the testing to ensure that the assessments were administered in accordance with standardized procedures. In total, there were six testing time points throughout the year.

For each unit, a pretest was administered before the start of the curriculum, and a posttest was administered immediately after its completion. Students in the Head Start control group followed the same testing schedule. Therefore, all students were assessed every eight weeks on author-created assessments.

The author-created measures were piloted prior to assessment with students from the University of Michigan laboratory preschool and a local Head Start program. The measures' final versions used selected items that demonstrated evidence of reliability (i.e., internal consistency) and validity (i.e., content validity) in our pilot studies. Next, we report reliability statistics, using Cronbach's , for each of the author-created measures for this sample population.

## Woodcock-Johnson Picture Vocabulary Subtest (Form A and B)

We used this subtest to assess students' expressive vocabulary. This measure was selected because its norming sample seemed to represent students of varying income and native language status better than other measures, such as the Peabody Picture Vocabulary Test–3 receptive measure or the Expressive One-Word Vocabulary Test. The subtest consists of 42 pictures of words that increase in difficulty. For each item, students are prompted to label the picture; administration is discontinued after a student fails to label six items in a row. Scaled scores were reported based on the Woodcock-Johnson norming sample. In the current sample, the reliability of the measure was = 0.80.

## Curriculum-Related Word Knowledge

We constructed a 40-item WOW receptive vocabulary task to measure the number of curriculum-specific words that students learned throughout each unit of instruction on four topics. Words were randomly selected from the corpora of target words taught throughout each unit. Students were shown three pictures and asked to point to the target word. Of the three pictures, one was the target (e.g., *eyebrows*), one was a thematically related out-of-category distractor (e.g., *glasses*), and one was a taxonomically related in-category distractor (e.g., *toes*). The ordering of picture type was counterbalanced across items, and the order of presentation of items was randomized across students. The total number correct was recorded for each student. Reliability of the measure was = 0.86, 0.90, and 0.92 for Units 1, 2, and 3, respectively.

## Conceptual Knowledge

We designed a 32-item task to measure growth in students' conceptual understanding of target vocabulary for each unit. Four conceptual properties from each topic were selected. Assessment questions were devised to include the target word in a sentence that was related to the concept (e.g., do our <u>legs</u> help our bodies move around?) or not related to the concept (e.g., does a <u>jacket</u> help our bodies move around?). Each conceptual property was tested using both in-category and out-of-category target words to measure students' understanding of when the concept property could be applied to the target vocabulary word and when the concept property could not be applied to the target word. The students heard an equal number of yes and no questions across the assessment, and the order of these questions was fully randomized. Students responded either "yes" or "no" to each question, and a total number of correct responses out of 32 were recorded. Reliability was = 0.81, 0.79, and 0.81 for Units 1, 2, and 3, respectively.

## Categories and Properties Knowledge

Informal observations indicated that many students struggled with the verbal demands in the conceptual development assessments. Therefore, we added an additional conceptual development measure starting in Unit 2. To examine students' conceptual knowledge in greater depth, we constructed a receptive task to identify categories and properties of target words. In this task, students were shown three pictures: a target picture (e.g., a katydid), a picture thematically related to the target (e.g., a twig), and an out-of-category but plausible distractor (e.g., a worm). Students were then asked to identify which item belonged to a particular category (e.g., "Which is an insect?") or identify the item that possessed a particular category attribute (e.g., "Which has three body segments?").

A total of four category-level questions, one for each topic, and a total of eight concept property questions, two for each topic, were assessed. Concept property questions were selected as most representative of the category. For example, students were assessed on the property "all insects have six legs," as it is true of all insects and therefore a critical and defining property of the insects category. Responses were tallied for accuracy on category and property questions and for the unit overall separately (total score possible = 12). Reliability was = 0.90. Following the same assessment protocol, a similar measure using category and property knowledge was also constructed for Unit 3 ( = 0.92), with a possible total score of 12.

**Table 3. Battery of Assessments and Timeline for the World of Words (WOW) Instructional Program**

| Assessments and timeline | Treatment group | Control group |
|---|---|---|
| Beginning in the fall (eight-week instructional session) | | |
| Woodcock-Johnson Picture Vocabulary Pretest | + | + |
| WOW word knowledge | | |
| • Unit 1 pretest | + | + |
| • Unit 1 posttest | + | + |
| Concepts | | |
| • Unit 1 pretest | + | + |
| • Unit 1 posttest | + | + |
| Beginning midyear (eight-week instructional session) | | |
| Woodcock-Johnson Picture Vocabulary Test–Interim | + | + |
| WOW word knowledge | | |
| • Unit 2 pretest | + | + |
| • Unit 2 posttest | + | + |
| Concepts | | |
| • Unit 2 pretest | + | + |
| • Unit 2 posttest | + | + |
| Categories and properties | | |
| • Unit 2 pretest | + | + |
| • Unit 2 posttest | + | + |
| Beginning in the spring (eight-week instructional session) | | |
| WOW word knowledge | | |
| • Unit 3 pretest | + | + |
| • Unit 3 posttest | + | + |
| Concepts | | |
| • Unit 3 pretest | + | + |
| • Unit 3 posttest | + | + |
| Categories and properties | | |
| • Unit 3 pretest | + | + |
| • Unit 3 posttest | + | + |
| End of school year | | |
| Woodcock-Johnson Picture Vocabulary Test | + | + |
| Inductive reasoning | + | + |
| November of the next school year (delayed posttest)[a] | | |
| WOW word knowledge | + | + |
| Concepts | + | + |
| Categories and properties | + | + |

[a]Only given to students who continued in Head Start.

### Inferences and Generalizations

Last, in Unit 3, we added one additional posttest measure to examine students' ability to make categorical generalizations and inductive inferences using familiar concepts applied to novel words. To examine students' ability to extend newly learned category properties to novel words, we designed an extension task. Six novel objects were introduced, two per topic in Unit 3 (i.e., decagon, trapezoid, one thousand, shifting spanner, backhoe, vise). Half of the objects were tested with a concept appropriate to the object's category. For example, "Can you use a shifting spanner to make things?" The other objects were tested using a concept property that was inappropriate for the category. For example, students were asked, "Can you use a decagon to count?"

There were three steps to this task. First, students were asked to identify a novel object from a set of three pictures. This step was to determine whether the object was, in fact, unfamiliar. Second, students were then told the name of the target object and its category membership. For example, a student would be shown a picture of a vise and told, "This is a vise. It's a tool." Third, on the next slide, the student was asked a property question about the category and object. For example, the student was shown the picture of the vise and asked, "Can you use this to make things?" There were an equal number of yes and no responses. Correct responses were tallied, and a total score was derived. A total score of 12 was possible. Reliability of the assessment was  = 0.80.

### Delayed Posttests

Six months later, we returned to Head Start to examine the delayed effects of the intervention on students' word knowledge and categorical and conceptual development. Approximately one third of the students had continued in the program.[2] Using similar design procedures as our previous measures, we randomly selected words and concepts from the entire corpus in the curriculum to construct three assessments: curriculum-based word knowledge, conceptual development, and categories and properties knowledge. The total number of items conformed with each of the previous assessments, and reliability estimates were calculated (0.92, 0.80, and 0.84 for word knowledge, conceptual development, and categories and properties, respectively).

## *Fidelity of Implementation*

Throughout the study, we examined the fidelity of implementation using the lesson plan as our guide. Researcher assistants on a weekly basis observed and examined the presence or absence of five features of the lesson: opening activity (tuning in), content (video and questions), information book reading, discussion and time for a challenge, and developmental writing and review. The final two lessons in the unit included an additional review feature covering words and concepts learned at the beginning of the unit. Teachers received 1 point for each component enacted, and conversely, 0 points if the component was not enacted. Points were tallied then averaged across observations for each teacher, and a percentage was derived to indicate the degree of fidelity for each teacher. Using this procedure, total fidelity ranged from 79% to 100%; average fidelity was 79% for 17% of the teachers, 85% for 33% of the teachers, 92% for 33% of the teachers, and 100% for 17% of the teachers. The average fidelity to the curriculum across teachers was 90%.

Monthly meetings were held with site leaders to provide updates, coordinate schedules, and discuss any challenges that might arise. In total, the intervention included 24 weeks of supplemental instruction.

# Results

We present our results in four parts to address our research questions. First, we examine the impact of the intervention on curriculum-related word knowledge. Next, we measure growth in concepts and knowledge of categories and properties within these concepts. Means and standard deviations are reported in raw scores and percentages for each assessment prior to conducting inferential statistics using raw scores. We then report on students' ability to make inferences and generalizations, using categorical properties to consider unfamiliar words. Finally, we examine Head Start students' knowledge of words, categories, and concepts six months later in a delayed posttest to examine the sustainability of the intervention. Effect sizes were calculated using Cohen's *d*, defined as the difference between the means (treatment versus control) divided by the pooled standard deviation (Hedges & Olkin, 1985). Correlation matrices of the measures in each unit were analyzed to examine the relationships among the author-created measures and the standardized assessment (see the Appendix).

Because of the multilevel nature of the data, we used hierarchical linear models with the treatment condition at the classroom level for the first three analyses. These analyses are more conservative than individual analyses, as they recognize that students are not independent from one another but are clustered within classrooms. To account for this, hierarchical linear models (HLMs) allowed us to partition the variance between students and between classrooms. For each outcome, we first determined whether there was statistically significant variability in the outcome between teachers and calculated the intraclass correlation, the amount of variance in the outcome that existed between students and between classrooms. Next, we estimated student-level effects by including covariates to predict variability between students. In each initial analysis, we used pretest scores as well as age (grand mean centered) and ethnicity (uncentered) as additional covariates. Neither ethnicity nor age was a significant predictor. These covariates, therefore, were eliminated from the subsequent analysis. Socioeconomic status was not entered because of lack of variability in this factor among students in the sample.

Finally, we created a fully conditional model with the pretest score as the covariate to estimate classroom- and student-level effects simultaneously. At the classroom level, treatment condition was our variable of interest and was included as the predictor of between-classroom variance (treatment = 1; control = 0). Fidelity to treatment was added as an additional predictor; however, it was insignificant. All continuous measures were $z$-scored to have a mean of 0 and a standard deviation of 1 so that all coefficients were comparable in size and could be interpreted as effect sizes.

Given that previous studies have reported Matthew effects in vocabulary interventions (Coyne et al., 2010; Marulis & Neuman, 2010), we also sought to examine the effects of students' incoming vocabulary knowledge. We wanted to determine whether the intervention might influence the Matthew effect. To do this analysis, we entered the Woodcock-Johnson Picture Vocabulary Pretest score (group mean centered) in the level 1 equation for each author-developed outcome, for which we had both pre- and posttests. Doing so allowed us to determine the relationship between incoming expressive language and learning outcomes for the sample. We then modeled this slope by including the treatment condition at level 2, creating a cross-level interaction. This analysis allowed us to determine if the relationship between incoming vocabulary and our outcomes differed across the two groups.

Since students were dispersed in different Head Start classrooms for our final analysis six months later, we used the ANCOVA. In this analysis, students' expressive language score on the Woodcock-Johnson Picture Vocabulary Posttest served as the covariate, with group (treatment or control) as the independent variable and individual posttest scores as dependent variables.

## *Impact on Word Knowledge*

Observed pre- and posttest means and standard deviations on curriculum-related word knowledge are reported in Tables 4 and 5 for the treatment and control groups. Prior to the start of the study, we examined differences

**Table 4. Comparisons of Pre- and Posttest Scores on Word Knowledge**

| | Treatment group | | Control group | |
|---|---|---|---|---|
| Unit (range) | Raw score (SD) | Percentage (SD) | Raw Score (SD) | Percentage (SD) |
| **Unit 1** | | | | |
| Pretest (5–39) | 25.2 (7.2) | 63 (18) | 24.4 (6.8) | 61 (17) |
| Posttest (4–40)*** | 30.8 (7.6) | 77 (19) | 27.6 (6.8) | 69 (17) |
| **Unit 2** | | | | |
| Pretest (0–39) | 31.2 (6.8) | 78 (17) | 31.2 (6.4) | 78 (16) |
| Posttest (4–40)*** | 35.2 (6.0) | 88 (15) | 31.6 (6.8) | 79 (17) |
| **Unit 3** | | | | |
| Pretest (1–39)* | 29.2 (7.2) | 73 (18) | 27.2 (8.0) | 68 (20) |
| Posttest (1–40)*** | 32.4 (7.2) | 81 (18) | 28.0 (8.0) | 70 (20) |

*Note*. SD = standard deviation.
*$p < .05$. ***$p < .001$.

**Table 5. Comparisons of Pretest, Midyear, and Posttest Scores on the Woodcock-Johnson Picture Vocabulary Subtest**

| | | Treatment group | | Control group | |
|---|---|---|---|---|---|
| Assessment measure | Range | Standard score | Standard deviation | Standard score | Standard deviation |
| Pretest | 37–156 | 98.4 | 14.29 | 97.5 | 12.73 |
| Midyear | 35–148 | 98.8 | 14.69 | 96.4 | 16.05 |
| Posttest | 33–153 | 100.2 | 13.81 | 98.4 | 12.46 |

between groups using a $t$-test. Although there were modest differences at pretest between treatment and control groups on word knowledge in Unit 1, these differences were insignificant: $t(600) = 1.49$, $p = .14$. Similarly, prior to instruction in Unit 2, there were no significant differences between groups: $t(1, 600) = 0.01$, $p = .93$; however, there were pretest differences in Unit 3 favoring the treatment group: $t(600) = 2.53$, $p = .05$.

Controlling for pretest, the HLM analyses revealed that for Unit 1, there were significant differences between the Head Start treatment and control groups (Cohen's $d = 0.44$; see summary Table 9 later in the article). There were also significant differences for Units 2 and 3. In these two units, however, we began to note an important pattern. In each case, effect sizes for the treatment group in Units 2 and 3 were more substantial than in Unit 1 (Cohen's $d = 0.56$ and 0.86 for Units 2 and 3, respectively).

Despite these gains in word knowledge, however, scores on the Woodcock-Johnson Picture Vocabulary Subtest remained stable throughout the experiment (see Table 5). At pretest, scores for treatment and control groups were statistically equivalent: $t(600) = 0.69$, $p = .49$. In midyear, there was still no substantial difference: $t(600) = 1.55$, $p = .12$. Although students in the treatment group gained slightly more than those in the control at posttest, the HLM analysis (see summary Table 9 later in the article) indicated no significant difference between groups.

## Impact on Conceptual and Categorical Development

Table 6 reports the pretest and posttest means and standard deviations from the analysis of students' developing concepts. Prior to treatment in Unit 1, there were once again small differences between the treatment and control groups, although not significant: $t(600) = 1.14$, $p = .24$. However, before instruction in Unit 2, the difference in means between the groups was significant: $t(600) = 1.95$, $p = .05$. In Unit 3, there were no initial differences between the Head Start treatment and control groups: $t(600) = 0.85$, $p = .40$.

Controlling for pretest, HLM results revealed that for Unit 1, the Head Start treatment group significantly outperformed the control group (Cohen's $d = 0.63$; see summary Table 9 later in the article). This pattern continued throughout Units 2 and 3, with the treatment group significantly exceeding the control group (Cohen's $d = 0.53$ and 0.41 for Units 2 and 3, respectively).

Table 7 presents the results of comparisons between groups on students' knowledge of categories and properties within concepts. This assessment was introduced in Units 2 and 3 and required students to make inferences about these categories and properties in new language contexts. Although there were modest differences between the Head Start treatment and control groups prior to instruction in Unit 2, these were insignificant: $t(600) = 1.17$, $p = .24$.

HLM analyses indicated that the treatment group significantly outperformed the control group on properties and categories in Unit 2 (Cohen's $d = 0.86$; see summary Table 9 later in the article). Similar to Unit 2, there were no significant differences between groups prior to instruction in Unit 3: $t(600) = 0.85$, $p = .40$. After instruction in Unit 3, the treatment group again significantly outperformed the control group on properties and categories. The effect size was still educationally meaningful for treatment students but lower than for the previous unit (Cohen's $d = 0.34$).

**Table 6. Comparisons of Pre- and Posttest Scores on Test of Conceptual Knowledge**

| | Treatment group | | Control group | |
|---|---|---|---|---|
| Unit (range) | Raw score (SD) | Percentage (SD) | Raw score (SD) | Percentage (SD) |
| **Unit 1** | | | | |
| Pretest (0–18) | 16.00 (3.52) | 50 (11) | 15.68 (4.48) | 49 (14) |
| Posttest (0–22)*** | 18.88 (4.16) | 59 (13) | 16.64 (3.32) | 52 (10) |
| **Unit 2** | | | | |
| Pretest (0–24)* | 17.28 (3.52) | 54 (11) | 16.32 (4.16) | 51 (13) |
| Posttest (0–25)*** | 19.84 (5.44) | 62 (17) | 17.28 (4.16) | 54 (13) |
| **Unit 3** | | | | |
| Pretest (0–24) | 17.60 (4.48) | 55 (14) | 16.96 (4.48) | 53 (14) |
| Posttest (0–26)*** | 19.52 (5.44) | 61 (17) | 17.28 (5.44) | 54 (17) |

*Note*. SD = standard deviation.
*$p < .05$. ***$p < .001$.

**Table 7. Comparisons of Pre- and Posttest Scores on Categories and Properties Within Concepts**

| | Treatment group | | Control group | |
|---|---|---|---|---|
| Unit (range) | Raw score (SD) | Percentage (SD) | Raw score (SD) | Percentage (SD) |
| **Unit 2** | | | | |
| *Overall* | | | | |
| Pretest (0–24) | 13.92 (5.28) | 58 (22) | 12.96 (5.04) | 54 (21) |
| Posttest (0–24)*** | 18.24 (5.28) | 76 (22) | 13.92 (5.76) | 58 (20) |
| *Properties* | | | | |
| Pretest (0–12) | 6.72 (2.88) | 56 (24) | 6.12 (2.88) | 51 (24) |
| Posttest (0–12)*** | 8.88 (2.88) | 74 (24) | 6.72 (2.88) | 56 (24) |
| *Category* | | | | |
| Pretest (0–12) | 7.32 (3.60) | 61 (30) | 7.08 (3.24) | 59 (27) |
| Posttest (0–12)*** | 9.24 (3.72) | 77 (31) | 7.08 (3.84) | 59 (32) |
| **Unit 3** | | | | |
| *Overall* | | | | |
| Pretest (0–24) | 12.96 (5.52) | 54 (23) | 12.24 (5.28) | 51 (22) |
| Posttest (0–24)*** | 14.64 (6.24) | 61 (26) | 12.48 (5.76) | 52 (24) |
| *Properties* | | | | |
| Pretest (0–12) | 5.52 (3.36) | 46 (28) | 5.16 (3.12) | 43 (26) |
| Posttest (0–12)*** | 6.60 (3.60) | 55 (30) | 5.16 (3.24) | 43 (27) |
| *Category* | | | | |
| Pretest (0–12) | 7.80 (3.96) | 65 (33) | 7.44 (3.84) | 62 (32) |
| Posttest (0–12)* | 8.52 (3.72) | 71 (31) | 7.56 (3.84) | 63 (32) |

*$p < .05$. ***$p < .001$.

Given the research on the Matthew effect, that is, the phenomenon in which those who are highest in initial vocabulary skills are most likely to benefit from vocabulary interventions (Stanovich, 1986; Walberg & Tsai, 1983), we sought to further understand whether there might be differential effects for students depending on their incoming vocabulary knowledge as measured by the standardized assessment measure. Tables 8 and 9 report the results from these analyses.

Results at level 1 indicated a Matthew effect (see Table 8). On six of the eight word, conceptual, and categorical development outcomes, initial expressive language significantly predicted student outcomes. However, the effects, although statistically significant, were small, ranging from 0.12 to 0.18. For every one-unit increase in initial expressive language, students' learning of words and concepts increased by an average of 0.13 standard deviation units. On two of the eight word, conceptual, and categorical knowledge measures, there were no significant effects.

The intervention, however, did not exacerbate the Matthew effect (see Table 9). On six of the eight author-developed outcomes, there were no significant differences in the relationship between incoming expressive language and outcomes for treatment and control groups. For two outcomes—knowledge of categories and properties in Unit 2 and conceptual knowledge in Unit 3—incoming expressive language was less predictive for treatment students, indicating a reduction in the Matthew effect as compared with the control group.

## *Ability to Make Inferences and Generalizations*

Our next analysis examined the potential impact of treatment on students' ability to make inductive inferences about the meaning of novel words. Although the tools category was taught in Unit 3, none of these words included in this assessment were introduced in the curriculum. Further, the initial step in the protocol determined that these words were unknown to the students. Consequently, the task required them to apply their knowledge of categories and concepts to reason to unfamiliar and novel objects. HLM analysis indicated a significant difference between groups (see Table 9). Results revealed that treatment students scored significantly higher than their control peers in using categories to identify the meaning of new words: 58% for the treatment group compared with 50% for the control group (Cohen's $d = 0.46$). In other words, categories appeared to bootstrap the ability to induce the meaning of novel words in a familiar concept. These results

**Table 8. Estimating the Effects of Incoming Expressive Language on Word, Conceptual, and Category and Property Knowledge**

| | Word knowledge | | Conceptual knowledge | | Category and property knowledge | |
|---|---|---|---|---|---|---|
| Variable | β | Standard error | β | Standard error | β | Standard error |
| **Unit 1** | | | | | | |
| Random effect (intercept) | 0.00 | 0.06 | –0.03 | 0.08 | | |
| *Fixed effect* | | | | | | |
| Pretest | 0.63*** | 0.05 | 0.38*** | 0.09 | | |
| Initial WJ vocabulary | 0.12* | 0.05 | 0.12* | 0.05 | | |
| ***Variance components for random effects***[a] | | | | | | |
| *Intercept* | | | | | | |
| Between-teacher SD | 0.26*** | | 0.35*** | | | |
| Between-teacher variance | 0.07*** | | 0.12*** | | | |
| Chi-square | 75.06*** | | 83.13*** | | | |
| *Initial WJ outcome slope* | | | | | | |
| Between-teacher SD | 0.03 | | 0.10 | | | |
| Between-teacher variance | 0.00 | | 0.01 | | | |
| Chi-square | 16.82 | | 33.73 | | | |
| **Unit 2** | | | | | | |
| Random effect | –0.16* | 0.08 | 0.05 | 0.08 | –0.08 | 0.09 |
| *Fixed effect* | | | | | | |
| Pretest | 0.63*** | 0.06 | 0.40*** | 0.07 | 0.43*** | 0.05 |
| Initial WJ vocabulary | 0.17*** | 0.05 | 0.14* | 0.05 | 0.18** | 0.05 |
| ***Variance components for random effects***[a] | | | | | | |
| *Intercept* | | | | | | |
| Between-teacher SD | 0.31*** | | 0.32*** | | 0.41*** | |
| Between-teacher variance | 0.10*** | | 0.10*** | | 0.17*** | |
| Chi-square | 124.87*** | | 67.73*** | | 104.04*** | |
| *Initial WJ outcome slope* | | | | | | |
| Between-teacher SD | 0.08* | | 0.07 | | 0.09 | |
| Between-teacher variance | 0.01* | | 0.01 | | 0.01 | |
| Chi-square | 38.54* | | 29.45 | | 17.35 | |
| **Unit 3** | | | | | | |
| Random effect | –0.24** | 0.07 | –0.03 | 0.07 | –0.16* | 0.07 |
| *Fixed effect* | | | | | | |
| Pretest | 0.70*** | 0.06 | 0.26*** | 0.04 | 0.39*** | 0.05 |
| Initial WJ vocabulary | 0.01 | 0.04 | 0.14** | 0.05 | 0.09 | 0.07 |
| ***Variance components for random effects***[a] | | | | | | |
| *Intercept* | | | | | | |
| Between-teacher SD | 0.24*** | | 0.25*** | | 0.21* | |
| Between-teacher variance | 0.06*** | | 0.06*** | | 0.04* | |
| Chi-square | 61.55 | | 49.01*** | | 36.16* | |
| *Initial WJ outcome slope* | | | | | | |
| Between-teacher SD | 0.07 | | 0.07 | | 0.11 | |
| Between-teacher variance | 0.01 | | 0.01 | | 0.01 | |
| Chi-square | 26.06 | | 12.69 | | 27.01 | |

*Note*. SD = standard deviation. WJ = Woodcock-Johnson.
[a]Variance components for random effects for the WJ Pretest outcome slope is reported in cases where there was statistically significant variability between teachers in the slope when the slope was unconditional. If there was significant variability between teachers, we allowed the slope to vary randomly. In all other cases, the variance components are not reported, because the slope did not significantly vary between teachers and was, thus, fixed and constrained.
*$p < .05$. **$p < .01$. ***$p < .001$.

**Table 9. Estimating the Effects of the World of Words Instructional Program on Word, Conceptual, and Category and Property Knowledge; Woodcock-Johnson (WJ) Picture Vocabulary; and Inferences and Generalizations**

| | Word knowledge | | Conceptual knowledge | | Category and property knowledge | | WJ picture vocabulary | | Inferences and generalizations | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | β | SE | β | SE | β | SE | β | SE | β | SE |
| **Unit 1** | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Base | −0.15 | 0.08 | −0.32** | 0.11 | | | | | | |
| Treatment | 0.29* | 0.11 | 0.57*** | 0.12 | | | | | | |
| *WJ Pretest outcome slope* | | | | | | | | | | |
| Base | 0.12* | 0.05 | 0.04 | 0.10 | | | | | | |
| Treatment | 0.01 | 0.07 | 0.17 | 0.10 | | | | | | |
| ***Variance components for random effects***[a] | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Between-teacher SD | 0.22*** | | 0.21** | | | | | | | |
| Between-teacher variance | 0.05** | | 0.04* | | | | | | | |
| Chi-square | 55.42 | | 40.37 | | | | | | | |
| **Unit 2** | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Base | −0.43*** | 0.07 | −0.22* | 0.11 | −0.46*** | 0.10 | | | | |
| Treatment | 0.50*** | 0.09 | 0.53*** | 0.11 | 0.72*** | 0.12 | | | | |
| *WJ Pretest outcome slope* | | | | | | | | | | |
| Base | 0.20** | 0.07 | 0.10 | 0.09 | 0.30*** | 0.05 | | | | |
| Treatment | −0.08 | 0.08 | 0.08 | 0.10 | −0.19* | 0.08 | | | | |
| ***Variance components for random effects***[a] | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Between-teacher SD | 0.18*** | | 0.19* | | 0.19* | | | | | |
| Between-teacher variance | 0.03*** | | 0.03* | | 0.04* | | | | | |
| Chi-square | 52.78 | | 34.26 | | 37.59 | | | | | |
| *WJ Pretest outcome slope* | | | | | | | | | | |
| Between-teacher SD | 0.09* | | | | | | | | | |
| Between-teacher variance | 0.01* | | | | | | | | | |
| Chi-square | 37.80 | | | | | | | | | |
| **Unit 3** | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Base | −0.48*** | 0.06 | −0.26** | 0.08 | −0.38*** | 0.05 | −0.09 | 0.11 | −0.21 | 0.04 |
| Treatment | 0.43*** | 0.09 | 0.42*** | 0.11 | 0.41*** | 0.10 | 0.07 | 0.17 | 0.40*** | 0.10 |
| *WJ Pretest outcome slope* | | | | | | | | | | |
| Base | 0.10* | 0.05 | 0.29*** | 0.05 | 0.17 | 0.10 | 0.54*** | 0.05 | 0.10* | 0.04 |
| Treatment | −0.13 | 0.07 | −0.22*** | 0.07 | −0.18 | 0.12 | 0.26** | 0.08 | −0.16 | 0.08 |
| ***Variance components for random effects*** | | | | | | | | | | |
| *Intercept* | | | | | | | | | | |
| Between-teacher SD | 0.10 | | 0.15 | | 0.05 | | 0.40*** | | 0.19** | 0.08 |
| Between-teacher variance | 0.01 | | 0.02 | | 0.00 | | 0.16*** | | 0.03 | 0.05 |
| Chi-square | 27.64 | | 29.85 | | 20.85 | | 232.72 | | 14.25 | |

*Note*. SD = standard deviation. SE = standard error.
[a]Variance components for random effects for the WJ Pretest outcome slope is reported in cases where there was statistically significant variability between teachers in the slope when the slope was unconditional. If there was significant variability between teachers, we allowed the slope to vary randomly. In all other cases, the variance components are not reported, because the slope did not significantly vary between teachers and was, thus, fixed and constrained.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

provided evidence that students were able to use their newly learned category information to make category generalizations and inductive inferences about novel words.

Table 9 reports the summary of all of our HLM analyses. The intraclass correlations exceeded 10% in all cases. After controlling for pretest differences at the student level, treatment classrooms significantly outperformed their counterparts in Head Start in scores on all of our author-created outcome measures. However, there was no significant difference between groups in scores on the Woodcock-Johnson Picture Vocabulary Subtest. Together, these results provide evidence for the effects of the intervention on students' word knowledge, category and concept development, and the ability to make inferences and generalizations beyond what was specifically taught. Developing word knowledge within taxonomies appeared to enable treatment students to create an interconnected knowledge of concepts. These skills are considered essential for later reasoning and comprehension development (Stahl & Nagy, 2006).

### *Sustainability Beyond the Immediate Treatment Period*

To determine whether the treatment might have advantaged students beyond the intervention period, we examined their word knowledge, concepts, and categories six months later. Students in Head Start were eligible for a second year only if they had entered the program at age 3; therefore, our sample of students in Head Start represented only an age-related analysis (ages 4.0–4.6 only). There was no supplemental treatment provided in this year. Table 10 describes the means and standard deviations of those students who remained in Head Start.

Since students had dispersed to different classrooms, we conducted an ANCOVA using the Woodcock-Johnson Picture Vocabulary Test (Form A) as the covariate to examine the retention of words, concepts, and categories after treatment. Results of this analysis indicated that the treatment group was significantly different from the control group on word knowledge: $F(1, 120) = 16.49$, $p < .001$. Knowledge of categories and properties was significantly different as well: $F(1, 120) = 6.17$, $p = .02$. In both cases, the results favored the treatment group. However, there were no significant differences between treatment and control on conceptual development: $F(1, 120) = 1.00$, $p = .75$. In summary, students in the treatment group appeared to retain their advantage in word knowledge and identifying categories and properties but did not retain their advantage in conceptual development.

## Discussion

The primary goal of the present study was to examine the hypothesis that helping preschoolers learn words through categorization and embedded multimedia might enhance their ability to retain these words and their conceptual properties, acting as a bootstrap for self-learning. We examined this hypothesis by investigating the effects of the WOW instructional program, a supplemental intervention for students in preschool designed to teach word knowledge and conceptual development through categorization and embedded multimedia. We subjected our hypothesis to a rigorous experimental trial, focusing not only on students' trajectory of growth but also on their ability to go beyond what was specifically taught, giving us some initial evidence of transfer.

Our focus was to promote oral language development and thinking in categories as a basic mental process known to support problem solving and reasoning (Gelman, 2003). Although students have been shown to use a variety of different types of category relationships to organize information, our focus was on taxonomic categorization. For these categories, items are grouped based on shared properties and are hierarchical, with principles of class inclusion that apply to lower and higher level categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). As a result, they have strong generative properties, ideal for cognitive tasks like

**Table 10. Delayed Posttest by Outcomes Six Months Later**

| | Treatment group ($N$ = 69) | | Control group ($N$ = 63) | |
|---|---|---|---|---|
| Assessment | Raw score (SD) | Percentage (SD) | Raw score (SD) | Percentage (SD) |
| Word knowledge*** | 28.40 (6.40) | 71 (16) | 25.60 (6.00) | 64 (15) |
| Categories and properties* | 14.64 (4.32) | 61 (18) | 13.44 (3.84) | 56 (16) |
| Concepts | 17.92 (4.48) | 56 (14) | 17.60 (4.48) | 55 (14) |

*Note*. SD = standard deviation.
$^*p < .05$. $^{***}p < .001$.

novel word learning and inductive reasoning. Studies by Brown and her colleagues (Brown et al., 1993; Smiley & Brown, 1979) have shown a shift in categorization behavior from thematic to taxonomic sorting as a result of instruction. Therefore, we hypothesized that teaching words in richly structured categories might not only improve word knowledge but also help students draw inferences beyond what was specifically taught.

The results of our study, replicated in each unit of instruction, supported our contention. Students who received the intervention not only learned curriculum-related vocabulary associated with each topic but were also better able to identify concepts and their conceptual properties and categories. In each unit of instruction, there were statistically significant differences and educationally meaningful gains reported between the treatment and control groups. Further, although incoming expressive language was significantly predictive of outcomes on six of the eight author-created measures, its impact was relatively modest in comparison with other vocabulary intervention studies (Coyne et al., 2010) and was not exacerbated by the intervention; in fact, for two outcomes—knowledge of categories and properties in Unit 2 and conceptual knowledge in Unit 3—incoming expressive language was less predictive for treatment students.

It was unsurprising that treatment students performed better on curriculum-related words than those in the control group. One might presume that students learn what is taught. Given the emphasis on these particular words in the curriculum compared with the more general literacy activities for the control group, one would assume that the treatment students certainly had an advantage on these word knowledge assessments. However, this was not the case with the other assessments, which were less susceptible to direct application. Concept, category, and property assessments all required students to apply knowledge in new contexts. In this case, the pattern was very clear and consistent: Students in the treatment group improved significantly in their ability to categorize and conceptualize as compared with their control group counterparts. These results demonstrate the potential of instruction for improving conceptual development among preschoolers.

Our work emphasized the relationship between words, categories, and concepts. Building on the work of Anderson and Freebody (1981), our study supports the knowledge hypothesis, which is the understanding that vocabulary terms may be surface representations of underlying concepts. Bos and Anders (1990), for example, conducted a study comparing the effectiveness of three interactive vocabulary strategies—semantic mapping, semantic feature analysis, and semantic-syntactic feature analysis derived from the knowledge hypothesis—with definition instruction based on the access and instrumental hypothesis. The researchers found that greater gains in comprehension, both short and long term, were associated with interactive strategies designed to emphasize conceptual understandings, lending support to the knowledge hypothesis and interactive learning. Bos and Anders's research further emphasized the importance of providing rich opportunities for students to learn the underlying concepts and their relation to one another.

If further research bears out our findings, this instructional design feature may have potential for structuring knowledge in such a way that could potentially accelerate vocabulary development while simultaneously building a rich network of knowledge that underlies reading comprehension and reasoning. As Stahl and Nagy (2006) have argued, it might not be the size of one's vocabulary per se that ultimately determines how well a person can understand what he or she reads, but rather the rich network of knowledge and concepts that these words represent. Consequently, by encouraging students to think in categories early on, teachers may be developing students' ability to comprehend, reason, and think on their own.

Further, categorization, which is an integral part of conceptual knowledge (Medin et al., 2000), may allow students to organize their knowledge and more efficiently process incoming information. Students in our study were able to use the inductive potential of categories to develop inferences about the meaning of novel words, as shown by the results of the inductive reasoning measure. Once students were given the category, they could use its properties to illuminate some basic understanding of a word not previously encountered. Knowing that a bulldozer is a building tool, for example, treatment students could hypothesize that it was a powerful machine that you could use to move things. Students who had received the intervention were significantly more successful than their control counterparts, providing some initial evidence of bootstrapping. Students used what they had learned about categories to induce knowledge of new words.

These findings substantiate the results of a previous design study with 322 Head Start students (Neuman & Dwyer, in press), focusing on the inductive potential of using categories to develop the meaning of new words. In this design experiment, we used a Picky Peter task that engaged students in sorting words not specifically taught in the curriculum into categories. For example, shown a picture card of an insect—in this case, a spider, a word that had not been taught—a student was asked, "How do you know that it is not an insect?" and asked to provide a justification, such as "because it doesn't have six legs." Results indicated significant quantitative differences and substantial qualitative differences between students in the treatment and comparison groups.

Despite students' gains in word and conceptual knowledge, our study did not show significant improvements in expressive language as measured by the Woodcock-Johnson Picture Vocabulary Test. In fact, students' scores remained rather stable throughout the year. These results need further exploration. One explanation could be that the measure was insufficiently sensitive to vocabulary growth for young students. That said, however, our selection of this particular standardized assessment was based on its norming sample (e.g., its greater sensitivity to diversity, native language status, and age range), relative to other standardized vocabulary measures (e.g., the Peabody Picture Vocabulary Test). It could also be that vocabulary interventions do not necessarily transfer to global vocabulary gains. Recent evidence from a meta-analysis by Elleman and her colleagues (2009), for example, found only a 0.13 effect size for standardized measures. The National Reading Panel (National Institute for Child Health and Human Development, 2000) also predicted a lower estimate of effect sizes when using standardized tests, leading researchers (e.g., Coyne, McCoach, & Knapp, 2007; Sénéchal, Ouellette, & Rodney, 2006) to increasingly rely on author-created measures to detect fine-grained and more comprehensive vocabulary growth.

Yet, the lack of improvement in expressive language might also reflect a limitation of the intervention itself and the context in which it was presented. It could be that students needed a greater dosage of instruction with opportunities to practice words and concepts throughout the day. Our intervention, after all, involved only a modest amount of time each day. Another potential explanation might relate to group configuration. Our intervention was provided in a whole-group setting. As a whole-group activity, teacher interactions may be more global and less responsive to individual students than in a small-group context. Our research team is currently conducting an experiment to examine the relationship between context and students' interactions.

In addition to the work on conceptual development, our research might also add to the growing literature on the selection of words to teach in the early years. As noted by Beck and McKeown (2007), the typical approach in vocabulary interventions for early learners has been to select words that are likely to be unfamiliar to students and that may be important to a story or text, or simply to choose words judged as likely unfamiliar. Recently, researchers have proposed more specific considerations. For example, Biemiller (2006) has advocated focusing on words that are partially familiar, those which 40–70% of a particular age group of students might know, because this is the area in which students might make the greatest gains. In contrast, Beck and her colleagues (Beck et al., 2002) have argued for sophisticated words—Tier 2 words of high utility for mature language users that are characteristic of written language.

Our approach in this study represented yet a third approach: emphasizing the semantic relatedness of words and their contribution to the category and conceptual framework within the topic of study. Although we cannot determine definitively whether this approach contributed to students' outcomes, the results of our delayed posttest suggested that young students both retained words and appeared to relate them to their categorical properties some six months later. In addition, recent evidence from two follow-up laboratory studies have found that semantic relatedness of words within known taxonomies influenced 3- and 4-year-old children's rate of learning (Kaefer, 2009).

Our study had a number of strengths. As a cluster-randomized experiment, schools throughout a countywide program were randomly assigned to treatment. Classroom resources and the staffing structure were similar among all classrooms. Because of the structure of the countywide program, both the timing and frequency of professional development and support were similar across sites and groups. In addition, the treatment and control groups received roughly a similar dosage of language and literacy instruction. All classrooms were supervised by a strong management team. Ongoing progress monitoring headed by a local site educational director ensured that all classrooms focused on areas of early literacy development, health, science, and math instruction as indicated in the Head Start early outcome standards. Our fidelity measure indicated that the teachers used similar practices across the different schools and classrooms and that they could effectively deliver the vocabulary intervention.

Our study also had significant limitations. Although the study was conducted in an economically distressed urban area, the program was well funded with an exceptionally well managed supervisory team, who ensured that teachers understood how to align their existing curricula and supplementary programs with prekindergarten guidelines and standards consistent with these goals. We cannot make a case that these results, therefore, could be generalized to other early childhood contexts or conditions. However, there is no reason to believe that its advantages would only be limited to particular types of programs (e.g., Head Start).

Second, our study would have benefited from an analysis of the active ingredients of the curriculum. The instructional design of WOW was based on several key principles: intentional word selection related to content-rich topics, words semantically clustered to support conceptual development, and the uses of embedded multimedia as a mechanism to instantiate words through dual coding of images as well as words. At this point, it is impossible to disentangle these instructional design features to determine which of them might be the strongest determinant of the effects. Future studies are needed to examine each of the instructional components in greater detail. It would also be useful to gather qualitative evidence of teachers' enactments of the intervention and how it might extend to other areas of the preschool curriculum. We intend to examine these issues in our future work.

Third, our work would have benefited from a more comprehensive set of standardized measures to complement our focus on conceptual development. However, these measures lacked content validity in relation to our instructional goals; further, we found very limited options for students at these age ranges. However, we recognize that it is important to explore the relationship of words and concepts with more conventional standardized measures in the future.

Finally, it would have been ideal to follow those students who went on to kindergarten in addition to those who remained in Head Start. Our initial analysis indicated that these kindergarten students went on to 65 different classrooms in many different private, charter, and public schools. Consequently, although ideal, it was unfeasible to conduct a longitudinal follow-up. As a result, evidence is lacking on the sustainability of word and conceptual knowledge for these kindergartners.

With these limitations in mind, this study provides substantial evidence for the improvement of content-rich vocabulary and conceptual development among pre-K students. Preschoolers learned these words within taxonomic categories and their conceptual properties, which appeared to act as a bootstrap for self-learning and inference generation.

## Notes

[1]All clips have been specially selected from the archives of *Sesame Street* and *Elmo's World*; clip length varies from 40 seconds to 1.5 minutes.

[2]Head Start is typically a one-year program.

## References


Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Beals, D.E., DeTemple, J.M., & Dickinson, D.K. (1994). Talking and listening that support early literacy development of children from low-income families. In D.K. Dickinson (Ed.), *Bridges to literacy: Children, families, and schools* (pp. 19–40). Cambridge, MA: Blackwell.

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

Beck, I.L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

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

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. doi:10.1037/0022-0663.98.1.44

Booth, A.E. (2009). Causal supports for early word learning. *Child Development, 80*(4), 1243–1250. doi:10.1111/j.1467-8624.2009.01328.x

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. doi:10.1017/S0305000906007574

Bos, C., & Anders, P. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly, 13*(1), 31–42.

Brown, A.L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J.C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188–228). New York: Cambridge University Press.

Bus, A.G., van Ijzendoorn, M.H., & Pellegrini, A.D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*(1), 1–21.

Cain, K., Oakhill, J.V., Barnes, M.A., & Bryant, P.E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850–859. doi:10.3758/BF03196414

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language, 3*(3), 167–181. doi:10.1111/j.1468-0017.1988.tb00141.x

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

Chambers, B., Cheung, A.C.K., Madden, N.A., Slavin, R.E., & Gifford, R. (2006). Achievement effects of embedded multimedia in a Success for All reading program. *Journal of Educational Psychology, 98*(1), 232–237. doi:10.1037/0022-0663.98.1.232

Chi, M.T., & Koeske, R.D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*(1), 29–39. doi:10.1037/0012-1649.19.1.29

Coyne, M.D., McCoach, D.B., & Knapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74–88.

Coyne, M.D., McCoach, D.B., Loftus, S., Zipoli, R., Jr., Ruby, M., Crevecoeur, Y.C., et al. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness, 3*(2), 93–120.

Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. doi:10.1037/0012-1649.33.6.934

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary: A national vocabulary inventory*. Chicago: World Book/Childcraft International.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, 28*(1), 125–127. doi:10.3758/BF03203646

DeBruin-Parecki, A., & Hohmann, M. (2006). *Growing readers early literacy curriculum* (2nd ed.) [Kit]. Ypsilanti, MI: HighScope Educational Research Foundation.

de Villiers, J., & Johnson, V. (2007). Implications of new vocabulary assessments for minority children. In R.K. Wagner, A.E. Muse, & K.R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 157–181). New York: Guilford.

Dickinson, D.K., McCabe, A., & Essex, M.J. (2006). A window of opportunity we must open to all: The case for preschool with high-quality support for language and literacy. In D.K. Dickinson & S.B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 11–28). New York: Guilford.
Dyson, A.H. (1993). *Social worlds of children learning to write in an urban primary school*. New York: Teachers College Press.
Elleman, A.M., Lindo, E.J., Morphy, P., & Compton, D.L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1–44. doi:10.1080/19345740802539200
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. doi:10.1016/j.ssresearch.2003.08.001
Finn, C.E., Jr., Julian, L., & Petrilli, M.J. (2006). *The state of state standards 2006*. Washington, DC: Thomas B. Fordham Foundation.
Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195154061.001.0001
Gelman, S.A., Coley, J.D., Rosengren, K.S., Hartman, E., Pappas, A., & Keil, F.C. (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, Serial No. 253).
Gelman, S.A., & Kalish, C.W. (2006). Conceptual development. In W. Damon, R.M. Lerner (Series Eds.), D. Kuhn, & R. Siegler (Vol. Eds.), *Handbook of child psychology. Vol. 2: Cognition, perception, and language* (6th ed., pp. 687–733). New York: John Wiley & Sons.
Gelman, S.A., & O'Reilly, A.W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development, 59*(4), 876–887.
Gershkoff-Stowe, L., & Hahn, E.R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research, 50*(3), 682–697. doi:10.1044/1092-4388(2007/048)
Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39*(2), 93–104. doi:10.1037/0003-066X.39.2.93
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*(6), 1523–1531.
Harris, J., Golinkoff, R.M., & Hirsh-Pasek, K. (2011). Lessons from the crib for the classroom: How children really learn vocabulary. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 49–65). New York: Guilford.
Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.
Hart, B., & Risley, T.R. (2003). The early catastrophe: The 30 million word gap. *American Educator, 27*(1), 4–9.
Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic.
Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368–1378. doi:10.1111/1467-8624.00612
Hohmann, M., & Weikart, D.P. (1995). *Educating young children: Active learning practices for preschool and child care programs*. Ypsilanti, MI: HighScope Educational Research Foundation.
Juel, C., Biancarosa, G., Coker, D., & Deffes, R. (2003). Walking with Rosie: A cautionary tale of early reading instruction. *Educational Leadership, 60*(7), 12–18.
Kaefer, T. (2009, December). *Conceptual support for vocabulary in preschool: Taxonomic category information scaffolds word learning for vulnerable children*. Paper presented at the 59th annual meeting of the National Reading Conference, Albuquerque, NM.
Kozma, R.B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179–211.
Levin, J.R. (1988). Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology, 13*(3), 191–205. doi:10.1016/0361-476X(88)90020-3
MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Vol. 1: Transcription format and programs* (3rd ed.). Mahwah, NJ: Erlbaum.
Marulis, L.M., & Neuman, S.B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. doi:10.3102/0034654310377087
Mayer, R.E. (2001). *Multimedia learning*. New York: Cambridge University Press.
Mayer, R.E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87–99. doi:10.1023/A:1013184611077
McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317*(5838), 631.
Medin, D.L., Lynch, E.B., & Solomon, K.O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121–147. doi:10.1146/annurev.psych.51.1.121
Metsala, J.L., & Walley, A.C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J.L. Metsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.
Miller, P.J., Cho, G.E., & Bracey, J.R. (2005). Working-class children's experience through the prism of personal storytelling. *Human Development, 48*(3), 115–135. doi:10.1159/000085515
Mol, S.E., Bus, A.G., & de Jong, M.T. (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. doi:10.3102/0034654309332561
Mol, S.E., Bus, A.G., de Jong, M.T., & Smeets, D.J.H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development, 19*(1), 7–26.
National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.
National Institute for 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 Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
Nelson, D.G.K., O'Neil, K.A., & Asher, Y.M. (2008). A mutually facilitative relationship between learning names and learning concepts in preschool children: The case of artifacts. *Journal of Cognition and Development, 9*(2), 171–193. doi:10.1080/15248370802022621
Neuman, S.B. (1995). *Literacy in the television age: The myth of the TV effect* (2nd ed.). Norwood, NJ: Ablex.
Neuman, S.B. (2009). *Changing the odds for children at risk: Seven essential principles of educational programs that break the cycle of poverty*. New York: Teachers College Press.
Neuman, S.B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-K. *The Reading Teacher, 62*(5), 384–392. doi:10.1598/RT.62.5.2
Neuman, S.B., & Dwyer, J. (in press). Developing vocabulary and conceptual knowledge for low-income preschoolers: A design experiment. *Journal of Literacy Research*.

Neuman, S.B., Dwyer, J., Koh, S., & Wright, T. (2007). *The World of Words (WOW): A vocabulary intervention for low-income preschoolers* [Curriculum]. Ann Arbor: University of Michigan.
Neuman, S.B., & Roskos, K. (1992). Literacy objects as cultural tools: Effects on children's literacy behaviors in play. *Reading Research Quarterly, 27*(3), 202–225.
Neuman, S.B., & Roskos, K. (with Wright, T., & Lenhart, L.). (2007). *Nurturing knowledge: Building a foundation for school success by linking early literacy to math, science, art, and social studies*. New York: Scholastic.
Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91*(2), 113–153. doi:10.1016/S0010-0277(03)00167-7
Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 438–439.
Scarborough, H.S. (2002). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York: Guilford.
Scarborough, H.S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245–302. doi:10.1006/drev.1994.1010
Scott, J.A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103*(3), 269–286.
Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D.K. Dickinson & S.B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 173–182). New York: Guilford.
Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101*(2), 305–314. doi:10.1037/a0014217
Smiley, S.S., & Brown, A.L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology, 28*(2), 249–257. doi:10.1016/0022-0965(79)90087-0
Spycher, P. (2009). Learning academic language through science in two linguistically diverse kindergarten classes. *The Elementary School Journal, 109*(4), 359–379.
Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72–110.
Stahl, S.A., & Nagy, W.E. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.
Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407. doi:10.1598/RRQ.21.4.1
Stockman, I.J. (2010). A review of developmental and applied language research on African American children: From a deficit to difference perspective on dialect differences. *Language, Speech, and Hearing Services in Schools, 41*(1), 23–38. doi:10.1044/0161-1461(2009/08-0086)
Storch, S.A., & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. doi:10.1037/0012-1649.38.6.934
Tizard, B., & Hughes, M. (1984). *Young children learning*. Cambridge, MA: Harvard University Press.
Walberg, H.J., & Tsai, S. (1983). Matthew effects in education. *American Educational Research Journal, 20*(3), 359–373.
Wellman, H.M., & Gelman, S.A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Series Ed.), D. Kuhn, & R.S. Siegler (Vol. Eds.), *Handbook of child psychology. Vol. 2: Cognition, perception, and language* (5th ed., pp. 523–573). New York: John Wiley & Sons.
Wilkinson, K.M., Ross, E., & Diamond, A. (2003). Fast mapping of multiple words: Insights into when "the information provided" does and does not equal "the information perceived." *Journal of Applied Developmental Psychology, 24*(6), 739–762. doi:10.1016/j.appdev.2003.09.006
Wright, T., & Neuman, S.B. (2010, April). *Explicit vocabulary instruction in kindergarten*. Paper presented at the annual meeting of the Michigan Reading Association, Grand Rapids.

## Literature Cited

Keats, E.J. (1999). *Over in the meadow*. New York: Puffin.

Submitted December 15, 2009
Final revision received February 2, 2011
Accepted February 16, 2011

***Susan B. Neuman*** is a professor in educational studies at the University of Michigan, Ann Arbor, USA; e-mail sbneuman@umich.edu.

***Ellen H. Newman*** is a lecturer at the IE University, Segovia, Spain; e-mail eehamilt@gmail.com.

***Julie Dwyer*** is an assistant professor in early childhood, Boston University, Massachusetts, USA; e-mail dwyerj@bu.edu.

## Appendix

# Intercorrelations Among Assessments in Units 1–3

| Assessment | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unit 1** | | | | | | | | | | | |
| 1. Word knowledge | — | | | | | | | | | | |
| 2. Conceptual development | 0.40** | — | | | | | | | | | |
| 3. Woodcock-Johnson Picture Vocabulary Pretest | 0.44** | 0.28** | — | | | | | | | | |
| **Unit 2** | | | | | | | | | | | |
| 4. Word knowledge | | | | — | | | | | | | |
| 5. Conceptual development | | | | 0.55** | — | | | | | | |
| 6. Categories and properties | | | | 0.54** | 0.45** | — | | | | | |
| 7. Woodcock-Johnson Picture Vocabulary Test–Interim | | | | 0.55** | 0.36** | 0.42** | — | | | | |
| **Unit 3** | | | | | | | | | | | |
| 8. Word knowledge | | | | | | | | — | | | |
| 9. Conceptual development | | | | | | | | 0.65** | — | | |
| 10. Categories and properties | | | | | | | | 0.41** | 0.24** | — | |
| 11. Woodcock-Johnson Picture Vocabulary Posttest | | | | | | | | 0.41** | 0.21** | 0.18** | — |

**$p < .01$.