Educational Effects of an Embedded Multimedia Vocabulary Intervention for Economically Disadvantaged Pre-K Children: A Randomized Trial

Susan B. Neuman

Ellen H. Newman

Julie Dwyer

University of Michigan

Educational Effects of an Embedded Multimedia Vocabulary Intervention for Economically Disadvantaged Pre-K Children: A Randomized Trial

The purpose of this study was to examine the effects of the World of Words (WOW), an embedded multimedia intervention designed to improve vocabulary and conceptual knowledge for low-income preschoolers. Based on the theoretical premise that taxonomic categories may a supportive instructional design feature for word and concept development, WOW is a 12-minute daily supplemental curriculum that teaches content-specific words related to preschool standards in health, science and mathematics. In Phase 1 of the study, Head Start classrooms were randomly assigned to either treatment and control groups (N=604); in phase II, two additional samples representing middle- and highly advantaged children were included to determine the extent to which the intervention might close the vocabulary gap. Children were assessed on word knowledge, expressive language, concept, categories and properties of concepts. Results indicated that Head Start treatment children consistently outperformed their Head Start control counterparts, and actually closed the gap on a number of assessments with their more highly advantaged peers. Further, treatment children were able to use their newly learned category information to make category generalizations and inductive inferences about novel words. These results suggest that a program targeted to conceptual learning may help bootstrap children’s development of knowledge and word learning.

Educational Effects of an Embedded Multimedia Vocabulary Intervention for Economically Disadvantaged Pre-K Children: A Randomized Trial

Children's knowledge of words and their meanings play an essential role in reading proficiency (Cain, Oakhill, Barnes, & Bryant, 2001; Durham, Farkas, Hammer, Tomblin, & Catts, in press; Farkas & Beron, 2004). A large and rich vocabulary is one of the strongest predictors of reading comprehension (Beck & McKeown, 2007). Longitudinal studies (Cunningham & Stanovich, 1997; Lee & Burkam, 2002) have demonstrated that the size of an individual's vocabulary knowledge is related not only to these skills in the elementary grades, but to fluency and comprehension in high school (Stahl & Nagy, 2006).

It is well established, however, that children from lower socioeconomic circumstances have less extensive knowledge of words and concepts before they enter school (Hart & Risley, 1995; Hoff, 2003). Vocabulary differentials lie at the heart of the achievement gap between poor and middle income children. By second grade, middle class children are likely to have acquired around 6,000 root word meanings, whereas children in the lowest quartile around 4,000 root words; a gap estimated to equal about two grade levels (Biemiller, 2006). Hart and Risley (2003) argue that the accumulated experiences with words for children who come from poor circumstances compared with children from professional families may constitute a 30 million word catastrophe that is difficult, if not impossible to close over time.

Compelling as these figures are, they may actually underestimate the problems associated with vocabulary differentials and school learning. As children get older, they will increasingly need academic vocabularies (Spycher, 2009) that consist of words and precise meanings that are often central to content area understanding and differ from general meanings of even the same terms (e.g. operation has a very specific meaning in mathematics) (Beck, McKeown, & Kucan, 2002; Neuman, 2006; Neuman & Celano, 2006). These academic terms, and their specialized meanings may pose the greatest challenges to children who lack a rich vocabulary and a network of concepts (Stahl & Nagy, 2006). Deficiencies in semantic development have been shown to lead to significant difficulties in reading comprehension for older readers (Vellutino et al., 1996).

This means that until instructional materials are prepared to emphasize vocabulary and conceptual knowledge early on, less advantaged children will continue to be handicapped even if they master reading the written words. Particularly disheartening, then, is the finding that there little emphasis on the acquisition of vocabulary in elementary school (Beck & McKeown, 2007) or preschool curriculum (Neuman & Dwyer, 2009). A recent content analysis of published early literacy programs (Neuman & Dwyer, 2009), for example, found little evidence of a deliberate effort to teach vocabulary to preschoolers. 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 children's progress. In short, current instructional materials appeared to offer little guidance to teachers who want to do a better job of teaching vocabulary to young children.

With the recognition of the vast differences in vocabulary evident early in children's lives and their consequences for subsequent literacy growth, there is an emerging consensus that efforts need to begin early on, prior to children entering formal schooling (Dickinson, McCabe, & Essex, 2006; Neuman, 2009). The question then becomes how do we effectively intervene with very young children who need more intensive vocabulary instruction? And moreover, how may we potentially accelerate its development? Recognizing that word learning—more than any other aspect of oral language development—falls at the intersection of cognitive and linguistic development (Wellman & Gelman, 1998) children will need a rich storehouse of words and concepts (Hirsch, 2003) in order to successfully advance in developing content knowledge and reading proficiency in later grades.

There is an emerging body of evidence indicating that the organization in which children learn words may support word learning (Booth, 2009; Chi & Koeske, 1983; Glaser, 1984). Recent research has shown that when children 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 a synergistic relationship between them. Borvsky 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. Their results were consistent with previous research by Gopnik and Meltzoff (1987), who have argued for the "bi-directional interaction" of categorization as a tool for learning language.

Richly-organized concepts are structured as taxonomies (groupings of like things, e.g., pets) (Markman & Callanan, 1984), a hierarchy in which successive levels refer to increasing generalizations. Taxonomies have similar properties (e.g. pets—dogs, cats are animals that live with people), and fall into an intermediate level of abstraction (Smith, 1995). In this respect, they are different than themes or thematic groupings (e.g. things you do in a grocery store--clusters of things that interact), which are based on associations and have a less clear-cut structure (Markman & Hutchinson, 1984). Specifically, it is the structure and the coherence of taxonomic categories that has been associated with improved word learning.

Studies (Gelman & Markman, 1986; Murphy & Lassaline, 1997) have documented that taxonomic categories help children store ideas efficiently, and are the building blocks of concepts. In maternal speech, for example, middle-class mothers often provide explicit information of domain specific properties (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998). As children become more familiar with single objects, they go beyond labeling, pulling together properties that tie the objects together ("These are dogs that like to play in the water.") to develop generalizations and beginning concepts. Further, existing evidence indicates that children use category membership to gain information about unfamiliar terms and to make inferences (Gelman & Kalish, 2004; Gelman & O'Reilly, 1988).

Consequently, the intervention described in this article known as the World of Words was based on the theoretical premise that learning words in taxonomies could enhance and potentially accelerate word learning and concept development. By explicitly teaching word meanings and category membership early on, our goal was to foster growth in vocabulary and conceptual knowledge for low-income 3- and 4-year olds in classroom settings.

The World of Words Vocabulary Intervention

The World of Words (WOW) curriculum (Authors, 2007) is a supplemental intervention to support vocabulary instruction and conceptual learning in pre-K. There are several key features that underlie its design. First, the curriculum is organized by topics that represent animate taxonomies (e.g. parts of the body) with properties identified for each taxonomic topic (e.g. body parts are attached to our body). Topics represent content standards in health, science and mathematics in states that received the highest quality ratings from the Fordham Foundation (Finn, Petrilli, & Julian, 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 them from animal life.

Second, 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. In addition, words that challenge children to think about the category structure are also included (e.g. hair; tears), along with words to support children's conversations about the taxonomies and their properties.

Third, the curriculum uses embedded multimedia, strategies in which animations and other video are woven into teachers' lessons. The use of embedded multimedia is based on two related theoretical models. One is that multimedia can support word learning and concept development through a synergistic relationship (Neuman, 1995). Supporting evidence 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 allow for better retention than information held in only one memory system. The second is Paivio's dual coding theory (Paivio, 1986), which posits that visual and verbal information are processed differently, creating separate representations for information processed in each channel. Together, these codes for representing information can be used to organize and create mental models of knowledge that can be stored, and retrieved for subsequent use. Chambers and her colleagues (Chambers et al., 2008; Chambers, Cheung, Madden, Slavin, & Gifford, 2006), for example, have shown that the use of embedded multimedia can enhance learning, reporting a moderate effect size when compared with instruction without media.

Taken together, these design principles form the basis of the World of Words, a supplemental vocabulary program for preschoolers. A sample of the curriculum matrix is provided in Appendix A.

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 8-day period. For example, consider the topic, "Insects." Each day begins with a "tuning-in"—a rhyme, song or word play video "clip" that is shown from a DVD[1] to bring children together to the circle. The "tuning-in" is followed by a 'content' video, introducing children to the definition of the category. The first video is designed to act a prototype of the category, a particularly salient exemplar of the topic (i.e., a katydid). After the video, the teacher engages the children, focusing on 'wh' questions. She might ask, "Where does a katydid live? What is an insect?" Words are then reinforced using an information book (i.e. in this case on insects) specially designed to review the words just learned (e.g. antennae; segments, camouflage; familiar, wings; outside), and to provide redundant information in a different medium.

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

[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 ½ minutes.

information book and picture cards to engage children in sorting tasks. Children are presented with "time for a challenge" items which require them to problem-solve about the category (e.g. is a bat an insect?). These challenge items are designed to encourage children to apply the concepts they have acquired to think critically about what may or may not constitutes category membership. Lastly, the children review their learning through journal writing activities that involve developmental (phonic) writing.

The 8-day instructional sequence is designed to help teachers scaffold children's learning of words and concepts. In the beginning, for example, the teacher lesson plan focuses on explicit instruction, helping children to 'get set'—providing background information—and "give meaning," to deepen their understanding of the topic. As the instructional sequence progresses, the teacher begins to "build bridges" to what children have already learned and what they will learn (establishing inter-textual linkages across media). Here, the teacher begins to release more control to the children during the teacher-child language interactions. Finally, the teacher is encouraged to "step back" giving children more opportunities for open-ended discussion. At the end of the instructional sequence, children are given a "take-home" book—a printable version of the information book used in the lesson. Throughout the sequence, familiar words are used for helping children talk about a topic, and for incorporating the approximately 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 Appendix B.

Overview of the Research Design

This study was designed to investigate the effects of WOW on word learning and concept development for low-income preschoolers. In Phase I of the study, our research design employed random assignment of teachers and classrooms to either treatment (WOW) or control conditions from a county-wide Head Start program in which all children were low-income and a small proportion were non-English speakers. These children were considered especially vulnerable to reading failure and other adverse outcomes without special language-specific intervention. In Phase II of the study, we added two additional control samples. Here, our goal was to compare growth in vocabulary and concept development for the Head Start groups with children from a state pre-kindergarten program representing children who were largely middle-class, and children from a university-based program who were highly advantaged. These additional samples were designed to look at the comparative advantages of WOW, and the extent to which the intervention might help level the playing field for low-income children.

We examined three primary hypotheses. First, we theorized that the WOW curriculum would produce greater gains in content-related vocabulary. Second, we proposed that words learned through categories would enhance children's ability to develop conceptual knowledge associated with these words. Third, we hypothesized that conceptual knowledge gains would improve children's ability to make inferences and generalizations about novel words and their meanings.

Method

Phase I

Research sites and participants

Centers selected for the research study were part of a county-wide Head Start program with five delegate agencies and 12 schools. The schools are located in an urban area that has been severely economically depressed, reporting over 15% unemployment. The Head Start program offers A.M. and P.M. classes, and full day programs for children ages three and four, four days a week, eight months a year. All classrooms served mixed-age groups.

Schools were randomly assigned to treatment or control groups. Fourteen classrooms (7 full-day; 7 half-day) were randomly assigned to use the WOW supplemental curriculum in addition to their traditional curriculum (High/Scope); 14 classrooms (7 full-day; 7 half-day) to the control group, who used the supplemental curriculum, Growing Readers (High/Scope, 2008) as well. Growing Readers is a story-based supplemental curriculum that provides small group activities in vocabulary, print knowledge, and phonological awareness skills. All programs adhered to the Early Learning Outcome Standards approved by the national office of Head Start.

School policy was set by the county Head Start program. Each classroom had teacher and an aide. Class size was limited to 18 children. Supports included mentor teachers to help teachers maintain the pace of instruction and to ensure coverage of all content areas related to the standards. Each school included a director, and a site manager responsible for educational activities. Monthly meetings with these staff and the director, educator coordinator, and parent educators of the county-wide program were held to provide consistency of services across sites.

Overall, 55% of the teachers were Caucasian; 38% African-American, and 7%, Middle-Eastern; 54% had a two-year degree; 25%, a bachelor's degree, and 21%, a Child Development Associates Degree (CDA). No significant differences between groups were reported on minority status, level of education and amount of previous professional development. All children in these classes were eligible for free- and reduced lunch. Approximately 50% of the children were African-American; 26% Caucasian, 8% Asian, 4% Middle-Eastern, and 12% multi-racial. There were no significant differences between groups for age, gender or minority status. A total of 604 children participated in the study.

Procedures

In early fall, all 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 before classes began in the school year. Similarly, control group teachers attended workshops on their supplemental curriculum as well. Both groups attended a 4-hour refresher workshop in early winter, and received targeted feedback on classroom instruction once per month during the academic year. For both trainings, district supervisors emphasized the alignment of the curriculum and the Head Start standards.

Child assessments

For phase 1 of the study, assessments were designed to measure growth in word knowledge and concept development for children in the treatment and control groups. Prior to the start of the intervention, children were individually assessed in a quiet section of the school environment. Graduate students in educational psychology and experienced researchers with advanced degrees, trained and certified prior to their work in the field, conducted the assessments. Throughout the data collection period, the site coordinator monitored the testing to ensure that the administration remained reliable.

Children were assessed with three different instruments: curriculum-specific word knowledge and concept development, and the Woodcock-Johnson Picture Vocabulary Test.

Word knowledge: We constructed a 40-item receptive vocabulary task to measure growth in word knowledge for the unit (4 topics). Words were randomly selected from the corpora of words taught throughout the unit. Children 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 randomized across students. The total number correct was recorded for each student. Reliability of the measure α= .86.

Conceptual knowledge: We designed a 32- item task to measure growth in children's conceptual understanding of target vocabulary. Four conceptual properties from each topic were selected, and assessment questions devised to include items in which the concept property could be applied to the target word (e.g. do our legs help our bodies move around?); or could not be applied to the target vocabulary (e.g. does a jacket help our bodies move around?). Children responded either yes or no to each question, and a total number of correct responses were recorded. Reliability was α=.86.

Expressive vocabulary: We used the Woodcock-Johnson Picture Vocabulary Subtest (Form A) to assess children's expressive vocabulary. This measure was selected because its norming sample seemed to better represent children of varying income and native language status. The subtest consists of 42 pictures of words that increase in difficulty. For each item, children are prompted to label the picture; administration is discontinued after failing to label six items in a row. Reliability of the measure is α=.80.

Pretests were administered to treatment and control groups prior to the intervention. Children were tested in a single session which lasted approximately 15-20 minutes. Following initial assessments, children in the treatment group received the first unit of instruction (4 topics) in WOW, 10-12 minutes per day, 4 days a week. The control group used its supplemental curriculum; and both groups continued with their regular curriculum (High/Scope). Although the control supplemental curriculum covered much of the same 'academic' content as WOW, it did not specifically focus on categorical learning or concept development. Posttests, using alternate forms of the pretests, were conducted eight weeks later, following the last topic of instruction.

Phase II

Research Sites and Participants

In phase II of the study, we added two additional control samples. The purpose was to examine the extent to which we might close the vocabulary and conceptual knowledge gap for children who were economically disadvantaged compared to children who attended a state pre-K program, or those who were highly advantaged in a University preschool program. Twenty-eight state pre-K classrooms from the same surrounding county as Head Start participants agreed to participate in the study, with an additional 11 classrooms from the university-based preschools.

Demographic characteristics of the children from Head Start, as shown in Table 1 compared to those in state pre-k and the university-based programs varied significantly. Approximately half of the Head Start population was African-American, whereas over half of the state pre-k and university-based children were Caucasian. Over a third of the sample from the university was Asian. The majority of children across the sample spoke English at home (96%, with 4% unreported). In total, the sample size included 1284 3-year and 4-year old children.

As shown in Table 2, demographic characteristic of the teachers varied across the settings as well with fewer minority teachers in the middle- and highly advantaged preschools. Teachers in the Head Start treatment group had significantly less formal education, but more years as a lead teacher than those in the other groups. Teachers in the state pre-k and university-based programs all had bachelor’s degrees.

Insert Table 1 about here

Insert Table 2 about here

Additional Child Assessments.

Prior to the second unit of instruction, children in all four groups were administered the following assessments.

Word knowledge. We constructed a similar version of the WOW receptive word knowledge measure as in the previous unit for Unit 2 Living Things. The 54-word assessment included items from each of the four topics in the unit. The order of pictures was counterbalanced across the assessment and the order of the items was randomized across children. The test had a reliability of $\alpha = .90$.

Categories and Properties Knowledge. To examine children's conceptual knowledge in greater depth, we constructed a receptive task to identify categories and properties of target words. In this task, children 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). Children were then asked to identify which item/object belonged to a particular category (Which is an insect?) or to identify the item/object that possessed a particular category attribute (which has three body segments?). A total of four category level questions were assessed (one for each topic), and a total of eight concept property questions (two for each topic). Concept property questions were selected as most representative of the category. For example, children 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 category, "Insects."   Responses were tallied for accuracy on category and property questions, and for the Unit overall separately (Total score possible=12). Reliability was α=.90.

Conceptual Knowledge: The WOW Concepts task design and administration was identical to Unit 1. Children were asked simple "yes-no" questions measuring conceptual understanding of target items in Unit 2, assessed in a different context than what was taught throughout the curriculum.

Expressive Vocabulary: Woodcock-Johnson Picture Vocabulary Subtest (Form B) (Woodcock & Mather, 2001). This assessment was administered to allow for a comparison between groups, taking into account our additional control groups.

Following Unit 2, eight weeks later, posttests on word knowledge, categories and properties, and conceptual knowledge were individually administered to children. After a week's break, a similar cycle occurred for Unit 3. Prior to instruction, pretests for words and concepts in Unit 3 were conducted using similar formats followed by posttests eight weeks later. However, in Unit 3, we added one additional measure to examine children's ability to make categorical generalizations and inductive inferences using familiar concepts applied to novel words.

Inductive reasoning. To examine children's 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 (decagon, trapezoid, one thousand, shifting spanner, backhoe, and vise) in this task. 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, children were asked, "Can you use a decagon to count?"

There were four steps to this task. First, children 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. Children were then told the name of the target object and its category membership. For example, a child would be shown a picture of a vise and told, "This is a vise. It's a tool." On the next slide the child was then asked a property question about the category and object. For example, the child 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. Reliability of the assessment α=.80.

At the conclusion of the study, Woodcock-Johnson Picture Vocabulary Subtest (Form A) was administered again.

Fidelity of implementation

Throughout the study, we examined the fidelity of implementation using the lesson plan as our guide. Observers on a weekly basis observed and examined the presence or absence of five features: opening activity (tuning in); content (video and questions); information book reading; discussion and time for a challenge; and developmental writing and review. A criterion of 90% was established as "full" implementation for each feature. On occasions when classrooms fell below the criterion, follow-up conversations were conducted to help teachers fully implement the key features.

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.

In summary, Head Start children in the treatment and control groups participated throughout the eight month study, with pre- and post assessments given at three time periods. Two additional control samples joined at mid-year. These control samples used a similar curriculum framework as Head Start (High/Scope) but without any additional supplemental program, and were assessed at two time periods.

Results

We present our results in three parts. First, we examine the impact of the intervention on word knowledge. Next, we measure growth in concepts, and knowledge of categories and properties within these concepts. Finally, we report on children's ability to make inferences and generalizations, using categorical properties to reason about unfamiliar words.

Two sets of analyses were conducted. First, effects on child outcomes were estimated through analyses of covariance (for differences between two groups in Unit 1; four groups in Units 2 and 3) with individual posttest score as dependent variable and group and pretest scores as independent variables.

Next, given the multi-level nature of the data, we also estimated hierarchical linear models with treatment condition at the classroom level. These analyses are more conservative as they recognize that children are not independent from one another but are clustered within classrooms. HLM models allowed us to partition the variance between children and between classrooms to take this into account. For each outcome, we first determined whether there was statistically significant variability in the outcome between teachers and calculated the intra-class correlation (ICC), the amount of variance in the outcome that existed between children and between classrooms. Next, we estimated child-level effects by including covariates to predict variability between children. Covariates that were not significant were eliminated from the subsequent analysis. Finally, we created a fully conditional model to estimate classroom-level and child-level effects simultaneously. At the classroom level, treatment condition was our variable of interest and was included as the predictor of between-classroom variance. Each control condition (Head Start control; state pre-K; university pre-K) was entered into the model as a dummy-coded variable with children in Head Start treatment classrooms as the comparison.

Both methods yield essentially the same estimated effects. In each analysis, we controlled for pretest; other predictors were not significant.

Word Knowledge

Table 3 presents the results of comparisons of the WOW curriculum and control classrooms. Prior to the start of the study, there were no significant differences between Head Start treatment and control groups on word knowledge in Unit 1 or on the Woodcock-Johnson Picture Vocabulary pretest. These results provide further reassurance that randomization worked to ensure comparability of groups.

Insert Table 3 about here

Following instruction, the analyses of covariance indicated statistically significant differences at Unit 1 posttest between groups, $F (2, 601) = 17.92$, $p < .001$. Treatment children significantly outperformed the controls, indicating improvements in word knowledge. The effect size using Cohen's *d*, was substantial at .44, suggesting an educationally meaningful difference (Cohen, 1988).

However, when we compared pretest scores for Unit 2, the disparities in vocabulary knowledge between Head Start groups and the two additional control samples became clearly evident. There were significant differences between groups on the Woodcock-Johnson Picture Vocabulary, with the middle-class and highly advantaged groups scoring significantly above those in the Head Start groups ().

Similarly, prior to treatment in Units 2 and 3, children in state pre-K programs and university-based preschool programs scored significantly higher on pretests than the Head Start children, F (3, 1278)= 61.70, $p <.001$, where there were no pretest differences between Head Start groups.

Analyses of covariance with pretests as covariates and posttest scores as dependent variables indicated significant improvements in word knowledge after each unit of instruction. Statistically significant gains were recorded for Unit 2 (F 4, 1279 = 27.94, $p < .001$) and Unit 3, (F 4, 1279) = 21.39, $p < .001$. Post-hoc analyses indicated that the Head Start treatment group significantly outperformed the Head Start control group. Effect sizes for Units 2 and 3 were even more substantial than in Unit 1 (Cohen's *d*= .56; and .86, for Units 2 and 3, respectively). Nevertheless, as shown in Figures 1a, b, c, while the trajectory of improvement for the treatment group was substantial, children still scored significantly below the more average and above average control groups in Unit 3, with a series of topics that focused on the more abstract mathematical concepts and words.

Woodcock-Johnson pre-, interim and posttests scores showed a similar pattern. Although children in the Head Start treatment group improved in expressive vocabulary, essentially closing the gap with the middle-income sample, statistically significant differences remained between groups, F (3, 1278)= 61.70, and the more advantaged sample ($p < .001$). These results indicate that although the instructional intervention significantly narrowed the gap, it did not close the gap in word knowledge.

Insert Figures 1a, 1b, and 1c

Conceptual Development

Table 4 reports the results of the analysis on children's developing concepts. Prior to treatment in Unit 1, there were no significant differences between Head Start treatment and control groups. Following instruction in Unit 1, the treatment group significantly exceeded the controls (F 2, 601) =37. 54, $p < .001$. This effect size was substantial (Cohen's *d*.= 64) and educationally meaningful.

A similar pattern is reported for Units 2 and 3 with the additional control samples. Before treatment, there were significant differences between Head Start groups and other controls, F (3, 1279)=27.81, $p <.001$, although in the case of Unit 2, Head Start treatment was significantly higher than Head State control (p. 05). However, for these Units the Head Start treatment group not only significantly outperformed the Head Start controls, (F (4, 1278)=29.6, $p < .001$, Cohen's *d*= .53 for Unit 2; F(4, 1278)=16.03, $p <.001$, Cohen's *d* =.41) , as shown in Figures 2a, b, and c, they were statistically undifferentiable (p=ns) from those children in the middle- and more advantaged samples. In essence, Head Start children starting from behind prior to instruction exceeded the middle-class sample in Unit 2, and caught up to their more advantaged peers in Unit 3.

Table 4 about here

Figures 2a,b,c

Knowledge of Categories and Properties

Table 5 presents the results of comparisons between groups on their knowledge of categories and properties within concepts. Recall that this assessment was introduced in Units 2 and 3 and required children to make inferences about these category and properties in new language contexts. Consistent with the randomization process, there were no significant differences between Head Start treatment and control groups prior to instruction in Unit 2 (p=ns). Both Head Start groups, however, scored significantly below the more advantaged samples F (3, 1279) = 40.65, p <.001.

Table 5 about here

Posttest scores for Unit 2 revealed statistically significant differences between groups, F (4, 1278)= p <.001. Children in the treatment group and their more advantaged peers were statistically equivalent (p=ns); in fact, treatment children's scores exceeded that of the middle-class sample. Head Start controls, however, scored statistically below other groups ($p < .001$). Effect size between the Head Start treatment and control groups was Cohen's *d* = .86.

A similar pattern occurred throughout Unit 3 with topics emphasizing more abstract concepts in mathematics. Prior to instruction, there were significant differences between the upper-income control groups and Head Start treatment and controls, $F(3, 1279)=30.59$, $p < .001$, but once again, no differences between Head Start groups (p= ns) Following intervention, Head Start treatment children were statistically on par with children in the middle-income group (p=ns), while the Head Start controls remained statistically below their peers in all three groups ($p < .001$). Effect size in this Unit was still educationally meaningful yet lower than in the other Units (Cohen's *d*= .34). Children in the university-based program scored significantly higher than all three groups in Unit 3 ($p < .001$). Figures 3a and 3b display these relationships.

Figures 3a and 3b

As graphically shown in Figures 3, children in the treatment group not only improved in conceptual, categorical and knowledge of properties compared to their equivalent control group, they essentially closed the gap between those children who were middle- and more advantaged in Unit 2, and were statistically equivalent in categorical development in Unit 3 with middle-class children. Developing word knowledge within taxonomies enabled treatment children to create an interconnected knowledge of concepts, critical for later reasoning and comprehension development (Stahl & Nagy, 2006).

Inductive Reasoning

Our final analysis was to examine the potential impact of treatment on children's ability to make inductive inferences about the meaning of novel words. Recall that while the category "tools" 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 children. Consequently, the task required children to apply their knowledge of categories and concepts to reason to unfamiliar and novel objects.

Results indicated significant differences between groups, $F(3, 1279) = 33.47$, $p < .001$. Post-hoc tukey analyses indicated that the middle-class and highly advantaged control groups, on par with one another, were significant different than the Head Start groups ($p < .001$).

Insert Figure 4 about here

However, our analysis revealed that Head Start treatment children scored significantly higher than Head Start controls in using categories to identify the meaning of new words ($p < .001$, Cohen's $d = .46$). In other words, categories appeared to bootstrap the ability to induce meaning of novel words in a familiar concept. Although the treatment group did not overcome the initial advantage of those children in the middle- and upper-middle class groups, results provided evidence that children were able to use their newly learned category information to make category generalizations and inductive inferences about novel words.

Insert Table 6 about here

Table 6 reports the HLM analyses. The intra-class correlations exceeded 10% in all cases, with the exception of the concepts analysis (8%), indicating variability in the outcome between teachers. Using the level 2 model, it is noteworthy that the estimated effects on word knowledge, concepts, categories and properties remained statistically significant at the .01 and .001 levels as in the previous analyses. After controlling for pretest differences at the child level, Head Start treatment classrooms significantly outperformed their counterparts in Head Start on all outcome measures. Together, these results provide powerful evidence for the effects of the World of Words (WOW) on low-income children's word knowledge and concept development.

Discussion

The World of Words (WOW) was constructed on the theoretical premise that categorization can act as a tool for word learning and the development of concepts. We hypothesized that the development of words within such a structure might enhance children's ability to learn and retain words and their conceptual properties, acting as a bootstrap for self-learning. We subjected our hypothesis to a rigorous experimental design to examine children's trajectory of growth, and to determine its potential to close the gap between low-income and more advantaged children.

The results of our study, replicated in each unit of instruction, indicated that children who received the World of Words (WOW) intervention learned content-rich vocabulary associated with each topic. Statistically significant differences were reported consistently between the Head Start treatment and control groups. These differences were educationally significant, not only in terms of the size of the effects, which were substantial. Targeted to Head Start early outcome and pre-k standards, they were also educationally-relevant and meaningful for content learning in science, math, and health as well. Given that the control group also used a supplemental curriculum, these results suggest that the words and the instructional design features of WOW were more effective in promoting word knowledge in these critical content areas.

Our use of two additional control groups, one middle-class, the other highly advantaged, however, provided sobering evidence of the stark vocabulary gap between low-income and middle-income children. In contrast to the Head Start children, those in the more advantaged groups knew the content-rich words without any additional instruction. At the same time, the results provide powerful evidence for the effects of quality instruction. Within eight weeks in one case, treatment children essentially closed the gap in words knowledge; within another case, they significantly narrowed the gap, demonstrating their ability to learn and retain these content-specific words.

. Nevertheless, these results are not to be unexpected. One might presume that children learn what is taught. Given the emphasis on these particular words in the curriculum compared to a more generic literacy supplemental curriculum for the control group, one would assume that the treatment children certainly had an advantage on these curriculum-based 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 children to apply knowledge in new contexts. And in this case, the pattern was very clear and consistent: Children in the treatment group improved significantly in their ability to categorize and conceptualize than their Head Start counterparts. Further, in several cases they actually exceeded their middle-class peers, and were statistically on par with the highly advantaged group. These results demonstrate the enormous potential of instruction for improving children's ability to develop a rich network of interconnections of knowledge about concepts and the meanings that words represent. It also shows that a program targeted to conceptual learning is highly appropriate for preschoolers.

Further, children used the inductive potential of categories to develop inferences about the meaning of novel words. Once children were given the category, they could use its properties to illuminate some basic understanding of a word not previously encountered. Knowing that ‘bulldozer’ was a building tool, for example, treatment children could hypothesize that it was “a powerful machine that you could use to move things.” Children who had received the intervention were significantly more successful than their control counterparts, demonstrating their ability to apply their categorical information to novel words. These results provide some initial evidence of bootstrapping; children used what they learned about categories to induce knowledge of new words. Future research is recommended to determine the extent to which these findings can be replicated, extended, and generalized to new contexts and content.

Although categorical formation has been known to be an important factor in word learning, to our knowledge this is the first attempt to develop an intervention based on the premise that word learning organized through categories can help support vocabulary development and to extend children’s knowledge beyond what is already known. Previous research (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998) with middle-income parents has shown that children as young as 2 ½ are capable classifiers, and that richly structured categories can have rich inductive potential. Yet, this is the first example indicating its efficacy as an instructional design feature. If further research bears out our findings, this cost-effective instructional design feature has great 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. Rather, it might be the rich network of knowledge and concepts that the words represent. Consequently, by encouraging children to think in categories, teachers may be developing their ability to comprehend, to reason and to think on their own.

Our study had significant strengths. As a randomized controlled trial, teachers throughout a county-wide program were randomly assigned to treatments. Classroom resources and the staffing structure were identical between groups. In addition, the study ensured that the timing and amounts of professional development and support were similar across sites and groups. Further, both groups received a supplemental curriculum, indicating 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 areas of early literacy development, health, science, and math instruction as indicated in the Head Start early outcome standards. Our fidelity measure indicated that teachers used similar practices across the different schools and classrooms. Finally, the study also benefited from the use of additional control groups which provided a valuable calibration for determining the extent to which treatment children needed to improve in order to reach benchmark. Since classroom instruction tends to be targeted to the average achiever (Hirsch, 2003), it allowed us to go beyond the typical treatment and control comparisons, to examine how treatment children might compare with those they are likely encounter in elementary school.

Our study also has significant limitations. It was conducted in a community-wide Head Start program in a highly economically distressed urban area. The population was entirely low-income, mostly Caucasian and African-American, with a small percentage of Middle-Eastern second language learners. The program was well-funded, and very well managed through careful attention to early outcome standards and ongoing monitoring. Therefore, we cannot make a case that these results could be generalized to other early childhood contexts and 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), although one might hypothesize that the advantages would be greatest for programs where a curriculum is aligned with prekindergarten guidelines and standards consistent with these goals.

The study also did not show dramatic improvements in expressive language as measured by the Woodcock-Johnson Picture Vocabulary Test. This assessment had been selected due to its higher congruence between our sample and the test's norming sample, both in income level and native language status, relative to other standardized vocabulary measures (e.g. the PPVT). Modest improvements were recorded for the treatment groups; little to lesser growth was recorded for all other three samples. These results need further exploration. However, we suspect that the measure may not be sufficiently sensitive to vocabulary growth for very young children. The National Reading Panel (2000), for example, predicted an underestimate of effect sizes when using standardized tests. As a result a number of researchers (e.g., Sénéchal, Ouelette & Rodney, 2006; Coyne et al, 2007) have increasingly relied on author-created measures in order to attain enough sensitivity to detect fine-grain and more comprehensive vocabulary growth. Future research on WOW is needed that combines both author-created and standardized measures to examine the extent to which overall language development and comprehension growth are acquired.

Lastly, 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 theoretical premises: the uses of taxonomic structures in vocabulary and concept development; the instruction of words in semantic clusters, and the uses of embedded multimedia as a mechanism to instantiate words through dual coding of images as well as words. At this point, we cannot 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.

With these limitations in mind, this study provided substantial evidence for the improvement of content-rich vocabulary and conceptual development among low-income children. With targeted instruction, these children not only made educationally meaningful gains, they often closed the gap, achieving at levels consistent with those of middle-class and highly advantaged children. Together, this evidence suggests that we have just begun to tap these children's potential.

References

Author. (2006). *Growing Readers*. Ypsilanti, MI: High/Scope Foundation.

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

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

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

Booth, A. (2009). Causal supports for early word learning. *Child Development, 80*, 1243-1250.

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

Cain, K., Oakhill, J., Barnes, M., & Bryant, P. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*, 850-859.

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

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

Chi, M., & Koeske, R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*, 29-39.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

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

Durham, R., Farkas, G., Hammer, C., Tomblin, J. B., & Catts, H. (in press). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*.

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

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

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

Gelman, S., & Kalish, C. (2006). Conceptual development. In W. Damon, R. Lerner, D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology* (pp. 687-733). New York: Wiley.

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

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

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

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

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

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

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, *27*, 10, 12-1316-1322,1328-1329, 1348.

Hoff, E. (2003). The specificity of environment influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368-1378.

Lee, V. & Burkam, D. (2002). Inequality at the starting gate. Washington D.C.: Economic Policy Institute.

Markman, E., & Callanan, M. (1984). An analysis of hierarchical classification. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Erlbaum.

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

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R.E. & Moreno, R. (2002). Animation as an aide to multimedia learning. *Educational Psychology Review, 14*, 87-99.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 1126*.

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

National Reading Panel Report (2000). *Teaching young children to read.* Washington, D.C.: The National Institute for Child Health and Development.

Neuman, S.B. (2009). *Changing the odds for children at risk.* New York: Teachers College Press.

Neuman, S. B. (1995). *Literacy in the television age: The myth of the TV effect* (2nd ed.). Norwood, NJ: Ablex.

Neuman, S. B. (2006). The knowledge gap: Implication for early literacy development. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of Early Literacy Research* (pp. 29-40). New York, NY: Guilford Press.

Neuman, S. B., & Celano, D. (2006). The knowledge gap: Implications of leveling the playing field for low-income and middle-income children. *Reading Research Quarterly, 41*, 176-201.

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

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Sénéchal, M., Ouelette, G. & Rodney,D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. Dickinson & S.B. Neuman (Eds). *Handbook of early literacy research, Vol.II* (pp. 173-182). New York: Guilford.

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

Spycher, P. (2009). Learning academic language through science in two linguistically diverse kindergarten classes. *Elementary School Journal, 109*, 359-379.

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

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

Wellman, H., & Gelman, S. (1998). Knowledge acquisition in foundational domains. In W. Damon, D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (pp. 523-574). New York: Wiley.

Woodcock, R. W., & Mather, N. (2001). *Woodcock Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Table 1. Demographic Characteristics of Children Across the Four Samples

	Head Start Treatment (N-294)	Head Start Control (N=310)	State Pre-K Control (N=508)	University Pre-K Control (172)
Age in Months	47	47	52	47
Woodcock-Johnson (Pre-)	98.4	97.5		
Woodcock-Johnson (Mid-year)	98.8	96.4	103.9**	111.8***
% Female	55	51	53	42
% Minority	74***	75***	39	45
% White	26	25	62**	56**
% Black	53***	46***	22	3
% Hispanic	1	2	8	0
% Asian	10	7	4	37***
% Middle Eastern	3	6	0	1
% Multi-racial	8	14	3	2
% English as primary language	96%	96%	94%	100%

* p < .05 ; ** p < .01; ***p < .001

Table 2. Demographic Characteristics of Teachers across the Four Samples

	Head Start Treatment (N=12)	Head Start Control (N=11)	State Pre-K (N=28)	University Pre-K (N=11)
% Minority	33	50	10	14
% High Professional Development (more than 10 hours)	58	71	62	75
% Associates Degree or Higher	67*	90	100	100
Mean Age	47	43	44	41
Mean Years as Lead Teacher in Preschool	12.7*	7.8	7.9	1.3
Mean Years as Assistant, Aide, or Volunteer in Preschool	2.4	1.3	2.8	.1

*$p < .05$

Table 3. Comparisons of Pre- and Posttest Scores by Treatment Group on Word Knowledge

	H. S. Treatment	H. S. Control	State Pre-K	University Pre-k
Unit 1				
Pretest, % Correct	.63	.61		
(SD)	(.18)	(.17)		
Posttest	.77	.69***		
(SD)	(.19)	(.17)		
Unit 2				
Pretest, % Correct	.78	.78	.90***	.92***
(SD)	(.17)	(.16)	(.12)	(.11)
Posttest	.88	.79***	.92**	.94**
(SD)	(.15)	(.17)	(.09)	(.09)
Unit 3				
Pretest, % Correct	.73	.68*	.85***	.88***
(SD)	(.18)	(.20)	(.13)	(.12)
Posttest	.81	.70***	.86**	.89***
(SD)	(.18)	(.20)	(.15)	(.12)
Woodcock-Johnson (Post)	100.15	98.35	101.87	109.60***

*p < .05; **p < .01; ***p < .001

Table 4. Comparisons by Treatment Group on Conceptual Knowledge

	H. S. Treatment	H. S. Control	State Pre-K	University Pre-k
Unit 1				
Pretest, % Correct (SD)	.50 (.11)	.49 (.14)		
Posttest, % Correct (SD)	.59 (.13)	.52*** (.10)		
Unit 2				
Pretest, % Correct (SD)	.54 (.11)	.51* (.13)	.58** (.13)	.63*** (.16)
Posttest, % Correct (SD)	.62 (.17)	.54*** (.13)	.60 (.12)	.65 (.17)
Unit 3				
Pretest, % Correct (SD)	.55 (.14)	.53 (.14)	.58** (.13)	.65*** (.14)
Posttest, % Correct (SD)	.61 (.17)	.54*** (17)	.63 (.15)	.65 (.22)

*p < .05; **p < .01; ***p < .001

Table 5. Comparisons by Treatment Group on Categories and Properties within Concepts

	H. S. Treatment	**H. S. Control**	**State Pre-K**	**University Pre-K**
Unit 2				
Overall Pretest, % Correct	.58	.54, ns	.69***	.74***
(SD)	(.22)	(.21)	(.19)	(.22)
Overall Posttest	.76	.58***	.74, ns	.78, ns
(SD)	(.22)	(.20)	(.19)	(.21)
Properties Pretest, % Correct	.56	.51, ns	.66***	.71***
(SD)	(.24)	(.24)	(.23)	(.24)
Properties Posttest	.74	.56***	.72, ns	.76, ns
(SD)	(.24)	(.24)	(.22)	(.23)
Category Pretest, % Correct	.61	.59, ns	.73***	.79***
(SD)	(.30)	(.27)	(.25)	(.26)
Category Posttest	.77	.59***	.74, ns	.78, ns
(SD)	(.31)	(.32)	(.28)	(.30)
Unit 3				
Overall Pretest, % Correct	.54	.51, ns	.64***	.71***
(SD)	(.23)	(.22)	(.23)	(.22)
Overall Posttest	.61	.52***	.66, ns	.73***
(SD)	(.26)	(.24)	(.23)	(.23)
Properties Pretest, % Correct	.46	.43, ns	.56***	.65***
(SD)	(.28)	(.26)	(.26)	(.27)
Properties Posttest	.55	.43***	.59, ns	.69***
(SD)	(.30)	(.27)	(.26)	(.26)
Category Pretest, % Correct	.65	.62, ns	.74***	.79***
(SD)	(.33)	(.32)	(.29)	(.27)
Category Posttest	.71	.63, ns	.74, ns	.78, ns
(SD)	(.31)	(.32)	(.30)	(.29)

Table 6.

HLM Analysis Estimating the Effects of WOW on Word Knowledge, and Conceptual Understandings

Variable	Labels	Concepts	Category	Properties	Overall	WJ
	β S.E.	β S.E.	β S.E.	β S.E.	β S.E.	β S.E.
Unit 1						
Fixed Effects[a]						
Teacher-Level						
Intercept	.14	.23**				
	.08	.07				
Head Start Control	-.30*	-.57***				
	.11	.12				
MSRP						
UMCC						
Child-Level						
Pretest Score	.66***	.44***				
	.04	.10				
Random Effects[b]						
Between-teacher variance	.05	.05				
Between-child variance	.43	.68				
Intra-class correlation	.20	.11				
Effect Size						
Unit 2						
Fixed Effects[a]						
Teacher-Level						
Intercept	.35***	.37***	.28***	.32***	.40***	
	.05	.07	.06	.06	.06	
Head Start Control	-.49***	-.48***	-.62***	-.68***	-.74***	
	.07	.12	.11	.12	.11	
MSRP	-.29***		-.26**	-.23**	-.35***	
	.06		.08	.07	.07	
UMCC	-.33***		-.19*	-.15	-.30**	
	.08		.09	.11	.10	
Child-Level						
Pretest Score	.74***	.44***	.30***	.33***	.47***	
	.02	.07	.04	.04	.03	
Random Effects[b]						
Between-teacher variance	.01	.03	.01	.02	.02	
Between-child variance	.27	.78	.84	.75	.64	
Intra-class correlation	.22	.11	.09	.14	.19	
Effect Size						
Unit 3						

Fixed Effects[a]						
Teacher-Level						
Intercept	.28***	.17**	.09	.06	.12	.13*
	.05	.07	.09	.08	.08	.06
Head Start Control	-.45***	-.45***	-.28*	-.47***	-.42***	-.09
	.09	.11	.11	.11	.10	.09
MSRP	-.27***	-.04	-.01	.03	-.04	-.16
	.06	.08	.12	.10	.10	.08
UMCC	-.19**	.03	.07	.28	.11	-.07
	.06	.18	.13	.15	.13	.07
Child-Level						
Pretest Score	.74***	.37***	.25***	.29***	.42***	.74***
	.03	.03	.03	.04	.03	.04
Random Effects[b]						
Between-teacher variance	.01	.05	.01	.04	.03	.03
Between-child variance	.31	.65	.88	.78	.71	.33
Intra-class correlation	.21	.18	.05	.15	.15	
Effect Size						

*p<.05; **p<.01; ***p<.001

a. Coefficient (S.E.)
b. Variance

Figures.

Figure 1a. Differences between Treatment and Control Groups on Word Knowledge: Unit 1

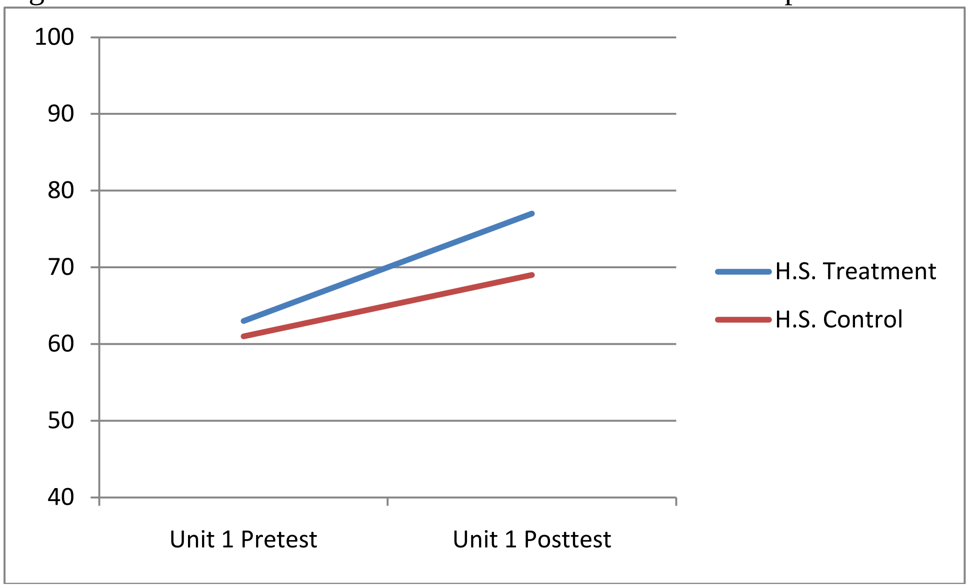


Figure 1b: Differences between Groups on Word Knowledge: Unit 2

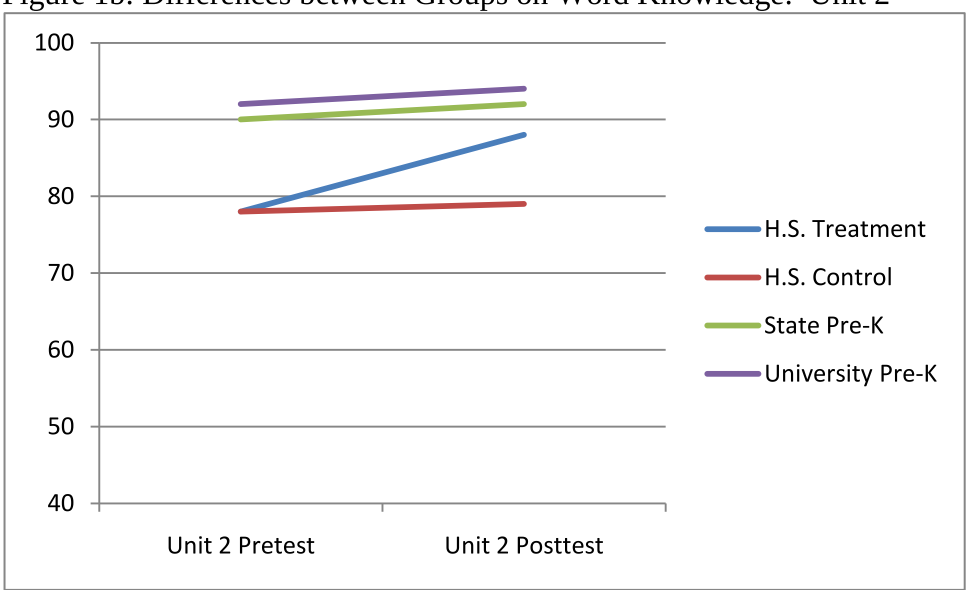
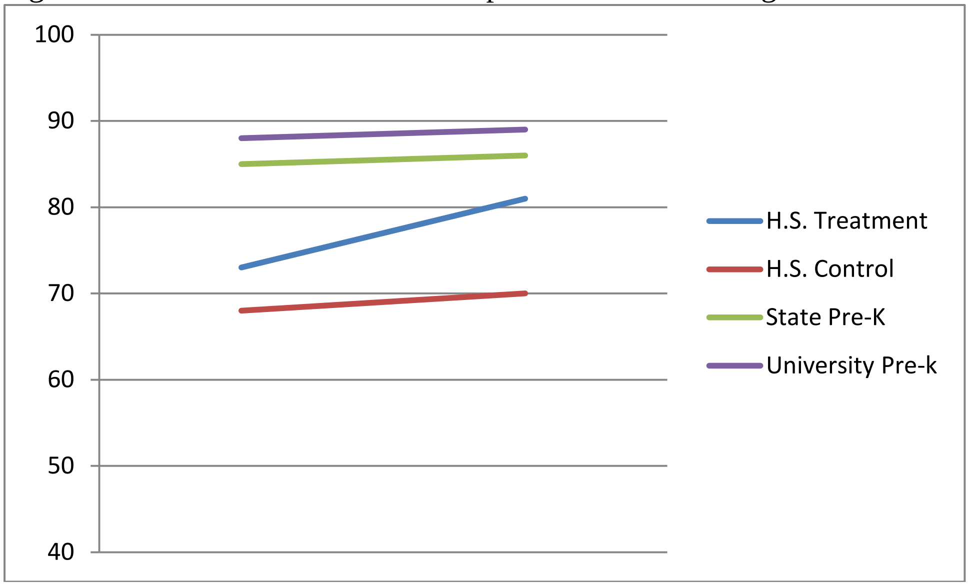


Figure 1c: Differences between Groups on Word Knowledge: Unit 3



Figures 2, a, b, c. Comparisons of Pre- and Posttest Differences on Concepts

Figure 2a: Comparisons of Head Start Treatment and Control Groups on Concepts: Unit 1

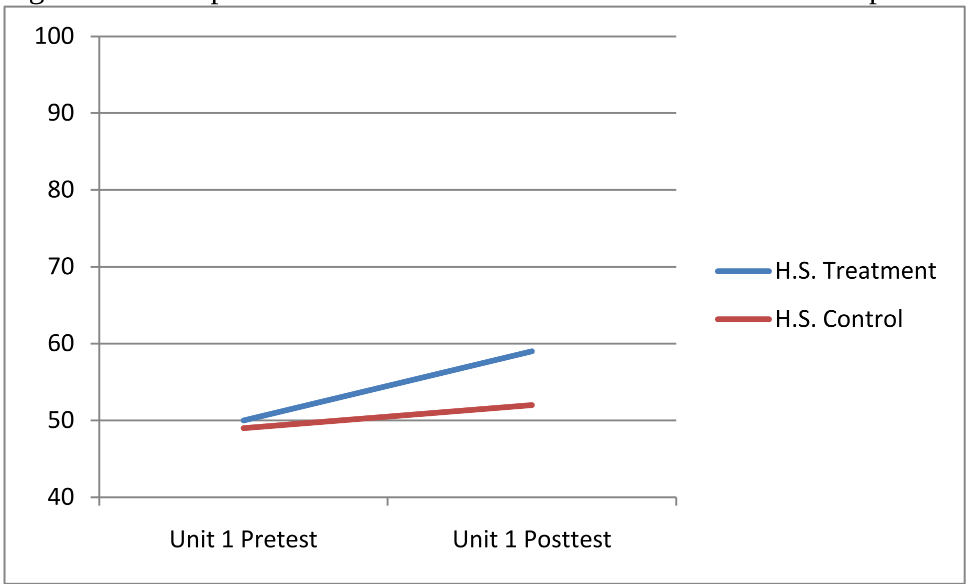


Figure 2b Comparisons of all groups on concepts—Unit 2

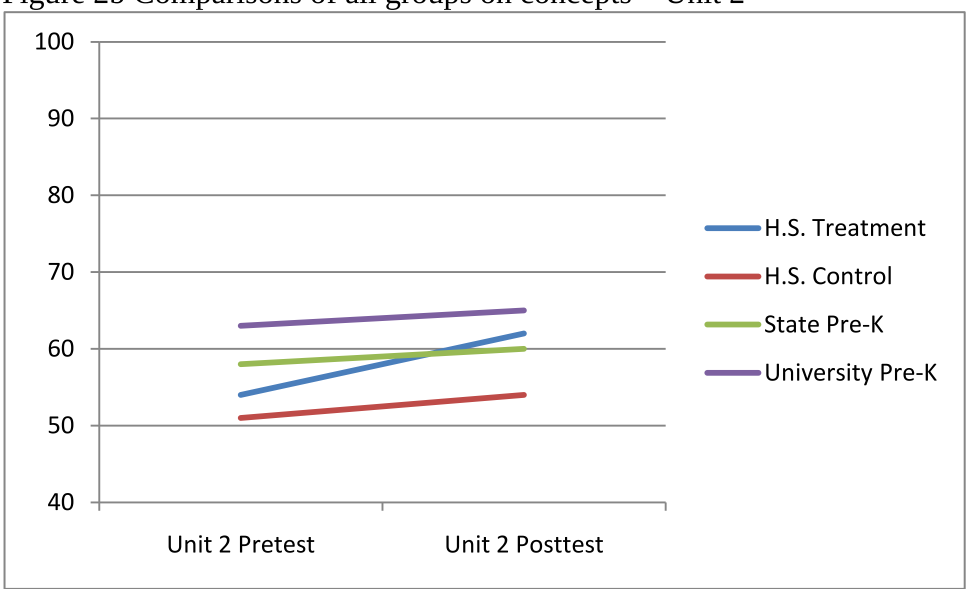
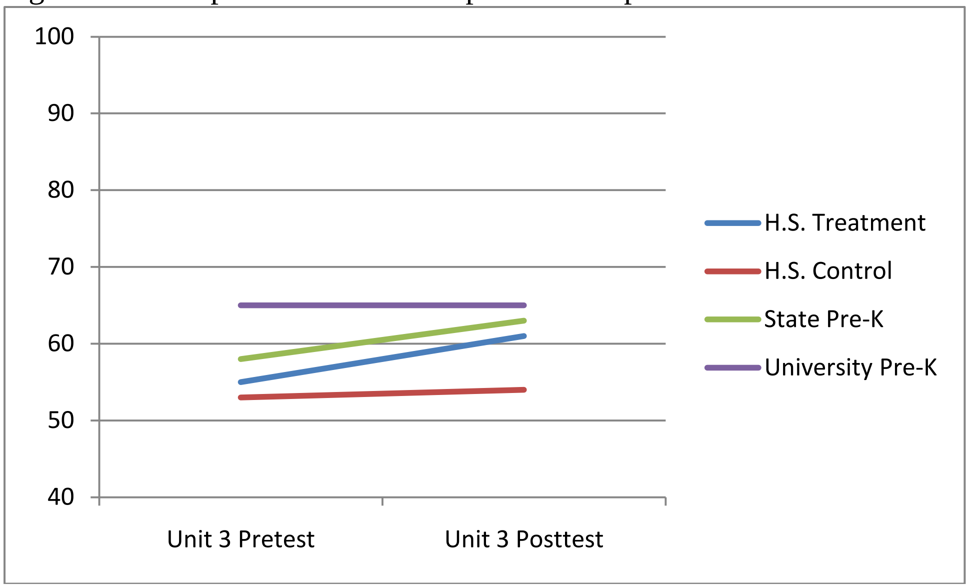


Figure 2c: Comparisons of all Groups on Concepts: Unit 3



Figures 3a, b. Comparisons of all four groups on Categories and Properties

Figure 3a. Comparisons of all four groups on Categories and Properties: Unit 2

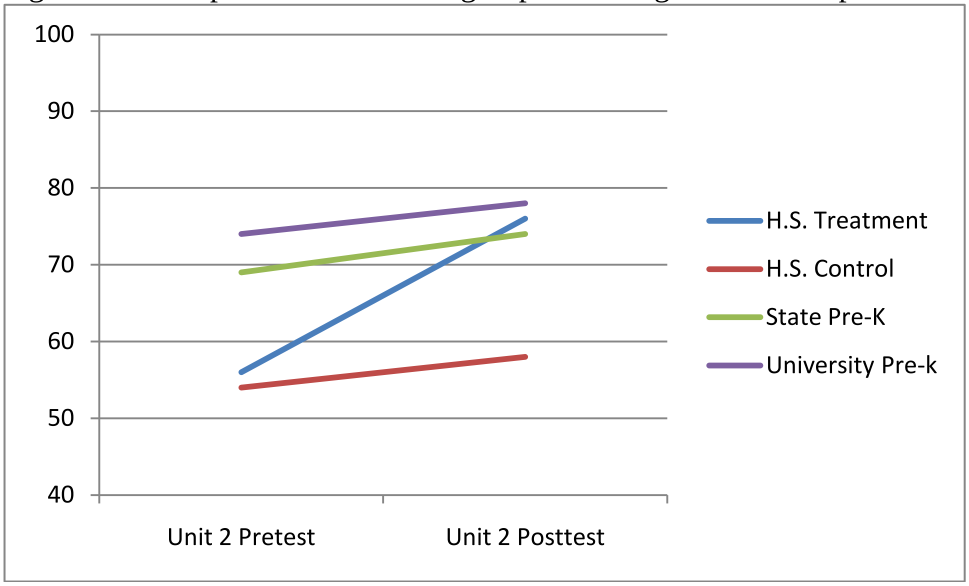


Figure 3b. Comparisons of all four groups on Categories and Properties: Unit 3

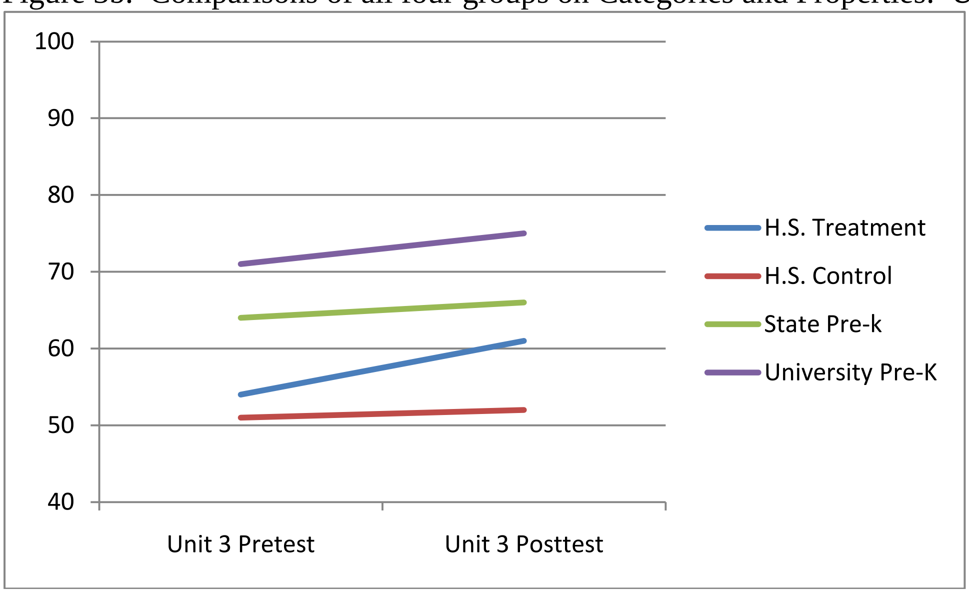
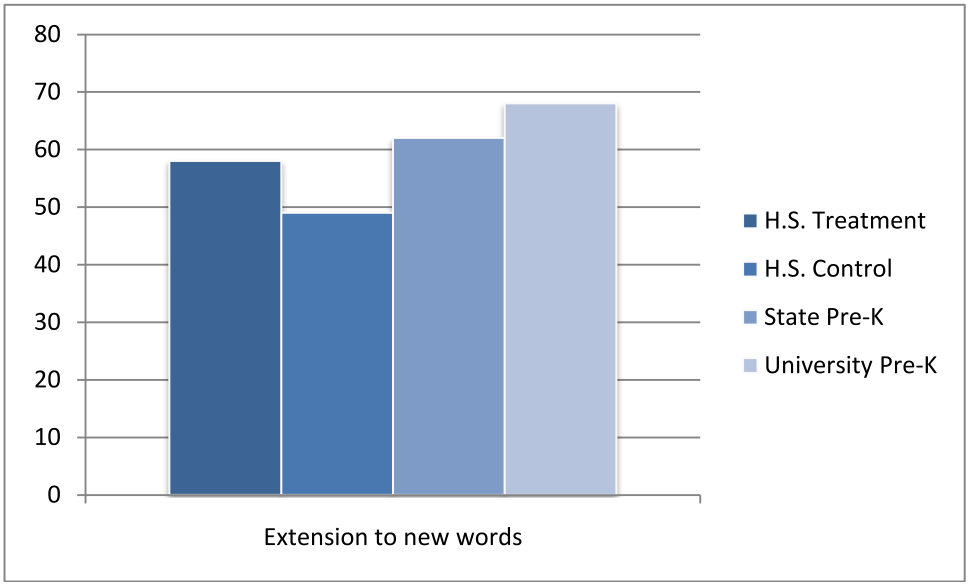


Figure 4. Comparisons Between Groups on Extension Task



Appendix A: Sample from the World of Words Curriculum Matrix

<table>
<tr><th>Topic</th><th>Phonological Awareness Skill</th><th>Tuning-in Clips</th><th>Main Concepts</th><th>Content Clips</th><th>Vocabulary</th></tr>
<tr><td>INSECTS</td><td>• Alphabet
• Alliteration</td><td>1. J.Fox (shortened to the alphabet song twice)
2. La La La Letter L</td><td>1. There are small creatures that live all around us that are called insects.
2. Insects are creatures that always have six legs
3. Insects also have three different parts to their body; these sections are called segments.
4. Insects often have wings, but sometimes they are hidden.
5. Insects also have antennae that they use to smell and feel things.
6. People sometimes call insects “bugs
7. Insects mostly live outside.</td><td colspan="2">Prototype: ant,
In-category: moth, bee, katydid, ladybug, butterfly, grasshopper,
Out-category: lizard, snail, worm, bat,
Challenge words: centipede, spider,
Supporting words: six legs, segments, antennae, wings, small creature, outside, bug, leaves, flowers, blend in, camouflage, protect, enemies, survive, head, thorax, abdomen, sting, honey, hive, nectar, cooperate, ant hill, warn, poisonous, habitat</td></tr>
<tr><td>PLANTS</td><td>• Alphabet
• Alliteration</td><td>1. P Pea Plant Pod Pick
2. P words Dance With Kids</td><td>1. Plants are living things that do not move around.
2. Plants need food, sunlight and water to grow.
3. Plants have important parts that help them live and grow.
4. Plants live in places with the kind of weather that they need.
5. Some plants can be food for people and animals.</td><td colspan="2">Prototype: flower,
In-category: grass, tree, cactus, bush, weed,
Out-category: pail, hose, cobweb, butterfly net,
Challenge words: bouquet of flower, leaf pile,
Supporting words: food, sunlight, water, grow, bloom, soil, stem, leaves, roots, ground, trunk, wood, branches, bark, spines, desert, seed, rain, sun, petals, garden, fruit, vegetable, pick, ripe</td></tr>
</table>

Appendix B: Example of the Instructional Design

	Component 1: Phonological awareness	Component 2: Content video	Component 3: Book-reading	Component 4: In-category picture cards	Component 5: Out-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