ENHANCING THE INTENSITY OF VOCABULARY INSTRUCTION FOR PRESCHOOLERS AT RISK

The Effects of Group Size on Word Knowledge and Conceptual Development

ABSTRACT

This study was designed to experimentally examine how supplemental vocabulary instruction provided in either whole-group or small-group settings influences low-income preschoolers' word knowledge and conceptual development. Using a within-subject design, 108 preschool children from 12 Head Start classrooms participated in an 8-week intervention, which included four topics of targeted vocabulary instruction counterbalanced in either a whole-group or small-group configuration. Pre- and posttest measures examined children's outcomes in word learning and in conceptual and categorical knowledge. Our results indicated that group size did not appear to be a powerful mechanism for intensifying instruction. Although children gained significantly in word knowledge, concepts, and categories, they did so regardless of whether they were in small or whole groups. Implications for these findings, as well as limitations of the research and directions for future research, are discussed.

Susan B. Neuman
UNIVERSITY OF MICHIGAN, NEW YORK UNIVERSITY

Tanya Kaefer
LAKEHEAD UNIVERSITY

RECENT studies reveal that vocabulary instruction can profoundly influence at-risk children's word knowledge, conceptual development, and comprehension (Beck & McKeown, 2007; Coyne et al., 2010; Marulis & Neuman, in press). Converging evidence indicates that intensive interventions that focus on explicit explanations of words, discussions of words in various contexts, review and interaction with others using these words on many occasions, and monitoring

THE ELEMENTARY SCHOOL JOURNAL VOLUME 113, NUMBER 4

children's progress on a frequent basis improve vocabulary development (Marulis & Neuman, 2010). Therefore, how to maximize children's opportunities for these kinds of intensive vocabulary experiences in the early years has received increasingly widespread attention.

Although extended school days with additional hours of instruction might be optimal (Tough, 2008), the most practical approach for enhancing instructional intensity is to organize classroom activities in ways that may best support children's vocabulary learning. Group setting can be viewed as an opportunity structure that supports different types of teacher behaviors and child engagement (Powell, Burchinal, File, & Kontos, 2008). In large-group settings, for example, children may have opportunities to develop a shared understanding of vocabulary and content through activities such as shared book reading (Holdaway, 1979). In small-group settings (Morrow & Smith, 1990), on the other hand, children may be able to engage in discourse patterns that are highly interactive and responsive to their questions and comments.

Surprisingly, there has been limited experimental research on grouping patterns in early childhood programs. Several recent correlational studies, however, have examined children's cognitive outcomes in relation to group configuration. A longitudinal, cross-national study of preprimary programs of education and care in 10 countries found that the amount of whole-group activities was negatively related to children's cognitive performance at age 7 years, and that children in preprimary settings in which free-choice activities predominated had significantly better language performance at age 7 than in settings dominated by personal care and group social activities (Montie, Xiang, & Schweinhart, 2006). Similarly, Powell and his colleagues (Powell et al., 2008), using a time-sampling method to examine 4-year-old children's active engagement in early childhood groups, found that children were more likely to be engaged when involved in a peer-group setting, and less likely to be engaged in a whole-group setting and when teachers were providing directions. These results support long-standing concerns about the developmental appropriateness of whole-group instruction (Bredekamp & Copple, 1997).

Nevertheless, an observational study by Connor and her colleagues (Connor, Morrison, & Slominski, 2006) reported that while instructional strategies carried out in small-group and individual settings were related to greater growth in children's alphabetics, whole-class instruction was associated with preschoolers' greater vocabulary growth. Although they recognized that this finding might be an artifact of their observational procedures, these colleagues suggested that whole-class instruction rather than small-group or individual instruction may have provided more opportunities to engage in hearing complex vocabulary and syntax.

These contrasting findings reflect different theoretical assumptions about how to promote instructional intensity for vocabulary learning. On the one hand, there is the view that smaller groupings (one-to-one instruction, small groups) as opposed to larger group instruction support a greater number of language interactions that can be more carefully attuned to children's instructional needs (Foorman & Torgesen, 2001). Small groupings can maximize a teacher's opportunity to provide individualized attention to a child, a process considered especially important in programs serving at-risk children. Torgesen and his colleagues (Torgesen et al., 2001), for example, have demonstrated very powerful instructional effects for one-to-one instruction compared to large-group instruction. Other studies, however, have re-

ported similar rates of growth using small groups of three or four at a time (Rashotte, MacPhee, & Torgesen, 2001). In fact, in one classic study (Morrow & Smith, 1990), as well as in more recent meta-analyses (Elbaum, Vaughn, Hughes, & Moody, 2000; Vaughn & Linan-Thompson, 2003), small-group instruction involved children in more complex interactions than individualized instruction, with subsequent improvements in learning outcomes.

On the other hand, there is evidence that whole-group instruction can support greater amounts of academically engaged time (Brophy & Good, 1986) and "time on task." For example, in his synthesis of research on teaching, Rosenshine (1986) found that teacher-directed whole-group instruction supported more time on task than small group, and subsequently was most effective for promoting gains in reading. Working with the whole class, teachers can introduce new vocabulary that can build common experiences and provide a shared basis for further exploration, problem solving, and skill development. Children are likely to hear more teacher talk and sophisticated language and may model their language interactions accordingly (Kameenui & Carnine, 1998). At the same time, however, there are trade-offs. Instruction tends to be teacher managed, rather than child initiated (Connor et al., 2009). Attention to children's individual needs may be minimal and, especially among second-language learners (Paez, Bock, & Pizzo, 2011), children may be less likely to participate than in smaller group or play-related activities. A National Research Council report on pedagogy in early childhood (Bowman, Donovan, & Burns, 2000), therefore, recommends a variety of classroom structures—including whole group, small group, and individual time with teachers.

Especially for children at risk for significant language delays, instructional intensity may also relate to the particular intervention implemented in these different group settings (Kaiser, 2011). Although general curriculum-level instruction and storybook reading may be adequate for improving vocabulary development for some children, these interventions are not likely to be sufficiently powerful for narrowing the gap for low-income, at-risk children (Institute of Education Sciences, 2008; U.S. Department of Health and Human Services, 2005). Based on two recent meta-analyses (Marulis & Neuman, 2010, in press), these children will need supplementary intervention that includes the identification of target words, explicit instruction, and frequent practice with words in meaningful settings.

In addition, these students will need more cognitive and emotional supports in the form of carefully scaffolded instruction (Foorman & Torgesen, 2001). One type of scaffolding involves an instructional regime with systematic instruction so that skills build upon one another with frequent review and practice (Neuman & Dwyer, 2011). Another type of scaffolding involves teacher-children interaction in which teachers provide modeling and support for children who are then encouraged to practice on their own with corrective feedback (Swanson, 1999). As Juel and her colleagues (Juel, Biancarosa, Coker, & Deffes, 2003) have reported, the ability to offer such scaffolded support while children are acquiring these skills takes on increasing importance for those with multiple risk factors.

Therefore, although it might appear that small-group instruction is ideally suited to these types of vocabulary interventions and teacher supports, recent research reports conflicting findings. For example, in terms of vocabulary growth in particular, the National Early Literacy Panel report (2008) and Marulis and Neuman (2010) in their meta-analyses of vocabulary interventions on word learning outcomes re-

ported no significant advantages for small-group compared to whole-group or one-on-one instruction on vocabulary outcomes for preschoolers (Marulis & Neuman, 2010, in press). These results stand in contrast to research by Elbaum and her colleagues (Elbaum, Hughes, Moody, & Vaughn, 1999), who reported positive effects for alternative grouping formats compared to whole-class instruction ($ES = .43$). However, the number of studies in her moderator analysis was small (e.g., one study of small-group instruction, three studies using multiple formats of instruction). Additionally, a minimum of four studies is typically recommended per group to reliably interpret contrasts (Bus, van IJzendoorn, & Mol, 2010). Consequently, additional research is clearly needed to determine the most effective ways to increase instructional intensity in vocabulary development for children at risk for failure to read.

Given that the question of how to strengthen children's instructional experiences is especially salient for policy makers and practitioners in current educational reform, this study was designed to experimentally examine how supplemental vocabulary instruction provided in either whole-group or small-group settings might impact word learning, concept development, and comprehension. To do so, we addressed the following questions: Are there differences between whole-class and small-group instruction in child vocabulary outcomes? Do teacher-child interactions vary by group size? And, if so, do these interactions influence vocabulary outcomes?

Method

Participants

Six Head Start programs located in public elementary schools were recruited to participate in the study. Schools were located in a severely economically depressed rural area in the rust-belt region of the United States reporting over 10% unemployment. From these programs, 12 classrooms, two from each program, were randomly selected. All teachers had their bachelor's degree in early childhood; one had a master's and early childhood state certification. All were Caucasian and had over 5 years of teaching experience. Classrooms enrolled an average of 18 3- and 4-year-old children for a 3.5-hour session, 4 days per week. There was a lead teacher and a full-time aide in each classroom. Teachers in the program reported using the High/Scope curriculum (Hohmann & Weikart, 1995), a child-centered educational approach that views children as active learners who learn best from activities they plan, carry out, and reflect upon.

A letter was sent to parents informing them of the study, and all agreed to have their child participate in the intervention and the assessments. Using a table of random digits, 10 children in each of the 12 treatment classes were randomly selected as participants in the study; however, all children participated in the supplemental vocabulary intervention (described below). These children received the 10–12-minute intervention in either a whole group or in small groups in addition to their regular instructional activities.

The 120 children (48% boys) were an average age of 4.25 years ($SD = .29$) prior to the beginning of the study. Across the 12 classrooms, demographic characteristics of the children reflected an average of 8% from racial and ethnic minorities, including

African American (3%), Hispanic (4%), and Middle Eastern (1%). Other students were Caucasian. All students were eligible for free or reduced-price lunch. Children's Peabody Picture Vocabulary Test (PPVT) scores were slightly below the national norm, with an average standard score of 93.92 ($SD = 13.88$).

Design

To examine the effects of group configuration on word learning and conceptual development, we used a within-subject design. In a within-subject design, each student receives both instructional conditions in a counterbalanced approach and serves as his/her own control. In our case, the within-subject factor was group configuration (e.g., small group, whole group). Therefore, in this study each participant received whole-group instruction on one set of words, and small group in another set of words. Knowledge of words from one set was compared with knowledge of words from another set within each individual participant.

There were a number of benefits in using this design. First, because each student received both small- and whole-group instruction, we were able to control for between-subject variability, reducing error and increasing our power to detect differences. Second, within-subject designs may control for threats to internal validity, since individuals essentially act as their own controls.

Instructional Intervention

Our instructional intervention included four 2-week topics from the World of Words (WOW) embedded multimedia curriculum (Neuman, Dwyer, Koh, & Wright, 2007), a supplemental intervention to support vocabulary instruction and conceptual development for pre-K children. The use of media is based on dual coding theory (Paivio, 1986), which posits that visual and verbal information are processed differently, creating separate representations for information processed in each channel. Vocabulary words are introduced first through video clips, then through information books, which allows children to develop an understanding of these words in multiple contexts; picture cards are used, as well, for gamelike activities. The curriculum is organized by topics that represent taxonomies (e.g., parts of the body) with properties identified for each taxonomic topic (e.g., each part of the body has a job to do). In this study, we focused on health-related topics (e.g., healthy foods, emotions, exercise, and parts of the body) representing content standards related to the Head Start outcomes framework.

Within our curriculum, words are selected that represent labels within the category structure (e.g., shoulder, eyebrows are a part of the body). We used two databases of children's early language development to calibrate the level of difficulty of words in the curriculum (Neuman & Dwyer, 2011): the MacArthur-Bates Communicative Developmental Inventories (MCDI; Dale & Fenson, 1996) and a collection of recordings of child-adult/parent interactions from the Child Language Data Exchange System (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 (Dale & Fenson, 1996). We also used a set of corpora

from the CHILDES database (MacWhinney, 2000). This database consists of transcriptions of adult-child spoken interactions in different home and laboratory settings around the world. We selected a combination of English-American corpora focusing on 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 a norming database 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 above corpora), with 56% of the primary words considered unfamiliar to preschoolers. For the second two topics, we increased the difficulty level to 60% unfamiliar words and 40% familiar words. Ten target words were taught each week, with an additional 20 words used to support their understanding.

Taught over an 8-day period (i.e., 2 weeks in Head Start), each lesson begins with a content video, introducing children to the definition of the category (e.g., healthy foods). The first video is designed to act as a prototype of the category, a particularly salient exemplar of the topic (e.g., a banana). After the video, the teacher engages the children, focusing on *wh* questions. She might ask, "Why is a banana good for you to eat?" Words are then reinforced using an information book (in this case on healthy foods) specially designed to review the words just learned and to provide redundant information in another medium. Here the teacher reads about the topic in a different, meaningful context. Based on research in multimedia (Mayer, 2001; Paivio, 1986), working memory can be increased by using dual modalities rather than just one. 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 (a total of four for each topic) that focus on new words in- and outside the category, helping to build the children's knowledge of the properties that are related to the category. New words and properties are introduced, and previous ones are reviewed. In addition, videos and teacher questions in the information book deepen children's knowledge of the concept by providing information about the topic. Picture cards are then used as a strategy for reviewing information and 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 cake a healthy food?). These challenge items are designed to encourage children to apply the concepts they have acquired to think critically about what may or may not constitute category membership (Wellman & Gelman, 1998).

Therefore, the primary mechanism for word learning and concept development in the curriculum is that categories 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 child may use information about the category to generalize to new instances and to make inferences (Rehder & Hastie, 2004). For example, when told that the novel word *lungs* refers to a part of the body, children can infer properties about *heart* or *abdomen* based on their knowledge about the human body. Children as young as 2 years of age have been shown to use category membership to make novel extensions and inferences (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998). Invoking category membership as part of word learning in previous re-

search (Neuman, Newman, & Dwyer, 2011) appeared to provide a rich background of conceptual and semantic scaffolding for new words.

The intervention was designed to be delivered in either whole-group or small-group instruction. A teacher's instructional manual accompanies each topic lesson. Teachers are encouraged to teach words and concepts explicitly and systematically, with guidance for asking children questions and comments. However, these supports are designed to illustrate how teachers may scaffold instruction to be responsive to children's inquiries and promote vocabulary development; they are not designed to be a script to be rigidly followed. Each topic lesson is intended to be approximately 10–12 minutes in length.

In the whole group, teachers engage children in a highly interactive set of responses, primarily through choral responses and recalling, at a brisk pace to support ongoing group interactions. For example, the teacher might involve children in the ongoing group conversation:

Teacher: What am I touching now? [shoulder]
Children (choral): Shoulder.
Teacher: And is my shoulder a part of my body?
Children: Yes.
Teacher: Now what am I touching? [glasses]
Children: Glasses.
Teacher: Are my glasses a part of my body?
Children: No.
Teacher: That's right. Glasses are not attached to my body.

Although teachers follow the same lesson sequence in small-group sessions, the small-group configuration allows for more tailored instruction, targeted to the children within the group. Here, teachers use more traditional scaffolding techniques during the learning process, tailored to the needs of the children with the intention of helping to resolve misconceptions and provide clarifications.

Teacher: What is a part of the body that we talked about before?
Child l: Lungs.
Teacher: That's right. Lungs help us breathe and are inside our bodies.
Child 2: Shoes.
Teacher: We talked about shoes, but is that a part of the body?
Child 3: I don't think so.
Teacher: Why not?
Child 3: Cause it's not attached.
Teacher: That's right. Shoes are something we wear to protect our body. So [to Child 2], what is another part of the body?
Child 2: Heart.
Teacher: Yes, the heart is another part of the body that's inside us; it's attached to our body.

Therefore, the lesson goals, specific target words, and strategies to link word knowledge to conceptual understandings are the same in both group settings. Both are designed to be highly interactive, involving children in ongoing exchanges with

continuous feedback supported in recent research (e.g., Corbett & Anderson, 2001; Kegal, Bus, & van IJzendoorn, 2011). However, in the small-group setting, teachers are given a greater opportunity to provide more specific feedback and to tailor responses to individual children.

Measures

The following measures were administered to the randomly selected children prior to the intervention. All assessments were given one-on-one with a trained assessor who had been certified by the second author.

Children's receptive language ability. To assess children's receptive vocabulary development, we administered the Peabody Picture Vocabulary Test—III (Dunn & Dunn, 1997) (Form A) as a pretest. This individually administered, norm-referenced test is designed for children as young as 2.5 years of age and has been shown to be a reliable and valid measure of receptive language skills. Raw scores were converted to age-referenced standard scores for the purposes of this study.

Head-to-toe task. Because children's ability to pay attention may influence responses to grouping patterns, we assessed children's behavioral self-regulation in the treatment group prior to the start of the study. In this task, children are asked to play a game where they are instructed to touch their head, then to do the opposite and touch their toes. A total of 10 commands are given verbally in random order, without feedback. Children receive 2 points for a correct response, 1 point for a self-correct, and zero points for an incorrect response. Total score possible is 20 points. Cronbach's alpha reported in a recent study was .95 (Ponitz et al., 2008); based on our data, it was .98.

Curriculum-Based Measures

We administered assessments used in our previous design-based experiment and cluster randomized controlled trial with 782 preschoolers to examine children's growth in vocabulary, concepts, and categorical knowledge (Neuman & Dwyer, 2011; Neuman et al., 2011). Cronbach's alpha across these studies ranged from .79 to .92 (vocabulary, .86–.92; concepts, .79–.81; categories, .90–.92). The following assessments were given prior to the start of the study (pretest) and after the conclusion of the intervention (posttest).

Curriculum-related word knowledge. We constructed a 40-item WOW receptive vocabulary task to measure the number of curriculum-specific words children learned throughout each unit of instruction (10 words $\times$ 4 topics). Words were randomly selected from the corpora of target words taught throughout each 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 for this sample was $\alpha = .81$.

Conceptual knowledge. The curriculum is designed to help children develop conceptual linkages among words. For example, teachers are encouraged to empha-

size how words are related (e.g., knees, ankles are attached to the body; eyeglasses are not). We designed a 32-item task to measure growth in children's conceptual understanding of target vocabulary for each topic. 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 *legs* help our bodies move around?), or not related to the concept (e.g., does a *jacket* help our bodies move around?). Each conceptual property was tested using both in-category and out-of-category target words in order to measure children's 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 questions' answers were divided equally between yes and no across the assessment, and the order of these questions was fully randomized. Children responded either yes or no to each question, and a total number of correct responses out of 32 were recorded. Reliability for this sample was $\alpha = .76$.

In the design phase of our research, children were asked to justify their yes or no responses on this conceptual measure, which provided information on the validity of their responses (see Neuman & Dwyer, 2011, for details). For this study, however, we used two sample items along with the 32-item measure in order to minimize testing time.

Categories and properties knowledge. To examine children's conceptual knowledge in greater depth prior to the start of the study, 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., fruit), a picture thematically related to the target (e.g., a scale), and an out-of-category but plausible distractor (e.g., french fries). Children were then asked to identify which item/object belonged to a particular category (Which is a healthy food?) or to identify the item/object that possessed a particular category attribute. Four category-level questions (one for each topic) and eight concept property questions (two for each topic) were assessed. Concept property questions were selected as most representative of the category. For example, children were assessed on the property "parts of the body mean that they are attached to the body," as it is a critical and defining property of the category "parts of the body." Responses were tallied for accuracy on category and property questions and for the overall assessment (total score possible = 12). Reliability in this sample was $\alpha = .88$.

Procedures

Six trained teachers, who served as assistants to classroom teachers, provided the intervention in this study. All six were former early-childhood teachers with at least 5 years or more of direct teaching experience; four had bachelor's degrees and two had master's degrees, one of which with an early childhood certification. In this respect, they were representative of the teachers in the project. Each teacher was responsible for two classrooms and for both whole-group and small-group instruction. Classrooms were randomly assigned to a whole-group/small-group rotation, counterbalanced so that each classroom received both whole-group and small-group instruction twice. Rotations were also counterbalanced so that each lesson had an equal number of whole and small groups. Throughout these lessons, the classroom teacher and aide remained in the room but did not participate in the intervention.

For each topic, classrooms were assigned to either a whole group or a small group of four or five children by their classroom teacher. Teachers were asked to group children heterogeneously, keeping in mind gender equivalence when possible. Group configurations were changed every 2 weeks, which allowed for counterbalancing of rotation patterns between classrooms across the four lessons. Those that had been in the small-group setting for the first 2-week topic lesson would then participate in a whole-group configuration for the next 2-week lesson; similarly, those that had been in the whole-group setting would then participate in the small-group setting. Children would remain in these groups throughout the entire 2-week lesson. In total, children received 8 weeks of daily supplementary vocabulary instruction (10–12 minutes per day), 4 weeks in whole-group and 4 weeks in small-group sessions.

Twelve children in the treatment did not complete posttests due to high mobility within the region. These 12 children did not significantly differ from the retained sample in terms of their PPVT scores, $t(118) = .85$, $p = .395$, or any of the target pretest measures (word knowledge, $t(118) = 1.73$, $p = .086$; conceptual knowledge, $t(118) = .74$, $p = .459$; category knowledge, $t(118) = 1.61$, $p = .110$). Our final sample included 108 children, indicating a 9% attrition rate.

Fidelity of Treatment

Two sessions per topic were videotaped for each teacher, staggered in such a way that each component of the 8-day lesson would be videotaped at least once. Videotaped visits were unannounced; teachers were not told beforehand of the anticipated visit. During the visit, the research assistant would station a video camera in an unobtrusive area that could focus on the teacher and the children, and record the lesson in its entirety. A total of eight videotaped lessons were recorded for each teacher, four in whole-group and four in small-group instruction.

Two trained research assistants coded each video for fidelity of treatment. Using a checklist, research assistants recorded the amount of time for instruction, content coverage, and the quality of the lesson. Content coverage included whether each part of the lesson was delivered (e.g., video clip, information book, picture cards, time for a challenge, discussion). Quality features included whether the lesson was well paced, engaging and facilitative of discussion, and responsive to children's comments and questions. Research assistants indicated the absence or presence of each feature and tallied for content, quality, and a total fidelity score. A second research assistant, blind to the treatment, viewed 10 lessons and coded for fidelity; reliability was 100%. Fidelity indicated 99% for content coverage and 98% for quality features across both small- and large-group configurations. Scores ranged from 97% to 100%, indicating high fidelity to treatment throughout the study.

Verbal Behaviors

Videos were transcribed verbatim into CLAN (Computerized Language Analysis) for analysis of adults' and children's verbal behaviors. CLAN is a computer program developed from the CHILDES database (MacWhinney, 2000) that facilitates the analysis of language, including frequency analyses, co-occurrence analyses, and interactional analyses. Consistent with the theoretical assumption that group config-

uration might enhance the amount and quality of interaction, we counted the total number of children's comments and questions throughout the whole-group and small-group sessions. Child language was counted at the group level. At the same time, we also analyzed the experimenters' verbal behaviors in these sessions. Using a coding scheme developed by Morrow and Smith (1990), we examined three types of adult interactive behaviors: managing (e.g., addressing disruptive behaviors, redirecting behavior), prompting (e.g., inviting children to ask questions or comment, scaffolding responses for the children), and supporting and informing (e.g., extending children's understanding of words and concepts). Two research assistants participated in a coding practice session in which the categories were described and examples were provided. Ten transcripts were randomly selected to be double-coded, and Cronbach's alpha was calculated to estimate the extent to which the ratings from the judges measured the same dimension (Crocker & Algina, 1986). The coefficients calculated in this method indicated $\alpha = .95$ for managing, $\alpha = .85$ for prompting, and $\alpha = .80$ for supporting and informing, indicating an acceptable level of interrater reliability.

Results

In this section, we address the effects of group size on children's word learning and conceptual development. To test our hypothesis, we first focus on the descriptive features of children's experiences within these group-size configurations. Each child's attendance record was compared to the coded video transcripts to determine the number of child questions/comments and teacher managing, prompting, and supporting behaviors that he or she might have experienced in the small- or whole-group setting. We use simple *t*-tests to analyze differences between children's experiences in these groupings. We then conducted repeated-measures analyses of variance using group size as our independent variable to analyze differences in child outcomes, followed by a series of multiple hierarchical linear regressions to examine our full model, taking into account differences in length of lessons as well as teacher and child exchanges.

Given that children were nested within classrooms, we employed two-level models with children nested within classrooms to examine differences for each child outcome. However, preliminary analyses revealed no significant intraclass correlation among classrooms for any of our dependent variables (ICC = 0.03 for word learning, 0.05 for concepts, 0.04 for categories; $p > .05$ for all measures). This analysis indicated that our participants varied more individually than they did by classroom, and therefore met the assumption of independence (Kenny & Lavoie, 1985). Consequently, subsequent analyses of the differences between groups were examined using analysis of variance.

The Effects of Group Size on Word Knowledge and Concept Development

Our first question was to examine the impact of group size on children's word knowledge and concept development. Descriptive statistics indicated that instruction varied by group size. As shown in Table 1, there were differences in the amount of instructional time, the number of children's comments and questions, and the

Table 1. Differences between Whole-Group and Small-Group Instruction

Characteristic	Whole Group ($N = 108$)	Small Group ($N = 108$)
Average amount of time of instruction***	14.67 (4.7)	12.96 (6.8)
Average number of child comments/questions per session***	78.56 (19.43)	63.73 (20.18)
Average number of verbal exchanges from teacher per session:		
Managing*	21.39 (10.46)	24.84 (15.45)
Prompting	76.13 (18.46)	76.89 (17.35)
Supporting	63.44 (15.38)	60.94 (12.32)

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

type of teacher's verbal exchanges in whole-group compared to small-group instruction.

Sessions were shorter in small-group settings than in whole-group settings. A paired sample *t*-test indicated that these differences were significant, $t(99) = 8.77$, $p < .001$, Cohen's $d = 1.76$. However, lessons in both configurations tended to be highly interactive, as evidenced by the number of child comments/questions as well as the number of teacher verbal exchanges per lesson. Although the number of teacher exchanges children experienced did not differ significantly across group settings, $t(109) = .58$, $p = .562$, there were differences in the quality of these exchanges. A repeated-measures analysis of variance (ANOVA) with type of verbal exchange as a three-level within-subject variable and whole or small group as a two-level within-subject variable revealed a significant interaction between the type of exchange and group-size configuration, $F(2, 218) = 6.53$, $p = .002$, partial $\eta^2 = .05$. Paired contrasts indicated a significantly greater number of managing exchanges in small-group settings, $t(109) = 2.74$, $p = .007$, Cohen's $d = .52$. No significant differences, however, were reported for the other types of teacher exchanges: $t(109) = .46$, $p = .666$ for prompting; $t(108) = 1.62$, $p = .094$, for supporting. Counter to previous research (Morrow & Smith, 1990), children engaged in significantly fewer comments and questions in small groups than in whole groups, $t(109) = 5.39$, $p < .001$, Cohen's $d = 1.03$.

Table 2 provides the means and standard deviations on child outcomes. To conduct this analysis, we first calculated whole- and small-group means for each participant, averaging the scores of the topic that the child had in each group size. We then averaged scores according to group configuration for each outcome. As shown in Table 2, mean scores indicated that pretest scores for word knowledge were relatively high prior to treatment; children appeared likely to know more than three-quarters of the words for each topic. Concepts and categories and their properties, however, were less familiar. Children appeared to know slightly more than half of the concepts and categories related to these words before treatment.

Repeated-measures ANOVA, with pretest and group size (whole vs. small group) as within-subject variables, showed a significant gain in word knowledge, $F(1, 109) = 190.26$, $p < .001$, Cohen's $d = 1.23$. However, there was no significant main effect of group size, $F(1, 109) = .02$, $p = .888$, or significant pretest by group size interaction, $F(1, 109) = .004$, $p = .950$. These results indicated that group configuration did not appear to affect word learning.

Table 2. Percent and Standard Deviations for Word Knowledge, Concepts, and Categories by Group Setting

	Whole Group		Small Group	
Outcome	Pretest	Posttest	Pretest	Posttest
Word knowledge:				
Healthy foods	.72 (.13)	.86 (.13)	.75 (.14)	.90 (.13)
Emotions	.71 (.20)	.91 (.14)	.73 (.19)	.87 (.16)
Exercise	.84 (.16)	.92 (.10)	.80 (.16)	.92 (.12)
Parts of the body	.75 (.16)	.87 (.12)	.74 (.19)	.88 (.14)
Total	.76 (.16)	.89 (.11)	.75 (.16)	.89 (.12)
Concepts:				
Healthy foods	.66 (.17)	.75 (.22)	.66 (.16)	.79 (.19)
Emotions	.52 (.15)	.66 (.18)	.53 (.15)	.63 (.19)
Exercise	.61 (.16)	.75 (.20)	.62 (.18)	.74 (.18)
Parts of the body	.61 (.16)	.78 (.18)	.59 (.15)	.74 (.20)
Total	.59 (.14)	.73 (.16)	.59 (.13)	.73 (.16)
Categories:				
Healthy foods	.54 (.23)	.69 (.26)	.57 (.23)	.76 (.22)
Emotions	.46 (.25)	.73 (.24)	.47 (.26)	.63 (.26)
Exercise	.51 (.22)	.71 (.23)	.42 (.21)	.71 (.25)
Parts of the body	.68 (.25)	.82 (.21)	.69 (.28)	.81 (.23)
Total	.55 (.20)	.74 (.22)	.53 (21)	.73 (.22)

A similar pattern was reported for concepts and categories. Repeated-measures ANOVAs for concepts and categories and properties indicated that children increased their knowledge of concepts and categories over the intervention period, $F(1, 109) = 100.31$, $p < .001$, Cohen's $d = 1.08$ for concepts; $F(1, 109) = 147.16$, $p < .001$, Cohen's $d = 1.05$ for categories and properties. However, there were no significant main effects of group size for concepts, $F(1, 109) = .04$, $p = .846$, or for categories, $F(1,109) = .27$, $p = .605$. Further, there was no significant pretest by group size interaction for either outcome: $F(1, 109) = .14$, $p = .708$, for concepts; $F(1, 109) = .061$, $p = .806$, for categories and properties. Together, these results indicated that gains in words, concepts, and categories were statistically equivalent regardless of whether students were in whole- or small-group settings.

Given the differences in the length of lessons, teacher comments, and student exchanges, we conducted three hierarchical linear regressions to determine whether children's outcomes were influenced by these contextual variables. For each, we first entered self-regulation as a potential covariate along with the differences between whole group and small group at pretest. In each case, self-regulation was not significant and was eliminated in all further analyses. After entering differences in pretest means at step 1, we then added differences in the length of lessons and differences in teacher exchanges that children experienced at step 2. At step 3, we added the mean difference in the amount of child language between whole and small groups.

As shown in Tables 3–5, differences in performance between topics taught in either whole- or small-group configurations did not appear to be influenced by the amount or type of teacher exchange or child language. For word knowledge, the overall model was not significant either at step 2, $F(5, 93) = 1.99$, $p = .086$, or step 3, $F(6, 92) = 1.72$, $p = .125$; similarly, for concepts, the overall model was not significant at step 2, $F(5, 93) = .90$, $p = .486$, or step 3, $F(6, 92) = .76$, $p = .603$. For categorical knowledge, however, there was a significant effect at step 1, with differences in pretest

Table 3. Hierarchical Regression Analysis Predicting Differences in Target Word Learning between Whole and Small Groups

Step	ΔR^2	β[a]	*t*-Value	*p*-Value	sr^{2},[b]
Step 1	.036				
Pretest		.18	1.89	.061	.04
Step 2	.061				
Pretest		.13	1.23	.223	.02
Length of lesson		.21	1.84	.069	.03
Teacher managing comments		<.01	<.01	1.000	<.01
Prompting comments		−.20	−1.36	.178	.02
Supporting comments		.09	.63	.534	<.01
Step 3	.006				
Pretest		.12	1.12	.226	.01
Length of lesson		.19	1.64	.104	.03
Teacher managing comments		<.01	<.01	1.000	<.01
Prompting comments		−.19	−1.30	.196	.02
Supporting comments		.11	.75	.454	<.01
Child comments		−.07	−.63	.528	<.01

[a] Standardized regression coefficient.
[b] Squared semipartial correlation.

scores, $t(105) = 3.22$, $p = .002$. In steps 2 and 3, although teacher comments did not account for a significant amount of additional variance, there was a significant effect of length of lesson, indicating that the difference in children's performance in whole and small groups could be partially attributed to the difference in average length of the lessons, $F(5, 93) = 3.05$, $p = .014$; $F(6, 92) = 2.66$, $p = .020$. The overall model was significant, $F(1, 97) = 10.34$, $p = .002$. Together, these results suggest that for categorical knowledge, length of lesson appeared to influence learning gains. However, differences in teacher and children's exchanges in these group configurations did not appear to impact gains in word knowledge, concepts, and categories.

Table 4. Hierarchical Regression Analysis Predicting Differences in Concepts between Whole and Small Groups

Step	ΔR^2	β[a]	*t*-Value	*p*-Value	sr^{2},[b]
Step 1	.002				
Pretest		.04	.38	.702	<.01
Step 2	.045				
Pretest		.05	.49	.622	<.01
Length of lesson		−.08	.74	.463	<.01
Teacher managing comments		.20	1.82	.071	.03
Prompting comments		.07	.50	.617	<.01
Supporting comments		−.12	−.79	.434	<.01
Step 3	.001				
Pretest		.06	.54	.589	<.01
Length of lesson		−.07	−.62	.536	<.01
Teacher managing comments		.20	1.77	.080	.03
Prompting comments		.07	.44	.659	<.01
Supporting comments		−.13	−.84	.405	<.01
Child comments		.04	.34	.737	<.01

[a] Standardized regression coefficient.
[b] Squared semipartial correlation.

Table 5. Hierarchical Regression Analysis Predicting Differences in Categorical Learning between Whole and Small Groups

Step	ΔR^2	β[a]	*t*-Value	*p*-Value	sr^{2}, [b]
Step 1	.096				
Pretest		.31	3.22	.002	.10
Step 2	.044				
Pretest		.32	2.95	.004	.09
Length of lesson		.22	2.07	.041	.04
Teacher managing comments		−.06	−.54	.592	<.01
Prompting comments		.09	.62	.535	<.01
Supporting comments		−.07	−.44	.659	<.01
Step 3	.007				
Pretest		.29	2.63	.010	.07
Length of lesson		.25	2.22	.029	.05
Teacher managing comments		−.07	−.61	.547	<.01
Prompting comments		.09	.62	.539	<.01
Supporting comments		−.11	−.68	.500	<.01
Child comments		.10	.87	.387	<.01

[a] Standardized regression coefficient.
[b] Squared semipartial correlation.

Discussion

Children enter preschool with significant differences in vocabulary knowledge, and these differences grow substantially throughout their schooling (Farkas & Beron, 2004; Hart & Risley, 1995). Converging evidence (National Early Literacy Panel, 2008) has suggested that general curriculum-level instruction is insufficient to meet the needs of children who are at risk for experiencing reading delays because of vocabulary difficulties. Rather, there is growing recognition that these children will need to accelerate their vocabulary development through more intensive instruction if they are to keep pace in reading growth with their age peers.

Increasing the intensity of instruction to prevent reading difficulties, therefore, has taken on a sense of urgency for children's very early years. Farkas and Beron (2004), in an analysis of the children of the NLSY79 data set, for example, found that the highest rate of vocabulary growth occurs during the preschool ages (birth through 5) and that this rate declines for each subsequent age period. Consequently, the preschool years play a critical role in the development of oral vocabulary knowledge (Dickinson & Porche, 2011) and for potentially accelerating its development (Copple & Bredekamp, 2009; Neuman et al., 2011).

However, strategies for intensifying instruction, particularly for very young children, are not all that straightforward. Increasing the intensity of instruction by extending the total time children spend in classrooms is often not feasible in many early childhood programs where teachers must juggle multiple goals within half-day, 4-day, or 9-month programs; further, increasing the timing, pacing, and frequency of practices—typical strategies for older students—may not be appropriate for young children who need to anchor their learning in playful contexts.

For early childhood programs, therefore, the most practical methods for intensifying instruction are to group children in ways that might maximize language learning and to provide targeted, content-rich interventions with special emphases on these vocabulary skills (Powell et al., 2008). Neither strategy for improving vocabu-

lary, however, has received wide-scale endorsement. For example, although Marulis and Neuman (in press) found in a recent meta-analysis of 51 studies that vocabulary interventions for children at risk were effective with an overall effect size of Hedges's $g = .88$, they were not sufficiently powerful to close the gap between those who were more average learners. Further, group configuration (e.g., small group, whole group, one-on-one) in this meta-analysis, as well as others (Marulis & Neuman, 2010; National Early Literacy Panel, 2008), has not yielded significant effect sizes. At least for children in the early years, group size has not been shown to moderate growth in vocabulary development.

As a mechanism for intensifying instruction, however, group size and its effects have often been confounded by the person who is delivering instruction. In these meta-analyses and others (Mol, Bus, & deJong, 2009), for example, interventions that were delivered by teachers or experimenters had large effect sizes, while those delivered by paraprofessionals had minimal to no effects on children's outcomes. Therefore, in this study we chose to conduct a more rigorous analysis of group size by involving highly trained teachers who would engage in the intervention with high fidelity. By holding the quality of teaching constant, and by using a within-subject design in which teachers taught children in both whole- and small-group settings, we could examine this factor more deliberately as a strategy for enhancing children's vocabulary outcomes.

Our results indicated that group size did not appear to be a powerful mechanism for intensifying instruction. Although children gained significantly in word knowledge, concepts, and categories, they did so regardless of whether they were in small or whole groups. In this respect, our findings substantiated what has been found in previous meta-analyses on the effects of vocabulary interventions. Given roughly the same amount of time for instruction and the same quality of instruction, children appeared to gain equally well in both learning contexts (Marulis & Neuman, 2010, in press).

These results are contrary to studies that have examined children's active engagement in learning activities in early childhood (Kontos, Burchinal, Howes, Wisseh, & Galinsky, 2002; Powell et al., 2008). Powell and his colleagues found that children were more likely to be actively engaged when involved in a peer group, and least likely to be actively engaged in a whole-group setting. However, their findings might reflect the particular task within the whole-group setting. For example, their research found that whole-group instruction was dominated by teacher talk with children passively listening. In contrast, our qualitative analysis of teacher exchanges and child comments and questions showed a highly interactive set of exchanges in both settings. These results suggest that the setting itself may not determine activity level; rather, it is the teachers' interactive style or the content of the lesson that influences active engagement.

To ensure a similar dosage of treatment, we involved children in daily small-group lessons; as a result, we found that lessons were somewhat shorter than in whole-group settings. Previous studies (Allington, 1977; Brophy & Alleman, 1991) have shown that transition time between activities can take time away from instructional time. Although not to the detriment of learning, we found this to be the case in this study. Videotape analysis indicated that these young children had difficulty transitioning from one activity to the next, often taking time to settle into the lesson. This may have been due to the brevity of the lesson, or the multiple transitions that were

inherent in our experimental design (e.g., regrouping, moving from one space to another, etc.). As a result, teachers engaged in more "management"-related exchanges than in the whole-group setting.

Our results also run counter to studies (Foorman & Torgesen, 2001) that have shown greater effects of small-group instruction for children at risk of later reading disabilities. Meta-analyses (Elbaum et al., 2000; Vaughn & Linan-Thompson, 2003) have reported that children with reading disabilities learn more rapidly in small homogeneous groups compared to larger group settings or even one-on-one. These studies, however, have involved older students who have been diagnosed with reading difficulties. To the contrary, in this study children were grouped heterogeneously, with lessons designed to be responsive to children's comments and questions, but not diagnostic or designed to target particular skill deficits, for example. For these reasons, we might speculate why group size did not affect children's outcomes: small-group size in and of itself does not influence learning. Rather, small-group instruction may only be beneficial if teachers take advantage of what the group configuration has to offer, such as a greater attention to children's individual needs. Responsive though undifferentiated instruction might be more efficiently provided in whole groups.

Grouping decisions might also be dependent on the teachers' instructional goals. In the case of vocabulary development, for example, a shared understanding of content-related words learned in whole-group settings might be to the benefit of all learners and serve as a foundation for children to engage more deeply in independent explorations or in small-group-inquiry activities (Bowman et al., 2000). Nevertheless, used in excess, teaching in whole groups might come at a cost, limiting children's spontaneous exploration and discovery (Bonawitz et al., 2010). Consequently, group setting might be viewed as an opportunity structure that supports different types of content learning and different modes of children's engagement.

There are, of course, important limitations in our study. First, even our expert teachers found it challenging to engage three small groups in vocabulary instruction per day. In this case, we needed to weigh the needs for an equal dosage of treatment within our experimental design; however, this would not be an optimal strategy for classroom teachers on a regular basis. Second, our small-group sessions may not have been representative of typical instruction within this group configuration. Even though we encouraged teachers to be spontaneous and responsive in their interactions with the children, they were working within an intervention with explicit learning goals. And, third, we engaged outside expert teachers to deliver the instruction; although this ensured the quality and the fidelity of the treatment, we cannot generalize our findings to typical instruction in the preschool classroom. It could be that these results overestimated the effects of a specialized intervention by including a novel treatment; on the hand, they might have underestimated the effects by involving outsiders who could not provide for further extensions to instruction in play centers or other activities throughout the children's day. Future research should consider a longer-term intervention in which classroom teachers deliver the instruction in a more classic randomized controlled experimental design to measure the generalizability of these effects.

In summary, our findings indicate that grouping patterns, in and of themselves, do not appear to ensure greater intensity of vocabulary instruction. Rather, it is what

teachers do within these group structures that may determine whether or not they lead to more intensive instruction for preschoolers at risk for failure in reading.

Note

The authors would like to acknowledge the generous support from the Institute of Education Sciences (R305A090013).

References

Allington, R. (1977). If they don't read much, how they ever gonna get good? *Journal of Reading*, **21**(1), 57–61.

Beck, I., & McKeown, M. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, **107**(3), 251–271. doi: 10.1086/511706

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N., Spelke, E., & Schulz, L. (2010). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, **120**(3), 322–330. doi:10.1016/j.cognition.2010.10.001

Bowman, B., Donovan, S., & Burns, M. S. (2000). *Eager to learn: Educating our preschoolers*. Washington, D.C.: National Academy Press.

Bredekamp, S., & Copple, C. (1997). *Developmentally appropriate practice—revised*. Washington, D.C.: National Association for the Education of Young Children.

Brophy, J., & Alleman, J. (1991). A caveat: Curriculum integration isn't always a good idea. *Educational Leadership*, **49**(2), 66.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (Vol. **3**, pp. 328–375). New York: Macmillan.

Bus, A., van IJzendoorn, M., & Mol, S. (2010). Meta-analysis. In N. Duke & M. Mallette (Eds.), *Literacy research methodologies* (Vol. **2**, pp. 270–300.). New York: Guilford.

Connor, C., Morrison, F., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology*, **98**(4), 6654–6689. doi:10.1037/0022-0663.98.4.665

Connor, C., Piasta, S., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child X instruction interactions on first graders' literacy development. *Child Development*, **80**(1), 77–100.

Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice in early childhood programs serving birth through eight*. Washington, D.C.: NAEYC.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In J. Jacko, A. Sears, M. Beaudouin-Lafon, & R. Jacob (Eds.), *CHI 2001 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 245–252). New York: ACM Press.

Coyne, M., McCoach, D., Loftus, S., Ruby, M., Creveceour, Y., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Educational Effectiveness*, **3**(2), 93–120. doi:10.1080/19345741003592410

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.

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

Dickinson, D., & Porche, M. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, **82**(3), 870–886. doi:10.1111/j.1467-8624.2011.01576.x

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

Elbaum, B., Hughes, M., Moody, S., & Vaughn, S. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children*, **65**(3), 399–415.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure: A meta-analysis of the intervention research. *Journal of Educational Psychology*, **92**(4), 605–619.

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

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, **16**(4), 203–212. doi:10.1111/0938–8982.00020

Gelman, S., Coley, J., Rosengren, K., Hartman, E., & Pappas, A. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, **63**(1), i–148. doi:10.2307/1166211

Gelman, S., & O'Reilly, A. (1988). Children's inductive inferences within super-ordinate categories: The role of language and category structure. *Child Development*, **59**(4), 876–887. doi:10.2307/1130255

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

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

Holdaway, D. (1979). *The foundations of literacy*. Portsmouth, NH: Heinemann.

Institute of Education Sciences. (2008). *Effects of preschool curriculum programs on school readiness: Report from the Preschool Curriculum Evaluation Research Initiative*. Washington, D.C.: U.S. Department of Education.

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. Retrieved from http://www.ascd.org/

Kaiser, A. (2011). Young children with language impairments: Challenges in transition in reading. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (Vol. **3**, pp. 153–171). New York: Guilford.

Kameenui, E., & Carnine, D. (1998). *Effective teaching strategies that accommodate diverse learners*. Upper Saddle River, NJ: Merrill.

Kegal, C., Bus, A., & van IJzendoorn, M. (2011). Differential susceptibility in early literacy instruction through computer games: The role of the dopamine D4 receptor gene (DRD4). *Mind, Brain, and Education*, **5**(2), 71–78. doi:10.1111/j.1751-228X.2011.01112.x

Kenny, D. A., & LaVoie, L. (1985). Interpersonal relations and group processes. *Journal of Personality and Social Psychology*, **48**(2), 339–348.

Kontos, S., Burchinal, M., Howes, C., Wisseh, S., & Galinsky, E. (2002). An eco-behavioral approach to examining the contextual effects of early childhood classrooms. *Early Childhood Research Quarterly*, **17**(2), 239–258. doi:10.1016/S0885-2006(02)00147-3

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

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary training on word learning: A meta-analysis. *Review of Educational Research*, **80**(3), 300–335. doi:10.3102/0034654310377087

Marulis, L., & Neuman, S. B. (in press). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Educational Effectiveness*.

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

Medin, D., Lynch, E., & Solomon, K. (2000). Are there kinds of concepts? *Annual Review of Psychology*, **51**, 121–147. doi:10.1146/annurev.psych.51.1.121

Mol, S., Bus, A., & deJong, M. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, **79**(2), 979–1007. doi:10.3102/0034654309332561

Montie, J., Xiang, Z., & Schweinhart, L. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, **21**(3), 313–331. doi:10.1016/j.ecresq.2006.07.007

Morrow, L., & Smith, J. (1990). The effects of group setting on interactive storybook reading. *Reading Research Quarterly*, **25**(3), 213–231. doi:10.2307/748003

National Early Literacy Panel. (2008). *Developing early literacy*. Washington, D.C.: National Institute for Literacy.

Neuman, S. B., & Dwyer, J. (2011). Developing vocabulary and conceptual knowledge for low-income preschoolers: A design experiment. *Journal of Literacy Research*, **43**, 103–129.

Neuman, S. B., Dwyer, J., Koh, S., & Wright, T. (2007). *The world of words: A vocabulary intervention for preschool children.* Ann Arbor: University of Michigan.

Neuman, S. B., Newman, E., & Dwyer, J. (2011). Educational effectiveness of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster randomized trial. *Reading Research Quarterly*, **46**, 249–272.

Paez, M., Bock, K., & Pizzo, L. (2011). Supporting the language and early literacy skills of English language learners: Effective practices and future directions. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (Vol. **3**, pp. 136–152). New York: Guilford.

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

Ponitz, C., McClelland, M., Jewkes, A., Connor, C., Farris, C., & Morrison, F. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, **23**(2), 141–158. doi:10.1016/j.ecresq.2007.01.004

Powell, D., Burchinal, M., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, **23**(1), 108–123. doi:10.1016/j.ecresq.2007.04.001

Rashotte, C., MacPhee, K., & Torgesen, J. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disabilities Quarterly*, **24**(2), 119–134. doi:10.2307/1511068

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

Rosenshine, B. (1986). Synthesis of research on explicit teaching. *Educational Leadership*, **43**(7), 60–69. Retrieved from http://www.ascd.org/

Swanson, H. L. (1999). Instructional components that predict treatment outcomes for students with learning disabilities: Support for a combined strategy and direct instruction model. *Learning Disabilities Research & Practice*, **14**(3), 129–140. doi:10.1207/sldrp1403_1

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, **34**(1), 33–58. doi:10.1177/002221940103400104

Tough, P. (2008). *Whatever it takes*. Boston: Houghton-Mifflin.

U.S. Department of Health and Human Services. (2005). *Head Start impact study: First year findings*. Washington, D.C.: Administration for Children and Families.

Vaughn, S., & Linan-Thompson, S. (2003). Group size and time allotted to intervention: Effects for students with reading difficulties. In B. R. Foorman (Ed.), *Interventions for children at-risk for reading difficulties or identified with reading difficulties*. Timonium, MD: York.

Wellman, H., & Gelman, S. (1998). Knowledge acquisition in foundational domains. In D. Kuhn & R. Siegler (Vol. Eds.), & N. Eisenberg (Series Ed.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (5th ed., pp. 523–573). New York: Wiley.