The Effects of Vocabulary Intervention on Young Children's Word Learning:

A Meta-analysis

Loren Marie Marulis

Susan B. Neuman

University of Michigan

The Effects of Vocabulary Intervention on Young Children's Word Learning:

A Meta-analysis

This meta-analysis examines the effects of vocabulary interventions on pre-K and kindergarten children's oral language development. The authors quantitatively reviewed 67 studies and 216 effect sizes to better understand the impact of training on word learning. Results indicated an overall effect size of .88, demonstrating on average, a gain of nearly one standard deviation on vocabulary measures. Moderator analyses reported greater effects for trained adults in providing the treatment, combined pedagogical strategies that included explicit and implicit instruction, and author-created measures compared to standardized measures. Middle and upper-income at-risk children were significantly more likely to benefit from vocabulary intervention than those students also at risk and poor. These results indicate that although they might improve oral language skills, vocabulary interventions are not sufficiently powerful to close the gap—even in the preschool and kindergarten years.

The Effects of Vocabulary Intervention on Young Children's Word Learning:

A Meta-analysis

Vocabulary—knowledge of word meanings—is a powerful predictor of reading proficiency (Beck & McKeown, 2007; Snow, Burns, & Griffin, 1998). The size of a person's vocabulary is strongly related to how well that person will understand what he or she reads (Stahl & Nagy, 2006). Logically, children will need to know the words that make up written texts in order to understand them, especially as the vocabulary demands of content-related materials increase in the upper grades. Consequently, the larger the reader's vocabulary (either oral or print), the easier it will be to make sense of the text. Studies (Cunningham & Stanovich, 1997; Scarborough, 2001) have shown a substantial relationship between vocabulary size in first grade and reading comprehension later on.

It is well established, however, that there are profound differences in vocabulary knowledge among children from different socioeconomic groups beginning in young toddlerhood through high school (Hart & Risley, 1995; Hoff, 2003). Moats (1999), for example, estimated the difference at school entry to be about 15,000 words, with linguistically disadvantaged knowing about 5,000 words compared to the more advantaged with 20,000 words. Further, children from low-income groups tend to build their vocabulary at slower rates than children from high SES groups (Anderson & Nagy, 1992), potentially creating a cumulative disadvantage over time. By grade 4, children with below-vocabulary levels, even if they have adequate word identification skills, are likely to "slump" in reading comprehension, unable to profit from independent reading of most grade level texts (Biemiller & Boote, 2006).

With the recognition of these vast differences and their consequences for subsequent reading achievement, there is an emerging consensus that intensive interventions are needed early on to focus on enhancing children's vocabulary (Author, 2009). Average children acquire many hundreds of word meanings each year during the first 7 years of vocabulary acquisition. In order to catch up, therefore, children with vocabulary limitations will need to acquire several hundred words <u>in addition</u> to what they would otherwise learn (Biemiller, 2006). In essence, interventions will have to accelerate, not only to improve children's vocabulary development, to narrow the achievement gap.

To date, however, little is known about the effectiveness of vocabulary training on changes in general vocabulary. Previous published meta-analyses (Elleman, Lindo, Morphy, & Compton, 2009; Stahl & Fairbanks, 1986) have addressed the impact of vocabulary interventions on reading comprehension. Stahl and Fairbanks (1986), for example, reported an average effect size of .97 of vocabulary instruction on comprehension of pages containing taught words; a more modest effect size of .30 for global measures of comprehension. By contrast, Elleman and her colleagues (2009), using a more restrictive criterion for their selection of studies, found less substantial effects: a positive overall effect size of .50 of training programs on comprehension using author-created measures, a .10 effect size for standardized measures. In addition to these meta-analyses, the National Reading Panel (2000) conducted a narrative review of published experimental and quasi-experimental studies evaluating vocabulary instruction on comprehension skills, and found that it was "generally effective" for improving comprehension.

Although elucidating, neither of these meta-analyses nor the narrative review addressed the effects of training on word learning, a more proximal measure of the impact of the interventions. Further, the majority of the studies focus on vocabulary training as it applies to printed text. For example, exemplary training strategies supported by the National Reading Panel report (2000) include text restructuring, repetition and multiple exposures to words in text, and re-reading, assuming that children are already reading at least at a rudimentary level. In fact, there is a curious logic in many of these vocabulary training studies. As noted in both recent and past meta-analyses, much of this research has emphasized building children's skills in vocabulary by increasing the amount of reading. Given that poor readers are likely to select less challenging texts than average or above readers, rather than closing the gap this strategy could have the unfortunate potential of exacerbating vocabulary differentials.

Moreover, the combination of both oral and print-based vocabulary training interventions makes it difficult to disentangle whether difficulties in comprehension lie within the word identification demands or in the vocabulary of the text. Mol and her colleagues (Mol, Bus, & deJong, 2009) avoided this potential confound by separating oral language outcomes and print-related skills. Focusing specifically on the impact of interactive story book reading, their recent meta-analysis reported a moderate effect size for expressive vocabulary (.28), and a slightly more modest effect size for print knowledge (.25). However, the largest effect sizes appeared to be present only in experiments that were highly controlled, and executed by the examiners. Teachers appeared to have difficulty with fostering the same growth in young children's language skills as researchers. Further, in another recent meta-analysis by Mol and her colleagues (Mol, Bus, deJong, & Smeets, 2008) examining parent-child storybook readings' on oral language development, strikingly two groups did not appear benefit from the intervention: those children at risk for language and literacy impairments, or kindergarten children.

Conceivably, if we are to substantially narrow the gap for children who have limited vocabulary skills, we need to better understand the potential impact of interventions specifically targeted to accelerate development, and the characteristics of those that may be most effective at increasing children's vocabulary. This meta-analysis was designed to build on the previous work (Elleman, Lindo, Morphy, & Compton, 2009; Mol, Bus, & deJong, 2009; Mol, Bus, deJong, & Smeets, 2008; National Early Literacy Panel, 2008) in several ways. First, since major vocabulary problems develop during the earlier years before children can read fluently, we examine vocabulary training interventions prior to formal reading instruction, in preschool and kindergarten. Farkas and Beron (2004), for example, in a recent analysis the children of the National Longitudinal Survey of Youth 1979 cohort (NLYDY79), found more than half of the social class effect on early oral language was attributable to the years before five, and that rates of vocabulary growth declined for each subsequent age period. Second, we elaborate on the work of Mol and her colleagues (Mol, Bus, & deJong, 2009; Mol, Bus, deJong, & Smeets, 2008) to include all vocabulary interventions in the early years in addition to interactive reading. Third, we examine the impact of these interventions to the more proximal variable of word learning to examine whether they might affect overall vocabulary growth. And finally, we examine specific characteristics that appear to influence word learning.

This meta-analysis, therefore, expands the current literature by addressing the following questions:

1) Are vocabulary interventions an effective method for teaching word learning to young children?
2) What methodological characteristics are associated with effect size?
3) Is there evidence that vocabulary interventions narrow the achievement gap?

To address these questions, it was essential to include all vocabulary interventions rather than a subset of the most common or most examined types. This approach allowed for a thorough and comprehensive meta-analysis, permitting us not only to identify and examine non-traditional interventions, but the wider range of variables associated with treatments. Based on previous research, we anticipated that our resulting sample would be highly heterogeneous. We viewed this as an inevitable compromise and therefore planned to focus much of our analysis on subgroup moderators, helping to explain these effects. This procedure, in addition to the use of the random effects model, is recommended in meta-analyses conducted on diverse literature such as ours (Cooper, Hedges, & Valentine, 2009; Landis & Koch, 1977; Raudenbush, 1994; Wood & Eagly, 2009)

Method

Search Strategy and Selection Criteria

This meta-analysis examines the effects of vocabulary training on the receptive and expressive language of children not yet reading conventionally. Studies were included when they met the following criteria: a) The study included a training, intervention or specific teaching technique to increase word learning; b) a (quasi-) experimental design was applied incorporating one or more of the following: a randomized controlled trial, a pretest-intervention-posttest with a control group, or a post-intervention comparison between pre-existing groups (e.g., two kindergarten classrooms); c) participants had no mental, physical, or sensory handicaps and were within ages birth through eight, to examine the potential effects of interventions prior to preschool through the early childhood years; d) the study was conducted with English words, excluding foreign language or nonsense words (to be able to make comparable comparisons across studies); e) outcome variables included a dependent variable that measured word learning, identified as either expressive or receptive vocabulary development or both. The measure could be standardized (e.g. PPVT) or an author-created measure.

Our goal was to obtain the corpus of vocabulary intervention studies that met our eligibility criteria including both published and unpublished studies. To do so, we developed comprehensive search terms to capture the various iterations and ways of describing relevant studies. In addition we consulted an Education Specialist Librarian to ensure that we included all different keywords and tags used by the various databases. We searched the following electronic databases: PsycINFO, ISI Web of Science, Education Abstracts ProQuest Dissertations and Theses and Educational Resources Information Center (ERIC; CSA; OCLC FirstSearch) through September, 2008 using the following search terms: word learning OR vocabulary AND intervention; OR instruction, training, learning, development, teaching. This search yielded 53,754 citations.

We imported all citations into the bibliographic software Endnote, to maintain and code our library of citations. We then performed preliminary exclusion coding on these citations; studies were excluded if children were older than the cut-off established in the study; off-topic and not related to word learning; study was conducted in a language other than English; and a conference proceeding that included no primary data. To be excluded at this phase, citations needed to meet the above criteria with 100% certainty. Twelve exclusion coders were trained by the first author and prior to beginning coding, inter-rater reliability was established (Cohen's kappas ranged from .9-1.0). In addition, once exclusion coding was completed, 25% of the citations were independently coded by two research assistants, which also resulted in high levels of agreement (k=.96). The exclusion coding revealed 3,586 relevant citations subsequently retrieved and read in full.

In addition, we also contacted experts and authors in the field for any published and unpublished data and other relevant references. We sent out a total of 95 emails and received 28 responses (29% response rate), which generated 12 manuscripts. In total, therefore, our process yielded a total of 3,598 manuscripts.

Inclusion Coding

Our next task was to examine papers with our eligibility criteria in mind. Four graduate students were trained in both meta-analytic coding procedures and those specific to our project. After sufficient training was completed, the four coders read ten studies together and discussed whether each should be included based on our inclusion criteria. All disagreements were resolved through discussion until 100% agreement was reached. Following this discussion, a training set of 50 studies was coded separately by all four coders. The level of agreement reached between the four raters on their inclusion determination (Fleiss' Kappa = .96) fell well within range (Landis & Koch, 1977). Subsequently, each coder individually coded the remainder of the studies. One hundred and eleven studies met all criteria and were set aside for the comprehensive study variable coding. These studies were divided into two groups: those that were targeted on oral language and word learning prior to conventional reading (birth through kindergarten), k=64 and interventions that focused on word learning in texts (grades 1-3), k=40.

Interventions for these two groups represented different foci of instruction and different goals in measurement. For example, interventions birth through kindergarten, focused on oral language development through listening and speaking, with concomitant changes in receptive and expressive language. Intervention for grades 1-3 emphasized the ability to identify and understand vocabulary words in print, and children’s subsequent understanding of these words in a text. Consequently, the decision was made to conduct this particular meta-analysis on studies targeting the very early years of vocabulary development (birth through age 6), considered to be a period of time when word learning accelerates to examine their potential effects on children’s growth and development.

Study Characteristics/ Potential Moderators

We coded important study characteristics as well as 'coder determined' variables for key characteristics. We identified nine characteristics for planned moderator analyses based on previous research and findings (Mol et al., 2009; National Early Literacy Panel, 2008; National Reading Panel Report, 2000): the adult who carried out the intervention, group size, duration of the intervention, type of training, at-risk status, socioeconomic status (SES), research design, and the type and focus of the dependent measure (s) used to determine changes in word learning. If study descriptions were unclear or key characteristics missing, authors were contacted to obtain the information necessary for coding. If this was not possible or if the information was unavailable, we coded the variable as missing. Studies were excluded from the particular analysis when variables used in specific moderator analyses were missing. For example, if it was unclear whether study participants were at-risk or not, the effect size data from that study would be excluded from the at-risk moderator analysis but included in the overall mean effect size calculation and in other moderator analyses to maintain statistical power (and to increase the comprehensiveness and precision of the research synthesis (Bakermans-Kranenburg, van Ijzendoorn, & Juffer, 2003).

Due to the large number of variables and the importance of accuracy, training for this coding process involved extensive tutorials on research design, variable coding, and practical coding techniques. At the conclusion of the training, all four coders coded five papers together. Following the coding, we discussed each study, and revised the coding manual and protocol sheets accordingly. Next, the coders coded five papers independently. Fleiss' kappa was calculated at .67, which while falling within the "substantial agreement" range, was not sufficiently high enough to allow for proper use of moderator analysis. Borenstein, Hedges, Higgins, and Rothstein (2005) and Lipsey and Wilson (2004) recommend an agreement level of at least .81. Therefore, we initiated a second round of coding and revisions to the coding sheets. We independently coded an additional 35 papers (over 60 studies; 150 effect sizes) and achieved an agreement level within the "almost perfect agreement" range (k=.89). Studies were then coded individually by one of the four trained coders.

Analytic Strategy

We entered the data into the *Comprehensive Meta-Analysis* software *(CMA*; Borenstein, Hedges, Higgins, & Rothstein, 2008). Through the use of the CMA program, we were able to enter various types of reported data, including means and standard deviations, mean gain scores, *F* or *t* statistic data, categorical data, odds ratios, chi-square data and so on. Due to the ability to enter data in over 100 formats, we were able to calculate effect sizes even when we were unable to compute the magnitude of the treatment groups' improvement through treatment and control group mean differences and standard deviations, the standard way to calculate effect sizes.

We estimated all effect sizes using the Hedges' *g* coefficient, a more conservative form of the Cohen's *d* effect size estimate. Hedges' *g* uses a correction factor J to correct for bias due to sample size, and weights each effect size by the standard error of the effect size. In this way, less precise estimates are given less weight in the analyses. Hedges *g* is computed as:

$$g = \frac{M_{\text{treatment}} - M_{\text{control}}}{\sqrt{\text{MSE}_{\text{within S's}}}}$$

The correction factor J is calculated: J = 1 - (3 / (4 * df - 1) where df = N_{TOTAL} - 2. To obtain Hedges *g* from Cohen's *d*, the following calculations can be made: *g* = *d* * J, StdErr(g) = StdErr(d) * J, Variance(g) = StdErr(g)^2.

We weighted the effect sizes by the inverse of their error variances ($1/SE^2$) in order to factor in the proportionate reliability of each one to the overall analysis (Shadish & Haddock, 1994). The resultant effect size gives the magnitude of the treatment effect, with an effect size of .20 considered small, .50 in the moderate range, and .80, large (Cohen, 1988). In addition, in order to avoid dependency in our effect size data (e.g., when a study used more than one outcome measure or treatment group resulting in multiple effect sizes per study), we randomly selected one effect size per study across the variable of interest (moderator variable) prior to calculating the overall effect size (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper & Hedges, 1994). Similarly, for the overall mean effect size calculation, one mean effect size was used per study so that there was one treatment group compared to one comparison group for each included study.

We used a random effects model for our overall effect sizes, 95% confidence intervals around the point estimates in order to address heterogeneity. Within a random effects model, the variance includes the within-study variance plus the estimate of the between studies variance (Borenstein, Hedges, Higgins, & Rothstein, 2009). Random effects models are used when there is reason to suspect that variability is not limited to sampling error (Lipsey & Wilson, 2001), which we believed was a good description of our sample of studies. Under this model, we assumed that the true population effect sizes might vary from study to study, distributed about a mean. However, since our sample of studies involved a larger corpus of vocabulary interventions than previous meta-analyses, we expected to have a dispersion of effect sizes. Therefore, in addition to the random effects model, we planned subgroup analyses on the characteristics we believed might moderate these effects.

Outliers and Publication Bias

Only one effect size was considered an outlier (i.e., 4 standard deviations above the sample mean; SD=.53). This outlying effect size was quite large (Lipsey & Wilson, 2001, mean effect size *g*=5.43, SE=.69). However, due to its low precision (large standard error), it was weighted the lowest in our analysis, and did not disproportionately influence the mean effect size. To substantiate this claim, we compared our analysis with this outlying value (*g*=.89, SE=.065, CI_{95}= .76, 1.00) and without (*g*=.86, SE=.062 CI_{95}=.74, .98), and found no significant difference. Nevertheless, in order to reduce the impact of this large outlier in the planned moderator analyses, we windsorized it by resetting this effect size to the next largest effects size of 2.13 (which was only two standard deviations from the mean). This also allowed for a smoother distribution of effect sizes and a less extreme upper limit. All subsequent analyses, therefore, were conducted using the windsorized value. Effect sizes ranged from -10 to 2.13 with 37 effect sizes below the sample mean and 29 above the sample mean.

Shown in the Stem and Leaf plot in Figure 1, effect sizes formed a relatively normal distribution, though there was a slight skewness toward the lower end. The largest number of effect sizes was within the 0.4, 0.5 and 0.6 stems, with the modal effect size reported at 0.45 and the median, 0.82.

Insert Figure 1 about here

In addition to the precautions described in our sampling strategies, we calculated a fail-safe N, which indicated that we would need to be missing 17,582 studies in order to potentially invalidate significant effect size results (rejecting the null hypothesis that an effect size is the same as 0.00). This number far exceeded the criterion number (i.e., 5k+10=345 where k=67 studies, (Rosenthal, 1991). We also calculated the Orwin fail-safe which allowed us to use a value other than null against an effect size criterion (i.e. rather than a p-value). This test addressed the possibility that missing studies, if included, would diminish, rather than invalidate, our findings. This allowed us to evaluate how many missing studies would need to exist to bring our mean below a moderate effect size (0.5). Even with this criterion value of 0. 5, our Orwin fail-safe number was 555, meaning that we would have had to locate 555 studies with mean effect sizes of 0.49 to bring our overall Hedge's *g* under a moderate 0. 5. We were confident, therefore, that we could proceed with our analysis, and not be overly concerned about publication bias.

Results

The final set of intervention studies targeting vocabulary training in educational settings for preschool and kindergarten aged children (through age 6 when grade was not specified) comprised of 64 papers which yielded a total of 67 studies and 216 effect sizes. In total, 5,929 children ($N_{\text{experimental group}}$= 3,202; $N_{\text{control group}}$=2,727) were studied. Seventy percent of the studies were published in peer-reviewed journals, and 60% of the children sampled were in pre-K classrooms. The typical study was quasi-experimental and used an alternative treatment control condition. The majority of studies used a standardized measure of receptive or expressive language, and about a third used author-created measures.

We used researcher-specifications to describe the interventions. As shown in the Table, storybook reading and dialogic reading were the most prevalent interventions. However, noted in both the National Reading Panel (2000) and the National Early Literacy Panel (2008), there were wide variations across interventions. For specific descriptive study characteristics, see Table 1.

Insert Table 1 about here

We expected our sample to be heterogeneous which was subsequently confirmed (Q (66)=551.54, p<.0001, I^2=.88). Total variability due to true heterogeneity or between-studies variability was 88%, indicating that 88% of the variance was between studies variance (e.g. could be explained by study-level covariates), and 12% of the variance was within-studies based on random error. Since a high degree of between-subjects heterogeneity was present, we further partitioned effect sizes by creating subgroups using study factors that have been shown in previous research to influence them (e.g., Elleman et al, 2009; Mol et al 2008; National Reading Panel, 2000; Stahl & Fairbanks, 1986). These key characteristics of instruction were addressed through moderator analysis to determine if the variance could be explained.

Overall Effect Sizes

To examine the benefit of vocabulary training on word learning, we first calculated an overall effect size for pre-K and K. The overall effect size was *g*=.88, SE=.06, CI_{95}=.76, 1.01, p<.0001. Vocabulary training demonstrated a large effect on word learning in pre-K, (*g*= .85, SE=.09, CI_{95}=.68, 1.01, p <.0001), and an even larger effect for kindergarten (*g*=.94, SE=.11, CI_{95}=.73, 1.14, p <.0001). Though the magnitude of the effect was slightly larger for kindergarten students, differences were not statistically significant, Q(1)=.48, p=.49. These effect sizes are considered both educationally relevant and large (Cohen, 1988). However, as shown in Figure 1, the dispersion of the effect sizes was substantial, reflecting the large heterogeneity of the interventions in the sample.

Not too surprisingly, we found that published studies had significantly higher effects sizes (*g*=.95, SE =.084, CI_{95}=.79, 1.11) than unpublished studies (*g*=.71, SE=.087, CI_{95}=.54, .88) Q(1)=4.53, p <.05). Therefore, our overall effect size could be considered conservative due to the inclusion of a considerable number of unpublished studies (20/67).

Eleven studies also reported a delayed posttest. To avoid dependency of effect sizes, we excluded the delayed posttests (e.g. defined as three or more days after the completion of the intervention) from our overall analysis (28 effect sizes). The mean effect size for the delayed posttests (*g*=1.06, SE=.22, CI_{95}=.58, 1.45) did not differ significantly from those immediate posttests (*g*=.88, SE=.06, CI_{95}=.76, 1.01). These results indicated that moderate effects persisted over time for these 11 studies (3-180 days post intervention; Mean=63.52, SD=57.61). Due to the small number of studies, however, we were not able to conduct further interactions between moderators.

Analysis of Moderator Effects

To examine the influence of key study variables on effect sizes, we conducted moderator analyses on nine (four intervention, two participant, one study level and two outcome measures) study characteristics. We were able to conduct all of the above planned moderator analyses using contrasts since each subgroup had more than four studies even after the removal of studies due to missing data.

We used a fixed effects model to examine the amount of variance explained by the moderators. In addition, we used a random effects model to combine studies within subgroups. This mixed model approach allowed us to partition the variance and examine the large heterogeneity found in our sample (Borenstein, Hedges, Higgins, & Rothstein, 2009; Wood & Eagly, 2009).

Context of the intervention. Among the most important characteristics of training was the person who provided the intervention. As shown in Table 2, a sizeable portion of the trainers were the experimenters themselves. The next highest category was classroom teachers, identified as an individual holding a bachelor's degree and state certification. Certified preschool teachers, therefore, were included in this category. Fewer instances of training were provided by parents and still fewer by child care providers, identified as an individual who taught in a community-based organization and did not hold a bachelor's degree or state certification.

Group size, as well, has been often regarded as a key contextual characteristic of training. Previous studies have argued for one-to-one instruction (Wasik & Slavin, 1993), and small group (Elbaum, Vaughn, Hughes, & Moody, 2000) compared to whole group instruction (Powell, Burchinal, File, & Kontos, 2008), particularly for young children. As shown in the Table, over 20% of the studies did not specify group size; the remaining studies included a relatively even number of small group, and large group intervention, with a somewhat larger number of individualized interventions.

Insert Table 2 about here

To examine these effects, we used a mixed effects model to conduct the moderator analysis on these contextual features. Our analysis indicated a significant effect for the adult who carried out the intervention. Training provided by the experimenter (*g*=.96, SE= .13,CI_{95}=.70, 1.22) or the teacher (*g*=.92, SE=.11, CI_{95}=.70, 1.15) appeared equally effective. Although interventions given by the parent were somewhat less effective, the effect size was still substantial (*g*=.76, SE=.18, CI_{95}=.41, 1.10). There were no significant differences among these trainers, Q(2)=.61, p=.44.

On the other hand, trainings given by child care providers were significantly less successful. Interestingly, these interventions were highly and significantly homogeneous, Q_W=1.9, p=.4, I^2=0.00; no between-studies variance was unexplained. Our analysis indicated that trainings given by child care providers yielded smaller effect sizes that were significantly lower than all others, *g*=.13, SE=.095, CI_{95}=-.06, .31; Q(3)=41.26, p<.0001. It should be noted, however, that there were substantially fewer studies in this group than others.

In addition to variations in the person who provided the treatment, the studies varied in terms of the number of participants that made up the intervention group. Whether children were taught in individualized settings (*g*= .99, SE=.14, CI_{95}=.72, 1.26), small groups of five children or less (*g*=.88, SE=.13, CI_{95}=.64, 1.12) or large groups of six children or more (*g*=1.04, SE=.21, CI_{95}=.64, 1.44) did not alter the effect size in either magnitude or significance, Q(2)=.56, p=.75. Rather, all group configurations appeared to benefit substantially from the vocabulary interventions.

Dosage of Instruction

Intensity of instruction or ‘dosage’ refers to the amount of training that is delivered to participants. However, the concept goes beyond answering the question of “how much” is provided. It involves duration (i.e. how long the intervention lasted from start to finish); frequency (i.e. how many sessions were delivered); and intensity (i.e. the amount of time within each session). For example, if an intervention was given for 20 minutes, 3 times a week for 5 weeks, the duration would be 35 days, frequency would be 15, and intensity would be 20.

As shown in Table 3, dosage of instruction and the characteristics within it varied dramatically across studies. Therefore, each aspect of dosage was examined to measure how these characteristics might influence word learning.

Insert Table 3 about here

Duration. The duration of intervention ranged broadly from 1 to 270 days, with a median of 42 days of instruction. Shown in Table 3, interventions lasting a week or less (g=1.35, SE=.18, CI_{95}=.99, 1.70), surprisingly resulted in significantly higher effect sizes than those lasting longer than a week, up to 270 days (g=.85, SE=.07, CI_{95}=.72, 1.02), $Q(1)=6.28$, $p < .05$. These results, of course, must be interpreted with caution for several reasons. First, studies with short term goals (e.g. specific words related to a storybook) may indicate that even a week of training can be highly effective in increasing young children's word learning. Second, there were only 7 studies that lasted a week or less in the sample. That said, however, we then conducted an analysis to examine whether the median of duration of instruction—42 days—would moderate the effect size. Our analysis indicated interventions lasting 42 days or less (g=.97, SE=.11, CI_{95}=.75, 1.20) had no less effect than those lasting more than 42 days (g=.87, SE=.10, CI_{95}=.67, 1.10), $Q(1)= .44$, p=.51. Taken together, these results suggest that interventions of brief duration can be associated with positive word learning outcomes.

Frequency. The number of interventions sessions within studies ranged from 1 to 180 sessions, with a median of 18 sessions. However, there was a substantial amount of studies in which the interventions included either a week or fewer sessions (k=12, 52 effect sizes). To examine whether these studies with fewer sessions differed from those with more, we conducted a moderator analysis. As shown in Table 4, we found that studies with less than a week of instruction had significantly higher effect sizes (g=1.42, SE=.22, CI_{95}=.98, 1.85) than those with more (g=.83, SE=.08, CI_{95}=.67, .99), $Q(1)= 6.06$, $p < .05$. The approximate equal number of studies and effect sizes within these two categories provided more confidence in this moderator analysis. As in the case of the duration analyses, these differences might reflect the goals of the intervention: More targeted goals and assessments would likely call for fewer training sessions than those with more global and broad objectives. To follow up, we once again split our sample by the median. Our analysis indicated that studies with fewer than 18 sessions had significantly higher effect sizes (*g*=1.13, SE=.13, CI_{95}=.87, 1.39 than those with 18 sessions or more (*g*=.80, SE=.11, CI_{95}=.58, 1.01), Q (1)=3.78, $p < .05$. Consequently, it suggests that studies with a smaller numbers of sessions could effectively improve children's word learning outcomes.

Intensity. We calculated the length of each individual training session as a final component of dosage. Length in our studies lasted from 7 to 60 minutes, with a median of 20 minutes. In cases where studies reported a range of time for each training session, we calculated an 'average' time. We then examined whether the length of the intervention sessions moderated effect sizes. Our analysis indicated no significant differences between effect sizes for the length of training. Sessions lasting less than 20 minutes (*g*=.97, SE=.12, CI_{95}=.74, 1.20) were not significantly different than those lasting more than 20 minutes (*g*=.91, SE=.15, CI_{95}=.62, 1.20), Q (1)= .11, $p < .05$. Given that these interventions were geared to young children, it is not surprising that longer sessions did not significantly affect word learning gains. In fact, shorter sessions were somewhat more effective than longer sessions.

Taken together, the analysis of dosage suggests that even smaller amounts of treatment can be associated with vocabulary gains. We can hypothesize that one explanation might hinge on the goals and scope of the intervention. Vocabulary training targeted to discrete set of skills (e.g. such as dialogic reading) may involve shorter-term intervention activities than those that are designed to enhance more global skills. This is an important area for further research.

Type of Instruction

Our sample included a large variety of instructional methods and independent treatments. Although we coded for the type of intervention the authors reported (see Table 1), we decided to follow Elleman and colleagues' meta-analytic strategy (2009), and focus on several key characteristics of the interventions in our meta-analysis. This decision was made because many of the interventions used different components within their treatments. Storybook reading, repeated readings, and dialogic reading, while identified under the rubric of “storybook reading intervention” were fairly different in their strategies for teaching vocabulary. Further, similar treatments often used different terms. For example, interactive storybook reading, and shared reading, although different in terms, shared many components of instruction. Especially important for vocabulary instruction, we decided to focus on the approach that was used in the vocabulary intervention: whether or not words were explicitly taught, implicitly taught through embedded activity, or whether both strategies were used to teach new words. The pedagogical approach represented a key component of instruction used by the National Reading Panel in their report of vocabulary training.

This variable was fairly easy to code. Explicit instruction emphasizes strategies for directly teaching vocabulary. An intervention was coded as explicit if detailed definitions and examples were given before, during or after a storybook reading with a follow-up discussion designed to review these words. Implicit instruction, on the other hand, involved teaching words within the context of an activity. An intervention was coded as implicit if words were embedded in an activity, such as storybook reading activity without intentional stopping or deliberate teaching of word meanings. In some cases, interventions used a combination of both strategies. Treatments in which the deliberate instruction of words was followed by implicit uses of the words in contexts were coded as combined instruction.

This distinction was useful because it allowed us to examine the pedagogical strategy within similarly-identified interventions. For example, one study examined implicit word learning through dialogic reading, while another intervention used direct instruction of words prior to dialogic reading. Similarly, in some cases, researchers would use implicit and explicit learning through interactive reading alouds, while others relied exclusively on implicit strategies through reading aloud.

Insert Table 4 about here

We tested what was more effective: to have explicit instruction, implicit instruction, or a combination of both methods. We found a significant effect for the type of instruction, $Q(2)=17.67$, $p<.0001$. Children made significantly higher gains with interventions that used an explicit method (*g*=1.01, SE=.13, CI_{95}=.81, 1.30) or a combination of explicit and implicit methods (*g*=1.21, SE=.13, CI_{95}=.94, 1.47) than those that employed an implicit method only (*g*=.62, SE=.084, CI_{95}=.46,.78), Q(2)=17.67, p<.0001. As shown in Table 4, interventions that used a combination of explicit and implicit methods resulted in a slightly higher magnitude of effect than explicit alone; however, there was no significant difference between the effects of these two treatment methods. These results indicate that interventions which provided explicit and implicit instruction through multiple exposures of words in rich contexts were most effective in supporting word learning for pre-K and kindergarten children.

Instruction for At-risk Children

Evidence of the substantial differences in vocabulary between at-risk and average children, and its concomitant effects on achievement has driven much of the research on vocabulary development (Hart & Risley, 1995). Conceivably, if interventions are designed to narrow the gap, they must not only improve vocabulary for at-risk children, but accelerate its development. This would seem to suggest that vocabulary interventions specially targeted for at-risk children must have a stronger and more powerful effects than those for average and above average learners.

Insert Table 5 about here

We conducted several moderator analyses to examine this question as can be seen in Table 5. First, we conducted a moderator analysis using coder- determined qualifications for at-risk participants. We compared studies with participants we considered to be at-risk in which at least 50% of the participant sample was within one risk category: low SES level (at or below the national poverty level of $22,000, parental education of high school graduation or below, qualification for free and reduced lunch), second language status, low academic achievement (as identified by a teacher report, standardized school assessment or adequate yearly progress (AYP) and/or special needs (as identified as having an IEP or Title 1 placement) to those that were not at risk. Our analysis indicated that there was no difference between gains on vocabulary measures for at-risk children (g=.85, SE=.081, CI_{95}=.69, 1.01) and all other children (g=.91, SE=.10, CI_{95}=.69, 1.12), Q(1)=.18, p=.67. Studies reportedly targeted to at-risk populations were no more effective than those designed for average and above average achievers.

In addition, we conducted a moderator analysis on SES status within our entire sample that included at-risk children (k=40), children not at-risk (k=18) and those whose at-risk status could not be determined (k=9). Though there was a magnitude difference favoring the middle to high SES children, no significant differences were found between vocabulary gains obtained by low SES children (g=.75, SE=.11, CI_{95}=.54, .96) and those by middle to high SES children (g=.99, SE=.11, CI_{95}=.79, 1.21), Q(1)=2.71, p=.10. As low SES children are likely to have lower baseline scores, even parallel gains would not substantially close the gap.

Next, in order to examine the at-risk population further, we conducted a moderator analysis comparing children who qualified as low-SES in addition to another risk factor as described above, to those children who were coded as at-risk but did not qualify as low-income. Within this category of at-risk, children with low SES status (g=.77, SE=.12 CI_{95}=.53, 1.01) received gains that were significantly lower than middle to high SES at-risk children (g=1.35, SE=.26, CI_{95}=.85, 1.85), Q(1)=4.19, $p<.05$ (see Table 5). When SES status was combined with at least one other risk factor, the difference in effect sizes was significant. Further, middle to high SES children who had an additional risk factor(s) gained more than low SES children with an additional risk factor (s). These results suggest that poverty was the most serious risk factor; additional risk factors appeared to compound the disadvantage. Vocabulary interventions, therefore, did not close the gap; in fact, given the differences in the effect sizes, they could potentially exacerbate already existing differentials.

Research Design

In a synthesis of over 300 social science intervention meta-analyses, Wilson and Lipsey (2001) reported that research methods accounted for almost as much variance as characteristics of the actual interventions. Associated with the largest proportion of variance was the type of research design, particularly random vs. non-random assignment. To examine this issue, we coded studies according to this distinction and found that 11 (16%) had true experimental designs while 56 (84%) employed quasi-experimental designs. Our analysis indicated, however, that although there was a difference in magnitude favoring the experimental studies (*g*=1.92, SE=.22, CI_{95}=.49, 1.35), these studies were not significantly different in the size of their effects than quasi-experimental studies (*g*=.88, SE=.07, CI_{95}=.74, 1.00), Q(1)=.04, p=.84.

In addition, in accordance with prior meta-analyses (Elleman et al, 2009; Mol et al, 2009) and best practices in meta-analysis (Cooper & Hedges, 1994; Lipsey & Wilson, 2001), we examined whether the type of control or comparison group used influenced effect sizes. We attempted to obtain information about the control group for all studies, and when insufficient data were provided, we contacted authors for the necessary data.

Our sample encompassed four types of comparison groups as shown in Table 6: a control group that received no treatment which included wait-list designs, a comparison group that received "business as usual" (e.g., same practices as usual), an alternate treatment (e.g., received a deliberately diluted version of the treatment with the hypothesized key ingredient missing), and a within-subject design where subjects acted as their own control groups. Since we conducted eight moderator analyses for the various control group types, we used a Bonferroni-corrected significance criterion of .008 (6 contrast comparisons were made). Within this criterion, only one moderator analyses was significant. Studies in which the controls used alternate treatment comparisons (*g*=1.03, SE=.13, CI_{95}=.78, 1.27) were significantly higher in effect sizes compared to studies in which the controls received no treatment at all (*g*=.51, SE=.13, CI_{95}=.25, .76), Q(1)=8.25, p<.005. However, these results should be interpreted with caution due to the differences in the number of studies in each group (e.g. control groups who received nothing k=7; alternate treatments k=22). Within-subject design studies also reported a significantly higher effect size (*g*=1.09, SE=17, CI_{95}=.75, 1.44) than the no treatment control group studies (g=.51, SE=.13, CI_{95}=.25, .76), Q(1)=7.25, p=.007. Although measures were taken to obtain equivalence for repeated measure designs (the standardized difference is computed, taking into account the correlation between measures) (Dunlap, Cortina, Vaslow, & Burke, 1996), these studies should be interpreted with caution when compared with others using more traditional experimental designs. However, these findings replicate Mol and colleagues (2009) who also found that no-treatment control groups had significantly lower effect sizes.

Together these results indicated that the particular design—whether or not it was an experimental or quasi-experimental did not influence effect sizes. However, it should be noted that sample sizes, even in randomized studies, were small (e.g. an average of 28 subjects). Interventions which compared treatment to alternative treatments appeared to show greater improvements over subjects receiving no treatment at all.

Insert Table 6 about here

The Measurement of Word Learning

In their narrative analysis, the National Reading Panel report (2000) raised important questions about the measurement of vocabulary development that have since been voiced by other researchers. Specific to vocabulary development, the panel recommended the development of more sensitive measures that could be used to determine whether an intervention might be effective. Ideally, they suggested, experimenters should use both author-created and standardized measures in order to best examine vocabulary gains. As a result of their recommendations, a number of researchers (e.g., Leung & Pikulski, 1990; Lonigan, et al., 1999) have included both author-created and standardized assessments in their studies. Others (e.g., Coyne, McCoach, & Kapp, 2007; Sénéchal & Cornell, 1993), however, have moved to relying on author-created measures to attain enough sensitivity to detect fine-grain and more comprehensive vocabulary growth.

We coded measures according to the type of measurement used to examine changes in word learning. Author-created measures examined gains in the vocabulary taught in the curriculum, and were likely to be more sensitive to the effects of intervention. Standardized measures, on the other hand, were more likely to measure growth in more global language development. They were unlikely to contain target words taught in the intervention. Some studies used both types and were coded as a combined set of measures.

We systematically examined the type of measurement through a moderator analysis with approximately equal numbers of effect sizes obtained for each type of test

(See Table 7). Our analysis revealed that effect sizes (e.g. vocabulary gains) on the standardized assessments were significantly lower (*g*=.71, SE=.072, CI_{95}=.57, .85) than those on author-created measures (*g*=1.21, SE=.18, CI_{95}=.86, 1.60), Q(1)=6.35, $p<.01$. These results provide support of the National Reading Panel's recommendation for using multiple measures to examine word growth. Taken together, these moderator analyses revealed that author-created measures appeared to be more proximal indicators of vocabulary improvement, and more targeted to what was to be learned in the interventions. Global measures were less sensitive to gains in vocabulary interventions. These results, however, could be affected by study designs, the specific goals of the vocabulary intervention, and other factors such as the features of the words in the intervention, which unfortunately could not be detected in this moderator analysis.

Insert Table 7 about here

Finally, we examined whether or not there were differences in receptive and expressive vocabulary as a result of the vocabulary interventions. The majority of the vocabulary intervention studies in our sample tested participants using more than one type of measure. Consequently, if we had used the same analysis method applied to our other moderators (using one mean effect size per study to avoid effect size dependency), we would have pooled our variable of interest. Rather than use a mixed model analysis, instead we compared the overall Hedge's *g* effect sizes for these categories and examined the 95% confidence intervals for overlap to determine significant differences. Also, we coded some measures (N=33) as combined. These were comprehensive measures that tested both receptive and expressive vocabulary such as the Preschool Language Assessment Instrument (PLAI).

Our results indicated that children made significantly higher gains on combined receptive and expressive vocabulary measures (*g*=1.11, SE=.14, CI_{95}=.84, 1.39) than on expressive vocabulary alone (*g*=.69, SE=.04, CI_{95}=.60, .77), Q(1)=6.81, $p<.01$, as can be seen by the non-overlapping confidence intervals. Given that the combined measures often included an author-created test among its dependent variables, these differences might reflect the recommendations of the NRP report, an issue we intend to pursue in the future. There were no differences between gains for receptive and expressive measures (g=.80, SE=.06, CI_{95}=0.68, 0.91).

Discussion

This meta-analysis examined the effects of vocabulary interventions on the growth and development of children's receptive and expressive language development. Results indicated that children's oral language development benefited strongly from these interventions. The overall effect size was .88, demonstrating on average a gain of nearly one standard deviation on vocabulary measures. If anything, this effect size may be somewhat conservative, given that a portion of the studies in the analysis were not published. Consequently, we conclude with a fair degree of certainty that vocabulary instruction does appear to have a significant impact on language development.

These results support Stahl and Fairbanks meta-analysis of vocabulary instruction (1986), which reported an average effect size of .97. By contrast, recent meta-analyses by Elleman and her colleagues (2009), and Mol and her colleagues (2009) found far smaller effects for global measures. In the case of Elleman's meta-analysis, differences clearly related to the selection of studies and its focus on passage-related comprehension. Their analysis included only one study in the pre-K age range, and emphasized print-related interventions. In this respect, it was not surprising to find differences in our results.

In contrast, meta-analyses by Mol and her colleagues (2009) focused on similar objectives and similar age ranges as in our meta-analysis, and therefore warrant further explanation. They reported effect sizes ranging from d =54-.57, resulting from interaction before, during and after shared reading sessions on oral language development. In a separate meta-analysis (2008), they found an average effect size of d=.59 for parent-child dialogic reading. In both cases, more moderate effects were reported than our overall mean size.

This divergence in findings might be due to differences in the selection criteria or the methods used to evaluate effects. For example, we excluded studies in a foreign language, and excluded studies that did not measure real word learning (i.e. pseudo-words). Our meta-analysis also included studies with multiple methods of vocabulary intervention. Many of the interventions, for example, used storybooks within more comprehensive programs. Inclusion of these additional elements to the traditional book reading interventions, therefore, might have accounted for the more potent effect sizes in our meta-analysis. These more powerful interventions were likely to be implemented by experimenters or teachers (not parents or child care providers). For example, in programs with effect sizes equal to or greater than 1.0 (N=79), 43% of the trainers were experimenters and 42% teachers; for effect sizes equal to or greater than 2.0 (N=24) 58% were experimenters and 33% teachers. We also included vocabulary interventions that were not related to storybook reading. Computer-based interventions, video-related, and technology-enhanced interventions were considered within our analysis. In this respect, our goal was to identify all possible vocabulary interventions, representing the corpus of experimental techniques targeted to enhancing children's oral language development to examine the average size of their effects.

Our strategy, therefore, was to better understand the potential overall effects of intervening in the early years to improve word learning. But the expansiveness in our inclusion strategy also came at a cost. By including many different instructional techniques, our meta-analysis could not specify the particular intervention that was most effective. Contrary to the wishes of many policymakers, we could find no specific intervention that worked more powerfully than others. To further complicate these wishes, there were dramatic variations across similarly-termed interventions, such as storybook reading. It was for this reason, some 10 years, the National Reading Panel report concluded that a meta-analysis could not be conducted in vocabulary because studies were too varied and too few for each of the different types of instruction to be examined.

In their recent report, the National Early Literacy Panel (2008) took an alternate approach in their meta-analysis than our strategy. Their goal was to isolate shared book reading to examine its effects. Contrary to Mol and her colleagues, however, the committee restricted its selection to published book reading studies only, and those not potentially confounded with any additional enhancements. Their inclusion criteria yielded a total of 19 studies (5 from a single intervention type and a common author). Nevertheless, even under these highly restrictive criteria, the authors acknowledged wide variations in procedures. Further, because of the limited number of studies they could not identify the impact of age, risk status or agent of intervention, arguing that at best, it appeared that “some kind of intensified effort to read” compared to a somewhat “less intensified effort” (p.154) might have a moderate effect on oral language skills (.73).

In contrast, we recognized the variation within similar types of vocabulary interventions and used a broader inclusion strategy, including both published and unpublished studies to examine oral language development for children in their early years. Our approach indicated heterogeneity of variances, yet at the same time, it provided us with the additional statistical power to conduct moderator analyses to detect potential differences among them (Lipsey & Wilson, 2001). Further, it allowed us to look beyond the most conventional interventions to examine vocabulary trainings that have not been previously reviewed or synthesized.

Based on the moderator analyses, in particular, we can make some propositions for interventions that appeared to work best. First, the person conducting the intervention appeared to matter. Largest effect sizes occurred when the experimenter conducted the treatment; the most negligible, when the intervention was given by the child care provider. Similar to the results from Mol and her colleagues (2009), providers appeared to be less effective in implementing vocabulary training with young children. These finding appeared not to be related to the age of the child, given the average effect sizes for young children and kindergarten were relatively similar. Rather, one could hypothesize that providers were not sufficiently trained to incorporate and internalize the strategies to implement the training materials with the intention and fidelity of the program developers. It could be that vocabulary interventions, in particular, may require additional professional development for providers. Previous studies of professional development (Author, in press) have found that, in contrast to areas such as phonological awareness or letter name knowledge, early childhood providers may need longer doses of training and support to improve vocabulary development.

Second, we did not find evidence for group size as an important contextual feature in vocabulary trainings. Traditionally, whole group instruction has been considered to limit children's use of language. Powell and his colleagues (2008), in their study of the ecology of early learning settings found support for longstanding concerns about whole group instruction. Whole group instruction appeared to support passive modes of child engagement: listening, and watching rather than talking and acting. At the same time, whole group instruction can provide an introduction to the shared understanding of the words and their meanings that may take place in the classroom. It could be hypothesized that certain activities in large group settings, like vocabulary training, are appropriate with smaller groups and individualized one-on-one instruction reserved for reviewing and practicing the words that have been introduced. Our results are further buttressed with findings from Mol and her colleagues' meta-analysis (2009), which demonstrated that children's oral language skills improved in whole group instruction. These results could make a case for a more differentiated model of organizing instruction, one that more clearly aligns the learning goals with the most promising organizational features.

Third, we did not find support for Ramey and Ramey's conclusion (Ramey & Ramey, 2006) that higher dosages of treatment lead to better effects. Longer, more intensive, and more frequent interventions did not yield larger effect sizes than treatment with smaller dosages. In fact, even brief doses of vocabulary intervention (e.g. Whitehurst et al.,1998) were associated with large effect sizes. For example, Wasik and Bond (2001) employed a four week intervention, not necessarily implemented in consecutive weeks on interactive book reading. The training resulted in significant effects on children's vocabulary and expressive language development. Yet at the same time, we must be cautious in interpreting these results, especially in relation to vocabulary training. Given that many of the interventions employed author-created measures, it would be impossible to disentangle whether or not these results might have been due to interventions that were tied to a more discrete set of skills than others with a broader focus. Halle and her colleagues (Halle, in press) have argued that interventions that are narrower in scope may only require short-term training; those with a broad focus may require a larger dosage. Clearly, our current finding that "more is not necessarily always better" raises important questions for further research.

Fourth, by analyzing the instructional approach to vocabulary training rather than the type of program (e.g. storybook reading), we were able to detect differences in the effectiveness of vocabulary instruction. Programs that used explicit instruction deliberately, either through explanation of words or key examples, produced larger effect sizes than those that taught words implicitly. In addition, programs that combined explicit and implicit instruction, enabling student to be introduced to words followed by meaningful practice and review demonstrated even larger effects. Implicit instruction, alone, was far less effective. Similar results have been reported in the training of phonological awareness (Lonigan, 2006) and word recognition skills in the early years (Foorman, Fletcher, Francis, Schatschneider, & Mehta, 1998), and follow on the recommendations of the National Reading Panel (2000) which called for providing direct instruction in vocabulary with multiple exposures in rich contexts. Given that the meta-analysis focused on many different training programs, these results should generalize across specific types of programs. Further, there is evidence to suggest the benefits of explicit instruction may not be limited to word learning (Klahr & Nigam, 2004; Rittle-Johnson, 2006; Star & Rittle-Johnson, 2008.)

Fifth, there are degrees of at-risk populations. Examining the effects of author-identified at-risk status versus not at-risk, we reported no differences in the effect sizes of vocabulary interventions. Vocabulary training interventions appeared to be equally effective for all children in these studies. However, when we used poverty as an additional risk factor, we found significant differences in the effect sizes between groups: Middle and upper-income at-risk children were significantly more likely to benefit from vocabulary intervention than those students, also at risk and poor. These results indicate that although they might improve oral language skills, vocabulary interventions are not sufficiently powerful to close the gap—even in the preschool and kindergarten years. These results also highlight the devastating effects that poverty may have on children's language development (Author, 2008). It may be that only high doses of treatment such as the Abecedarian project (Campbell & Ramey, 1995)—an intensive intervention starting at birth through age 6, may begin to ameliorate these differences.

And sixth, it was clear that the type of measurement makes a difference in determining the effects of an intervention. Like Elleman and her colleagues (2009), we found support that author-created measures were more sensitive in detecting improvements in language development than standardized measures. These author-created measures could reflect more proximal learning outcomes than the more distal measures of language development. Since author-created assessments may be more closely tied to the vocabulary training in the intervention, these measures may answer a basic question: did children learn what was taught?

That author-created tests, targeting the content and specifics of their intervention programs, show gains is not surprising, nor necessarily newsworthy. Without standardized measures for confirmatory evidence, these author-created measures may provide an inflated portrait of the vocabulary gains made in studies. Relying solely on curriculum-based assessments may overestimate the ability of children to transfer their new vocabulary to other contexts. Consequently, it may be best to interpret vocabulary learning effect sizes in tandem with standardized measures, being cognizant of the differences between the types of tests and their ability to report learning growth. Ideally, we would recommend multiple measures—author created and standardized, to provide strong evidence of the malleability of vocabulary use in different settings. Further, our meta-analyses suggests that even though the standardized measures may not be as sensitive or able to measure nuanced and specific vocabulary acquisition, and are associated with significantly lower growth than author-created measures, children were still able to make moderate gains (g=.71). This could reflect the most conservative end of the spectrum of vocabulary acquisition for young children, with the growth on author-created measures (g=1.30, SE= .19) reflecting the other end of the spectrum in which young children are able to make very large vocabulary gains. These moderator analyses however, should not be interpreted as testing causal relationships (Cooper, 1998; Viechbauer, 2007). Rather, these results should be verified through experimental manipulations that vary these factors systematically.

Limitations

The findings of this study must be considered within the limitations of meta-analysis. A meta-analysis can only generalize from the characteristics of existing studies. Many of the studies we included lacked details in their descriptions of the interventions, the specific materials used, the amount of professional development training provided, and the fidelity of implementation. Specific to the vocabulary intervention, many studies did not include details on the difficulty level of words, how many words were actually taught, the rationale for the selection of words, and whether or not it supplemented an existing curriculum. Further, very few details were provided about the control conditions and their exposure to the vocabulary in the intervention. Given the unconstrained nature of vocabulary development (Paris, 2005), these details are particularly important if we are to understand the extent to which vocabulary interventions may improve language development over time.

In addition, there were a number of potential confounds within our moderator analyses that should be the subject of future research. For example, we suspect that the type of measure (author-created, or standardized) may be confounded with the goals of the instruction (e.g. number of words taught) along with the number of sessions. It seems quite plausible, for instance, that studies with fewer sessions and more narrow goals might use more author-created measures than standardized. Further, the word selection in vocabulary training may influence the type of instruction (e.g. explicit and implicit). For example, there are words that can be easily be taught explicitly (e.g. camouflage; habitat) but there are also numerous words that are hard to teach explicitly (e.g. before; after) without contextualization. Therefore, word features may represent a potential confound with the type of instruction. Unfortunately, too few studies identified the words in the vocabulary trainings to allow us to examine these issues thoroughly.

Finally, it could be argued that we could have split storybook reading by the type of instruction (e.g. explicit/ implicit or combined) to further disentangle the type of instruction that is most effective. Unfortunately, we believe that this would have only introduced additional confounds. Storybook reading, from our perspective, did not represent a single clearly-defined intervention. For example, some storybook reading interventions included extended and purposeful dialogue; others included extension activities; others used play objects to encourage retellings. Further, most of the storybook reading interventions did not detail the names, or genres of the books, another potential confound.

We would have wished that more studies conducted delayed posttests to examine the sustainability of treatment. Our sample included only 11 studies will delayed posttests. While our results were encouraging, we need additional experimental studies to examine the longer-term impact of word learning interventions.

Statistically our sample remained largely heterogeneous, which is not uncommon to meta-analysis. In a review of 125 medical meta-analyses, over 50% were found to be largely heterogeneous (Engel, Schmid, Terrin, Olkin, & Lau, 2000). Due to our heterogeneous sample, however, we were unable to identify a set of homogeneous practices that systematically lead to higher gains in vocabulary.

Finally, our use of the random effects model fit our heterogeneous distribution of effect sizes but did not fully explain the variance in effect sizes. It is possible that there were systematic ways in which our studied differed that we did not address in our moderator analyses. For example, it is possible that within explicit instruction other factors such as expressive or receptive vocabulary may have helped to reduce the heterogeneity and explain the variance in effect sizes. We intend to pursue these potential relationships in future analyses.

Implications for Practice and Future Research

Results from this meta-analysis of the impact of vocabulary interventions on the word learning skills of young children indicate positive effects. These effects were robust across variations in the type of intervention for children in prekindergarten and kindergarten. These results highlight the importance of teaching vocabulary in the early years.

Still, there is much work to be done. Although this meta-analyses detailed a number of instructional features that seem to support stronger effects, further work is needed to help design more effective vocabulary interventions. It did not yield recommendations for how to promote quality instruction in vocabulary. For example, we still need better information on what words should be taught, how many should be taught, and what pedagogical strategies are most useful for creating conceptually sound and meaningful instruction. Further, although author-created measures appeared to demonstrate more powerful effects, evidence is missing on the quality of these measures, and their reliability and validity. If we believe that author-created measures are more sensitive and able to detect growth in vocabulary, we need better assurances that they are, in fact, predictive of greater proficiency in oral language. Given that vocabulary is a strong predictor of comprehension of text, and later achievement (Cunningham & Stanovich, 1997), until such evidence, researchers should consider multiple measures, author-created and standardized, to examine achievement.

The good news about the overall positive effects of vocabulary instruction must be tempered by the not-so-good news that children who are at-risk and poor are not faring as well as we would hope. Vocabulary interventions did not close the gap; in fact, given that middle- and upper-middle class children identified at risk are gaining substantially more than their counterparts, there is the possibility that such interventions might exacerbate vocabulary differentials. Therefore, it is imperative that we continue to work toward developing more powerful interventions to enhance their skills. Researchers will need to better understand the environmental and participant factors that place these children at risk to more fully develop interventions that are better targeted to their needs and can potentially accelerate their language development.

References

*Indicates meta-analyzed studies

*Ammon, P. R., & Ammon, M. S. (1971). Effects of training black preschool children in vocabulary versus sentence construction. *Journal of Educational Psychology, 62*(5), 421-26.

Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator, Winter*, 14-18, 44-46.

*Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology, 86*(2), 235-243.

Author (Ed.). (2008).

Author (2009).

Author, in press.

Bakermans-Kranenburg, M., van Ijzendoorn, M., & Juffer, F. (2003). Less is more: meta-

analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*, 195-215.

*Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal, 107*(3), 251-273.

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.

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

*Bonds, L. T. G. (1987). A comparative study of the efficacy of two approaches to introducing reading skills to kindergarteners. (Unpublished doctoral dissertation).

University of South Carolina, South Carolina.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (Version 2)*. Englewood, New Jersey: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons, Ltd.

*Bortnem, G. M. (2005). The effects of using non-fiction interactive read-alouds on expressive and receptive vocabulary of preschool children. (Unpublished doctoral dissertation). University of South Dakota, South Dakota.

*Brackenbury, T. P. (2001). Quick incidental learning of manner-of-motion verbs in young children: Acquisition and representation. (Unpublished doctoral dissertation). University of Kansas, Kansas.

*Brickman, S. O. (2002). Effects of a joint book reading strategy on Even Start. (Unpublished doctoral dissertation). The University of Oklahoma, Oklahoma.

*Brooks, R. T. (2006). Becoming acquainted with the faces of words: Fostering vocabulary development in kindergarten students through storybook readings. (Unpublished doctoral dissertation). Auburn University, Alabama.

Campbell, F., & Ramey, C. (1995). Cognitive and school outcomes for high-risk African-American study at middle adolescence: Positive effects for early intervention. *American Educational Research Journal, 32*, 743-772.

*Christ, T. (2007). Oral language exposure and incidental vocabulary acquisition: An exploration across kindergarten classrooms. (Unpublished doctoral dissertation). State University of New York at Buffalo, New York.

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

Cooper, H., & Hedges, L. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Cooper, H., Hedges, L., & Valentine, J. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.

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

*Coyne, M. D., McCoach, D. B., Loftus, S., Jr., R. Z., & Kapp, S. (in press). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal.*

*Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145-162.

*Coyne, M. D., Zipoli, R., & McCoach, D. (2008). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically-related and embedded word review. Paper presented at the American Educational Research Association Conference, New York, NY.

*Crevecoeur, Y. C. (2008). Investigating the effects of a kindergarten vocabulary intervention on the word learning of English-language learners. (Unpublished doctoral dissertation). University of Connecticut, Connecticut.

*Cronan, T. A., Sonia G. Cruz, Arriaga, R. I., & Sarkin, A. J. (1996). The effects of a community-based literacy program on young children's language and conceptual development. *American Journal of Community Psychology, 24*(2).

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

*Danger, S. E. (2003). Child-centered group play therapy with children with speech difficulties. (Unpublished doctoral dissertation). University of North Texas, Texas.

*Daniels, M. (1994). The effect of sign language on hearing children's language development. *Communication Education, 43*(4), 291-298.

*Daniels, M. (1994). Words more powerful than sound. *Sign Language Studies, 83*, 156-166.

*Daniels, M. (2004). Happy hands: The effect of ASL on hearing children's literacy. *Reading Research and Instruction, 44*(1), 86-100.

Dunlap, W., Cortina, J., Vaslow, J., & Burke, M. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170-177.

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method for accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89- 98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455-463.

Elbaum, V., 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.

*Eller, R. G., Pappas, C. C., & Brown, E. (1988). The lexical development of kindergartners: Learning from written context. *Journal of Reading Behavior, 20*(1), 5-24.

Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Educational Effectiveness, 2*, 1-44.

Engel, E., Schmid, C., Terrin, N., Olkin, I., & Lau, J. (2000). Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. *Statistics in Medicine, 19*, 1707-1728.

*Ewers, C. A., & Brownson, S. M. (1999). Kindergartners' vocabulary acquisition as a function of active vs. passive storybook reading, prior vocabulary, and working memory. *Reading Psychology, 20*(1), 11-20.

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

Foorman, B., Fletcher, J., Francis, D., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37-55.

*Freeman, L. (2008). A comparison of the effects of two different levels of implementation of read-alouds on kindergarten students' comprehension and vocabulary acquisition. (Unpublished doctoral dissertation). The University of Alabama, Alabama.

Halle, T. (in press). Dosage of professional development: What have we learned. In S. B. Neuman (Ed.), *Professional Development for Early Childhood Educators: Principles and Strategies for Improving Practice*. Baltimore, MD: Brookes.

*Hargrave, A. C., & Senechal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75-90.

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

*Harvey, M. F. (2002). The impact of children's audiobooks on preschoolers' expressive and receptive vocabulary acquisition. (Unpublished doctoral dissertation). Walden University, Minnesota.

*Hasson, E. A. (1981). The use of aural cloze as an instructional technique for the enhancement of vocabulary and listening comprehension of kindergarten children. (Unpublished doctoral dissertation). Temple University, Pennsylvania.

Higgins, J.P.T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. International Journal of Epidemiology, 37(5), 1158-1160.

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

*Huebner, C. E. (2000). Promoting toddlers' language development through community-based intervention. *Journal of Applied Developmental Psychology, 21*(5), 513-535.

Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. (2006). Assessing Heterogeneity in Meta-Analysis: Q Statistic or I^2 Index? *Psychological Methods, 11* (2), 193–206.

*Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *American Speech-Language-Hearing Association,36*(1), 17-32.

*Karweit, N. (1989). *The effects of a story reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students. Report No. 39*. Retrieved from ERIC database. (ED313655)

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: The effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661-667.

*Lamb, H. A. (1986). The effects of a read-aloud program with language interaction (early childhood, preschool children's literature). (Unpublished doctoral dissertation). The Florida State University, Florida.

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

*Leung, C. B. (2008). Preschoolers' acquisition of scientific vocabulary through repeated read-aloud events, retellings, and hands-on science activities. *Reading Psychology, 29*(2), 165-193.

*Leung, C. B., & Pikulski, J. J. (1990). Incidental learning of word meanings by kindergarten and first-grade children through repeated read aloud events. *National*

Reading Conference Yearbook, 39, 231-239.

*Lever, R. (2008). *Discussing stories: Using a dialogic reading intervention to improve kindergartner's oral narrative construction.* (Unpublished master's thesis). Carleton University, Canada.

*Levinson, J. L., & Lalor, I. (1989). *Computer-assisted writing/reading instruction of young children: A 2-year evaluation of "writing to read."* Retrieved from ERIC database. (ED310390)

*Light, J., Drager, K., McCarthy, J., Mellott, S., Millar, D., Parrish, C., et al. (2004). Performance of typically developing four- and five-year-old children with AAC systems using different language organization techniques. *AAC: Augmentative and alternative communication, 20*(2), 63-88.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.

Lipsey, M. W. & Wilson, D. B. (2001). The way in which intervention studies have "personality"

and why it is important to meta-analysis. *Evaluation & the Health Professions,* 24(3),

236-254.

*Loftus, S. M. (2008). Effects of a Tier 2 vocabulary intervention on the word knowledge of kindergarten students at-risk for language and literary difficulties. (Unpublished doctoral dissertation). University of Connecticut, Connecticut.

Lonigan, C. (2006). Conceptualizing phonological processing skills in pre-readers. In D.Dickinson & S. B. Neuman (Eds.), *Handbook of Early Literacy Research: Volume II* (Vol. II, pp. 77-89). New York: Guilford.

*Lonigan, C. J., Anthony, J. L., Bloomfield, B. G., Dyer, S. M., & Samwel, C. S. (1999). Effects of two shared-reading interventions on emergent literacy skills of at-risk preschoolers. *Journal of Early Intervention, 22*(4), 306-322.

*Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly, 13*(2), 263-290.

*Lowenthal, B. (1981). Effect of small-group instruction on language-delayed preschoolers. *Exceptional Children, 48*(2), 178-179.

*Lucas, M. B. (2006). The effects of using sign language to improve the receptive vocabulary of hearing ESL kindergarten students. (Unpublished doctoral dissertation). Widener University, Pennsylvania.

*McConnell, B. B. (1982). *Bilingual education: Will the benefits last? Bilingual Education Paper Series,5*(8).

*Meehan, M. L. (1999). *Evaluation of the Monongalia County schools Even Start program child vocabulary outcomes*. Charleston, WV: AEL.

*Mendelsohn, A. L., Mogilner, L. N., Dreyer, B. P., Forman, J. A., Weinstein, S. C., Broderick, M., et al. (2001). The impact of a clinic-based literacy intervention on language development in inner-city preschool children. *Pediatrics, 107*(1), 130-134.

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

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.

Mol, S., Bus, A., deJong, M., & Smeets, D. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education and Development, 19,* 7-26.

*Murphy, M. M. (2007). Enhancing print knowledge, phonological awareness, and oral language skills with at-risk preschool children in Head Start classrooms. (Unpublished doctoral dissertation). The University of Nebraska - Lincoln, Nebraska.

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

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

*Nedler, S., & Sebera, P. (1971). Intervention strategies for Spanish-speaking preschool children. *Child Development, 42*(1), 259-267.

*Neuman, S. B. (1999). Books make a difference: A study of access to literacy. *Reading Research Quarterly, 34*(3), 286-311.

*Neuman, S. B., & Gallagher, P. (1994). Joining together in literacy learning: Teenage mothers and children. *Reading Research Quarterly, 29*(4), 382-401.

*Notari-Syverson, A., & et al. (1996). *Facilitating language and literacy development in preschool children: To each according to their needs*. Retrieved from ERIC database. (ED395692)

Paris, S. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184-202.

*Peta, E. J. (1973). *The Effectiveness of a Total Environment Room on an Early Reading Program for Culturally Different Pre-School Children*. Retrieved from ERIC database. (ED090514)

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*, 108-123.

*Rainey, E. W. (1968). *The Development and Evaluation of a Language Development Program for Culturally Deprived Preschool Children*. Retrieved from ERIC database. (ED038412).

Ramey, S. L., & Ramey, C. (2006). Early educational interventions: Principles of effective and sustained benefits from targeted early education programs. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. II, pp. 445-459). New York, NY: Guilford.

Raudenbush, S. (1994). Random effects models. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-322). New York: Russell Sage Foundation.

Rittle-Johnson, B. (2006). Promoting transfer: The effects of direct instruction and self-explanation. *Child Development. 77*, 1-15.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Scarborough, H. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of Early Literacy Research* (pp. 97-110). New York: Guilford.

*Schetz, K. F. (1994). An examination of software used with enhancement for preschool discourse skill improvement. *Journal of Educational Computing Research, 11*(1), 51-71.

*Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language, 24*(1), 123-138.

Sénéchal, M., & Cornell, E. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, *28*(4), 360-374.

* Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual-differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87*(2), 218-229.

Shadish, W., & Haddock, C. (1994). Combining estimates of effect size. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). New York: Russell Sage Foundation.

*Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *Elementary School Journal, 108*(2), 97-113.

*Silverman, R. D. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *Elementary School Journal, 107*(4), 365-384.

*Simon, K. L. K. (2003). Storybook activities for improving language: Effects on language and literacy outcomes in Head Start preschool classrooms. (Unpublished doctoral dissertation). University of Oregon, Oregon.

Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

*Smith, M. (1993). *Characteristics of picture books that facilitate word learning in preschoolers.* (Unpublished doctoral dissertation). Tate University of New York at Stony Brook, New York.

Stahl, S., & Fairbanks, M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110.

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

Star, J. R. & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction,* 18, 565-579.

*Terrell, S. L., & Daniloff, R. (1996). Children's word learning using three modes of instruction. *Perceptual and Motor Skills, 83*(3), 779-787.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44,* 423-432.

Viechtbauer, W. (2007). Moderator analysis in meta-analysis: A tutorial. *Zeitschrift für Psychologie (Journal of Psychology), 215,* 104-121.

*Walsh, B. A., & Blewitt, P. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal, 33*(4), 273-278.

*Warren, S. F., & Kaiser, A. P. (1986). Generalization of treatment effects by young language-delayed children: A longitudinal analysis. *Journal of Speech & Hearing Disorders, 51*(3), 239-251.

*Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*(2), 243-250.

*Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*. *98*(1), 63-*74*.

Wasik, B., & Slavin, R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28*, 178-200.

*Watson, B. (2008). *Preschool book reading: teacher, child and text contributions to vocabulary growth.* (Unpublished doctoral dissertation). Vanderbilt University, Tennessee.

*Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day-care and home for children from low-income families. *Developmental Psychology, 30*(5), 679-689.

*Whitehurst, G. J., Falco, F. L., Lonigan, C., Fischel, J. E., Valdez-Menchaca, M. C., DeBaryshe, B. D., et al. (1988). Accelerating language development through picture-book reading. *Developmental Psychology, 24*(4), 552-55

Table 1. Study characteristics included in the meta-analysis

Author(s) name	Publication Status	Grade	At risk[a]	Trainer	Condition[b]	Type of Testing[cd]	Control Group[e]	ES (g)
Ammon & Ammon, 1971	publ	Pre-K	No	Experimenter	PPVT training	S	n	1.00
Arnold et al, 1994	publ	Pre-K	No	Parent	DR	S	tu	0.51
Beck & McKeown, 2007a	publ	K	Yes	Teacher	SB	A	tu	1.54
Beck & McKeown, 2007b	publ	K	Yes	Teacher	SB	A	ws	1.96
Bonds, 1987	not publ	K	-	Teacher	Basal	S	at	0.71
Bortnem, 2005	not publ	Pre-K	Yes	Experimenter	SB	S	tu	0.97
Brackenbury, 2001	not publ	Pre-K	No	Experimenter	LIP	A	at	1.63
Brickman, 2002	publ	Pre-K	Yes	Parent	SB	A&S	at	0.31
Brooks, 2006	not publ	K	No	Teacher	SB	A	at	0.48
Christ, 2007	not publ	K	-	Teacher	SB	A	n	0.26

Coyne 2008	not publ	K	Yes	Experimenter	SB	S	ws	0.84
Coyne et al in press, 2009	not publ	K	Yes	Teacher	SB	A	ws	0.85
Coyne et al, 2007a	publ	K	Yes	Experimenter	SB	A+	ws	2.13
Coyne et al, 2007b	publ	K	Yes	Experimenter	SB	A+	ws	1.64
Coyne et al, 2004	publ	K	Yes	Experimenter	SB	A	at	0.96
Crevecoeur, 2008	not publ	K	Yes	Teacher & Experimenter	SB	S+	tu	1.20
Cronan, 1996	publ	Pre-K	No	Parent	SB	A&S	n	0.22
Danger, 2003	publ	Pre-K & K	Yes	Experimenter	Play	S	tu	0.52
Daniels, 1994a	publ	Pre-K	Yes	Teacher	SL	S	at	0.70
Daniels, 1994b	publ	K	-	Teacher	SL	S	tu	1.88
Daniels, 2004	publ	K	No	Teacher	SL	S	n	1.26
Eller et al, 1988	publ	K	-	Experimenter	SB	A	n	1.28
Ewers & Brownson, 1999	publ	K	No	Experimenter	SB	A & S	at	1.06
Freeman, 2008	not publ	K	Yes	Teacher	SB	A & S	at	1.87
Hargrave & Senechal,	publ	Pre-K	Yes	Teacher	DR	A & S	at	0.71

2000								
Harvey, 2002	not publ	Pre-K	No	Parent	AB	S	at	1.63
Hasson, 1981	publ	K	Yes	Teacher	Cloze	A & S	at	0.67
Huebner, 2000	publ	Pre-K	No	Parent	DR	S	at	0.82
Justice, 2005	publ	K	Yes	Experimenter	SB	A+	n	1.49
Karweit, 1989	publ	K	Yes	Teacher	SB	S	m	0.35
Lamb, 1986	not publ	Pre-K	Yes	Experimenter	SB	S	n	0.14
Leung & Pikulski, 1990	publ	K	-	Child care teacher & Experimenter	SB	S	tu	0.69
Leung, 2008	publ	Pre-K	No	Experimenter	SB	A&S	at	0.55
Lever & Senechal, 2008	not publ	K	No	Experimenter	DR	A	at	0.95
Levinson, 1989	publ	K	-	Teacher	Writing	S	at	0.66
Light et al, 2004	publ	Pre-K	-	Experimenter	AAC	A	at	5.43
Loftus, 2008	not publ	K	Yes	Experimenter	VOC	A+	ws & pp	0.45
Lonigan & Whitehurst, 1998	publ	Pre-K	Yes	Experimenter	DR	S	n	-0.10

Lonigan et al, 1999	publ	Pre-K	Yes	Parent	SB	S	tu	0.59
Lowenthal, 1981	publ	Pre-K	Yes	Teacher	LT	S	n	1.04
Lucas, 2006	not publ	K	Yes	Teacher	LT	S	at	0.16
McConnell, 1982	publ	K	Yes	Parent	IBI	S	tu	0.58
Meehan, 1999	not publ	Pre-K	Yes	Parent & Specialist	SB	S	na	0.52
Mendelsohn, 2001	publ	Pre-K	Yes	Parent	SB +	S	n	0.45
Murphy, 2007	not publ	Pre-K	Yes	Experimenter	DR	S	tu	1.50
Nedler & Sebera, 1971a	publ	Pre-K	Yes	Child care teacher	BEEP	S	n	0.43
Nedler & Sebera, 1971b	publ	Pre-K	Yes	Child care teacher	BEEP	S	n	0.17
Neuman & Gallagher, 1994	publ	Pre-K	Yes	Parent	SB + Play	S	n	1.43
Neuman, 1999	publ	Pre-K	Yes	Child care teacher	SB	S	tu	0.06
Notari-syverson et al, 1996	not publ	Pre-K	Yes	Teacher	LTL	A & S	ws	0.46
Peta, 1973	publ	Pre-K	Yes	Teacher	TER	A & S	at	1.75
Rainey, 1968	not publ	Pre-K	Yes	Teacher	SV	S	tu	0.24

Schetz, 1994	publ	Pre-K	-	Experimenter	CAI	S	tu	0.45
Senechal et al, 1995	publ	Pre-K	No	Experimenter	SB	A+	at	1.55
Senechal, 1997	publ	Pre-K	No	Experimenter	SB	A+	at	0.71
Silverman, 2007a	publ	K	Yes	Teacher	MDV	A & S+	n	0.92
Silverman, 2007b	publ	K	No	Teacher	SB +	A & S+	n	1.43
Simon, 2003	not publ	Pre-K	Yes	Teacher	SB +	A & S	tu	0.42
Smith, 1993	not publ	Pre-K	No	Experimenter	Augmented SB	A	ws	0.67
Terrell & Daniloff, 1996	publ	K	No	Experimenter	Book/Video	A	n	0.59
Walsh & Blewitt, 2006	publ	Pre-K	No	Experimenter	SB	A	tu	2.04
Warren & Kaiser, 1986	publ	Pre-K	Yes	Experimenter	LIP	S	n	1.50
Wasik & Bond, 2001	publ	Pre-K	Yes	Teacher	IR	A&S	tu	1.53
Wasik et al, 2006	publ	Pre-K	Yes	Teacher	SB +	S	at	1.47
Watson, Betsy, 2008	publ	Pre-K	Yes	Teacher	SB	A&S	ws	0.64
Whitehurst et al, 1988	publ	Pre-K	No	Parent	DR	S+	tu	0.96
Whitehurst et al, 1994	publ	Pre-K	Yes	Child care teacher &	DR	A&S+	at	0.63

Parent

[a] Sample was coded at risk if at least 50% of the participant sample was within one risk category: low SES (at or below the national poverty level of $22,000); parental education of high school graduation or below; qualification for free and reduced lunch; second language status; low achievement (as identified by teacher report, achievement or AYP); IEP or Title I placement.

[b]AB=Audio Books, AAC=Augmentative and Alternative Communications system, BEEP=Bilingual Early Childhood Educational Program, CAI=Computer-assisted instruction, DR=Dialogic Reading, IBI=Individual Bilingual Instruction, IR=Interactive Reading, LIP=Language Intervention Program, LT=Language Training, LTL=Ladders to Literacy, MDV=Multi-Dimensional Vocabulary, SB=Storybook, SL=Sign Language, SV=Sight Vocabulary, TER=Total Environment Room, VOC=General Vocabulary Intervention

[c]A=Author-created, S=Standardized;

[d] + included a delayed posttest

[e]n=Received no treatment (includes wait list), tu=treatment as usual, at=alternate treatment, ws=within subject, pp=pre-post design

Table 2. Mean Effect Sizes of Contextual Characteristics

Characteristics	*k*	*g*	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Intervener					41.26	
Experimenter	24	0.96***	0.70, 1.22	84.94		77.63
Teacher	22	0.92***	0.70, 1.15	249.89		91.60
Parent	11	0.76***	0.41, 1.11	44.94		82.20
Child care provider	8	0.13	-0.06, .31	1.91		0.00
Group Size					0.56	
Individual	21	1.02***	0.76, 1.28	122.99		84.55
5 or less	15	.88***	0.64, 1.12	61.68		77.30
6 or more	16	1.04***	0.64, 1.44	87.91		89.76

*** $p < .0001$
[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)
[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Table 3. Mean Effects by Dosage

Characteristics	*k*	*g*	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Duration of Instruction					183.82	
1 week or less	7	1.35***	0.99, 1.70		19.90	69.81
2 weeks	10	1.12***	0.79, 1.46	34.56		73.96
More than 2 weeks	49	0.87***	0.72, 1.02	444.79		89.21
Less than 6 weeks	30	0.97***	0.75, 1.19	246.62		88.24

More than 6 weeks	29	0.87***	0.67, 1.10	244.10		88.00
Frequency					6.06	
5 sessions or less	12	1.42***	0.98, 1.85	75.31		85.39
More than 5 sessions	30	0.83***	0.67, 0.99	106.40		72.75
					3.78	
18 sessions or less	22	1.13***	0.87, 1.39	139.97		84.99
More than 18 sessions	21	0.80***	0.58, 1.01	90.79		77.97
Intensity					0.11	
Less than 20 minutes	19	0.97***	0.74, 1.20	83.94		78.56
More than 21 minutes	17	0.91***	0.62, 1.20	124.96		87.20

***$p < .0001$

[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)

[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Table 4. Mean Effect Sizes by Type of Instruction

Characteristic	***k***	***g***	**95% CI**	Q_{within}[a]	$Q_{between}$[b]	I^2
Type of Instruction					17.67	
Explicit	15	1.10***	0.81, 1.30	67.89		79.38
Implicit	25	0.62***	0.46, 0.78	95.63		74.90
Combination	17	1.21***	0.94, 1.47	85.70		81.33

*** p < .0001

[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)
[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Table 5. Mean Effect Sizes for At-risk and Average and Above Average Learners

Characteristic	*k*	*g*	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Type of learner (Experimenter-Identified)					0.18	
At-risk	40	0.85***	0.69, 1.01	245.86		84.14
Average and above						
Average learners	18	0.91***	0.69, 1.12	69.51		5.54
At-risk and SES status					4.19	
At-risk low SES	25	0.77***	0.53, 1.01	158.06		84.82
At-risk middle to high SES	9	1.35***	0.85, 1.85	137.33		94.17

*** $p < .0001$
[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)
[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Table 6. Mean Effect Sizes of Research Design Characteristics

Characteristics	***k***	***g***	**95% CI**	Q_{within}[a]	$Q_{between}$[b]	I^2
Type of Control group					21.44	
Received nothing (includes wait list)	7	0.51***	0.30, 76	25.60		76.55
Alt. treatment	22	1.03***	0 .78, 1.27	106.95		80.36
Treatment as usual	17	0.78***	0.49, 1.08	98.50		83.76
Within-subjects	8	1.09***	0.75, 1.44	54.45		87.76
Design					0.04	
Quasi-Experimental	56	0.88***	0.74, 1.00	434.32		87.34
Experimental	11	0.92***	0.49, 1.35	116.90		91.45

***p<.0001

[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)

[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Table 7. Mean Effect Sizes for Author-Created Assessments and Standardized Assessments

Characteristic	***k***	***g***	**95% CI**	Q_{within}[a]	$Q_{between}$[b]	I^2
Assessments					6.35	
Author-created	19	1.21***	0.86, 1.60	257.42		93.01
Standardized	36	0.71***	0.57, 0.85	168.07		79.18
Type of Vocabulary Measure					9.76	
Receptive vocabulary	97	0.80***	0.68, 0.91	812.30		88.18
Expressive vocabulary	86	0.69***	0 .60, 0.78	1066.99		92.03
Combination	33	1.11***	0.84, 1.39	404.02		92.08

*** $p < .0001$

[a]Q_{within} refers to the homogeneity of each subgroup (df = k-1)

[b]$Q_{between}$ refers to the moderator contrast (df =number of subgroups–1)

Figure 1. Stem-and-leaf display of the effect sizes per study on all outcome measures

Stem	Leaf
-0.	1
0.0	6
0.1	4, 6, 7
0.2	2, 4, 6
0.3	1, 5
0.4	2, 3, 5, 5, 5, 6, 8
0.5	1, 2, 5, 8, 9, 9
0.6	3, 4, 6, 7, 7, 9
0.7	0, 1, 1, 1
0.8	2, 4, 5
0.9	2, 5, 6, 6, 7
1.0	0, 4, 6
1.2	0, 6, 8
1.4	3, 3, 7, 9
1.5	0, 0, 3, 4, 5
1.6	3, 3, 4
1.7	5
1.8	7, 8
1.9	9
2.0	4
2.1	3 3

Key: 0.4| 5 is an effect size of 0.45