Review of Educational Research
September 2010, Vol. 80, No. 3, pp. 300–335
DOI: 10.3102/0034654310377087

The Effects of Vocabulary Intervention on Young Children's Word Learning: A Meta-Analysis

Loren M. Marulis and Susan B. Neuman
University of Michigan

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.

Keywords: achievement, achievement gap, early childhood, literacy, meta-analysis, vocabulary

Learning the meanings of new words is an essential component of early reading development (Roskos et al., 2008). Vocabulary is at the heart of oral language comprehension and sets the foundation for domain-specific knowledge and later reading comprehension (Beck & McKeown, 2007; Snow, Burns, & Griffin, 1998). As Stahl and Nagy (2006) report, the size of children's vocabulary knowledge is strongly related to how well they will come to understand what they read. Logically, children will need to know the words that make up written texts to understand them, especially as the vocabulary demands of content-related materials increase in the upper grades. 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 significant differences in vocabulary knowledge among children from different socioeconomic groups beginning in young toddlerhood through high school (Hart & Risley, 1995; Hoff, 2003). Extrapolating to the first 4 years of life, Hart and Risley (2003) estimate that the average child from a professional family would be exposed to an accumulated experience of about 42 million words compared to 13 million for the child from a poor family.

Moats (1999) estimated the difference at school entry to be about 15,000 words, with linguistically disadvantaged children knowing about 5,000 words compared to the more advantaged children knowing 20,000 words. Furthermore, children from low income groups tend to build their vocabulary at slower rates than children from high socioeconomic status (SES) groups (Anderson & Nagy, 1992), potentially creating a cumulative disadvantage over time. By fourth grade, children with vocabulary below grade level, 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 significant 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 (Neuman, 2009). Average children acquire many hundreds of word meanings each year during the first 7 years of vocabulary acquisition. To catch up, therefore, children with vocabulary limitations will need to acquire several hundred words in addition to what they would otherwise learn (Biemiller, 2006). In essence, interventions will have to accelerate—not simply improve—children's vocabulary development to narrow the achievement gap.

To date, however, little is known about the effectiveness of overall vocabulary training on changes in general vocabulary knowledge. Previously published meta-analyses (Elleman, Lindo, Morphy, & Compton, 2009; Stahl & Fairbanks, 1986) have addressed the impact of different vocabulary interventions on reading comprehension.

As can be seen in Table 1, Stahl and Fairbanks (1986) reported an average effect size of 0.97 of vocabulary instruction on comprehension of passages containing taught words; a more modest effect size of 0.30 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 0.50 of training programs on comprehension using author-created measures and a 0.10 effect size for standardized measures. In addition to these meta-analyses, the National Reading Panel (2000) conducted a narrative review of published experimental and quasiexperimental studies evaluating vocabulary instruction on comprehension skills and found that it was "generally effective" for improving comprehension.

Although informative, neither of these meta-analyses nor the narrative review addressed the effects of training on learning the meanings of words, a more proximal measure of the impact of the interventions. Furthermore, 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 (2000) report include text restructuring, repetition and multiple exposures to words in text, and rereading, 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.

TABLE 1
Previous vocabulary meta-analyses

Author(s)	Grade	Intervention(s)	Skill assessed	Effect size (ES) author-created (words taught)	ES standardized (global)	ES both
Stahl and Fairbanks, 1986	2–college	Any	Comprehension	.97	.30	
Elleman et al., 2009	Pre-K–12	Any (beyond read-alouds)	Comprehension	.50	.10	
Mol et al., 2009	Pre-K–1	Interactive book reading	Oral language			.28
Mol et al., 2009	Pre-K–1	Interactive book reading	Print knowledge			.25
Mol et al., 2008	Pre-K	Dialogic parent–child readings	Oral language		.50	
Mol et al., 2008	K	Dialogic parent–child readings	Oral language		.14	
National Early Literacy Panel, 2008	Pre-K	Interactive book reading	Oral language			.75
National Early Literacy Panel, 2008	K	Interactive book reading	Oral language			.66
National Early Literacy Panel, 2008	Pre-K	Code focused	Oral language			.32
National Early Literacy Panel, 2008	K	Code focused	Oral language			.13

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 the vocabulary of the text. Mol, Bus, and deJong (2009) avoided this potential confound by separating oral language outcomes and print-related skills. Focusing specifically on the impact of interactive storybook 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 fostering the same growth in young children's language skills as researchers did when implementing interventions. Furthermore, in another recent meta-analysis by Mol and her colleagues examining the effects of parent–child storybook readings on oral language development (Mol, Bus, deJong, & Smeets, 2008), two groups did not appear to benefit from the intervention: those children at risk for language and literacy impairments and kindergarten children. Using a more rigorous set of screening criteria (e.g., published in peer-reviewed journals studies, randomized controlled trials or quasiexperimental studies only), the National Early Literacy Panel (2008) essentially confirmed Mol et al.'s findings. They, too, reported only moderate effects of storybook reading interventions on oral language and print knowledge, with smaller effect sizes reported for children at risk of reading difficulties compared to those not at risk. Furthermore, their analyses of code-focused interventions and pre-K and kindergarten programs found negligible effects (0.32, 0.13, respectively) on oral language skills.

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 previous work (Elleman et al., 2009; Mol et al., 2008; Mol et al., 2009; National Early Literacy Panel, 2008) in several ways. First, because 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 of the children of the National Longitudinal Survey of Youth 1979 cohort, found more than half of the social class effect on early oral language was attributable to the years before 5 and that rates of vocabulary growth declined for each subsequent age period. Second, we extend the work of Mol and her colleagues (2008, 2009) to include all vocabulary interventions in the early years in addition to interactive or shared book reading. Third, we examine the impact of these interventions on growth of general vocabulary knowledge. 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 words to young children?
2. What methodological characteristics of vocabulary interventions 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 to identify and examine not only nontraditional interventions but also 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 literatures 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. 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 postintervention comparison between preexisting groups (e.g., two kindergarten classrooms); (c) participants had no mental, physical, or sensory handicaps and were within ages birth through 9; (d) the study was conducted with English words, excluding foreign language or nonsense words (to be able to make comparable comparisons across studies); and (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., Peabody Picture Vocabulary Test) or an author-created measure (e.g., Coyne, McCoach, & Kapp's, 2007, Expressive and Receptive Definitions).

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 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 program 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 our established cutoff, if they were off topic and not related to word learning, if the study was conducted in a language other than English, or if the citation referred to 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 interrater reliability was established (Cohen's $\kappa = .9–1.0$). 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 ($\kappa = .96$). The exclusion coding revealed 3,586 relevant citations, which were subsequently retrieved and read in full.

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 e-mails and received 28 responses (29% response rate), which generated 12 manuscripts. In total, our process yielded 3,598 articles and manuscripts.

Inclusion Coding

Our next task was to examine the research 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 10 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's $\kappa = .96$) fell well within range (Landis & Koch, 1977). Subsequently, each coder individually coded the remainder of the studies. In all, 111 studies met all criteria and were set aside for the comprehensive study variable coding. These studies were divided into two groups: those that targeted 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 for birth through kindergarten focused on oral language development through listening and speaking, with concomitant changes in receptive and expressive language. Intervention for Grades 1 to 3 emphasized the ability to identify and understand vocabulary words in print and children's subsequent understanding of these words in a text. Consequently, our focus was 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 and Potential Moderators

To address our second and third research questions, we identified 10 characteristics for planned moderator analyses based on previous research and findings (Mol et al., 2009; National Early Literacy Panel, 2008; National Reading Panel, 2000): four intervention (the adult who conducted the intervention, group size, dosage of the intervention, and type of training), two participant (at-risk status and SES), two measurement outcome (the type and focus of the dependent measures used to determine changes in word learning), and two study level (research design and nature of the control group). If study descriptions were unclear or key characteristics were 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 particular study were 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).

Because of 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 5 studies together. Following the coding, coders discussed each study and revised the coding manual and protocol sheets accordingly. Next, the coders coded five studies independently. Fleiss's kappa was calculated at .67, which, although falling within the "substantial agreement" range, was not sufficiently high enough to allow for proper use of moderator analysis. Borenstein, Hedges, Higgins, and Rothstein (2009) and Lipsey and Wilson (2001) 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 articles (more than 60 studies, 150 effect sizes) and achieved an agreement level within the "almost perfect agreement" range ($\kappa = .89$). Studies were then coded individually by one of the four trained coders.

Analytic Strategy

To calculate effect size estimates, we entered the data into the Comprehensive Meta-Analysis (CMA) program (Borenstein, Hedges, Higgins, & Rothstein, 2005) and standardized by the change score standard deviations (SDs). 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. Because of the ability to enter data in more than 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. As this standard procedure yields the most precise estimates, we would have been concerned if a large proportion of the effect sizes were calculated using an alternate method. However, this was not necessary for the large majority of studies; nearly 88% of the effect sizes were calculated using means and standard deviations.

We estimated all effect sizes using Hedges's *g* coefficient, a more conservative form of the Cohen's *d* effect size estimate. Hedges's *g* uses a correction factor J to correct for bias from sample size, which is calculated as follows: $J = 1 - (3 / (4 \times df - 1))$, where $df = N_{TOTAL} - 2$. To obtain Hedges's *g* from Cohen's *d*, the following calculations can be made: $g = d \times J$, $StdErr(g) = StdErr(d) \times J$.

We then weighted the effect sizes by the inverse of their error variances ($1/SE^2$) to factor in the proportionate reliability of each one to the overall analysis (Shadish & Haddock, 1994). In this way, less precise estimates are given less weight in the analyses. 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). Specific to psychological, behavioral, and educational interventions, an effect size of .30 fell in the bottom quartile, .50 at the median, and .67 in the top quartile in an examination of more than 300 meta-analyses (Lipsey & Wilson, 1993).

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 used the mean effect size for each study across conditions while not pooling the variable of interest (moderator variable) in conducting the moderator analyses (Borenstein et al., 2009; Cooper & Hedges, 1994). In other words, in analyzing the effectiveness of an intervener, we averaged the effect sizes per study across the other multiple conditions (e.g., multiple outcome measures) and used one mean effect size per study per moderator analysis. 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 and 95% confidence intervals around the point estimates to address heterogeneity. Within a random effects model, the variance includes the within-study variance plus the estimate of the between studies variance (Borenstein et al., 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, because our sample of studies involved a larger corpus of vocabulary interventions than previous meta-analyses, we expected to have a large 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; $SD = 0.53$). This outlying effect size was quite large (Lipsey & Wilson, 2001, mean effect size $g = 5.43$, $SE = 0.69$). However, because of 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 ($g = 0.89$, $SE = 0.065$, $CI_{95} = 0.76, 1.00$) and without ($g = 0.86$, $SE = 0.062$, $CI_{95} = 0.74, 0.98$) and found no significant difference ($p > .05$). Nevertheless, to reduce the impact of this large outlier in the planned moderator analyses, we Winsorized it by resetting this effect size to the next largest effect size of 2.13 (which was only two standard deviations from the mean). This allowed for a smoother distribution of effect sizes and a less extreme upper limit without losing data (Lipsey & Wilson, 2001). All subsequent analyses, therefore, were conducted using the Winsorized value. Effect sizes ranged from –10 to 2.13, with 37 effect sizes below the sample mean and 29 above the sample mean.

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 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 *N* (Orwin, 1983), which allowed us to use a value other than null against an effect size criterion (i.e., rather than a *p* value) addressing the possibility that file-drawer studies could have a nonzero mean effect. 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 a mean effect size of 0.49 to bring our overall Hedges'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 64 articles, 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. Of the studies, 70% were published in peer-reviewed journals, and 60% of the children sampled were in pre-K classrooms. The typical study was quasiexperimental 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 can be seen in Table 2, storybook reading and dialogic reading were the most prevalent interventions. However, as noted in both National Reading Panel (2000) and National Early Literacy Panel (2008), there were wide variations across interventions. For specific descriptive study characteristics, see Table 2.

We expected our sample to be heterogeneous which was subsequently confirmed, $Q_w(66) = 551.54$, $p < .0001$, $I^2 = .88$. Total variability that could be attributed 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. For the range of associated effect sizes and the precision of each estimate, see Figure 1.

Overall Effect Sizes

To examine the benefit of vocabulary training on word learning, we first calculated an overall effect size for pre-K and kindergarten. The overall effect size was $g = 0.88$, $SE = 0.06$, $CI_{95} = 0.76, 1.01$, $p < .0001$. Vocabulary training demonstrated a large effect on word learning in pre-K ($g = 0.85$, $SE = 0.09$, $CI_{95} = 0.68, 1.01$, $p < .0001$) and kindergarten ($g = 0.94$, $SE = 0.11$, $CI_{95} = 0.73, 1.14$, $p < .0001$). Although the magnitude of the effect was slightly larger for kindergarten students, differences were not statistically significant, $Q_b(1) = 0.48$, $p = .49$. These effect sizes are considered both educationally significant (Lipsey & Wilson, 1993) and large (Cohen, 1988).

We found that published studies had significantly higher effect sizes ($g = 0.95$, $SE = 0.084$, $CI_{95} = 0.79, 1.11$) than unpublished studies ($g = 0.71$, $SE = 0.087$, $CI_{95} = 0.54, 0.88$), $Q_b(1) = 4.53$, $p < .05$. Therefore, our overall effect size could be considered conservative because of the inclusion of a considerable number of unpublished studies (20 of 67).

A total of 11 studies also reported a delayed posttest. To avoid dependency of effect sizes, we excluded the delayed posttests (e.g., defined as posttests given more than 24 hours after the completion of the intervention) from our overall analysis (28 effect sizes). The mean effect size for the delayed posttests ($g = 1.02$, $SE = 0.22$, $CI_{95} = 0.58, 1.45$) did not differ significantly from that of immediate posttests ($g = 0.88$, $SE = 0.06$, $CI_{95} = 0.76, 1.01$), $Q_b(1) = 0.30$, $p = .58$. These results indicated that moderate effects persisted over time for these 11 studies (2–180 days

TABLE 2
Key characteristics and effect sizes of meta-analyzed studies

Author(s)	Publication status	Grade	At risk[a] (coder determined)	Trainer	Condition[b]	Type of testing[c]	Control group[d]	Effect size (*g*)
Ammon and Ammon, 1971	Publ.	Pre-K	No	Experimenter	Peabody Picture Vocabulary Test training	S	n	1.00
Arnold et al., 1994	Publ.	Pre-K	No	Parent	DR	S	tu	0.51
Beck and McKeown, 2007 Study 1	Publ.	K	Yes	Teacher	SB	A	tu	1.54
Beck and McKeown, 2007 Study 2	Publ.	K	Yes	Teacher	SB	A	ws	1.96
Bonds, 1987	Unpubl.	K	—	Teacher	Basal	S	at	0.71
Bortnem, 2005	Unpubl.	Pre-K	Yes	Experimenter	SB	S	tu	0.97
Brackenbury, 2001	Unpubl.	Pre-K	No	Experimenter	LIP	A	at	1.63
Brickman, 2002	Publ.	Pre-K	Yes	Parent	SB	A and S	at	0.31
Brooks, 2006	Unpubl.	K	No	Teacher	SB	A	at	0.48
Christ, 2007	Unpubl.	K	Both	Teacher	SB	A	n	0.26
Coyne et al., 2004	Publ.	K	Yes	Experimenter	SB	A	at	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., 2008	Unpubl.	K	Yes	Experimenter	SB	S	ws	0.84
Coyne et al., in press	Unpubl.	K	Yes	Teacher	SB	A	ws	0.96
Crevecoeur, 2008	Unpubl.	K	Yes	Teacher and experimenter	SB	S +	tu	1.20
Cronan et al., 1996	Publ.	Pre-K	No	Parent	SB	A and S	n	0.22
Danger, 2003	Publ.	Pre-K and K	Yes	Experimenter	Play	S	tu	0.52
Daniels, 1994a	Publ.	Pre-K	Yes	Teacher	SL	S	at	1.88

(continued)

TABLE 2 (CONTINUED)

Author(s)	Publication status	Grade	At risk[a] (coder determined)	Trainer	Condition[b]	Type of testing[c]	Control group[d]	Effect size (*g*)
Daniels, 1994b	Publ.	K	—	Teacher	SL	S	tu	1.26
Daniels, 2004	Publ.	K	No	Teacher	SL	S	n	0.70
Eller et al., 1988	Publ.	K	—	Experimenter	SB	A	n	1.28
Ewers and Brownson, 1999	Publ.	K	No	Experimenter	SB	A and S	at	1.06
Freeman, 2008	Unpubl.	K	Yes	Teacher	SB	A and S	at	1.87
Hargrave and Sénéchal, 2000	Publ.	Pre-K	Yes	Teacher	DR	A and S	at	0.71
Harvey, 2002	Unpubl.	Pre-K	No	Parent	AB	S	at	1.63
Hasson, 1981	Publ.	K	Yes	Teacher	Cloze	A and S	at	0.67
Huebner, 2000	Publ.	Pre-K	No	Parent	DR	S	at	0.82
Justice et al., 2005	Publ.	K	Yes	Experimenter	SB	A +	n	1.49
Karweit, 1989	Publ.	K	Yes	Teacher	SB	S	m	0.35
Lamb, 1986	Unpubl.	Pre-K	Yes	Experimenter	SB	S	n	0.14
Leung, 2008	Publ.	Pre-K	No	Child care teacher and experimenter	SB	A and S	at	0.69
Leung and Pikulski, 1990	Publ.	K	No	Experimenter	SB	A and S	tu	0.55
Lever and Sénéchal, 2009	Unpubl.	K	No	Experimenter	DR	A	at	0.95
Levinson & Lalor, 1989	Publ.	K	—	Teacher	Writing	S	at	0.66
Light et al., 2004	Publ.	Pre-K	—	Experimenter	AAC	A	at	5.43
Loftus, 2008	Unpubl.	K	Yes	Experimenter	VOC	A +	ws	0.45
Lonigan et al., 1999	Publ.	Pre-K	Yes	Experimenter	SB	S	tu	–0.10
Lonigan and Whitehurst, 1998	Publ.	Pre-K	Yes	Parent	DR	S	n	0.59
Lowenthal, 1981	Publ.	Pre-K	Yes	Teacher	LT	S	n	1.04
Lucas, 2006	Unpubl.	K	Yes	Teacher	LT	S	at	0.16
McConnell, 1982	Publ.	K	Yes	Parent	IBI	S	tu	0.58

(continued)

TABLE 2 (CONTINUED)

Author(s)	Publication status	Grade	At risk[a] (coder determined)	Trainer	Condition[b]	Type of testing[c]	Control group[d]	Effect size (*g*)
Meehan, 1999	Unpubl.	Pre-K	Yes	Parent and specialist	SB	S	na	0.52
Mendelsohn et al., 2001	Publ.	Pre-K	Yes	Parent	SB +	S	n	0.45
Murphy, 2007	Unpubl.	Pre-K	Yes	Experimenter	DR	S	tu	1.50
Nedler and Sebera, 1971a	Publ.	Pre-K	Yes	Child care teacher	BEEP	S	n	0.43
Nedler and Sebera, 1971b	Publ.	Pre-K	Yes	Child care teacher	BEEP	S	n	0.17
Neuman, 1999	Publ.	Pre-K	Yes	Child care teacher	SB	S	tu	0.06
Neuman and Gallagher, 1994	Publ.	Pre-K	Yes	Parent	SB + play	S	n	1.43
Notari-Syverson et al., 1996	Unpubl.	Pre-K	Yes	Teacher	LTL	A and S	ws	0.46
Peta, 1973	Publ.	Pre-K	Yes	Teacher	TER	A and S	at	1.75
Rainey, 1968	Unpubl.	Pre-K	Yes	Teacher	SV	S	tu	0.24
Schetz, 1994	Publ.	Pre-K	—	Experimenter	CAI	S	tu	0.45
Sénéchal, 1997	Publ.	Pre-K	No	Experimenter	SB	A +	at	1.55
Sénéchal et al., 1995	Publ.	Pre-K	No	Experimenter	SB	A +	at	0.71
Silverman, 2007a	Publ.	K	Yes	Teacher	MDV	A and S +	n	0.92
Silverman, 2007b	Publ.	K	No	Teacher	SB +	A and S +	n	1.43
Simon, 2003	Unpubl.	Pre-K	Yes	Teacher	SB +	A and S	tu	0.42
Smith, 1993	Unpubl.	Pre-K	No	Experimenter	Augmented SB	A	ws	0.67
Terrell and Daniloff, 1996	Publ.	K	No	Experimenter	Book, video	A	n	0.59
Walsh and Blewitt, 2006	Publ.	Pre-K	No	Experimenter	SB	A	tu	2.04
Warren and Kaiser, 1986	Publ.	Pre-K	Yes	Experimenter	LIP	S	n	1.50
Wasik and Bond, 2001	Publ.	Pre-K	Yes	Teacher	IR	A and S	tu	1.47
Wasik et al., 2006	Publ.	Pre-K	Yes	Teacher	SB +	S	at	1.53
Watson, 2008	Publ.	Pre-K	Yes	Teacher	SB	A and S	ws	0.64

(continued)

TABLE 2 (CONTINUED)

Author(s)	Publication status	Grade	At risk[a] (coder determined)	Trainer	Condition[b]	Type of testing[c]	Control group[d]	Effect size (g)
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 and parent	DR	A and S +	at	0.63

Note. — indicates missing, insufficient, or unclear information.
[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 or reduced-price lunch, second language status, low achievement (as identified by teacher report, achievement, or adequate yearly progress), individualized education program or Title I placement.
[b]AAC = augmentative and alternative communications system; AB = audio books; 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 = multidimensional vocabulary; SB = storybook; SL = sign language; SV = sight vocabulary; TER = total environment room; VOC = general vocabulary intervention.
[c]A = author created; S = standardized. + includes a delayed posttest.
[d]N = received no treatment (includes wait list); TU = treatment as usual; AT = alternate treatment; WS = within subject.

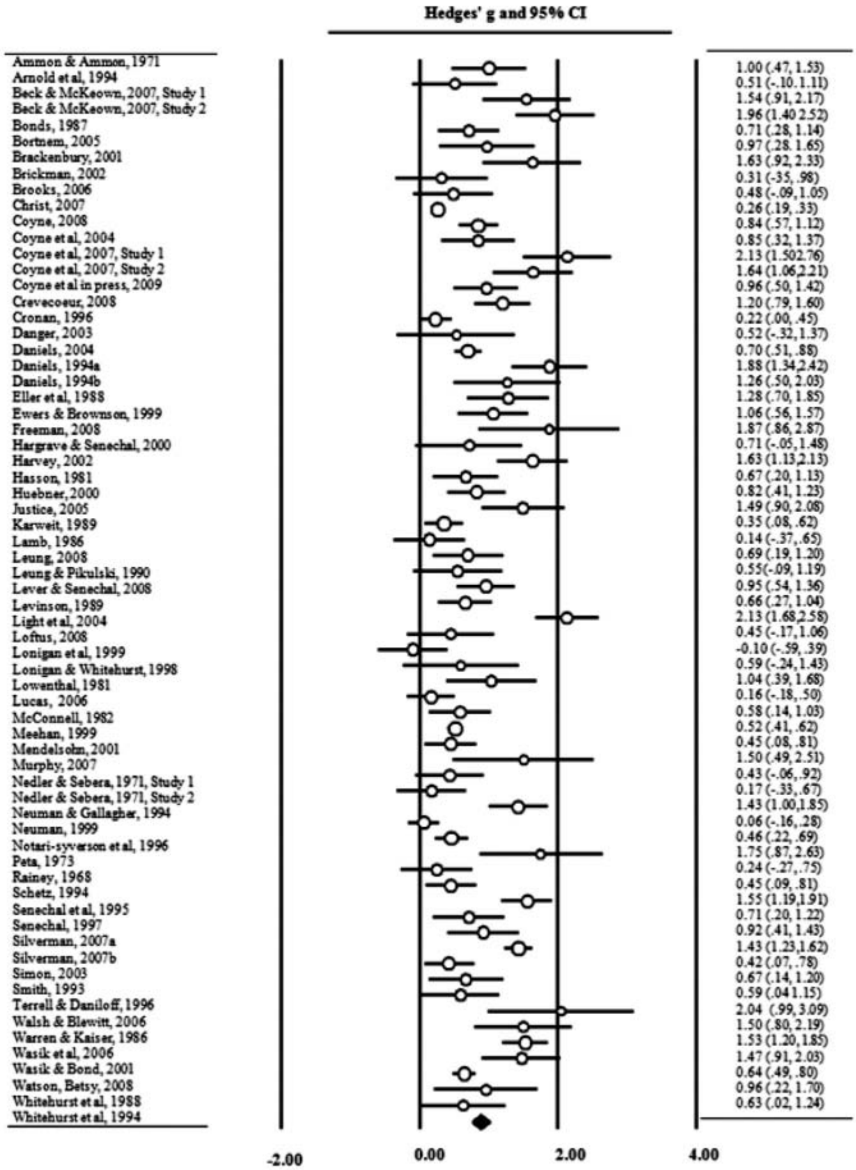

Figure 1. *Forest plot of effect sizes (k = 67). Each cirle represents each individual study and is proportionate to its weight in the overall analysis; the circles and confidence interval bars illustrate the estimated precision of each study. The diamond represents the mean effect size for the entire sample.*

postintervention; $M = 63.52$, $SD = 57.61$). Because of the small number of studies, however, we were not able to conduct further interactions between moderators.

Analysis of Moderator Effects

To attempt to explain variance, we examined the influence of key study variables on effect sizes by conducting moderator analyses within 10 (4 intervention, 2 participant, 2 outcome measure, and 2 study level) categories of study characteristics. We were able to conduct all of the above-planned moderator analyses using contrasts because each subgroup had more than four studies even after the removal of studies because of missing data.

We used a fixed effects model to combine subgroups and 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 et al., 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 3, 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 often been 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) as compared to whole group instruction (Powell, Burchinal, File, & Kontos, 2008), particularly for young children. As shown in Table 3, more than 20% of the studies did not specify group sizes; the remaining studies included a relatively equal number of small and large group interventions, with a somewhat larger number of individualized interventions.

Our moderator analysis on these contextual features analysis indicated a significant effect for the adult who carried out the intervention, $Q_b(3) = 41.26$, $p < .0001$. Training provided by the experimenter ($g = 0.96$, $SE = 0.13$, $CI_{95} = 0.70, 1.22$) or the teacher ($g = 0.92$, $SE = 0.11$, $CI_{95} = 0.70, 1.15$) resulted in equal magnitude of growth. Although interventions given by the parent appeared to have a lower magnitude of growth, the effect size was still substantial ($g = 0.76$, $SE = 0.18$, $CI_{95} = 0.41, 1.11$). There were no significant differences in the effect sizes associated with these trainers, $Q_b(2) = 0.61$, $p = .44$.

On the other hand, trainings given by child care providers were significantly less successful. Our analysis indicated that trainings given by child care providers yielded smaller effect sizes that were significantly lower than all others, $g = 0.13$, $SE = 0.095$, $CI_{95} = -0.06, 0.31$; $Q_b(3) = 41.26$, $p < .0001$. It should be noted, however, that there were substantially fewer studies in this group than in others. In contrast to others, these interventions were highly and significantly homogeneous, $Q_w = 1.9$, $p = .4$, $I^2 = 0.00$; no between-studies variance was unexplained.[1]

TABLE 3

Mean effect sizes for contextual characteristics of interventions

Characteristic	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	0.98***	0.73, 1.22	122.99		84.55
5 or fewer	15	0.88***	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 contrasts (df = number of subgroups – 1).

In addition to variations in the person who provided the treatment, the studies varied in terms of the number of participants who made up the intervention group. Whether children were taught in individualized settings (g = 0.98, SE = 0.14, CI_{95} = 0.73, 1.22), small groups of five children or fewer (g = 0.88, SE = 0.13, CI_{95} = 0.64, 1.12), or large groups of six children or more (g = 1.04, SE = 0.21, CI_{95} = 0.64, 1.44) did not affect the effect size, $Q_b(2) = 0.56$, $p = .75$. Rather, all group configurations benefited equally and 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, the frequency would be 15, and the intensity would be 20.

As shown in Table 4, 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.

Duration. The duration of intervention ranged broadly from 1 to 270 days, with a median of 42 days of instruction. Shown in Table 4, interventions lasting a week or less (g = 1.35, SE = 0.18, CI_{95} = 0.99, 1.70) resulted in significantly higher effect sizes than those lasting longer than a week, up to 270 days (g = 0.85, SE = 0.07, CI_{95} = 0.71, 1.00), $Q_b(1) = 6.28$, $p < .05$. 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 seven studies that lasted a week or less in the sample. However, we then conducted an analysis to examine whether the median of duration of instruction—42 days—would

TABLE 4
Mean effects for dosage of intervention

Characteristic	k	g	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Duration of training					183.82	
1 week or less	7	1.35***	0.99, 1.70	19.90		69.81
More than 1 week	52	0.85***	0.71, 1.00	445.27		88.55
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 42 days	30	0.97***	0.75, 1.19	246.62		88.24
More than 42 days	29	0.87***	0.67, 1.10	244.10		88.53
Frequency					6.06	
5 sessions or fewer	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 fewer	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
20 minutes or more	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 contrasts (df = number of subgroups – 1).

moderate the effect size. Our analysis indicated interventions lasting 42 days or less ($g = 0.97$, $SE = 0.11$, $CI_{95} = 0.75, 1.19$) had no less effect than those lasting more than 42 days ($g = 0.87$, $SE = 0.10$, $CI_{95} = 0.67, 1.10$), $Q_b(1) = 0.44$, $p = .51$. Taken together, these results suggest that interventions of brief duration can be associated with positive word learning outcomes.

Frequency. The number of intervention sessions within studies ranged from 1 to 180 sessions, with a median of 18 sessions. However, there were a substantial number of studies in which the interventions included five 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 five or fewer intervention sessions had significantly higher effect sizes ($g = 1.42$, $SE = 0.22$, $CI_{95} = 0.98, 1.85$) than those with more than five sessions ($g = 0.83$, $SE = 0.08$, $CI_{95} = 0.67, 0.99$), $Q_b(1) = 6.06$, $p < .05$. The approximately 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 ($g = 1.13$, $SE = 0.13$, $CI_{95} = 0.87, 1.39$) than those with 18 sessions or more ($g = 0.80$, $SE = 0.11$, $CI_{95} = 0.58, 1.01$), $Q_b(1) = 3.78$, $p < .05$. Consequently, this suggests that studies with a smaller number of sessions can 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 the effect sizes. Our analysis indicated no significant differences between effect sizes for the length of training. Sessions lasting less than 20 minutes ($g = 0.97$, $SE = 0.12$, $CI_{95} = 0.74, 1.20$) were not significantly different from those lasting 20 minutes or more ($g = 0.91$, $SE = 0.15$, $CI_{95} = 0.62, 1.20$), $Q_b(1) = 0.11$, $p = .74$. Given that these interventions were geared toward 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 a discrete set of skills (e.g., dialogic reading) might 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 Training

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 2), we decided to follow Elleman and colleagues' (2009) meta-analytic strategy 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—although identified under the rubric of "storybook reading intervention"—were fairly different in their strategies for teaching vocabulary. Furthermore, 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 or 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 (2000) in its report on vocabulary training.

This variable was straightforward 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 a 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

TABLE 5
Mean effect sizes for type of training in interventions

Characteristic	k	g	95% CI	Q_{within}[a]	Q_{between}[b]	I^2
Type of training					18.36	
Explicit	15	1.11***	0.83, 1.40	82.02		82.93
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 contrasts (df = number of subgroups – 1).

implicit word learning through dialogic reading, whereas 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, whereas others relied exclusively on implicit strategies through reading aloud.

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. Children made significantly higher gains with interventions that used an explicit method ($g = 1.11$, $SE = 0.13$, $CI_{95} = 0.83, 1.40$) or a combination of explicit and implicit methods ($g = 1.21$, $SE = 0.13$, $CI_{95} = 0.94, 1.47$) than those that employed an implicit method only ($g = 0.62$, $SE = 0.084$, $CI_{95} = 0.46, 0.78$), $Q_b(2) = 18.36$, $p < .0001$. As shown in Table 5, interventions that used a combination of explicit and implicit methods appeared to have 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 that provided explicit or 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 and Low SES 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 also accelerate its development. This would seem to suggest that vocabulary interventions specially targeted for at-risk children must have stronger and more powerful effects than those for average and above average learners.

We conducted several moderator analyses to examine this question, as can be seen in Table 6. 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 less, qualification for free or reduced-price lunch), second language status, low academic achievement (as identified by a teacher report, standardized school assessment, or adequate yearly progress), and/or special needs (as identified as having an individualized education program

TABLE 6
Mean effect sizes for participant characteristics

Characteristic	k	g	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Type of learner (coder 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		75.54
SES status					2.71	
Low SES	28	0.75***	0.54, 0.96	172.75		84.37
Middle to high SES	25	0.99***	0.79, 1.21	297.75		91.94
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 contrasts (df = number of subgroups – 1).

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 ($g = 0.85$, $SE = 0.081$, $CI_{95} = 0.69, 1.01$) and all other children ($g = 0.91$, $SE = 0.10$, $CI_{95} = 0.69, 1.12$), $Q_b(1) = 0.18$, $p = .67$. 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 ($k = 9$). Although 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 ($g = 0.75$, $SE = 0.11$, $CI_{95} = 0.54, 0.96$) and those by middle to high SES children ($g = 0.99$, $SE = 0.11$, $CI_{95} = 0.79, 1.21$), $Q_b(1) = 2.71$, $p = .10$. As low SES children are likely to have lower baseline scores, even parallel gains would not substantially close the gap.

Next, 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 at-risk category, children with low SES status ($g = 0.77$, $SE = 0.12$, $CI_{95} = 0.53, 1.01$) received gains that were significantly lower than those of middle to high SES at-risk children ($g = 1.35$, $SE = 0.26$, $CI_{95} = 0.85, 1.85$), $Q_b(1) = 4.19$, $p < .05$ (see Table 6). In other words, middle to high SES children who had at least one risk factor gained more than low SES children with at least one additional risk factor. 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.

Type of Word Learning Measurement

In their narrative analysis, the National Reading Panel (2000) report raised important questions about the measurement of vocabulary development that have since been

TABLE 7
Mean effect sizes for outcome measure characteristics

Characteristic	k	g	95% CI	Q_{within}[a]	$Q_{between}$[b]	I^2
Type of assessment					6.35	
Author created	19	1.21***	0.85, 1.57	257.42		93.01
Standardized	36	0.71***	0.57, 0.85	168.07		79.18
Focus of vocabulary measure						
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 contrasts (df = number of subgroups – 1).

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 to best examine vocabulary gains. As a result of their recommendations, a number of researchers (e.g., Leung & Pikulski, 1990; Lonigan, Anthony, Bloomfield, Dyer, & Samwel, 1999) have included both author-created and standardized assessments in their studies. Others (e.g., Coyne et al., 2007; Sénéchal & Cornell, 1993) 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 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 = 0.71, SE = 0.072, CI_{95} = 0.57, 0.85) than those on author-created measures (g = 1.21, SE = 0.18, CI_{95} = 0.85, 1.57), $Q_b(1) = 6.35$, $p < .01$. These results provide support of the National Reading Panel's (2000) recommendation of 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 taught 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.

Focus of Word Learning Measurement

We also 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, we compared the overall Hedges'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.

Our results, shown in Table 7, indicated that children made significantly higher gains on combined receptive and expressive vocabulary measures ($g = 1.11$, $SE = 0.14$, $CI_{95} = 0.84, 1.39$) than on expressive vocabulary alone ($g = 0.69$, $SE = 0.04$, $CI_{95} = 0.60, 0.78$), shown by the nonoverlapping 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 National Reading Panel report, an issue we intend to pursue in the future. There were no differences between gains for receptive ($g = 0.80$, $SE = 0.06$, $CI_{95} = 0.68, 0.91$) and expressive measures ($g = 0.69$, $SE = 0.04$, $CI_{95} = 0.60, 0.78$) or between gains for receptive ($g = 0.80$, $SE = 0.06$, $CI_{95} = 0.68, 0.91$) and combined measures ($g = 1.11$, $SE = 0.14$, $CI_{95} = 0.84, 1.39$), shown by nonoverlapping confidence intervals.

Experimental Design

In a synthesis of more than 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 versus nonrandom assignment. To examine this issue, we coded studies according to this distinction and found that 11 (16%) had true experimental designs whereas 56 (84%) employed quasiexperimental designs. Our analysis indicated, however, that although there was a slight difference in magnitude favoring the experimental studies ($g = 0.92$, $SE = 0.22$, $CI_{95} = 0.49, 1.35$), these studies were not significantly different in the size of their effects than quasiexperimental studies ($g = 0.88$, $SE = 0.07$, $CI_{95} = 0.74, 1.00$), $Q_b(1) = 0.04$, $p = .84$ (see Table 8).

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 the authors for the necessary data.

Control Group

Our sample encompassed four types of comparison groups, as shown in Table 8: a control group that received no treatment (which included wait-list designs), a comparison group that received "business as usual" (e.g., same vocabulary instruction

TABLE 8
Mean effect sizes for study design characteristics

Characteristic	*k*	*g*	95% CI	Q_{within}[a]	Q_{between}[b]	I^2
Design					0.04	
Quasiexperimental	56	0.88***	0.74, 1.00	434.32		87.34
Experimental	11	0.92***	0.49, 1.35	116.90		91.45
Type of control group					21.44	
Received nothing (includes wait list)	7	0.51***	0.25, 0.76	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		87.14

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

or practices as usual), an alternate treatment (e.g., a deliberately diluted version of the treatment with the hypothesized key ingredient missing), and a within-subject design where participants acted as their own control groups. Because we conducted eight moderator analyses for the various control group types, we used a Bonferroni-corrected significance criterion of .008 (six contrast comparisons were made). Within this criterion, only two moderator analyses were significant. Studies in which the controls used alternate treatment comparisons (g = 1.03, SE = 0.13, CI_{95} = 0.78, 1.27) were significantly higher in effect sizes compared to studies in which the controls received no treatment at all (g = 0.51, SE = 0.13, CI_{95} = 0.25, 0.76), $Q_b(1) = 8.25$, $p < .005$. However, these results should be interpreted with caution because of 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} = 0.75, 1.44) than the no-treatment control group studies (g = 0.51, SE = 0.13, CI_{95} = 0.25, 0.76), $Q_b(1) = 7.25$, $p = .007$. Although measures were taken to obtain equivalence for repeated measure designs (the standardized difference is computed, taking into account the correlation between measures; e.g., 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 experimental or quasiexperimental—did not influence effect sizes and that, in most cases, the nature of the control group also did not affect the effect size.

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 0.88, demonstrating, on average, a gain of nearly 1 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's (1986) meta-analysis of vocabulary instruction, which reported an average effect size of 0.97. 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 et al.'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.

On the other hand, the meta-analysis by Mol and her colleagues (2009) focused on similar objectives and similar age ranges to our meta-analysis and therefore warrants further explanation. They reported effect sizes ranging from $d = 0.54$ to $d = 0.57$, 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 $d = 0.59$ for parent–child dialogic reading. In both cases, more moderate effects were reported than our overall mean size.

This divergence in findings might be because of 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 elements additional 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 interventions, 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 vocabulary development. But the expansiveness of 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. Furthermore, there were dramatic variations across similarly termed interventions such as storybook reading. It was for this reason, some 10 years ago, that 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 its recent report, the National Early Literacy Panel (2008) took a different approach in its meta-analysis than we did in ours. Its goal was to isolate shared book reading to examine its effects. Contrary to Mol and her colleagues (2009), however, the committee restricted its selection to published book reading studies only and those not potentially confounded with any additional enhancements. Its 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. Furthermore, 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 (0.73).

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). It also 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 suggestions for interventions that appeared to work best. Clearly, among the most important factors is the person who delivers the instruction. Larger effect sizes occurred when the experimenter conducted the treatment, the most negligible when the intervention was given by the child care provider. Like Mol and her colleagues (2009), we found that child care providers seemed to have difficulty in enacting vocabulary training with pre-K and kindergarten children. It could be hypothesized 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. Previous studies of professional development (e.g., Neuman & Wright, in press), for example, have shown that early childhood providers need greater doses of training and supports to improve oral language in comparison to the more discrete skills such as letter knowledge.

A contextual factor that appeared less important in terms of its association with effect sizes was group configurations. Traditionally, early childhood educators have tended to favor small group and one-on-one interactions instead of whole group instruction (Bredekamp & Copple, 1997). Powell and his colleagues (2008), for example, in their study of the ecology of early learning settings found that whole group instruction appeared to support passive modes of child engagement: listening and watching rather than talking and acting.

However, we did not find support for the claim that whole group was less beneficial than small group instruction. In contrast, whole group vocabulary instruction had the largest effect sizes, although not statistically differentiable from the other configurations. Our results, therefore, confirm Mol and her colleagues' (2009) meta-analysis findings that demonstrated that children's oral language skills improved in whole group instruction. They suggest that certain foundational activities, such as vocabulary training, may be appropriately and perhaps more efficiently taught in whole group instruction settings, with smaller groups and individualized one-on-one instruction reserved for reviewing and practicing skills already introduced. These results could make a case for a more differentiated model of organizing instruction, one that more clearly aligns learning goals with the most promising organizational features.

Turning to the instructional features of these vocabulary interventions, our analysis revealed that pedagogical approach appeared to make a difference. Programs that used explicit instruction deliberately either through explanation of words or key examples were associated with larger effect sizes than those that taught words implicitly. In addition, programs that combined explicit and implicit instruction, enabling students to be introduced to words followed by meaningful practice and review, demonstrated even larger effects. These studies stand in contrast to those that used implicit instruction alone, which was found to be less effective. They confirm the recommendations of the National Reading Panel (2000), which called for providing direct instruction in vocabulary with multiple exposures in rich contexts. Given that this meta-analysis focused on many different training programs, these results should generalize across specific types of programs. Furthermore, 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).

Our analysis of the amount of exposure, however, did not reveal such clear-cut instructional recommendations. Importantly, we did not find support for Ramey and Ramey's (2006) conclusion that higher dosages of treatment lead to better effects. Longer, more intensive, and more frequent interventions did not yield larger effect sizes than smaller dosages. In fact, even brief doses of vocabulary intervention (e.g., Whitehurst et al., 1988) were associated with large effect sizes. Halle and her colleagues (in press) have argued that interventions narrower in scope may require only short-term training; those with a more global focus may require larger dosages.

The scope of the treatment may also relate to how these interventions were assessed. Like Elleman and her colleagues (2009), we found support for author-created measures being more sensitive in detecting improvements in language development than standardized measures. Yet at the same time we must be cautious in interpreting these results, especially in relation to vocabulary training. Given that many of the shorter interventions employed author-created measures, it was impossible to disentangle whether or not these results might have been because of interventions that were tied to a more discrete set of skills than others with a broader focus. Future research that systematically manipulates these factors could address this. These author-created measures probably reflected more proximal learning outcomes than the more distal measures of language development. Because 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, however, is not particularly surprising or newsworthy. Without standardized measures for confirmatory evidence, these author-created measures may provide an inflated portrait of the vocabulary gains made in studies. For example, our meta-analysis revealed moderate gains on standardized assessments ($g = 0.71$, *SE* = 0.07), essentially confirming vocabulary growth, although to a lesser extent. Standardized assessments, therefore, could reflect the most conservative end of the spectrum of vocabulary acquisition for young children, with the growth on author-created measures (*g* = 1.21, *SE* = 0.18) reflecting the other end of the spectrum.

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. This would suggest that multiple measures—author created and standardized—be used to provide strong evidence of the malleability of vocabulary use in different settings. Both types of measures, however, must be appropriate for the knowledge learned and the tasks to which the knowledge is to be applied—a high bar for many standardized assessments in vocabulary development (Pearson, Hiebert, & Kamil, 2007).

Finally, a primary goal of our research endeavor was to determine whether these interventions might work to close the gap in vocabulary development for young children. Although many of the interventions purported to be targeted to at-risk populations, it was clear that there are different degrees of risk. Examining the effects of coder-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, the vocabulary interventions were not sufficiently powerful to close the gap—even in the preschool and kindergarten years. It also highlights the effects that poverty may have on children's language development (Neuman, 2008). Extensive early intervention starting at birth through age 6 as in such projects as the Abecedarian program (Campbell & Ramey, 1995) may be needed to help ameliorate these differences.

Taken together, our meta-analysis provides some promising recommendations for classroom settings. However, these moderator analyses should not be interpreted as testing causal relationships (Cooper, 1998; Viechtbauer, 2007). Rather, our 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 generalize only 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, the number of words taught, the rationale for the selection of words, or the relationship of the intervention to the existing curriculum. Researchers need to make the decisions about their choices of vocabulary more transparent in the future; similarly, editors and reviewers of peer-reviewed journals must become more diligent in requiring such information. In addition, in some cases, 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.

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. Furthermore, the word selection in vocabulary training may influence the type of instruction (e.g., explicit and implicit). For example, there are words that easily can 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 introduced only 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. Furthermore, most of the storybook reading interventions did not detail the names or genres of the books, another potential confound.

We wish that more studies had conducted delayed posttests to examine the sustainability of treatment. Our sample included only 11 studies with delayed posttests. Although our results are 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, more than 50% were found to be largely heterogeneous (Engel, Schmid, Terrin, Olkin, & Lau, 2000). Because of 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-analysis 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. Furthermore, 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 is present, 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 as at risk are gaining substantially more than their at-risk peers living in poverty, 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.

Notes

This research was partially supported by the Ready to Learn Grant funded by the Corporation for Public Broadcasting/Public Broadcasting System through the Office of Innovation and Improvement, U.S. Department of Education. We thank the Ready to Learn project members and staff for their important contributions to this study and gratefully acknowledge Adriana Bus and Suzanne Mol for their helpful comments on an earlier draft of our article. We wish to thank anonymous reviewers for their valuable suggestions that greatly improved the article. Finally, we would like to thank the editor, who provided valuable insights and remarks.

[1] The I^2 value reflects the proportion of the total variability across studies because of heterogeneity, rather than chance, that could be explained through moderator analyses. When I^2 is zero as in this subgroup, all of the observed variance among studies is spurious and does not reflect real differences in effect sizes; there is nothing left to explain with further moderator analyses.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Ammon, P. R., & Ammon, M. S. (1971). Effects of training Black preschool children in vocabulary versus sentence construction. *Journal of Educational Psychology*, *62*, 421–426.

Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator*, *16*, 14–18, 44–47.
*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*, 235–243.
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*, 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. 2, pp. 41–51). New York, NY: 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, Columbia.
Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 2). Englewood, NJ: Biostat.
Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
*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, Vermillion.
*Brackenbury, T. P. (2001). *Quick incidental learning of manner-of-motion verbs in young children: Acquisition and representation* (Unpublished doctoral dissertation). University of Kansas, Lawrence.
Bredekamp, S., & Copple, C. (1997). *Developmentally appropriate practice–revised*. Washington, DC: National Association for the Education of Young Children.
*Brickman, S. O. (2002). *Effects of a joint book reading strategy on Even Start* (Unpublished doctoral dissertation). University of Oklahoma, Norman.
*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, Auburn, AL.
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.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage.
Cooper, H., & Hedges, L. (Eds.). (1994). *The handbook of research synthesis*. New York, NY: Russell Sage.
Cooper, H., Hedges, L., & Valentine, J. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage.

*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*, 74–88.

*Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & 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*, 145–162.

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

*Cronan, T. A., Cruz, S. G., 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*, 251–272.

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, Denton.

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

*Daniels, M. (1994b). 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*, 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.

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*, 605–619.

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.

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

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*, 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.

*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). University of Alabama, Tuscaloosa.

Halle, T., Zaslow, M., Tout, R., Starr, J., & McSwiggan, M. (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., & Sénéchal, 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*, 75–90.

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

Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap. *American Educator*, *27*, 4–9.

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

*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, Philadelphia, PA.

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*, 513–535.

*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*, 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. *Early Education and Development*, *1*, 105–114.

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). Florida State University, Tallahassee.

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*, 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. & Sénéchal, M. (2009). *Discussing Stories: On How a Dialogic Reading Intervention Improves Kindergartners' Oral Narrative Construction*. Manuscript submitted for publication.

*Levinson, J. L., & Lalor, I. (1989, March). *Computer-assisted writing/reading instruction of young children: A 2-year evaluation of "writing to read."* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from ERIC database (ED310390).

*Light, J., Drager, K., McCarthy, J., Mellott, S., Millar, D., Parrish, C., . . . Welliver, M. (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*, 63–88.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from a meta-analysis. *American Psychologist*, *48*, 1181–1209.

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

*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, Storrs.

*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*, 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*, 263–290.

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

*McConnell, B. B. (1982). *Bilingual education: Will the benefits last?* (Bilingual Education Paper Series 5(8)). Los Angeles: California State University, Evaluation, Dissemination and Assessment Center.

*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., & . . . Napier, C. (2001). The impact of a clinic-based literacy intervention on language development in inner-city preschool children. *Pediatrics*, *107*, 130–134.

Moats, L. (1999). *Teaching reading IS rocket science*. Washington, DC: 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*, 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). University of Nebraska-Lincoln.

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

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

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

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

Neuman, S. B. (Ed.). (2008). *Educating the other America*. Baltimore, MD: Brookes.
Neuman, S. B. (2009). *Changing the odds for children at risk*. New York, NY: Teachers College Press.
*Neuman, S. B., & Gallagher, P. (1994). Joining together in literacy learning: Teenage mothers and children. *Reading Research Quarterly*, *29*, 382–401.
Neuman, S. B., & Wright, T. (in press). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *Elementary School Journal*.
*Notari-Syverson, A., O'Connor, R., & Vadasy, P. (1996, April). *Facilitating language and literacy development in preschool children: To each according to their needs*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. Retrieved July 29, 2009 from ERIC database. (ED395692)
Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics*, *8*, 157–159.
Paris, S. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*, 184–202.
Pearson, P. D., Hiebert, E., & Kamil, M. (2007). Theory and research into practice: Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*, 282–296.
*Peta, E. J. (1973). *The effectiveness of a total environment room on an early reading program for culturally different pre-school children* (Unpublished doctoral dissertation). Lehigh University, Bethlehem, PA.
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* (Unpublished doctoral dissertation). University of Mississippi, Oxford.
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. 2, 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, NY: Russell Sage.
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.
Roskos, K., Ergul, C., Bryan, T., Burstein, K., Christie, J., & Han, M. (2008). Who's learning what words and how fast? Preschoolers' vocabulary growth in an early literacy program. *Journal of Research in Childhood Education*, *22*, 275–290.
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, NY: Guilford.
*Schetz, K. F. (1994). An examination of software used with enhancement for preschool discourse skill improvement. *Journal of Educational Computing Research*, *11*, 51–71.
*Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, *24*, 123–138.

Sénéchal, M., & Cornell, E. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, *28*, 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*, 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, NY: Russell Sage.

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

*Silverman, R. (2007b). Vocabulary development of English-language and English-only learners in kindergarten. *Elementary School Journal*, *107*, 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, Eugene.

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

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

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: Lawrence 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*, 779–787.

Viechtbauer, W. (2007). Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *60*, 29–60.

*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*, 273–278.

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

*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*, 679–689.

*Whitehurst, G. J., Falco, F. L., Lonigan, C., Fischel, J. E., Valdez-Menchaca, M. C., DeBaryshe, B. D., & Caulfield, M. (1988). Accelerating language development through picture-book reading. *Developmental Psychology*, *24*, 552–558.

Wilson, D., & Lipsey, M. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, *6*, 413–429.

Wood, W., & Eagly, A. (2009). Advantages of certainty and uncertainty. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 456–471). New York, NY: Russell Sage.

Authors

LOREN M. MARULIS is a doctoral student in the Combined Program in Education and Psychology at the University of Michigan, SEB 1400C, 610 E. University, Ann Arbor, MI 48109; e-mail: marulisl@umich.edu. Her research interests focus on the malleability of cognitive development. She serves as a research assistant in the Michigan Research Program on Ready to Learn. She has previously been a research associate with Larry Hedges and specializes in meta-analysis.

SUSAN B. NEUMAN is a professor in educational studies specializing in early literacy development at the University of Michigan, 1905 Scottwood, Ann Arbor, MI 48104; e-mail: sbneuman@umich.edu. She previously served as the U.S. Assistant Secretary for Elementary and Secondary Education. In her role as Assistant Secretary, she established the Early Reading First program, developed the Early Childhood Educator Professional Development Program, and was responsible for all activities in Title I of the Elementary and Secondary Act. She has directed the Center for the Improvement of Early Reading Achievement and currently directs the Michigan Research Program on Ready to Learn. Her research and teaching interests include early childhood policy, curriculum, and early reading instruction, pre-K to third grade, for children who live in poverty.