To read or not to read

Mol, S.E.

1

General Introduction

Parts of this chapter are based on:
Bus, A. G., Van IJzendoorn, M. H., & Mol, S. E. (in press). Meta-Analysis. In N. K. Duke & M. H. Mallette (Eds.), *Literacy research methods, volume 2*. New York: Guilford Press.

Entering search terms as "storybook reading" or "shared book reading" into electronic databases such as PsycInfo, ERIC, and ProQuest result in more than 300 hits of peer-reviewed journal articles, book chapters, and dissertations. At least another 100 studies focus on "leisure time reading" or "recreational reading" in conventionally reading students. In one way or another, this ever increasing set of correlational, longitudinal, and experimental studies tries to operationalize, prove, and/or nuance the appealing assumption that reading (story)books has a long-lasting impact on our language and reading abilities as well as our academic success. Do quantitative integrations of the research base thus far corroborate with the general belief in society that reading is as a miracle drug for the prevention and treatment of reading problems?

Meta-analysis is the empirical analysis of empirical studies – that is, the quantitative analysis and synthesis of a set of related empirical studies in a well-defined domain (Cooper, Hedges, & Valentine, 2009). The three meta-analyses in this thesis comprise 146 studies (N = 10,308) that address the role of book reading in language and reading development from infancy to early adulthood. For the group of pre-conventional readers, we also examine the effect of interventions that improve the quality of shared book reading such as dialogic reading programs (e.g., parents are trained to ask questions about words and story events; Whitehurst et al., 1988). Hereafter, I first introduce the importance of developing a book reading routine. Second, I elaborate on the meta-analytic approach we applied. In the third and final section, I present the aims and outline of this thesis.

The Quantity and Quality of Book Reading

We view parent-child book sharing as part of a continuum of leisure-time reading experiences that facilitate and influence language and reading skills throughout development (see chapter 2). Developing a book reading routine before the age of two may set in motion a causal spiral, in which language skills develop as a result of shared book reading and in which children's vocabulary size determines whether they comprehend storybooks and whether they enjoy being read to (Fletcher & Reese, 2005). For conventional readers, this spiral continues to determine their reading behavior: Reading books is seen as both a consequence of reading ability and a contributor to further reading development (Stanovich, 2000). More skilled readers are more likely to choose to read more frequently which, in turn, will improve their knowledge of word forms and semantics, and enhance their vocabulary size and text comprehension abilities. In contrast, readers with small vocabularies or word reading difficulties may not succeed in comprehending text, become less eager to read, and as a result, show stagnation in their reading development (Kush, Watkins, & Brookhart, 2005). Because of growing individual differences in leisure-time reading activities, we expect that the relationship between book reading and reading skills will strengthen from infancy to early adulthood as frequent readers will be more motivated and better skilled readers than infrequent book readers.

Young children need caregivers who bridge the gap between the world of the book and their own world so reading storybooks is not only a source of entertainment but also is a means to get familiar with the structure and syntax of written language and learn about the purpose and function of reading (Bus, 2003; DeTemple & Snow, 2003; Heath, 1982; Scheele, Leseman, & Mayo, 2010; Sulzby, 1985; Watson, 2001). The richness of children's learning experiences is thought to depend not only on the frequency of book sharing but also on the quality of reading sessions and parents' sensitivity towards children's cognitive abilities and interests: More new words are learned when young children are actively involved in storybook reading, for instance when questions are asked about pictures, difficult words, and story events, and informative feedback is provided on children's answers (e.g., Whitehurst et al., 1988). Because observational studies have suggested that most parents – and especially parents in low socioeconomic status groups – do not apply such interactive reading techniques spontaneously (e.g., Britto, Brooks-Gunn, & Griffin, 2006; Bus & Van IJzendoorn, 1995; Heath, 1982; Laakso, Poikkeus, & Lyytinen, 1999; Ninio, 1980; Silvén, Ahtola, & Niemi, 2003), attempts are made to train caregivers in techniques such as "Dialogic Reading" that may enhance their quality of reading interactions. We integrate effects of training studies to estimate overall effects of such interventions as well as effects of age and risk status (see chapter 3).

At the start of formal schooling, children from low socioeconomic status homes are more likely to lag behind in their language and reading skills than their peers with a middle- or high socioeconomic status (Dickinson & McCabe, 2001; Hart & Risley, 2003). As such an achievement gap is likely to widen in the course of primary school (Alexander, Entwisle, & Olson, 2007; Foster & Miller, 2003), it seems especially important for children who are at risk for language and literacy impairments to have frequent encounters with storybooks from an early age onwards. However, because children with a low socioeconomic status are more likely to grow up in home environments that are less stimulating (e.g., parents read less frequently, own few (children's) books) than the home environments of their peers from higher socioeconomic backgrounds (Christian, Morrison, & Bryant, 1998; Morrow, 1983; Van Steensel, 2006), the quantity and quality of storybook reading in their preschool and kindergarten classrooms is deemed essential for enhancing the language and reading skills they need to benefit from formal reading instruction in primary school. This has resulted in interactive reading techniques as a means for teachers to attract the attention of at-risk children, to stimulate them to use book-related language, and to check their story understanding. Chapter 4 presents an integration of studies that test effects of interactive reading techniques applied in classroom settings.

Steps in a Meta-Analysis

Meta-analysis can be applied most fruitfully within research programs in which studies with similar designs or measures accumulate over the years. In primary studies, data are collected to test a hypothesis derived from a well-articulated theory; the hypothesis often will be stated in the form: variable *X* is associated with variable *Y*, or *X* is causally related to *Y*. In correlational or experimental designs, measures prototypical to assess *X* and *Y* are being used, and the results are therefore comparable across studies. If the number of replications increases, and if characteristics of replication studies vary, the meta-analytic approach is feasible to synthesize the literature and to test the effects of variations in study characteristics on the outcome of the studies.

A common and defining characteristic of all meta-analytic approaches is the use of a specific set of statistical methods compared to the methods used in primary research. The reason is simple: In primary research the unit of analysis is the individual participant (or class, or other group), whereas the unit of meta-analysis is the study result. Study results are usually based on different numbers of participants, and they are, therefore, point estimates with different precision and confidence boundaries (Mullen, 1989). It would be incorrect to give a significant correlation of .30 in a sample of 50 participants (confidence interval: .02, .53) the same weight as a correlation of .30 in a sample of 500 participants (confidence interval: .22, .38). Basically, however, meta-analytic research follows the same steps and standards as empirical research (Cooper, 1982).

Step 1: Hypothesis Formulation

The meta-analysis should start with the formulation of a specific, theoretically relevant conceptual framework. Its domain should be clearly defined, and the central meta-analytic question should be theoretically derived and meaningful. When a meta-analyst is not sensitive to such substantive issues, a meta-analysis can become a pointless, merely statistical exercise (Littell, Corcoran, & Pillai, 2008). For example, when synthesizing the effects of interventions on struggling readers' reading comprehension (see Edmonds et al., 2009), the validity and/or practical use of the summary effects can be questioned when interventions with a focus on fluency, decoding, comprehension, and multiple components are heaped together. That is, even though the dependent measure is comparable and the target groups are similar, it is difficult to disentangle the kind of intervention that might support the comprehension skills of children with reading disabilities when the content of interventions varies extensively.

Step 2: Retrieval and Coding of Studies

In the next stage, the meta-analyst should systematically collect the relevant published as well as unpublished literature from at least three different sources.

The "snowball" method (using references lists from key articles in the field), the "invisible college" approach (using key figures in the field to collect recent or unpublished materials), and computer searches of subject indexes such as ERIC, PsycInfo, Medline, Proquest UMI Dissertations, and Google Scholar, or citation indexes such as SSCI or SCI may be used in a multimethod combination.

In some meta-analytic approaches, selection of studies is based on the idea that only randomized experimental designs produce valid findings to be taken seriously. "What Works Clearinghouse" (WWC; see http://ies.ed.gov/ncee/wwc/), that is set up to support educators and the U.S. Department of Education in making evidence-based recommendations about the effectiveness of programs, policies, and practices in a wide range of areas, uses eligibility screens in which randomization, level of attrition, and equivalence of treatment and control groups are taken into account to select studies that meet the evidence standards fully or with reservations (WWC, 2008). The National Reading Panel (NRP) also objects to inclusion of all studies regardless of design features. Restrictions of the type of papers to be included, however, may imply an untenable reduction of the available evidence. For instance, the NRP discards the many correlational investigations in the area of reading research (NRP, 2000; Williams, 2001) which means a loss of potentially important information. In this respect, the meta-analytic method is basically indifferent: The central hypothesis should decide about the feasibility of selection criteria, and when this hypothesis is not stated in strictly causal terms there is no reason to leave correlational studies aside. Furthermore, the impact of the quality of research on effect sizes can be examined by testing whether the overall effect is influenced by the presence of studies with other designs than randomized controlled trials (Rosenthal, 1995).

The exhaustive search for pertinent literature is preferred compared with the best evidence approach (Slavin, 1986), in which only the qualitatively sound studies would be allowed to enter a meta-analysis. Because of their emphasis on explanation of variability in effect sizes, in recent meta-analytic approaches it is preferred to test whether quality of research (which always is a matter of degree and a matter of different strengths and weaknesses) explains variation in study results in order to make the process of study selection and evaluation transparent and to maximize the power of the analyses. We, for instance, created a scale to score whether the researchers checked the use of trained techniques in the experimental group, the quality of reading sessions within the control group, and the actual frequency of book reading in the experimental and the control groups (see chapter 4). Experimental designs outperformed quasi-experiments on the scale, but intervention outcomes were not affected by the experiment fidelity score nor did quasi-experiments reveal higher effect sizes than true experiments for children's language and literacy measures.

The basic problem to be faced in the second stage of the meta-analysis is the "file-drawer" problem (Rosenthal, 1991). Primary researchers know that it is easier to get papers published in which they report significant results than to guide papers into print with null results; regardless of the quality of the study (Begg, 1994). This publication bias may even lead to the unfortunate situation that the majority of papers remain in the file drawers of disappointed researchers, whereas only a minority of papers with significant results are published (Cohen, 1990). Average or combined effect sizes of published papers may, therefore, present an inflated picture of the real state of the art in a specific field. The number of unpublished papers with null findings that are needed to make the meta-analytic outcome insignificant can be estimated (the "fail-safe number"; Rosenthal, 1991). A publication bias can be visually inspected by a funnel plot, which is a scatter plot of the effect size against sample size that will be skewed and show asymmetry (i.e., due to a lack of small effect sizes) when a publication bias is present (Lipsey & Wilson, 2001). The "file-drawer" problem may suggest that *a priori* selection of only published papers is not always warranted. Although published studies have been subjected to more or less thorough reviewing procedures and therefore seem to carry more quality weight than unpublished studies, the reasons for remaining unpublished may be unrelated to quality. In many cases, it is, therefore, better to collect all studies regardless of origin or status, and to analyze *post hoc* whether publication status makes a difference in combining effect sizes. To assess the likely impact of a publication bias, the "trim and fill" method can be used to estimate the unbiased effect size, by estimating the number of missing studies from an asymmetrical funnel plot and calculating an adjusted point estimate and variance (Duval & Tweedie, 2000a, 2000b).

The retrieved articles, dissertations, and unpublished documents are considered to be the raw data to which a coding system is applied to produce the variables to be used in the meta-analysis. The application of the coding system should be tested for intercoder reliability. The coding system contains potential moderator variables that can be used to explain the variability of the effect sizes in the specific set of studies. The variables in the coding system should therefore be theoretically relevant and constitute pertinent moderator hypotheses. In view of the relatively small number of studies included in most meta-analyses, the coding system should not be too extended. If potential moderators exceed the number of studies, inflated meta-analytic outcomes may be the nonreplicable result. On the other hand, if the number of studies per moderator is too small, the power will be too low to detect meaningful differences in effects across subgroups (Hedges & Pigott, 2004). We suggest that a minimum of four (Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003) or five (Seidel & Shavelson, 2007) studies per subgroup is needed to reliably interpret any contrast.

Studies may report effects on several dependent measures for similar outcome measures. To avoid a situation in which studies with more results have a greater impact, the effect sizes should be aggregated within studies and domains. For example, when a study reports outcomes for one receptive and two expressive vocabulary measures, the expressive outcomes are aggregated in a first step (domain) and combined with the effect size of the single receptive measure in a second step in order to calculate a vocabulary composite per study. Separate meta-analyses can be conducted to examine differential treatment effects per outcome measure (e.g., Does dialogic reading affect expressive vocabulary more strongly than receptive vocabulary skills?; see chapter 3) as long as each study contributes one effect size to each analysis. Creating a rather broad composite such as "academic achievement", in which a variety of reading, mathematical, and/or grade-related measures are aggregated, might limit the interpretation of specific treatment effects. Another complicating factor may be that experiments include two or more interventions but only one control group. Effect sizes are dependent if the same control group is used to calculate the effect sizes for each treatment (e.g., Ehri, Nunes, Willows, et al., 2001). Gleser and Olkin (1994) state that in multiple-treatment studies, "the treatments may all be regarded as instances or aspects of a common treatment construct." Furthermore they state that "there is strong reason a priori to believe that a composite effect size of treatment obtained by combining the end point effect sizes would adequately summarize the effect of treatment" (p. 351). Another, more pragmatic solution of the multiple-interventions problem is to divide the sample size of the control group in the same number of subgroups as there are interventions in order to avoid the situation in which control subjects count for more than one unit of analysis.

Step 3: Analysis of Study Results and Characteristics

Data analysis often consists of three steps (Mullen, 1989): First, the central tendency of the study results is computed (i.e., the combined effect size). Because *p* values heavily depend on the number of observations, recent meta-analyses focus on the combined standardized differences between the means of the experimental and the control groups. The statistic used to assess the effectiveness of a treatment or other variable is the effect size, *d*, which measures how much the mean of the treatment group exceeds the mean of the control group in standard deviation units. Effect size expresses how many standard deviation units treatment groups differ from control groups without treatment. An effect size of 1.0 indicates that the treatment group mean is one standard deviation higher than the control group mean, whereas an effect size of 0 indicates that treatment and control group means are identical. A mean effect, of which the precision is addressed by the 95% confidence interval (CI), is considered significant if the CI does not include zero. Differences between estimates can be interpreted as significant when the CIs do not overlap. According to Cohen (1988), an effect size of $d = .20$ is considered small, an effect size of $d = .50$ moderate, and an effect size of $d = .80$ or above large ($r = .10$ is small, $r = .30$ is moderate, $r = .50$ is large). Translated into percentiles, $d = .20$ indicates that the treatment has moved the average child from the 50th to the 58th percentile; $d = .50$ indicates that the treatment has moved the child, on average, to the 69th percentile; $d = .80$ indicates that the treatment has moved the child, on average, to the 79th percentile. As an alternative, Rosenthal and Rubin (1982) suggested the binominal effect size display (BESD), which indicates the change in predictive accuracy attributable to the relationship in question and is computed from the formula $.50 \pm (r/2)$. The BESD shows the extent to which prediction is enhanced (i.e., the percentage increase in prediction) with the use of intervention *X* to predict reading skill *Y* (for details, see step 4).

A weighted effect size is mostly used to adjust for the bias resulting from small sample sizes (i.e., the tendency of studies with small samples to overestimate effects). Unweighted *d*'s are sometimes presented to provide information about the direction of biases related to sample size. The effect size can be computed on the basis of the standard deviations of the control group (Glass, 1976), the pooled standard deviations (Rosenthal, 1991), or the pooled variance (Hedges & Olkin, 1985). Cohen's *d*, for instance, is calculated as the difference between control and experimental treatment posttest mean scores (partialed for the influence of pretest scores if information is available) divided by the pooled standard deviation. Alternatively, the test statistics (F, t, χ^2) can be transformed into an effect size (Rosenthal, 1991). In practice, different strategies do not seem to make a substantial difference (Johnson, Mullen, & Salas, 1995).

Second, the variability of the results around this central tendency is assessed, and outliers as well as homogeneous subsets of studies are identified. To determine whether a set of *d*'s shares a common effect size, a homogeneity statistic (Q) which approximates chi-square distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes, can be computed. Homogeneity analysis compares the amount of variance exhibited by a set of effect sizes with the amount of variance expected if only sampling error is operating. *I*-squared (I^2), another indicator of homogeneity, describes the impact of heterogeneity on a meta-analysis by measuring the degree of inconsistency between studies. Values that exceed 70% should invite caution about the homogeneity of the mean effect (Petticrew & Roberts, 2006). If sets of study results remain heterogeneous, combined effect size computed on the basis of the fixed model may be biased estimates, that is, it cannot be concluded that they are a sample from the same population, and a random model should be preferred (Hedges, 1994). In the random-effects model, studies are also weighted by the inverse of its variance, but, in addition, it accounts for within-study error as well as between-study variation in true effects (Borenstein, Hedges, Higgins, & Rothstein, 2009). If a distribution of study results is extremely skewed and shows several outlying values, the average effect size does not adequately represent the central tendency. Inflated meta-analytic findings may result from ignoring heterogeneity in study outcomes, and the (more conservative) random model may lead to lower estimates for the combined effect size as well as larger confidence boundaries (Hedges, 1994).

Third, through a moderator analysis, the meta-analysts try to explain the variability on the basis of study characteristics. A significant chi-square indicates that the study features significantly moderate the magnitude of effect sizes. For example, intervention studies with randomized designs may, on average, yield smaller effects than those without randomization. Mostly the analyses do not include tests of interactions between moderator variables because the number of comparisons is insufficient in many cases. It should be noted that in meta-analytic as well as in primary studies every subject or sample should be counted independently from each other and only once. That is, if a study presents more than one effect size for the same hypothesis, these effect sizes should be combined within the study before it is included in the overall meta-analysis.

Step 4: Interpretation of Meta-Analytic Outcomes

The interpretation of the size of the combined effects is a matter of much debate (McCartney & Rosenthal, 2000). In a meta-analysis, Van IJzendoorn and Bus (1994) showed that a powerful explanation of dyslexia, the phonological deficit hypothesis, explains only 6% of the variance in dyslexia (d = .48) which is about half a standard deviation difference between the experimental and the control groups. Bus, van IJzendoorn, and Pellegrini (1995) showed that the association between preschool storybook sharing and later literacy was even stronger (d = .59) explaining about 8% of the variance in children's literacy skills. A correlation of .28 between book sharing and reading may seem a rather modest outcome. However, in terms of the BESD (Rosenthal, 1991), this effect is sizable. The BESD is defined as the change in success ratio because of an intervention. The BESD shows the extent to which prediction is enhanced (i.e., the percentage increase in prediction) with the use of intervention X to predict reading skill Y. If we equal the combined effect size d = .59 with an r = .28, the success ratio in the experimental group would be: .50 + (.28/2) = .64; the success ratio in the control group would be .50 – (.28/2) = .36. It should be noted, therefore, that it certainly can make a tremendous difference in the lives of young children whether or not they are read to by their parent. The difference between the experimental and the control groups would amount to a substantial difference if we translate this outcome to the millions of children who may profit from book reading (Rosenthal, 1991). Taking into account that experimental studies revealed outcomes similar to correlational/longitudinal/retrospective studies, this meta-analysis provides a clear and affirmative answer to the question of whether or not storybook reading is one of the most important activities for developing the knowledge required for eventual success in reading. Therefore, parental storybook reading should be recommended because in terms of BESD it makes a difference for many thousands of preschoolers. In the same vein, phonological deficit is correctly considered as a main cause of dyslexia.

Aims and Outline Thesis

A previous meta-analysis (Bus et al., 1995), comprising 33 studies between 1951 and 1993, showed that reading storybooks to pre-conventional reading children explained about 10-12% of children's oral language skills (i.e., passively comprehending words and/or actively producing words) and 8% of the variance in children's basic reading skills such as knowledge of letter names, how letters relate to sounds in spoken words, and how to write your own name. The meta-analyses in this thesis do not only include a more recent set of studies than in Bus et al. (1995), but also extend the age range from infancy to early adulthood and examine more closely the impact of qualitative aspects of shared book reading. The latter meta-analyses are a critical test of early interventions that are designed to improve pre-conventional readers' literacy experiences at home and at school. Among the questions that guided our attempt to synthesize the available literature were:

a) Does exposure to (story)books affect language proficiency and does it get even stronger in adulthood?
b) Does book sharing already stimulate technical reading skills in pre-conventional readers and do these skills improve beyond the earliest stages of reading acquisition?
c) Do students' ability to spell words correctly depend on exposure to print?
d) Is it advisable to promote reading outside school in low-ability readers?
e) Does a reading habit enhance intellectual abilities and later success in society?
f) Do young children whose caregivers (i.e., parents, teachers) ask questions about words and story events during shared book reading learn more words from storybooks than children who are only read the story text?

In **chapter 2**, we relate leisure-time reading activities of (1) preschoolers and kindergartners, (2) children in grades 1 to 12, and (3) college and university students to indicators of comprehension and technical reading and spelling skills. We examine (a) whether the pattern of associations gets stronger across the age span from early childhood to early adulthood, and (b) the extent to which low-ability readers benefit from independent book reading.

In **chapter 3**, we meta-analyze the effects of Dialogic Reading interventions for parents of 2- to 6-year-old children. We expected that children's expressive vocabulary skills (i.e., producing words) are especially affected, because those skills are particularly emphasized within the Dialogic Reading format. Furthermore, we tested (a) whether children at risk for language and literacy impairments (e.g., due to a low socioeconomic status) benefited less from the intervention, because their parents may be less responsive to training in book-sharing skills, and (b) whether older children with more linguistic skills were affected less than younger children, because older children may be less dependent on book reading quality.

In **chapter 4**, we quantitatively summarize (quasi-)experiments that examined the effects of interactive reading in preschool and kindergarten classrooms. We expected that (a) children's oral language skills would improve as a result of the intervention, and (b) basic reading skills would be especially affected in older, kindergarten children who may be capable of interacting with the adult and simultaneously process other features of the printed text. We also examined whether intervention success was moderated by children's risk status, the size of the interactive reading groups, and whether the implementer was a researcher or children's own teacher.

In the **chapter 5**, a meta-analysis of meta-analyses will be presented and related to questions for future research.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72,* 167-180.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129,* 195-215.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 399-409). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, U.K.: Wiley.

Britto, P. R., Brooks-Gunn, J., & Griffin, T. M. (2006). Maternal reading and teaching patterns: Associations with school readiness in low-income African American families. *Reading Research Quarterly, 41,* 68-89.

Bus, A. G. (2003). Social-emotional requisites for learning to read. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 3-15). Mahwah, NJ: Lawrence Erlbaum Associates.

Bus, A. G., & van IJzendoorn, M. H. (1995). Mothers reading to their three-year-olds: The role of mother-child attachment security in becoming literate. *Reading Research Quarterly, 30*, 998-1015.

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

Christian, K., Morrison, F. J., & Bryant, F. B. (1998). Predicting kindergarten academic skills: Interactions among child care, maternal education, and family literacy environments. *Early Childhood Research Quarterly, 13,* 501-521.

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

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52*, 291–302.

DeTemple, J., & Snow, C.E. (2003). Learning words from books. In A. van Kleeck, S. A. Stahl & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 16-36). Mahwah, NJ: Lawrence Erlbaum.

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice, 16*, 186-202.

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

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

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K. et al. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79,* 262-300.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287.

Fletcher, K. L., & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review, 25,* 64-103.

Foster, W.A., & Miller, M. (2007). Development of the literacy achievement gap: A longitudinal study of kindergarten through third grade. *Language, Speech, and Hearing Services in Schools, 38,* 173-181.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Research, 5*, 3–8.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339–355). New York: Russell Sage.
Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap. *American Educator, 27,* 4-9.
Heath, S. B. (1982). What no bedtime story means: Narrative skills at home and school. *Language in Society, 11,* 49-76.
Hedges, L. V. (1994). Statistical considerations. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29–38). New York: Russell Sage.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.
Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426-445.
Johnson, B. T., Mullen, B., & Salas, E. (1995). Comparison of three meta-analytic approaches. *Journal of Applied Psychology*, *80*, 94–106.
Kush, J. C., Watkins, M. W., & Brookhart, S. M. (2005). The temporal-interactive influence of reading achievement and reading attitude. *Educational Research and Evaluation, 11,* 29-44.
Laakso, M. L., Poikkeus, A. M., &Lyytinen, H. (1999). Shared reading interaction in families with and without genetic risk for dyslexia: Implications for toddlers' language development. *Infant and Child Development, 8,* 179-195.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
Littell, J. H., Corcoran, J., & Pillai, V. K. (2008). *Systematic reviews and meta-analysis.* New York: Oxford University Press.
McCartney, K., & Rosenthal, R. (2003). Effect Size, practical importance, and social policy for children. *Child Development, 71,* 173–180.
Morrow, L. M. (1983). Home and school correlates of early interest in literature. *Journal of Educational Research, 76,* 221-230.
Mullen, B. (1989). *Advanced basic meta-analysis*. Hillsdale, NJ: Erlbaum.
National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Rockville, MD: NICHD Clearinghouse.
Ninio, A. (1980). Picture-book reading in mother-infant dyads belonging to two subgroups in Israel. *Child Development, 51,* 587-590.
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences; A practical guide.* Oxford: Blackwell Publishing.
Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183-192.
Rosenthal, R., & Rubin, D. B. (1982). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology, 74*, 708–712.
Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics, 31*, 117-140.
Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77,* 454-499.
Silvén, M., Ahtola, A., & Niemi, P. (2002). Early words, multiword utterances and maternal reading strategies as predictors of mastering word inflections in Finnish. *Journal of Child Language, 30,* 253-279.
Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*, 5–11.
Stanovich, K. E. (2000). *Progress in understanding reading; Scientific foundations and new frontiers.* New York: The Guilford Press.

Sulzby, E. (1985). Children's emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly, 20,* 458-481.
Van IJzendoorn, M. H., & Bus, A. G. (1994). Meta-analytic confirmation of the nonword reading deficit in developmental dyslexia. *Reading Research Quarterly, 29,* 266–275.
Van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading, 29*, 367-382.
Watson, R. (2001). Literacy and oral language: Implications for early literacy acquisition. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 43-53). New York: Guilford Press.
What Works Clearinghouse (2008, December). *Procedures and standards handbook (version 2.0).* Retrieved November 28, 2009, from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf
Whitehurst, G. J., Falco, F. L., Lonigan, C. J., & Fischel, J. E. et al. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24,* 552-559. Doi: 10.1037/0012-1649.24.4.552
Williams, J. (2001). Commentary: Four meta-analyses and some general observations. *The Elementary School Journal, 101,* 349–354.