



Review

Book reading and vocabulary development: A systematic review



Barbara A. Wasik*, Annemarie H. Hindman, Emily K. Snell

Temple University, United States

ARTICLE INFO

Article history:

Received 20 September 2014

Received in revised form 22 March 2016

Accepted 19 April 2016

Available online 20 July 2016

Keywords:

Vocabulary

Book reading

ABSTRACT

This paper reviews high-quality empirical studies on book reading practices in early childhood that have resulted in increases in child vocabulary. The overarching purpose of this work is twofold: first, to tease apart the myriad ways in which effective book readings can be delivered; and second, to identify questions that remain about book reading and vocabulary learning. We examine various aspects of effective book readings, including the contexts in which the book reading was conducted, the words that were taught through the book reading, the dosage of reading that children received, and the outcome measures used. Findings reveal that six strategies—reading and re-reading texts, explicitly defining words, encouraging dialogue about book-related vocabulary through questions and discussion, re-telling, using props, and engaging children in post-reading activities—are consistently implemented across the studies; however, they are used in widely varying combinations. There is great variability across studies in the number of words taught, the criteria for word selection, and the measures used to assess word learning. Moreover, in many studies, children learn only a small proportion of the number of words taught. Finally, this review identifies critical remaining questions about how to optimize vocabulary learning through book reading that require systematic investigation in order to inform effective practice.

© 2016 Elsevier Inc. All rights reserved.

Contents

1. Introduction.....	40
1.1. Vocabulary development and reading.....	40
1.2. Theoretical and empirical approach.....	40
1.3. Book reading as a context for vocabulary development.....	40
1.4. Aim of the current review.....	41
2. Method.....	41
2.1. Literature search strategy and selection criteria.....	42
2.1.1. Downs and Black criteria.....	42
3. Results.....	44
3.1. Quality of the studies reviewed.....	44
3.2. Context of book reading.....	44
3.3. Words during book reading.....	44
3.3.1. Number of words.....	44
3.3.2. Number of exposures to words.....	44
3.3.3. Criteria for word selection.....	45
3.3.4. Grammatical class of words.....	45
3.4. Study dosage.....	45
3.5. Word learning strategies implemented in the studies.....	46
3.6. Outcome measures and outcomes.....	46
3.6.1. Project-specific measures.....	46
3.6.2. Standardized measures and outcomes.....	52

* Corresponding author.

E-mail address: bwasik@temple.edu (B.A. Wasik).

3.6.3.	Delayed post-testing	52
4.	Discussion	52
4.1.	Adult-child interaction during book reading boosts word learning	52
4.2.	Effects of interventions featuring book reading are modest	53
4.2.1.	Measuring words taught	53
4.2.2.	Standardized vocabulary growth	53
4.3.	Wide variation in the quality of the studies	54
4.4.	Wide variation in the number of words targeted via book reading	54
4.5.	Nature (part of speech, morphology) of target words remains unclear	54
4.6.	Frequency of word exposures and dosage of strategies are not clear	54
4.7.	Most research has focused on classrooms	55
5.	Limitations	55
6.	Conclusion	55
	References	55

1. Introduction

Book reading is widely identified as an important activity in the development of children's oral language and vocabulary skills (Marulis & Neuman, 2010; Mol, Bus, & deJong, 2009; Whitehurst & Lonigan, 1998). Books provide opportunities for exposure to vocabulary that is not often found in everyday conversations (Mol, Bus, deJong, & Smeets, 2008), and book reading is a common practice in early childhood settings as a means of introducing and teaching vocabulary words. With the increased emphasis on developing language and vocabulary in young children, especially children in poverty, it is important to examine what we know from the research on book reading and vocabulary development in order to inform practices in classrooms and households. This review attempts to complement recent meta-analyses by unpacking main effect sizes; we attempt to identify what is known about optimum book reading conditions to teach words to young children. We also describe the remaining gaps in the research that must be filled in order to develop more effective, practical strategies to support vocabulary development in young children's homes and classrooms.

1.1. Vocabulary development and reading

Although many early skills support later reading success, shoring up early vocabulary knowledge improves reading in several ways, including (a) supporting comprehension of words that children decode, (b) helping children more rapidly recognize words they are decoding, (c) fostering phonological awareness skills that also support reading, and (d) increasing children's understanding of teachers' instruction in reading and other areas (National Early Literacy Panel, 2008). Rigorous, longitudinal studies clearly indicate that stronger vocabulary skills in preschool are linked to better reading outcomes over time (Dickinson & Tabors, 2001; Hart & Risley, 1995). A related finding from these and other studies is that children from low-income households often enter kindergarten 12–18 months behind middle-class children in vocabulary learning (Farkas & Beron, 2004; Layzer & Price, 2008), undermining their later reading and school success (Beck & McKeown, 2007; Biemiller, 2006; Duncan et al., 2007; National Early Literacy Panel, 2008). Indeed, about half of children in poverty cannot meet basic reading proficiency criteria in grade 4 (U.S. Department of Education, 2015). Thus, improving early vocabulary is important for all children, but especially those in poverty.

1.2. Theoretical and empirical approach

Although no single, widely accepted, comprehensive theory of word learning describes the mechanisms that underlie this process, a cognitive processing approach to learning guides this review.

The cognitive processing approach to learning (Brown, Bransford, Ferrara, & Campione, 1983; McKeown & Beck, 2014; Miller, 2003; Sternberg, 1979, 1982) asserts that active processing—defined as deliberate, attentive mental manipulation of ideas—is necessary if the learner is to understand, remember, and later use and apply new information. In the case of word learning, children need to actively process the association between a word (i.e., the lexical label) and its referent (i.e., the conceptual idea that the word represents), ultimately forging a strong and refined link in their memories (Akhtar, Jipson, & Callanan, 2001).

This broad, theoretical framework describing how word learning occurs is complemented by empirical evidence regarding the strategies or factors that best help children to successfully undertake this cognitive processing of word labels and meanings (and, thus, learn words). One essential strategy that supports the formation of mental representations of words is a clear definition, explanation, or example that helps children understand word meanings (Beck & McKeown, 2007). These are most often provided implicitly (i.e., in passing, casually) but could certainly be explicit. Connections to content that children are already familiar with may be especially beneficial (Pashler et al., 2007). Also, empirical evidence from language research suggests that some parts of speech, particularly nouns, might be somewhat easier for young children to learn, but this finding may depend on the language in question (Xuan & Dollaghan, 2013).

A second factor that likely matters for word learning is the dosage of opportunities to process words, with multiple exposures to words often needed in order to build flexible, enduring mental representations (Clay, Gill, Glynn, McNaughton, & Salmon, 2007; Goldin-Meadow et al., 2014). With the exception of instances of fast mapping (Carey, 1978), repeated interactions with a word are critical for word learning; estimates regarding the necessary dosage of exposures range from 40 exposures (McGregor, Sheng, & Ball, 2007) to 200 exposures (Clay et al., 2007), depending on the learning context, child, and word type.

A third factor involves opportunities for children to use words as they are processing and encoding them. Indeed, research indicates that, beyond simply hearing words several times in meaningful contexts, children need to use the words to communicate ideas and get feedback from adults on their use of the words (Tamis-LeMonda, Kuchirko, & Song, 2014). Taken together, this foundational theory and the empirical evidence suggest that optimally effective vocabulary interventions will offer children systematic exposures to words, as well as opportunities for a range of active-processing tasks.

1.3. Book reading as a context for vocabulary development

Interactions that foster this essential processing are many and varied; however, book reading represents a particularly potent con-

text for learning words. Book reading offers access to a wide range of words, including those that children would not typically encounter in day-to-day conversations (Snow, 1991), and explains novel words' meanings through text and illustrations (Ganea, Picard, & DeLoache, 2008; Mol et al., 2009). Moreover, these exposures to new words through book readings can be accompanied by parent or teacher explanations, as well as props and materials that further clarify word meanings (Wasik & Bond, 2001). Another benefit of book reading for word learning is that books can be re-read and referred to over time, allowing multiple exposures to new words and ideas. In addition, in book reading, words are generally introduced as part of a story (i.e., narrative text) or within a body of rich information about a specific topic (i.e., informational text), providing a meaningful and motivating context. Finally, books can help children use words, as they can serve as springboards for discussion in which children can learn more about the pronunciation and meaning of a new word and practice using it. These multiple assets of book reading for word learning may be particularly powerful for children from high-poverty homes, where language and vocabulary can be limited (Fernald, Marchman, & Weisleder, 2013; Goldin-Meadow et al., 2014).

To date, several meta-analyses have provided much insight into the role that book reading plays in vocabulary development (Marulis & Neuman, 2010, 2013; Mol et al., 2008, 2009; Scarborough & Dobrich, 1994). These meta-analyses are critically important, showing with precision and rigor that book reading reliably provides support for children's vocabulary development. Yet this research also shows that effectiveness varies across practices. To explore this variation, meta-analyses have generally created categories that allow comparison of book reading techniques across studies, drawing distinctions regarding who reads the book, how children are grouped for the reading, the types and number of words targeted, whether books were re-read or re-told, whether particular strategies were used to define words, and whether any activities extended the book into other parts of the classroom.

However, it was beyond the scope of these prior examinations—and quite difficult in light of the methodological architecture of meta-analytic techniques—to explore the details of these features at the level of specificity necessary for informing classroom decision making. A key issue involves determining how these features are interconnected. For example, Mol et al. (2009) examined the effects of book reading on children's oral language, vocabulary, and print knowledge development across 31 studies. Findings indicated that oral language effects were greater when teachers paired book readings with additional activities to support the book vocabulary and concepts. For example, children benefited from follow-up activities that allowed them to learn more about the words presented in the reading. However, the researchers acknowledged that additional studies would be needed to map out precisely how teachers could optimally use strategies in classrooms. This next step would require, for example, determining whether certain kinds of activities (e.g., review vs. extension) emerged as helpful, how often they should be provided, and for which children. This level of precision is necessary to align what is known in research with the daily decisions that teachers need to make.

As another example, Marulis and Neuman (2010) found that effects of book readings were greater when adults provided children with explicit and implicit instruction in new words during reading. However, the universe of techniques for providing explicit and implicit instruction is vast, and it was beyond the aims of that work to disentangle which instructional strategies are preferable. Thus, follow-up endeavors are needed to distill exactly which techniques were used in effective studies, as well as how frequently and with what effects on child outcomes. This information is critical if we hope to translate research into practice and impact classroom

instruction. Further, where specific, classroom-focused conclusions are concerned, it is essential to articulate what issues require additional research before clear conclusions can be drawn in order to move the field forward.

1.4. Aim of the current review

This paper complements extant meta-analyses by reviewing and unpacking the potential “active ingredients” that meta-analyses have highlighted, and carefully profiling the range of ways in which they operate in classrooms and homes. Teachers and administrators face many questions when they design their lesson plans, and families have many options when they consider how best to support children's vocabulary learning; these include what context to use to deliver the reading, which and how many words to teach, and what strategies to use. Thus, the goal of the current paper is to scrutinize the experimental literature to summarize what is known about these practical features and, where necessary, highlight questions that remain to be answered.

The research questions asked and the dimensions along which we compare existing studies to infer broader conclusions for the field were informed both by a top-down perspective, drawing on existing theory and empirical research (including previous meta-analyses) on word learning, and a bottom-up approach based on an examination of the studies in this review. Ultimately, five critical dimensions of the studies were identified, a priori, for close examination and comparison: (a) contexts in which the book reading was conducted, (b) selection and nature of the words taught, (c) the dosage of word exposure, (d) the intervention strategies, and (e) the outcome measures and findings.

To create a descriptive portrait of the field, this systematic review was guided by several specific research questions:

1. What is the methodological quality and comprehensiveness of the studies, after stringent selection criteria are applied?
2. In what contexts are effective book reading studies implemented, including who read the book and how children were grouped together?
3. What is known about the words that are taught during these effective studies, including the type of words, the number of exposures to words, and the criteria used by the researchers or other readers to select the words?
4. What instructional strategies are implemented during reading or extension activities to build vocabulary?
5. What is known about the dosage of the interventions, particularly regarding how long they are implemented for?
6. What measures are used to assess vocabulary learning, and how many words do children learn?

Moreover, to the extent that there is a lack of clarity and specificity for the families and educators who have the best opportunity to build vocabulary knowledge among young children, findings can also identify points worthy of future investigation (Currie & Neidell, 2007; Durlak & DuPre, 2008; Heckman, Pinto, & Saveljev, 2013).

2. Method

To augment important recent meta-analyses, this review examines the literature on book reading and vocabulary development among young children with a particular focus on teasing apart the practical contexts and components of effective studies. Like other recent reviews (Hammer et al., 2014), this review follows standard key steps: locating all possible studies, screening potential studies for inclusion using preset criteria, coding all qualified studies based

on their methodological and substantive features, and calculating effect sizes for all qualified studies for further combined analyses.

However, relative to a meta-analysis, little attention is devoted to determining *whether* the technique under examination works through calculating an effect size across a body of relevant studies of that technique (indeed, for research on book reading in early childhood, the review studies and meta-analyses above have provided excellent estimates of this information). Instead, attention is focused upon discerning *why* or *under what conditions* a technique works by coherently describing and comparing the methods or nature of studies.

2.1. Literature search strategy and selection criteria

This critical review of research examines the strategies and characteristics of vocabulary studies that showed significant impacts on the receptive and expressive vocabulary of children. In order to be included in this review, studies had to meet inclusion and exclusion criteria concerning type of treatment, sample, outcome measures, study design, peer review, and outcomes. First, studies had to examine an adult implementing the book reading (as opposed to a technology-based intervention) and explore its effects on children's word learning. We excluded studies of more comprehensive curricula where it was difficult to discern the discrete effect of a storybook reading component (e.g., Landry, Anthony, Swank, & Monseque-Bailey, 2009; Weiland & Yoshikawa, 2013). Second, samples had to include children ages 3–6 years (e.g., preschool and kindergarten age) who were typically developing monolingual learners of English. Although examining vocabulary development and book reading in DLLs is important, it raises other nuanced issues that are particular to DLLs and requires separate analyses (Hammer et al., 2014). Third, studies had to include a child outcome measure of English (rather than nonsense word) vocabulary and a measure of expressive or receptive vocabulary, whether using tools specific to the studies' target words or standardized tools (e.g. the PPVT). Fourth, the study had to employ an experimental or quasi-experimental design, incorporating one or more of the following: a randomized controlled trial, a pre-test–treatment–post-test comparison with a control group, group comparisons with random assignment or pre- and post-testing, or within-subject designs.

Fifth, studies had to be published in peer-reviewed outlets. We did not review unpublished dissertations or conference proceedings to ensure that work had undergone rigorous peer review. Finally, because we were interested in untangling the nature of effective practices in book reading related to vocabulary development, studies had to demonstrate statistically significant impacts on at least one child vocabulary outcome. These constraints helped to ensure that we could address the specific role of book reading in language acquisition during the early childhood years, and that we could draw reasonable comparisons across studies.

To obtain the body of vocabulary studies that met our inclusion criteria, the three authors developed a list of comprehensive key search terms. The terms were divided into four primary sets: terms that focused on the targeted population (preschool*, kindergarten, early childhood, Head Start), intervention and treatment (book-reading, book, reading), outcome (vocabulary, language, word*) and research design (experiment, quasi-experiment, treatment, control, pretest, posttest, intervention).

Multiple searches were run across the major academic databases: Elsevier, ERIC, Google Scholar, PsychInfo, PubMed, and Scopus. The search was limited to journals published in English that were available through these search engines between 1990 and 2016. The searches returned 1529 unique citations published in English-language journals. We also reviewed the studies included in meta-analyses on book reading to ensure that our search was inclusive.

Each citation was passed through multiple levels of review. Abstracts were read by the third author and marked for inclusion or exclusion using the team's criteria. Articles that clearly did not meet the criteria were excluded (N = 1469); for example, we excluded studies without a control group (either between or within children), studies focused on non-English speaking populations, studies of DLL or special education children, or studies of infants or children beyond kindergarten. The second and third authors reviewed a randomly selected subsample (10%) of abstracts to ensure that this exclusionary process was done accurately.

Following this, 60 articles remained. Publications of each were obtained and sent to members of the research team. Each co-author read the papers and made further exclusions using the review criteria as necessary. Co-authors reached 93.3% pairwise agreement on the 60 papers. For those papers without 100% agreement for inclusion or exclusion, papers were discussed, and a final decision was made by consensus about whether or not the article should be included. During this final selection process, 31 articles were selected. Five of these articles had two studies included in the paper, for a total of 36 studies.

To assist with the critical review, information from each article was coded and entered into a Table by the first and third authors. Information extracted from the articles included: the description of the study (including a description of the sample and the study research design), the book-reading context (home or school, implementer, and size of group), dosage, nature of the targeted words (number of words per book, number of books, total number of words taught, estimate of number of exposures to each word, dosage of treatment, types of words, words' part of speech), and nature of and results on the outcome measures (i.e., for standardized measures, any impacts as measured using effect sizes; or for a target-word measure, any impacts as measured using effect sizes as well as number of words actually learned).

2.1.1. Downs and Black criteria

To gauge the quality of each research study, we used the Downs and Black (1998) criteria. These criteria were organized across four domains of quality: reporting, external validity, internal validity-bias, and internal validity-confounding. All papers were coded across 25 indicators. A study received a "1" for an indicator if it was satisfactorily fulfilled according to the criteria described in Downs and Black (1998). If the study did not satisfy an indicator, or if it was not possible to determine whether an indicator was met, the study received a "0" for that indicator.

Indicators were aggregated into four conceptual clusters. The *reporting quality* indicators were comprised of items concerned with whether the study provided a clear description of its: (a) hypothesis, aims, and objectives, (b) main study outcomes, (c) participant characteristics, (d) experimental and comparison conditions, (e) distribution of confounders in each participant group, (f) main findings, (g) estimates of random variability in data for main outcomes, (h) characteristics of participants lost to attrition, and (i) actual probability values for main outcomes. Together, these 9 indicators resulted in a maximum score of "9" for the reporting quality indicators. The *external validity quality* indicators were comprised of individual indicators linked to whether the study reached standards in: (a) representativeness of recruited participants, (b) representativeness of actual participants, and (c) representativeness of study context (e.g., staff, facilities), for a maximum score of "3." The *internal validity-bias quality* indicators assessed presence of: (a) participants blinded to experimental condition, (b) assessors blinded, (c) no data dredging, (d) time between experimental condition and outcome same for all groups, (e) statistical tests appropriate for main study outcomes, (f) compliance with experimental condition, and (g) reliability and validity of main outcomes measures, for a maximum score of "7." The *internal*

Table 1
Study participants and design.

Study	Treatment of Interest	Sample	Research Design
Arnold et al. (1994)	Dialogic reading	64 young children, middle to upper-income, and mothers	RA (child)
Beck and McK. S1 (2007)	Defining, questioning	98 children in 8 classes (4 kindergarten and 4 first grade); low-SES children; all African American (AA)	Quasi (class) to treatment or control
Beck and McK. S2 (2007)	Defining, questioning	76 children in 3 kindergarten and 3 first grade classes; All children AA	Quasi: Within-subject design
Biemiller and B. S1 (2006)	Defining, questioning, re-reading	112 kindergarten-second grade working class-children	Quasi (class)
Biemiller and B. S2 (2006)	Defining, re-reading (enhanced condition)	107 from same school as in Study 1	Quasi (child)
Blewitt et al. S1 (2009)	Defining, questioning, re-reading	50–60 3-year-old children, middle- to upper-income	RA (child)
Blewitt et al. S2 (2009)	Defining, questioning, re-reading (scaffolded)	50–60 3-year-old children, middle- to upper-income	RA (child)
Coyne et al. (2004)	Defining, rereading, retelling, props	64 at-risk kindergarten students from 7 schools	RA (child)
Coyne et al. S1 (2007)	Defining, questioning, re-reading, activities	32 mid-SES kindergarten students	Quasi: Within-subject design
Coyne et al. S 2 (2007)	Defining, questioning, re-reading, activities	Same as in Study 1	Quasi: Within-subject design
Coyne et al. (2009)	Defining, questioning, re-reading, activities, props	42 kindergarten students from low-income school	Quasi: Within-subject design
Coyne et al. (2010)	Defining, questioning, re-reading, activities, props	124 kindergarten students from 3 low-income schools	Quasi: (class or student)
Ewers and B. (1999)	Questioning	66 suburban kindergarten students	RA (child)
Gonzalez et al. (2010)	Defining, questioning, re-reading, activities, props	21 pre-kindergarten and Head Start teachers from mixed SES	RA (teacher)
Hargrave and S. (2000)	Dialogic reading	36 day care children (ages 3–5), low-income	Quasi (child)
Justice et al. (2005)	Defining, re-reading	57 kindergarten students in low-income schools	RA (child)
Leung (2008)	Questioning, re-telling, re-reading, activities, props	32 3- and 4-year-old children at a child care; diverse	RA (child)
Lever & Sénéchal (2011)	Dialogic reading	40 English-speaking kindergarten children; diverse SES	RA (child)
Loftus et al. (2010)	Defining, questioning, re-reading, activities, props	20 at-risk kindergarten students	Quasi: Within-subject design
Lonigan and W. (1998)	Dialogic reading	91 day care children (ages 3–4 years old), low-income	RA (child)
Lonigan et al. (2013)	Dialogic reading	324 prekindergarten, low-income	RA (child)
McKeown and B. (2014)	Defining, questioning, re-reading, activities	131 working-class kindergartners	Quasi: Within-subject design
Penno et al. (2002)	Defining, re-reading	47 mixed-ethnic group of New Zealand children	Counter-balanced Latin-square design
Robbins & Ehri (1994)	Re-reading	45 non-reading kindergarten children	Counter-balanced design
Sénéchal et al. S1 (1995)	Questioning, re-reading, props	48 4- to- 5-year-old children from upper-middle-class homes	RA (child)
Sénéchal et al. S2 (1995)	Questioning with pointing, re-reading, props	Same as in Study 1	RA (child)
Senechal (1997)	Questioning and rereading (separately)	30 3- and 4-year-old children; middle-class day cares	RA (child)
Silverman (2007)	Defining, questioning, re-telling, re-reading	94 kindergartners in 6 classrooms; diverse	RA (class)
Silverman et al. (2013)	Defining, questioning, re-reading, props, activities	26 Head Start classrooms	RA (teacher)
Sim et al. (2014)	Questioning, re-reading	80 prekindergarten children, from 3 schools in Australia	RA (child)
Wasik & Bond (2001)	Defining, questioning, re-reading, props, activities	127 at-risk four-year-olds in public 4 public prekindergarten classrooms	RA (class)
Wasik et al. (2006)	Defining, questioning, re-reading, props, activities	207 children in 16 Head Start classrooms	RA (class)
Weisberg et al. (2015)	Defining, questioning, re-reading, props, activities	154 preschool children	RA (child)
Whitehurst et al. (1994)	Dialogic reading	73 day care children (aged 3), low-income	RA (child)
Zipoli et al. (2011)	Defining, questioning, re-reading, activities	80 K from 3 low-income schools	Quasi: Within-subject design
Zucker et al. (2013)	Defining, questioning, re-reading, props, activities	39 prekindergarten programs serving low-income children; 5% ELL	RA (teacher)

validity-confounding quality indicators assessed whether: (a) participants recruited from same population, (b) participants recruited during the same period of time, (c) participants randomized to treatment conditions, (d) random assignment concealed to participants and staff, (e) statistical adjustments for confounding, and (f) attrition addressed in analysis, for a maximum score of “6.” Finally, we also examined whether the study included a measure of implementation fidelity conducted by a trained observer (not participant self-report).

The first and third authors both read and coded 20 of the 31 papers across the above criteria (i.e., both the study characteristics as well as the Downs and Black quality codes). Comparisons of their codes on these 20 papers reached 91.4% pair-wise agreement, and any disagreements were discussed to reach 100% agreement for all codes. The third author then read and scored the remaining 11 papers independently, discussing issues with the first author when questions arose.

3. Results

Thirty-one articles are included in the review. Of these 31, five included two studies, resulting in a total of 36 studies reviewed. See Table 1 for an overview of each study, including the book reading treatment studied, the sample, and the research design. The studies included diverse samples of children, with some including participants of low-income backgrounds (15 studies), some from a mix of SES backgrounds (seven studies), and some from higher SES backgrounds (nine studies). The research designs also varied, with 24 random assignment experimental studies, one counter-balanced Latin Square experimental study, and 11 quasi-experimental studies.

3.1. Quality of the studies reviewed

All studies were scored using the Downs and Black (1998) scoring system (see Table 2). The average score for reporting was 8.1 out of 9. The average score for external validity was 1.9 out of 3. The average score for internal validity-bias was 4.9 out of 7. The average score for internal validity-confounding was 4.2 out of 6. Overall, the 36 studies' overall quality was 19.0 out of a total possible score of 25. There was variability in the overall quality, with 12 studies scoring below 70%, and 5 studies scoring above 85%. Half (50%) of the 36 studies contained an independent (non self-report) measure of fidelity of implementation.

3.2. Context of book reading

Of the 36 studies, the research was conducted in various settings, including the home, the school, and combinations of these two settings (see Table 3). Related to context are group size (one-to-one, and small- and whole-group settings) and who read the story. Of the 36 studies, two (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Sim, Berthelsen, Walker, Nicholson, & Fielding-Barnsley, 2014) were conducted entirely in home settings with parents implementing the treatment. In both of these studies, parents read one-to-one with their child.

Thirteen studies were conducted in classrooms with the teachers implementing the study. Of these, 10 were conducted in whole groups, and three were conducted in small groups (see Table 3).

Sixteen studies were conducted by researchers in schools. Of these 16 studies, seven had researchers conducting the book reading one-to-one with the child, whereas nine had researchers conducting the book reading in small groups.

Five of the studies used combinations of parents, teachers and/or researchers to deliver readings. In two studies (Lonigan & Whitehurst, 1998; Whitehurst et al., 1994), there was one condition

in which book reading was conducted in the classroom (teachers implementing dialogic reading) as well as a second home condition (parents implementing dialogic reading) to assess the added value of a home-plus-school condition. In these two studies, teachers read to the children in small groups in the classrooms, and parents were also trained to read with their children in one-on-one formats at home. In another two of these five studies (Coyne et al., 2010; Zipoli, Coyne, & McCoach, 2011), book reading was conducted both by the researchers in small groups and by teachers in whole-group settings in school. In the final one of these five studies (Leung, 2008), teachers read in small groups and researchers conducted story re-tells one-to-one with children at school.

3.3. Words during book reading

Several issues related to the number and types of words focused on during these studies were examined (see Table 4).

3.3.1. Number of words

We first examined the number of words presented during a single book reading. Of the 36 studies reviewed, seven, all of which implemented *dialogic reading*, did not focus on specific target words during the book reading (Arnold et al., 1994; Hargrave & Sénéchal, 2000; Lever & Sénéchal, 2011; Lonigan & Whitehurst, 1998; Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013; Sim et al., 2014; Whitehurst et al., 1994). The remaining 29 studies focused on teaching children specific words during book reading. Two studies did not report the specific number of words targeted per book reading (Wasik & Bond, 2001; Wasik, Bond, & Hindman, 2006). Among the 27 others, we examined the number of words presented during a single book reading and across the total number of books read. The number of words targeted during a single reading ranged from two or three words (e.g. Gonzalez et al., 2010; Silverman, Crandell, & Carlis, 2013) to 15 words (e.g. Penno, Wilkinson, & Moore, 2002), to 28 or 29 words (Weisberg et al., 2015). Most studies, though, fell on the lower end of this distribution; indeed, 21 of these studies (78%) targeted between two words and 10 words per a single book reading (See Table 4). The remaining six studies focused on 11–29 words per book reading.

We also examined the total number of words presented during an entire study (i.e., across all book readings in the study). Results revealed great variation. Of the 27 relevant studies, the total number of words targeted during the duration of the study ranged from six (Coyne, McCoach, & Kapp, 2007, Studies 1 & 2, in which one book was read over one week) to 120 (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004, which described a 7-month-long treatment including 40 books, each of which targeted three words). Overall, longer durations were related to more total words (see Table 4).

In sum, there was a remarkable amount of variation in the number of words taught, both during single book readings and across treatments as a whole. Although a clear descriptive trend emerged, such that more words were taught via longer treatments, there was substantial variation around this trend.

3.3.2. Number of exposures to words

The number of times children were exposed to individual target words was also examined. Only eight studies provided enough information that the total number of exposures per word could be calculated. Among these studies, the number of exposures to a word that occurred during any part of the treatment ranged from two to 15. In general, more exposures occurred among treatments with more sessions distributed over longer spans of time. No study specifically examined the number of exposures to words as a variable.

Table 2
Study Quality Indicators.

Study	Reporting (Max 9)	External Validity (Max 3)	Internal Validity-Bias (Max 7)	Internal Validity – Confounding (Max 6)	Total Score	Imple-mentation Measure
Arnold et al. (1994)	7	1	5	4	17	No
Beck and McK. S1 (2007)	6	3	5	3	17	Yes
Beck and McK. S2 (2007)	7	2	5	2	16	Yes
Biemiller and B. S1 (2006)	6	3	4	4	17	No
Biemiller and B. S2 (2006)	5	3	5	3	16	No
Blewitt et al. S1 (2009)	8	2	4	5	19	No
Blewitt et al. S2 (2009)	8	1	5	5	19	No
Coyne et al. (2004)	6	2	5	5	18	No
Coyne et al. S1 (2007)	9	3	4	4	20	Yes
Coyne et al. S 2 (2007)	8	2	5	4	19	Yes
Coyne et al. (2009)	8	2	5	4	19	Yes
Coyne et al. (2010)	9	3	5	3	20	Yes
Ewers and B. (1999)	8	1	3	5	17	No
Gonzalez et al. (2010)	9	2	5	5	21	Yes
Hargrave and S. (2000)	6	1	5	2	14	Yes
Justice et al. (2005)	9	2	5	5	21	No
Leung (2008)	9	3	5	5	22	No
Lever & Sénéchal (2011)	8	2	6	5	21	No
Loftus et al. (2010)	9	2	5	4	20	Yes
Lonigan and W. (1998)	9	1	5	5	20	Yes
Lonigan et al. (2013)	9	2	6	5	22	No
McKeown and B. (2014)	8	3	5	3	19	Yes
Penno et al. (2002)	7	1	5	4	17	No
Robbins & Ehri (1994)	8	1	4	4	17	No
Sénéchal et al. S1 (1995)	8	0	4	4	16	No
Sénéchal et al. S2 (1995)	8	0	4	4	16	No
Senechal (1997)	9	1	4	2	16	No
Silverman (2007)	9	2	5	5	21	Yes
Silverman et al. (2013)	9	2	5	5	21	Yes
Sim et al. (2014)	9	3	5	6	23	Yes
Wasik & Bond (2001)	8	1	5	5	19	No
Wasik et al. (2006)	9	1	5	4	19	Yes
Weisberg et al. (2015)	8	2	6	5	21	Yes
Whitehurst et al. (1994)	9	2	6	5	22	No
Zipoli et al. (2011)	9	3	5	4	21	Yes
Zucker et al. (2013)	9	2	6	5	22	Yes

3.3.3. Criteria for word selection

Of the 29 studies, six reported that they selected words that were tier 2 words, based on Beck's tier taxonomy (Beck, McKeown, & Omanson, 1987; Beck, McKeown, & Kucan, 2002). Eleven studies reported that words were selected because pretesting or pilot studies found that they were unfamiliar or unknown to the children. Six other studies indicated that the words were selected because they were likely to be unfamiliar to children; however, no empirical data to support this point were collected (e.g., Blewitt, Rump, Shealy, & Cook, 2009). Nine studies reported that words were selected because they were considered both important for story comprehension and likely unfamiliar to the children. For example, Silverman et al. (2013) followed the recommendations outlined by Beck et al. (2002) to choose words that were important for understanding the text, useful for children to know, and that could help build conceptual understanding. Silverman et al. (2013) also chose unfamiliar words, as these types of words are more likely to evidence growth over time. One study reported selecting low-frequency words that were easy for children to understand (Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010). Two studies identified words that were unfamiliar to the children but also had synonyms that were familiar to the children (Sénéchal, Thomas, & Monker, 1995). Three studies focused on theme-related words (e.g. science words; Gonzalez et al., 2010). Finally, one study used a combination of tier 2 and easier words (Zucker, Solari, Landry, & Swank, 2013).

3.3.4. Grammatical class of words

Twenty studies reported the grammatical class (e.g., nouns, verbs, etc.) of the words they targeted. Two – Justice, Meier, and

Walpole (2005) and Robbins and Ehri (1994) – explicitly categorized the words by parts of speech, with Justice and colleagues reporting 16 adjectives, 16 nouns, and 28 verbs while Robbins and Ehri included two adjectives, one noun, and eight verbs. Eighteen other studies listed the specific words that were targeted, allowing our team to retrospectively categorize the words (see Table 4), and revealing that most studied a mix of parts of speech, while three examined only noun learning. No study specifically examined grammatical word class as a factor in word learning.

3.4. Study dosage

The components of dosage include length of treatment (number of weeks), total number of sessions, and length of each book reading session (how long it took to read the book as well as any additional activity included as part of the book reading) (see Table 4). The length of the treatments varied from two sessions over two days (Robbins & Ehri, 1994; Sénéchal et al., 1995; Senechal, 1997) to 108 sessions over seven months (Coyne et al., 2004) or nine months (Wasik et al., 2006); the longest featured more than 150 sessions over a full school year (Lonigan et al., 2013). Less information was available regarding the amount of time spent on book reading. Of the 36 studies, 20 indicated an average range for book reading. Eight studies reported 30 min, two studies reported 20–30 min, six studies reported 20 min, one study reported 10–20 min and three studies reported 10 min. In sum, the number of sessions varied widely, with a great deal of unavailable information regarding the length, in minutes, of each book reading session.

Table 3
Contexts of book-reading studies.

Authors	Home	School					
		Teacher			Researcher		
		1–1	Whole group	Small group	1–1	Whole group	Small group
Arnold et al. (1994)	X						
Beck and McK. S1 (2007)		X					
Beck and McK. S2 (2007)		X					
Biemiller and B. S1 (2006)		X					
Biemiller and B. S2 (2006)		X					
Blewitt et al. S1 (2009)							X
Blewitt et al. S2 (2009)							X
Coyne et al. (2004)			X				
Coyne et al. S1 (2007)						X	
Coyne et al. S2 (2007)						X	
Coyne et al. (2009)						X	
Coyne et al. (2010)		X ^a				X ^a	
Ewers and B. (1999)							X
Gonzalez et al. (2010)			X				
Hargrave & S. (2000)			X				
Justice et al. (2005)						X	
Leung (2008)			X ^b				X ^b
Lever & Sénéchal (2011)						X	
Loftus et al. (2010)						X	
Lonigan and W. (1998)	X ^c		X ^c				
Lonigan et al. (2013)						X	
McKeown and B. (2014)		X					
Penno et al. (2002)						X	
Robbins & Ehri (1994)							X
Sénéchal et al. S1 (1995)							X
Sénéchal et al. S2 (1995)							X
Senechal (1997)							X
Silverman (2007)		X					
Silverman et al. (2013)		X					
Sim et al. (2014)	X						
Wasik & Bond (2001)		X					
Wasik et al. (2006)		X					
Weisberg et al. (2015)						X	
Whitehurst et al. (1994)	X		X ^d				
Zipoli et al. (2011)		X ^e				X ^e	
Zucker et al. (2013)		X	X				

^a Delivery was different at different schools.

^b Half of the children had retelling with researchers.

^c RA to home, school, or home and school.

^d RA to school and H+S.

^e Delivery was different at different schools.

3.5. Word learning strategies implemented in the studies

Although all studies targeted children's vocabulary as an outcome, they varied in the strategies used to teach vocabulary (see Table 5). Of the 36 studies, seven examined the effectiveness of dialogic reading. Although dialogic reading is presented by authors, and often thought of by teachers, as a discrete strategy, it actually includes a constellation of techniques. In particular, adults prompt the child to say something about the book, evaluate the child's response, expand on the child's response, and repeat the prompt (i.e., the PEER actions). In addition, open-ended and distancing questions prompt the child to relate the book to events or situations in his or her own life.

In the remaining 29 studies, six strategies were examined (alone or in conjunction with others) to determine their impact on vocabulary growth (Table 5). The identified strategies are (a) defining words, (b) questioning as a means to promote discussion of vocabulary and comprehension of the book, (c) re-telling, (d) re-reading, (e) using props to illustrate word meanings, and (f) providing extension activities that promote exploration and discussion of vocabulary. Twenty-six studies (90%) used multiple strategies as part of their treatment, whereas only three studies examined one strategy: Ewers and Brownson (1999) explored the effects of questioning on vocabulary development, Robbins and Ehri (1994)

examined the effect of re-reading, and Senechal (1997) compared the effects of single-reading to re-reading and questioning conditions.

The 26 studies employing multiple strategies wove them together in various combinations, and the exact amount of each strategy was not reported. For example, when questioning was used as a strategy, studies rarely reported how many questions were asked, how many children responded to each question, or how many children answered the questions using target words. Thus, as is apparent in Table 5, at this time, it is impossible to reliably explore and compare the effectiveness of different combinations of techniques in the current literature.

3.6. Outcome measures and outcomes

Of the 36 studies, nine used both project-specific (proximal) and standardized (distal) measures (see Table 6). Five studies used only standardized measures to assess outcomes; four of these five implemented dialogic reading. The remaining 22 used only project-specific measures.

3.6.1. Project-specific measures

Thirty-one of the studies used project-specific measures, or those featuring words that were targeted by the treatment, in order

Table 4
Words and dosage.

Study	# of words per book	# of books	Total # words	Estimate of # of exposures to each word	Dosage – length and intensity of treatment	Type of words	Examples of words	Parts of Speech
Arnold et al. (1994)	Utbd	Utbd	Utbd	Utbd	35 sessions over 5 weeks, Utbd minutes per session	Utbd	NA	NA
Beck and McK. S1 (2007)	3	36	108	Utbd	50 sessions over 10 weeks, Utbd minutes per session	Tier 2	feast, exhausted, cautiously, gazing, dazzling	Utbd
Beck and McK. S2 (2007)	6	7	42	<20 vs. 20	35 sessions over 7 weeks, 20 min per session	Tier 2	astonished, enormous extraordinary, commotion, impressed, inseparable	Adject (15), Adverb (2), Noun (6), Verb (19)
Biemiller and B. S1 (2006)	12 or 24 (but 4–6 words focused on per session)	3	48	Utbd (~3 or 5)	15 sessions over 3 weeks, Utbd minutes per session	Unfamiliar but important for story	not described	Utbd
Biemiller and B. S2 (2006)	7–9 words	2	42–55	Utbd (~5)	10 sessions over 2 weeks, 30 min a session	Unfamiliar but important for story	circus, sniff, bullseye, sprain, straight, lead caliper, davenport, grapple, lynx, pagoda, phlox	K words: Adj (4), Noun (19), Verb (19) Noun (9)
Blewitt et al. S1 (2009)	6	3	9	12 vs. 6	4 sessions over 6 weeks, Utbd minutes per sessions	Unfamiliar but important for story	same as above	Noun (9)
Blewitt et al. S2 (2009)	6	3	9	12	4 sessions over 6 weeks, Utbd minutes per session	Unfamiliar	same as above	Noun (9)
Coyne et al. (2004)	3	40	120	Utbd (many)	108 sessions over 7 months, 30 min per session	Unfamiliar but important for story	not described	Utbd
Coyne et al. S1 (2007)	6	1	6	1X vs. Utbd (many)	3 sessions over 1 week, 20–30 min per session	Unfamiliar but important for story	sow, cauldron, loitering, approaching, sturdy, scrumptious	Adj (2), Noun (2), Verb (2)
Coyne et al. S2 (2007)	6	1	6	Utbd (many) vs. Utbd (fewer)	3 sessions over 1 week, 20–30 min per session	Unfamiliar but important for story	same as above	Adj (2), Noun (2), Verb (2)
Coyne et al. (2009)	9	1	9	3X vs. Utbd (many)	3 sessions over 1 week, 30 min a session	Unfamiliar but important for story	weald, duvet, domicile, parlor, lass, shards	Adj (1), Noun (6), Verb (2)
Coyne et al. (2010)	3	18	54	Utbd (15+)	36 sessions over 18 weeks, 30 min a session	Tier 2	peculiar, apex, plummet, slender, drenched, sprawl	Adj (19), Noun (15), Verb (20)
Ewers and B. (1999)	10	1	10	Utbd	1 day, Utbd # minutes	Unfamiliar but important for story	not described	Utbd
Gonzalez et al. (2010)	2–3	36	94	Utbd	90 sessions over 18 weeks, 20 min a session	Theme words (science)	not described	Utbd
Hargrave and S. (2000)	NA	14	NA	Utbd	20 sessions over 4 weeks, 10 min per at school and 10 min per at home	NA	antlers, bat, clover, colt, croquet, fountain	Utbd
Justice et al. (2005)	6	10	60	4 vs. ~12	20 sessions total over 10 weeks, 20 min a session	Tier 2	heaved, pouted, ruffle, gaze, ripples, surface, discovered	Adj (16), Noun (16), Verb (28)
Leung (2008)	6–15	3	32	Utbd	12 sessions across 4 weeks of reading and retelling activities; 3 sessions of other activities, Utbd # minutes	Theme words (science)	prism, color wheel, shadow, radio waves, microwaves, infrared light,	Noun (32)
Lever & Sénéchal (2011)	NA	8	Utbd	Utbd	16 sessions across 8 weeks, 20 min per session	NA	grocer, fawn, carton, pelican, beret, partridge	Utbd

Table 4 (Continued)

Study	# of words per book	# of books	Total # words	Estimate of # of exposures to each word	Dosage – length and intensity of treatment	Type of words	Examples of words	Parts of Speech
Loftus et al. (2010)	4	2	8	Utbd (6+)	8 sessions across 2 weeks, 30 min per session	Low-frequency words easy to understand	apex, peculiar, plummet, immense, saunter, drenched	Adj (2), Noun (2), Verb (4)
Lonigan and W. (1998)	NA	6	Utbd	NA	30 sessions across 6 weeks, 10 min per session	NA	NA	Utbd
Lonigan et al. (2013)	NA	NA	NA	NA	150+ sessions across school year, 10–20 min per session	NA	NA	Utbd
McKeown and B. (2014)	10	3	30	12	21 sessions, Utbd minutes per session	Tier 2	glee, shrieked, clutch, curious, stunned, perfectionist	Adj (10), Adv (3), Noun (2), Verb (15)
Penno et al. (2002)	15	2	30	3–9 vs. Utbd (9+)	6 sessions over 6 weeks, Utbd minutes per session	Unfamiliar	not described	Utbd
Robbins & Ehri (1994)	11	1	11	2 or 4 (mostly 4)	2 sessions over 2 days, Utbd minutes per session	Unfamiliar	irate, duped, clamor, chortle, extricate, escorted	Ad (2), Noun (1), Verb (8)
Sénéchal et al. S1 (1995)	13	1	13	2–4	2 sessions over 2 days, Utbd minutes per session	Unfamiliar words with familiar synonyms	angling, ascending, elderly, fangs, fedora, gazing	Adj (1), Noun (9), Verb (3)
Sénéchal et al. S2 (1995)	10	1	10	4	2 sessions over 2 days, Utbd minutes per session	Unfamiliar words with familiar synonyms	Same as above	Adj (1), Noun (7), Verb (2)
Senechal (1997)	10	1	10	~3	2 sessions over 2 days, 20 min per session	Unlikely to be known	angling, fang, fedora, gazing, goblet, infant flutter, peer, silk, solitary, vanish, bicker	Noun (8), Verb (2)
Silverman (2007)	5	6	30	Utbd	18 sessions over 6 weeks, 30 min per session	Tier 2	field, harvest, sour, prepare, cargo, partrol	Adj (4), Noun (9), Verb (17)
Silverman et al. (2013)	2	24	48	~6–8 vs 12–24	48 sessions over 12 weeks, 30 min a session	Unfamiliar but important for story	field, harvest, sour, prepare, cargo, partrol	Adj (3), Noun (24), Verb (21)
Sim et al. (2014)	NA	8	NA	NA	~24 sessions over 8 weeks, Utbd minutes per session	NA	NA	Utbd
Wasik & Bond (2001)	Utbd (~10)	22	100	Utbd	44 sessions over 11 weeks, Utbd minutes per session	Unfamiliar but important for story	not described	Utbd
Wasik et al. (2006)	Utbd	44	Utbd	Utbd	180 sessions over 9 months, Utbd minutes per session	NA	not described	Utbd
Weisberg et al. (2015)	20	2	40	4+	8 sessions over 2 weeks, 10 min per session	Theme words	rung, weeds, lane, talons, scales, nostrils	Realistic theme: Adj (3), Adverb (1), Noun (7), Preposition (2), Verb (7); Fantasy theme: Adj (2), Noun (10), Preposition (2), Verb (6)
Whitehurst et al. (1994)	NA	6–8	Utbd	NA	30 sessions over 6 weeks, 10 min per session	NA	NA	NA
Zipoli et al. (2011)	3	18	54	At least 4–5 times	36 sessions over 18 weeks, Utbd minutes a session	Unlikely to be known	not described	Utbd
Zucker et al. (2013)	6–9	7	51	Utbd – at least 3 times	20 sessions over 4 weeks, 30 min per session	Mix of tier 2 and easier words	crab, clumsy, climb, shatter, catch, fins	Adj (11), Noun (15), Preposition (3), Verb (21)

Table 5
Strategies utilized.

Study	Dialogic Reading	Defining words	Questioning	Retelling	Rereading	Props	Add'l activities
Arnold et al. (1994)	Yes	Yes – # UtbD	Yes – # UtbD	No	UtbD	No	No
Beck and McK. S1 (2007)	No	Yes, more than once	Yes – # UtbD	No	No	No	No
Beck and McK. S2 (2007)	No	Yes – # UtbD	Yes – # UtbD	No	No	No	No
Biemiller and B. S1 (2006)	No	Yes – 1X	Yes – # UtbD	No	Yes – 2 or 4X	No	No
Biemiller and B. S2 (2006)	No	Yes – 1-2X	No	No	Yes – 4X	No	No
Blewitt et al. S1 (2009)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 3X	No	No
Blewitt et al. S2 (2009)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 3X	No	No
Coyne et al. (2004)	No	Yes – # UtbD	UtbD	Yes – # UtbD	Yes – 2X	During re-tell, not reading	No
Coyne et al. S1 (2007)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 3X	No	Yes
Coyne et al. S2 (2007)	No	Yes – at least 3X	Yes – # UtbD	No	Yes – 3X	No	Yes
Coyne et al. (2009)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 3X	Yes (one arm)	Yes
Coyne et al. (2010)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 2X	Yes (one arm)	Yes
Ewers and B. (1999)	No	No	Yes – # UtbD	No	No	No	No
Gonzalez et al. (2010)	No	Yes – # UtbD	Yes – 2X per word	No	Yes – 2X	Yes – picture cards	Yes
Hargrave and S. (2000)	Yes	No	Yes – # UtbD	No	Yes – 2X	No	No
Justice et al. (2005)	No	Yes – 4X or 0X	No	No	Yes – 4X	No	No
Leung (2008)	No	No (but defined in text)	Yes – # UtbD	Yes (1/2 children – 3X)	Yes – 3X	Yes	Yes
Lever & Sénéchal (2011)	Yes	Yes – # UtbD	Yes	No	Yes – 2X	No	No
Loftus et al. (2010)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 2X	Yes	Yes
Lonigan and W. (1998)	Yes	UtbD	Yes – # UtbD	No	UtbD	No	No
Lonigan et al. (2013)	Yes	UtbD	Yes – # UtbD	No	UtbD	UtbD	No
McKeown and B. (2014)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – 3 or 1X	No	Yes
Penno et al. (2002)	No	Yes – 0X, 3X, or 9X	No	No	Yes – 3X	No	No
Robbins & Ehri (1994)	No	No	No	No	Yes – 2X	No	No
Sénéchal et al. S1 (1995)	No	No	Yes – 0X or 2X	No	Yes – 2X	Yes	No
Sénéchal et al. S2 (1995)	No	No	Yes – 2X	No	Yes – 2X	Yes	No
Senechal (1997)	No	No	Yes – # UtbD	No	Yes – 3X	No	No
Silverman (2007)	No	Yes – # UtbD	Yes – # UtbD	Yes	Yes – 2X	No	No
Silverman et al. (2013)	No	Yes – # UtbD	Yes – # UtbD	UtbD	Yes – 2X	Yes (one arm)	Yes
Sim et al. (2014)	Yes	No	Yes – # UtbD	No	Yes – # UtbD	No	No
Wasik & Bond (2001)	No	Yes – # UtbD	Yes – # UtbD	No	Yes – half read 2X	Yes	Yes
Wasik et al. (2006)	No	Yes – # UtbD	Yes	No	Yes – 2X	Yes	Yes
Weisberg et al. (2015)	No	Yes – # UtbD	Yes – # UtbD	UtbD	Yes – 4X	Yes	Yes
Whitehurst et al. (1994)	Yes	No	Yes – # UtbD	No	Yes – UtbD	No	No
Zipoli et al. (2011)	No	Yes – # UtbD (a lot)	Yes – # UtbD	No	Yes – 2X	No	Yes
Zucker et al. (2013)	No	Yes – # UtbD	Yes – # UtbD	No	Yes	Yes – for Tier 2	Yes

Table 6
Outcome Measures.

Study	Standardized Measures Used	Standardized Measures – Outcomes	Project Specific Measures Used	Project Specific Measures – Outcomes
Arnold et al. (1994)	PPVT, EOWPVT, and ITPA	Video group scored better than control on EOWPVT (0.92) and ITPA (0.88); Direct training group scored better than control on ITPA (0.75)	Not measured	NA
Beck and McK. S1 (2007)	Not measured	NA	Receptive picture	~0.7–1.2 ES (Receptive measure: learned about 4 words out of 22 measured vs. 1/22 in control group)
Beck and McK. S2 (2007)	Not measured	NA	Receptive picture and Yes/No questions about meaning	~1–2 ES (Receptive measure: ~3 words for rich instruction; ~7–8 words for more rich instruction out of 42)
Biemiller and B. S1 (2006)	Not measured	NA	Expressive definition	~1.2 ES (~5 words for non-instructed condition; ~11 for instructed condition out of 48)
Biemiller and B. S2 (2006)	Not measured	NA	Same as above	d = ~2.5 (~16–24 words out of 42–55 words)
Blewitt et al. S1 (2009)	PPVT	No effects	Receptive and expressive picture	Receptive measure d = 1.7 (Questioning condition learned 6.5 words and no questioning learned 4 out of 9); Expressive measure d = 0.8 (Questioning condition learned 2.5 words and no questioning learned 1 out of 9)
Blewitt et al. S2 (2009)	Not measured	NA	Receptive picture and expressive definition	Receptive measure d = 0.36 and Expressive measure d = 0.44 for scaffolding compared to standard questioning (Receptive measure: learned ~4 out of 9 words)
Coyne et al. (2004)	Not measured	NA	Expressive definition	d = 0.85 (# words UtbD out of 20)
Coyne et al. S1 (2007)	Not measured	NA	Expressive definition; receptive definition; receptive measure of story context	d = 1–2.3 depending on measure (# words UtbD out of 6)
Coyne et al. S2 (2007)	Not measured	NA	Same as above	d = 0.84–2.6 for extended vs. incidental; 0.2–0.9 for embedded vs. incidental (# words UtbD out of 9)
Coyne et al. (2009)	Not measured	NA	Expressive definition; receptive context (2); receptive definition	d = 1.7 (# words UtbD out of 54)
Coyne et al. (2010)	PPVT	No effects	Expressive definition	ES not calculated (Active condition learned 4 words; passive condition learned 2.5 words out of 10)
Ewers and B. (1999)	Not measured	NA	Receptive picture	Receptive d = 1.0; Expressive d = 1.4 (# words UtbD out of 94)
Gonzalez et al. (2010)	PPVT and EOWPVT	Effects on EOWPVT only (d = 0.9)	Receptive and expressive picture	ES not calculated (students learned about 2/16 words)
Hargrave and S. (2000)	PPVT and EOWPVT measured;	Effects only EOWPVT d = 0.21	Expressive picture	
Justice et al. (2005)	Not measured	NA	Expressive definition	d = 0.85 (# words UtbD out of 60)
Leung (2008)	EVT and PPVT	No effects	Expressive definition receptive picture	Expressive measure: d = 0.44 for retelling and rereading (vs. just re-reading) (# words UtbD out of 32); no impacts on receptive measure
Lever & Sénéchal (2011)	Not measured	NA	Expressive picture	d = 0.66 (learned 1 out of 16 words)
Loftus et al. (2010)	Not measured	NA	Word recognition, receptive picture, context, expressive definition	d = 0.4–0.7 (# words UtbD out of 8)
Lonigan and W. (1998)	PPVT, EOWPVT, ITPA-VE	School+ home condition in “high compliance” EOWPVT: d = 0.41; other arms no effect	Not measured	NA

Lonigan et al. (2013)	EOWPVT and CELF	EOWPVT d = 0.2; CELF d = ~0.2	Not measured	NA
McKeown and B. (2014)	Not measured	NA	Receptive Meaning recognition (yes/no), context integration, listening comprehension, expressive picture	d = 0.3–0.7 (Receptive measure – learned about 4–5 words vs. 3–4 in control condition out of 30)
Penno et al. (2002)	Not measured	NA	Receptive picture, retelling	Receptive picture measure d = 0.9 (learned about 2.5 words vs. 0.5 in control out of 10)
Robbins & Ehri (1994)	Not measured	NA	Receptive picture	d = 0.74 (learned 1.2 words out of 11)
Sénéchal et al. S1 (1995)	Not measured	NA	Receptive picture, expressive picture	d = 0.6–0.7 Receptive measure: Listening condition learned 0.7 words, while labeling condition learned 2.1. Expressive measure: listening condition = 0.7 and labeling condition = 2.2 out of 13.
Sénéchal et al. S2 (1995)	Not measured	NA	Same as above	d = 1.0–1.1 (learned less than 1 word out of 10)
Senechal (1997)	Not measured	NA	Receptive picture, expressive picture	d = 0.5–1.8 (learned about 1–3 words out of 10)
Silverman (2007)	Not measured	NA	Receptive picture, expressive definition	d = 0.7–1.2 (learned about 6–7 words in treatment vs. 2–3 words in control out of 30)
Silverman et al. (2013)	PPVT	No effects	Receptive picture	d = 0.8 (treatment learned 3–4 words vs. 1 in control out of 48)
Sim et al. (2014)	Picture naming test and PPVT	Picture naming test d = 0.2	Not measured	NA
Wasik & Bond (2001)	PPVT	d = 0.7	Receptive picture, expressive picture	d = 1.6–2.1 (Receptive measure: Experimental group knew 38/44 vs. Control group 27/44; Expressive measure: Experimental group knew 7/44 vs. Control group 3/44)
Wasik et al. (2006)	PPVT, EOWPVT	PPVT d = 0.7; EOWPVT d = 0.4	Not measured	NA
Weisberg et al. (2015)	Not measured	NA	Receptive picture, expressive definition	d = 0.6–1.3 (# words UtbD out of 57)
Whitehurst et al. (1994)	EO, PPVT, and ITPA	EO has effects d = 0.3 for school + home condition	Expressive picture	School condition d = 0.3; school + home d = 0.6 (control learned 2 words, school learned 4 words, school + home learned 6 words out of 37)
Zipoli et al. (2011)	PPVT measured	No effects	Expressive definition	d = 0.7–1.0 (# words UtbD out of 54)
Zucker et al. (2013)	Not measured	NA	Receptive picture, expressive definition	d = 0.8 (Expressive measure: learned about 7 words vs. 3 out of 26)

to assess whether children learned the words that they were taught in the study (see Table 6). Studies used different types of measures to assess learned vocabulary. In 17 studies, a PPVT-like assessment was administered in which one of the four pictures on each template depicted a target word from that study's books. In 10 studies, an EOWPVT-like assessment was used in which a child had to say the name of a picture. In 15 studies, children were asked to define a target word. Other types of measures included yes/no questions related to word meanings and questions that tested students' understanding of a target word in other contexts.

Of the 31 studies that used project-specific measures to assess word learning (either receptive or expressive or both), 20 included enough information to allow estimation of the number of words learned compared to the number of words assessed. The other 11 studies used multiple measures and/or partial points for scoring word knowledge, which made it difficult to determine the words assessed and learned at post-test.

In the 20 studies, the total number of words assessed at post-test ranged from 9 words, in two studies (Blewitt et al., 2009; Study 1 & 2) each employing 4 sessions over 6 weeks, to 42–55 words, in a study employing 10 sessions over two weeks (Biemiller & Boote, 2006; Study 2).

At post-test, 12 of these studies showed that treatment children learned from 1 to 4 words on the project-specific measure (either receptive or expressive) (Beck & McKeown, 2007; Study 1; Blewitt et al., 2009; Study 2; Ewers & Brownson, 1999; Hargrave & Sénéchal, 2000; Lever & Sénéchal, 2011; McKeown & Beck, 2014; Penno et al., 2002; Robbins & Ehri, 1994; Senechal, 1997; Study 1 & 2; Sénéchal et al., 1995; Silverman et al., 2013). For example, in Beck and McKeown (2007) Study 1, children in the experimental group learned about 4 words out of 22 on a receptive measure, while children in the control group learned 1 out of 22; and in Study 2, children in the Rich Instruction group learned 3 out of 42 words on a receptive measure, while the More Rich Instruction group learned 7–8 words out of 42. Six studies (Beck & McKeown, 2007; Study 2; Biemiller & Boote, 2006, Study 1; Blewitt et al., 2009; Study 1; Silverman, 2007; Whitehurst et al., 1994; Zucker et al., 2013) showed that treatment children ultimately knew between 5 and 11 words. Two studies showed treatment children learned more than 12 words at post-test (Biemiller & Boote, 2006; Study 2; Wasik & Bond, 2001).

Although effect sizes (ES) were appropriately calculated in these studies and were found to be statistically significant, the aforementioned results show that children did not learn many target words. For example, in Beck and McKeown Study 1 (2007), while children learned about 4 out of the 42 target words on the receptive measure (vs. 1 in the control), the ES = 1.0. Similarly, in Lever and Sénéchal (2011), children learned 1 out of the 16 target words on an expressive picture measure, yielding an ES = 0.66. Similarly, in Silverman et al. (2013), children in the intervention learned 3–4 words out of 48 on a receptive picture, while peers in the control condition learned only one word, resulting in an ES = 0.80 (see Table 6).

3.6.2. Standardized measures and outcomes

Of the 36 studies, 14 studies used standardized measures to assess vocabulary learning. Of these 14 studies, 13 studies used the PPVT to assess receptive word learning. Two of the 13 studies found positive effects on the PPVT (Wasik & Bond, 2001; Wasik et al., 2006). Nine studies used a standardized expressive measure (e.g., Expressive One-Word Picture Vocabulary Test, Brownell, 2000; or Expressive Vocabulary Test-2, Williams, 2007). Eight of these nine studies found positive effects on the measures of expressive vocabulary assessments.

3.6.3. Delayed post-testing

Eleven studies measured a delayed post-test or follow-up post-test. Three of these studies did a 1-week follow-up, and the remaining 8 studies conducted delayed post-tests ranging from 6 weeks to 3 months after the study. Sim et al. (2014) and Whitehurst et al. (1994) measured and showed sustained effects on a standardized instrument at follow-up. The remaining studies assessed the target words in the study. Seven of these studies showed sustained effects on children's word knowledge.

4. Discussion

Young children's vocabulary skills represent a critical predictor of their success in reading and in school more broadly (Marchman & Fernald, 2008). Therefore, it is important to understand the specific practices and strategies that parents and teachers can implement to support young children's word learning. Book reading is one of the most widely examined activities for building early vocabulary. However, given that studies employ different samples and methods, the specific characteristics of effective ways to read books to promote word learning are not always clear. The primary goal of this review was to closely examine rigorous, empirical studies that demonstrated positive relations between book reading and vocabulary to elucidate trends regarding who reads the book, what words were targeted, the dosage of exposure to words, what instructional strategies were implemented during or after the reading, what measures were used to assess their effectiveness, and what evidence emerged from those measures. Below, we describe themes that emerged, and we outline questions that remain so as to guide research to support more specific implications for practice.

4.1. Adult-child interaction during book reading boosts word learning

As perhaps the primary take-away message for practice, this systematic review and previous meta-analyses revealed that adult-child interaction during book reading is critical for vocabulary learning to occur. In several studies, book reading without adult input was used as the control group, resulting in limited vocabulary learning – almost at the level of chance. Interestingly, when adults do not provide any input during book reading, children may try to engage the adult by asking them questions about the book (Mautte, 1990). This finding implies that children may perceive book reading as a natural opportunity for interaction, and that advising adults to systematically foster interaction in strategic ways takes advantage of this opportunity. More specifically, across these studies, strategies for engaging children in the narrative and/or the target words involved (a) defining words, (b) questioning as a means to promote discussion of vocabulary and comprehension, (c) re-telling, (d) re-reading, (e) using props to illustrate word meanings, and (f) providing extension activities that promote exploration and discussion of vocabulary. Broadly speaking, word learning was enhanced when adults asked questions and engaged children in discussion about target vocabulary words, relative to simply recasting the meaning of the words (e.g., Ewers & Brownson, 1999; Sénéchal et al., 1995). This principle appears to extend to repeated readings (Robbins & Ehri, 1994), with one study showing that simply re-reading without interaction is not as effective as combining re-reading with other strategies. This pattern of results is congruent with cognitive principles (Pashler et al., 2007), suggesting that children learn the content that they have explicit opportunities to rehearse and practice (i.e., meanings of particular words).

However, questions abound regarding the micro-level processes by which these strategies are delivered. With the exception of two studies, these interventions implemented combinations of

strategies, making it difficult to tease out the effects of individual strategies on word learning. In fact, most studies implemented a combination of three strategies: (a) defining words, (b) questioning as a means to promote discussion of vocabulary and comprehension, and (c) re-reading. We tried to determine if any strategy combination was more effective than others, but this was difficult to discern because several studies implemented similar combinations of strategies and found different results. For example, Coyne et al. (2007, 2009, 2010), Wasik and Bond (2001), Wasik et al. (2006), and Silverman et al. (2013) implemented similar strategies yet found different amounts of word learning in children. This suggests that some nuances concerning implementation, teacher training, dosage of treatment, and/or variability in how word learning was measured could explain the differences in outcomes.

Also, descriptions of the strategies implemented in the studies are often described very broadly, using phrases such as, “children were provided with definitions of words.” We know from Booth (2009) that when children are provided with definitions that include the function of the word, they are more likely to learn the word. Yet, the studies reviewed do not provide enough detail at this level to afford replication of the interventions in classrooms. If fidelity of implementation measures were available for these studies, researchers could perform more cross-study comparisons that might provide insight into what strategies are most effective.

Studies examining the nuances of strategy use [e.g., the print referencing study of McGinty, Breit-Smith, Fan, Justice, and Kaderavek (2011)] may find that there are multiple combinations of strategy use and dosage that will result in children’s word learning. Confirming this finding in additional studies would make a significant contribution to teachers and parents who want to promote word learning.

4.2. Effects of interventions featuring book reading are modest

Although meta-analyses clearly show that book reading makes a reliable contribution to children’s vocabulary development (explaining about 8–10% of the variance), it is important to note that, across all of the high-quality studies with positive effects of book reading examined in this critical review, children are still not learning all of the words that they were taught, and in some instances they are learning relatively few words.

4.2.1. Measuring words taught

Of the 36 studies, 31 used project-specific measures of the words taught to assess children’s vocabulary growth, many of which were similar to the PPVT. Results showed that, in many instances, they are learning less than 25% of the words regardless of the strategies that adults used, the number of books to which children were exposed, and the duration of the intervention (see Table 6). For example, in Lever and Sénéchal (2011), children on average learned 1 out of the 16 words. That said, word learning was greater in some other studies: In Biemiller and Boote (2006), for example, children learned almost half of the words with which they were presented. Two studies (Blewitt et al., 2009 Study 1; Wasik & Bond, 2001) at first glance appear to stand out because they showed children in the intervention condition learning most of the words to which they were exposed through book reading (i.e., 72% and 86%, respectively). However, children in the control groups in these two studies also knew many words (i.e., 44% and 61%, respectively), which clearly qualifies the strength of the intervention. However, among most studies that reported delayed post-testing, children’s knowledge of target words was either maintained or only slightly diminished. This suggests that, among these studies, when words were learned (even in modest numbers), knowledge was maintained over several weeks or months.

Although reporting effect sizes is very important, given these relatively small gains in absolute numbers of words learned, effect sizes may not present the entire picture of word learning. For example, in Penno et al. (2002), an effect size of 0.9 (which is large) is reported; yet, this value emerges because treatment children had learned 2.5 out of 10 words on the expressive measure, as compared to the control children who learned 0.5 out of 10 words. The difference, only two additional words learned, may lead to statistical significance but may not be practically substantial. In fact, in all studies that measured target word learning, children learned only a small percentage of words tested, and in general, children in the experimental condition were not very different from peers in a control condition. Thus, the field needs to investigate what these modest gains mean, whether they are educationally significant, and what can be done to increase children’s word learning. Perhaps reporting the number of words learned compared to the number of words taught, along with effect sizes, may result in a more sensitive measure of the treatment.

Finally, assessing taught words using receptive measures is highly feasible for young children and easy for researchers to score, but this approach (particularly when using a PPVT-like model involving choosing one of four options) includes the drawback of a 25% chance of a correct response, leaving some uncertainty in results about children’s deep knowledge of the word. However, expressive measures (especially non-standardized tools) can introduce subjectivity in scoring the words learned. For example, in the studies we reviewed, some researchers gave full points for children who had a semi-understanding of a word (Biemiller & Boote, 2006), whereas others gave only partial points for such answers (McKeown & Beck, 2014). How word knowledge is measured and scored will affect how outcomes are analyzed and reported. Perhaps book reading interventions should assess taught word learning – both at an immediate and a delayed post-test – using a combination of receptive and expressive measures to triangulate data.

Note that these findings support our earlier contention that much more needs to be learned about the nature of vocabulary learning and effective interventions. To a large extent, the findings reflect the need for revision and refinement in contemporary theories of word learning. If theories were sufficiently accurate and refined, mapping interventions onto these theories would be relatively straightforward. The fact that well-conceived approaches to book reading only had modest effects suggests that something is clearly missing in the current understanding of how and why children learn (or do not learn) words presented to them in meaningful and repeated contexts.

4.2.2. Standardized vocabulary growth

Studies also used more distal measures to assess word learning. Inherent in this approach is the idea, and perhaps hope, that learning words during book reading teaches children to pay attention to more words and therefore supports generalized word learning which, in turn, raises performance on standardized measures of words, closing the socioeconomic word gap. Fourteen of these studies assessed children on the PPVT, and only two found significant effects (Wasik & Bond, 2001; Wasik et al., 2006). Conversely, nine studies assessed children using expressive standardized measures, and eight found significant effects (ranging from $+0.9$ to 0.2 ; See Table 6). Interestingly, some studies that found no effects on standardized receptive measures (generally the PPVT) did find effects on standardized expressive measures (e.g., Gonzalez et al., 2010; Hargrave & Sénéchal, 2000).

One potential implication of this pattern of inconsistent improvement on standardized measures is that, if a primary goal of book reading is to build vocabulary, it may be necessary to consider book reading as one beneficial part of a larger system of instruction, including other activities (e.g., small groups, play-based centers,

coordinated home and school exposures) that work together to foster learning of target words. Indeed, it may be unrealistic to think that young children could learn all of the words they are presented with during one, or even three, book readings; consequently, to get children beyond learning just one out of 10 words, or just two out of 16 words, exploring how book reading operates as one part of a larger network of experiences will be key. Currently, 14 studies considered book reading in this way, and more research on this front is needed (see below). However, as above, variation in some key aspects of the studies may attenuate effects of reading on vocabulary; if the field could clarify precisely which strategies were best, effects might further increase. The points below endeavor to respond to this latter issue.

4.3. *Wide variation in the quality of the studies*

Among the most fundamental findings from this work is that, although all of the studies included in this systematic review were initially selected based on rigorous design criteria (i.e., a randomized controlled trial and/or a pretest–treatment–posttest comparison with a control group), there was variation in quality, particularly around internal validity. These findings are similar to those reported by Justice, Nye, Schwarz, McGinty, and Rivera (2008), who also found a high level of variability and scores on the four Downs and Black (1998) indicators in similar ranges. A key implication for future research is the need to improve levels of internal validity. Studies that, when feasible, use random assignment, keep assessors blind to treatment condition, keep teachers blind to condition, and take attrition into account during data analysis would improve internal validity.

In addition, only half of the studies reviewed included a fidelity of implementation measure, making it difficult to know how the identified strategies were implemented. This may be among the most important points for future work: because nuances in how treatments are delivered may attenuate (or strengthen) child outcomes, book reading researchers need to document and disseminate information about the nature and extent of implementation fidelity in book reading treatments, as well as how many people in the treatment achieved fidelity. This content will help researchers and practitioners understand which specific practices represent essential pathways through which a treatment increases children's vocabulary outcomes.

4.4. *Wide variation in the number of words targeted via book reading*

A second main finding that sheds new light on the nature of the research in this area is the discovery that studies varied greatly in the number of target words for a book reading (i.e., from 2 per book to 20 per book). Put another way, some studies endeavored to teach approximately three times as many words through the same (approx. 10–20 min) experience as others. The number of words targeted may be one of the most important variables to understand more fully, because identifying an optimal number, or range, that supports child learning would inform helpful guidance for teachers and parents.

4.5. *Nature (part of speech, morphology) of target words remains unclear*

Criteria for selecting words appeared to be consistent across studies, with many studies focused on words children may not be familiar with and/or words that fit Beck's word tiers. However, we know from language research that factors such as parts of speech and morphology can affect children's learning of that word (Byrnes & Wasik, 2009; Guo, Roehrig, & Williams, 2011). To date, we found

no studies that specifically examined these facets of the target words, even though all but two studies included a mix of parts of speech, and all studies included words with a mix of morphological features. By extension, no studies examined whether these facets were linked to which words children actually learned. It is possible that, given that nouns are typically easier for young children to learn, whereas verbs and adverbs may be more abstract and thus more difficult (Byrnes & Wasik, 2009; Fenson, Dale, Bates, Reznick, & Thal, 1994), these programs are more successful with nouns. Similarly, no attention was devoted in these studies to the morphology of the words taught or learned, even though there is research suggesting that the morphological structure of a word can affect word learning (Byrnes & Wasik, 2009). Remaining questions include: Should teachers and parents systematically select target words for book readings that represent all parts of speech, or include more nouns because they are easier to learn, or focus on verbs because they are harder to learn? Knowing this would help guide word choice and support word learning.

4.6. *Frequency of word exposures and dosage of strategies are not clear*

Across this set of studies, details were largely absent regarding how often children were exposed to particular words. There were some indications that more dosage was beneficial: all studies that impacted standardized measures had generally greater dosage and duration, as they ranged from being implemented for 5 weeks to a full school year. However, studies with similar levels of dosage did not always produce similar effects on standardized measures. One likely reason is that there were very different approaches to word exposures across studies. For example, in most studies, a set of words were selected for a particular book, and then all of those words were targeted each time the book was read. For example, Gonzalez et al. (2010) focused on 2–3 words per book reading and read a total of 36 books in their intervention. Biemiller and Boote (2006; Study 1), however, took a different approach, as they focused on 12–24 words per book and targeted 4–6 words for each reading of the book. Although these two approaches may appear to differ in only subtle ways, the opportunity to emphasize the target words is greater in the former and less in the latter, which could have an impact on children's learning. This issue may interact in meaningful ways with word type and with strategy use; for example, one could ask: might children need twice the exposure to verbs in order to learn them, relative to nouns, particularly in the presence of a prop that clearly depicts the noun? Although cognitive development has found individual differences in the number of times a child needs to be exposed to a word before he or she learns it (Tomasello, 2009), it would be helpful to identify the range of dosage needed for word learning, and especially how this may differ systematically across different types of words.

Also important, particularly for practitioners, is how often particular strategies were used to provide those exposures. For example, research on print referencing (McGinty et al., 2011) has shown that fewer print referencing exposures per session, along with higher dosages of sessions, were associated with larger gains in children's learning about print. No book reading study addressed strategy dosage, which may be part of why results from this review showed varied effects of the same strategies across studies. For future research, pressing questions about strategies include: Is questioning during each book reading as effective as questioning only on the 2nd and 3rd readings of a book? How many questions during a reading are optimal, and can asking too many questions begin to interfere with word learning and children's overall comprehension of the book?

4.7. Most research has focused on classrooms

This review found that the majority of rigorous book reading studies have been done in classroom settings with teachers and researchers implementing the book reading (see Table 3). In fact, only two studies in the corpus of high-quality empirical projects that we reviewed were conducted entirely in homes by parents. Given what we know about the importance of what children bring to preschool, more high-quality research needs to examine the role of parents in book reading and vocabulary development. Some research has attempted to examine this issue (Hindman, Skibbe, Miller, & Zimmerman, 2010), but most studies use parent report to determine book reading experience. Consequently, more observational measures, especially for use in the context of family-focused book reading interventions, are needed. Finally, early research has examined parent behaviors and reading, but without a particular focus on vocabulary (Aram & Levin, 2002). Given the importance of vocabulary and the unique benefits of book reading for this outcome, more research on this front is warranted. In turn, studies of this nature could help expand the relatively small number of interventions that rely on parents to enhance children's vocabulary development (e.g., PALS; Landry, Smith, Swank, & Guttentag, 2008).

5. Limitations

There are several limitations to this systematic review. One is that, although one of our aims was to inform practice, the nature of the studies we reviewed – especially their variation on multiple dimensions including the strategies used, the dosage provided to children, and the duration of the treatment – impeded our ability to distill clear guidance for practice and further research. Second, this paper focused narrowly on book reading and vocabulary development, rather than on reading comprehension and other outcomes, about which more guidance is needed to inform both research and practice. Third, we only selected papers that had resulted in positive effects on a measure of vocabulary. Although we chose to do this to facilitate an understanding of strategies that positively impact language development in children, only including “successful” studies could increase the risk for over-selling of results. However, since we found smaller effects across the studies than expected, the risk of overselling is largely mitigated. Fourth, in order to decrease the heterogeneity of the study samples, we chose to focus more narrowly on native English speaking children and children without diagnosed language disorders. Unfortunately, this means that our findings may not be generalized to these other, highly significant subgroups of children. We note, however, that there was a fairly wide variation in children's initial language proficiency among those studies that provided standardized scores.

6. Conclusion

Book reading is widely acknowledged in the field as an important vehicle for developing vocabulary in young children. Our systematic review of the research, aimed at unpacking what we know about conditions and strategies that lead to helping children learn words, suggests that more fine-grained analysis of book reading would help provide guidance to parent and teachers who want to know what, specifically, to do with children in order to increase vocabulary learning. This work also highlights the need for additional research to answer some critical questions about what parents and teachers should do during readings to increase children's vocabulary. In particular, we found that children learned relatively few of the total number of words upon which they were post-tested, even when treatments ran for several months. More

intensive, strategic approaches, informed by new research, are clearly needed to accelerate vocabulary learning, especially for children whose language skills lag behind. To a large extent, the pattern of results reflects two factors: the lack of a sufficiently accurate theoretical model of vocabulary acquisition and the lack of coordination among teams of researchers. Theoretical shortcomings could probably be best addressed in systematic lines of laboratory research in which key variables (dosage, frequency and manner of exposure, etc.) are strategically varied across experimental conditions. The lack of coordination could be addressed via funding agencies that recruit teams of researchers to define the landscape of needed studies and then focus on different aspects of the problem at multiple sites. It is only then that we will have the kinds of clear information needed to inform practice.

References

- Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416–430.
- Aram, D., & Levin, I. (2002). Mother-child joint writing and storybook reading: relations with literacy among low SES kindergartners. *Merrill-Palmer Quarterly, 48*(2), 202–224.
- Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: replication and extension to a videotape training format. *Journal of Educational Psychology, 86*(2), 235–243.
- Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251–271.
- Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown, & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147–163). NJ: Erlbaum: Hillsdale.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life*. New York, NY: Guilford Press.
- Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. <http://dx.doi.org/10.1037/00220663.98.1.44>
- 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: Guilford.
- Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: when and how questions affect young children's word learning. *Journal of Educational Psychology, 101*(2), 294–304. <http://dx.doi.org/10.1037/a0013844>
- Booth, A. E. (2009). Causal supports for early word learning. *Child Development, 80*(4), 1243–1250.
- Brown, A., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering and understanding. In J. H. Flavell, & E. M. Markman (Eds.), *Handbook of child psychology* (Vol. 3) (pp. 77–166). New York: John Wiley.
- Brownell, R. (2000). *Expressive one-word picture vocabulary test* (3rd ed.). Academic Therapy Publications.
- Byrnes, J. P., & Wasik, B. A. (2009). *Language and literacy development: what educators need to know*. New York, NY: Guilford Press.
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Clay, M., Gill, M., Glynn, T., McNaughton, T., & Salmon, K. (2007). *Record of oral language: observing changes in the acquisition of language structures* (3rd ed.). New Zealand: Heinemann.
- Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: an examination of differential effects. *Exceptionality, 12*(3), 145–162.
- Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74–88.
- Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: teaching for breadth versus depth. *The Elementary School Journal, 110*(1), 1–18. <http://dx.doi.org/10.1086/598840>
- Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., Ruby, M., Crevecoeur, Y. C., et al. (2010). Direct and extended vocabulary instruction in kindergarten: investigating transfer effects. *Journal of Research on Educational Effectiveness, 3*(2), 93–120. <http://dx.doi.org/10.1080/19345741003592410>
- Currie, J., & Neidell, M. (2007). Getting inside the black box of Head Start quality: what matters and what doesn't. *Economics of Education Review, 26*(1), 83–99.
- Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language: young children learning at home and school*. Baltimore, MD: Paul H Brookes Publishing.
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*(6), 377–384.

- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350.
- Ewers, C. A., & Brownson, S. M. (1999). Kindergarteners' 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(3), 464–497.
- Fenson, L., Dale, P. A., Bates, E., Reznick, J. S., & Thal, D. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59. Serial No. 242, No. 5.
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248.
- Ganea, P. A., Pickard, M. B., & Deloache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, 9(1), 46–66.
- Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S., & Small, S. (2014). New evidence about language and cognitive development based on a longitudinal study: hypotheses for intervention. *American Psychologist*, 69(6), 588–599. <http://dx.doi.org/10.1037/a0036886>
- Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., et al. (2010). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness*, 4(1), 25–52.
- Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, 43(2), 159–183.
- Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: a critical review. *Early Childhood Research Quarterly*, 28, 843–857.
- 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(1), 75–90.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of American children*. Baltimore: Paul H Brookes Publishing.
- Heckman, J., Pinto, R., & Saveliev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052–2086.
- Hindman, A. H., Skibbe, L. E., Miller, A. L., & Zimmerman, M. (2010). Ecological contexts and early learning: contributions of child, family, and classroom factors during Head Start to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, 25, 235–250.
- Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: an efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools*, 36(1), 17–32.
- Justice, L. M., Nye, C., Schwarz, J., McGinty, A., & Rivera, A. (2008). Methodological quality of intervention research in speech–language pathology: analysis of 10 years of group-design studies. *Evidence-Based Communication Assessment and Intervention*, 2(1), 46–59.
- Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. (2008). A responsive parenting intervention: the optimal timing across early childhood for impacting maternal behaviors and child outcomes. *Developmental Psychology*, 44(5), 1335–1353.
- Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, 101(2), 448.
- Layzer, J., & Price, C. (2008). Closing the gap in the school readiness of low-income children. In *Paper presented at the a working meeting on recent school readiness research: guiding the synthesis of early childhood research*.
- Leung, C. B. (2008). Preschoolers' acquisition of scientific vocabulary through repeated read-aloud events, retellings, and hands-on science activities. *Reading Psychology*, 29(2), 165–193.
- Lever, R., & Sénéchal, M. (2011). Discussing stories: on how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, 108(1), 1–24.
- Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, 25(3), 124–136.
- Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, 13(2), 263–290.
- Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology*, 114(1), 111–130.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3), F9–F16.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Review of Educational Research*, 80(3), 300–335.
- Marulis, L., & Neuman, S. (2013). How vocabulary interventions affect young children at risk: a meta-analytic review. *Journal of Research on Educational Effectiveness*, 6(3), 223–262.
- Mautte, L. A. (1990). The effects of adult-interactive behaviors within the context of repeated storybook readings upon the language development and selected prereading skills of prekindergarten at risk students. *Florida Educational Research Council Research Bulletin*, 22(3), n3.
- McGinty, A. S., Breit-Smith, A., Fan, X., Justice, L. M., & Kaderavek, J. N. (2011). Does intensity matter? Preschoolers' print knowledge development within a classroom-based intervention. *Early Childhood Research Quarterly*, 26(3), 255–267.
- McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools*, 38(4), 353–364.
- McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: comparing two approaches. *Early Childhood Research Quarterly*, 29(4), 520–530.
- Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Science*, 7, 141–144.
- Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent–child book readings: a meta-analysis. *Early Education and Development*, 19(1), 7–26.
- Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: a tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, 79(2), 979–1007.
- National Early Literacy Panel. (2008). *Developing early literacy: report of the national early literacy panel*. Washington, DC: National Institute for Literacy. Available at: <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., & McDaniel, M. (2007). *Organizing instruction and study to improve student learning. IES practice guide. (NCER 2007-2004) U.S. department of education*. Washington DC: National Center for Education Research, Institute of Education Sciences.
- Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: do they overcome the Matthew Effect? *Journal of Educational Psychology*, 94(1), 23–33.
- Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, 86(1), 54–64.
- Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, 87(2), 218–229.
- Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, 24(1), 123–138.
- Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14(3), 245–302.
- Silverman, R., Crandell, J. D., & Carlis, L. (2013). Read alouds and beyond: the effects of read aloud extension activities on vocabulary in Head Start classrooms. *Early Education & Development*, 24(2), 98–122.
- Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *The Elementary School Journal*, 108(2), 97–113.
- Sim, S., Berthelsen, D., Walker, S., Nicholson, J. M., & Fielding-Barnsley, R. (2014). A shared reading intervention with parents to enhance young children's early literacy skills. *Early Child Development and Care*, 184(11), 1531–1549.
- Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education*, 6(1), 5–10.
- Sternberg, R. J. (1979). The nature of mental abilities. *American Psychologist*, 34(3), 214–230.
- Sternberg, R. J. (1982). A componential approach to intellectual development. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum Associates.
- Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, 23(2), 121–126.
- Tomasello, M. (2009). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). http://www.nationsreportcard.gov/reading_math_2015/#reading?grade=4
- Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, 93(2), 243–250.
- Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, 98(1), 63–74.
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, 84(6), 2112–2130.
- Weisberg, D. S., Ilgaz, H., Hirsh-Pasek, K., Golinkoff, R., Nicolopoulou, A., & Dickinson, D. K. (2015). Shovels and swords: how realistic and fantastical themes affect children's word learning. *Cognitive Development*, 35, 1–14.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69(3), 848–872.

- Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, 30(5), 679–689.
- Williams, K. T. (2007). *Expressive vocabulary test* (2nd ed (evt-2)). San Antonio, TX: Pearson.
- Xuan, L., & Dollaghan, C. (2013). Language specific noun bias: evidence from bilingual children. *Journal of Child Language*, 40, 1057–1075.
- Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: strategic integration of semantically related and embedded word review. *Remedial and Special Education*, 32(2), 131–143.
- Zucker, T. A., Solari, E. J., Landry, S. H., & Swank, P. R. (2013). Effects of a brief tiered language intervention for prekindergartners at risk. *Early Education & Development*, 24(3), 366–392.