LSHSS

Research Article

Interactive Book Reading to Accelerate Word Learning by Kindergarten Children With Specific Language Impairment: Identifying an Adequate Intensity and Variation in Treatment Response

Holly L. Storkel,[a] Krista Voelmle,[a] Veronica Fierro,[a] Kelsey Flake,[a] Kandace K. Fleming,[a] and Rebecca Swinburne Romine[a]

Purpose: This study sought to identify an adequate intensity of interactive book reading for new word learning by children with specific language impairment (SLI) and to examine variability in treatment response.
Method: An escalation design adapted from nontoxic drug trials (Hunsberger, Rubinstein, Dancey, & Korn, 2005) was used in this Phase I/II preliminary clinical trial. A total of 27 kindergarten children with SLI were randomized to 1 of 4 intensities of interactive book reading: 12, 24, 36, or 48 exposures. Word learning was monitored through a definition task and a naming task. An intensity response curve was examined to identify the adequate intensity. Correlations and classification accuracy were used to examine variation in response to treatment relative to pretreatment and early treatment measures.
Results: Response to treatment improved as intensity increased from 12 to 24 to 36 exposures, and then no further improvements were observed as intensity increased to 48 exposures. There was variability in treatment response: Children with poor phonological awareness, low vocabulary, and/or poor nonword repetition were less likely to respond to treatment.
Conclusion: The adequate intensity for this version of interactive book reading was 36 exposures, but further development of the treatment is needed to increase the benefit for children with SLI.

Specific language impairment (SLI) affects 7.4% of kindergarten children (Tomblin et al., 1997). Children with SLI experience significant deficits in language acquisition relative to their peers in the absence of any obvious causal factor. Although many aspects of language can be affected in SLI, word learning deficits are frequently observed (Alt & Plante, 2006; Alt, Plante, & Creusere, 2004; Dollaghan, 1987; Gathercole, 1993; Gray, 2003, 2004, 2005; Nash & Donaldson, 2005; Rice, Buhr, & Nemeth, 1990; Rice, Oetting, Marquis, Bode, & Pae, 1994). In fact, children with SLI appear to need two to three times as many exposures as their peers to learn a new word (Gray, 2003; Rice et al., 1994). This is troubling given that spoken language vocabulary is a strong predictor of later success in reading (Catts, Fey, Tomblin, & Zhang, 2002; Scarborough, 1998). Thus, ameliorating vocabulary deficits by accelerating word learning may assist in preventing future reading deficits and subsequent academic failure. Although there are numerous empirical studies of word learning by children with SLI, there are few, if any, interventions with demonstrated effectiveness (Cirrin & Gillam, 2008; Steele & Mills, 2011). To illustrate, a systematic review by Cirrin and Gillam (2008) located only six intervention studies of vocabulary, and all were nonrandomized comparison studies.

One effective treatment for word learning generally is interactive book reading, which involves an adult reading a storybook to a child and deviating from the text to provide additional explicit instruction (e.g., define the new word). It is important to note that randomized clinical trials, meta-analyses, and systematic reviews show that interactive book reading has moderate to large effects on word learning by

[a]University of Kansas, Lawrence
Correspondence to Holly L. Storkel: hstorkel@ku.edu
Editor: Shelley Gray
Associate Editor: Margaret Kjelgaard
Received February 23, 2016
Revision received May 24, 2016
Accepted May 27, 2016
DOI: 10.1044/2016_LSHSS-16-0014

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

typically developing children and children with low vocabulary due to environmental differences in input (i.e., children from low-income families; Justice, Meier, & Walpole, 2005; Marulis & Neuman, 2010; Mol, Bus, & de Jong, 2009; Mol, Bus, de Jong, & Smeets, 2008; Whitehurst et al., 1988). Moreover, the intervention can be administered by a variety of adults (e.g., parents, teachers) with minimal training—a desirable quality given the shortage of speech-language pathologists (American Speech-Language-Hearing Association [ASHA], 2011). Although interactive book reading demonstrates compelling results in these groups of children and has particularly desirable delivery features, initial tests of this intervention with children with different types of language impairments show less clear and compelling results (Crain-Thoreson & Dale, 1999; Dale, Crain-Thoreson, Notari-Syverson, & Cole, 1996; Pile, Girolametto, Johnson, Chen, & Cleave, 2010; Whitehurst et al., 1991). As a consequence, interactive book reading needs to be optimized in a Phase I/II preliminary clinical trial before it will be effective in addressing the word learning deficits of children with SLI. A Phase I/II preliminary clinical trial is a preliminary study addressing core design and clinical issues (e.g., intensity, dosing regimen, magnitude, and extent of benefit) necessary to establish the feasibility of a Phase III/IV definitive clinical trial that, in turn, tests the efficacy and effectiveness of the intervention (Fey & Finestack, 2009; Robey, 2004). The current study has elements of both Phase I and Phase II preliminary clinical trials. In terms of Phase I, the study used a specialized research design to establish the adequate intensity of interactive book reading for children with SLI. This does not establish the efficacy or effectiveness of the treatment but rather uses data to empirically establish one parameter of the treatment: treatment intensity. In terms of Phase II, this study explored the extent of benefit of the treatment across children. Again, this does not establish the efficacy or effectiveness but rather determines whether the treatment needs further modifications to improve the extent of benefit before moving to a large-scale and costly definitive clinical trial.

Treatment Intensity

To apply interactive book reading to the treatment of word learning deficits experienced by children with SLI, it is crucial to consider whether the intensity of prior interactive book reading treatments is likely to be adequate for children with SLI. Warren, Fey, and Yoder (2007) defined *intensity* as the total number of teaching episodes accumulated at the end of treatment. This definition is an improvement over past definitions, which have focused on proxy measures such as the number of treatment sessions, which do not precisely reflect the amount of teaching. For interactive book reading, *intensity* is defined as the total number of exposures to each word being taught. An intensity that is effective for typically developing children is unlikely to be adequate for children with SLI. Children with SLI learn significantly fewer words than typically developing children when given the same intensity of training (Alt & Plante, 2006; Alt et al., 2004; Dollaghan, 1987; Gray, 2003, 2004, 2005; Rice et al., 1990, 1994). Moreover, intensities that are effective for children with low vocabulary due to differences in input (i.e., children from low-income homes) are unlikely to be adequate for children with SLI. Low-income children do not appear to have deficits in learning characteristic of SLI (Horton-Ikard & Weismer, 2007). Taken together, it is unclear what intensity of interactive book reading would be adequate for children with SLI. Thus, the current research used a Phase I escalation strategy previously used in drug trials (Hunsberger, Rubinstein, Dancey, & Korn, 2005) to hone in on an adequate treatment intensity of interactive book reading for word learning by children with SLI. The hypothesis was that children with SLI will need an intensity that is two or three times greater than the effective intensity for typically developing children.

Phase I trials that address issues of treatment intensity are standard in the drug treatment literature but surprisingly rare in the development of behavioral treatments. Identifying an adequate treatment intensity is a critical early step in clinical research. That is, when a treatment is given at too low of an intensity, minimal improvement will be observed on outcome measures, making it impossible to evaluate the efficacy or effectiveness of the overall treatment or of specific manipulations of treatment components. This study adapted an intensity escalation strategy used in nontoxic Phase I drug trials (Hunsberger et al., 2005) to interactive book reading. There are two basic tenets of this design. The first is to use as few participants as possible, limiting the number of children who potentially receive ineffective intensities of the treatment. The second is to identify an adequate—rather than optimal—intensity. An adequate intensity is an intensity in the plateau of an intensity–response curve. That is, it is the point where simply providing more of the intervention fails to produce greater benefit. The basis of this tenet is that finding an optimal intensity typically requires many more participants than finding an adequate intensity. The overall goal of these designs is to move as quickly as possible from Phase I intensity-finding studies to trials that evaluate efficacy. In this study, we attempted to identify the point where simply providing more exposures to the words during interactive book reading fails to produce more children with SLI responding to the treatment and/or more words being learned during treatment.

Variability in Response to Treatment

An important related issue when developing interactive book reading for children with SLI is the uniformity of treatment response across children. There is controversy in interactive book reading research regarding which typically developing children benefit the most from the intervention. Some studies show that children with poorer pretreatment vocabulary test scores benefit more than children with better pretreatment scores (Elley, 1989; Justice et al., 2005), whereas other studies report the opposite pattern (Blewitt, Rump, Shealy, & Cook, 2009; Penno, Wilkinson, & Moore, 2002; Robbins & Ehri, 1994). It is notable that only the study

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

by Justice et al. (2005) tested children with vocabulary scores in a clinically significant range (i.e., scores 1 *SD* or more below the mean). Thus, it is unclear how variation in language characteristics of children with SLI would affect treatment outcomes. It is useful to explore variability in treatment response early in developing an intervention (i.e., in Phase II) so that components of the intervention can be modified to increase the benefit across the population prior to conducting a large-scale clinical trial. Therefore, this study explored how pretreatment or early treatment characteristics were related to treatment response to facilitate further development of interactive book reading for children with SLI.

Method

Additional details about the method, as referenced below, are available in the KU ScholarWorks archive at http://hdl.handle.net/1808/20313. Additional details relevant to the participant characteristics and results, as referenced below, are available in the online supplemental materials.

Participants

Twenty-seven kindergarten children with SLI (age: M = 5;8 [years;months], SD = 0;6, range = 5;0–6;5) participated. Of these children, one child missed posttesting. Thus, this child's data appear in analyses using treatment data but not in analyses that are based on posttreatment data. Children were recruited by language screenings (52%), referral from speech-language pathologists or teachers (41%), or public announcement (7%). This mix of recruitment sources yielded a relatively equal sample of girls (52%) and boys (48%), which differs from prevalence rates reported in epidemiology studies (e.g., 8% vs. 6% prevalence of SLI in boys vs. girls according to Tomblin et al., 1997). Although it is thought that differing prevalence of SLI by gender is not due to referral bias (Tomblin et al., 1997), our screenings yielded greater recruitment of girls (57%) than boys (43%), and referrals by speech-language pathologists or teachers yielded greater recruitment of boys (55%) than girls (45%).

Children were required to (a) be enrolled in or eligible for kindergarten; (b) pass a hearing screening (ASHA, 1997); (c) score at or above the 16th percentile for nonverbal cognition as measured by the Reynolds Intellectual Assessment Scale (Reynolds & Kamphaus, 2003); (d) have a Core Language Score at or below the 10th percentile on the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003); and (e) score at or below the 10th percentile on at least one of three vocabulary measures: Diagnostic Evaluation of Language Variation (DELV) Semantic subtest (Seymour, Roeper, de Villers, & de Villers, 2005), Clinical Evaluation of Language Fundamentals–Fourth Edition Word Classes subtest (Semel et al., 2003), or Comprehensive Receptive and Expressive Vocabulary Test–Third Edition (Wallace & Hammill, 2013). Multiple measures of vocabulary were used to identify word learning deficits due to the inherent difficulty in identifying word learning deficits (e.g., Gray, Plante, Vance, & Henrichsen, 1999) and due to variability in the measures that predict word learning success (Gray, 2003, 2004). Table 1 shows the participant characteristics on these inclusionary measures as well as additional measures intended to further characterize the abilities of these children. The majority of participants (67%) qualified as having poor word learning on the basis of one vocabulary test. Fewer children qualified as having poor word learning on the basis of two (26%) or all three (4%) vocabulary tests. Table 1 shows the percentage of children who qualified on each vocabulary test (i.e., the percentage of children with a percentile rank at or below 10), with the majority of children qualifying on the basis of the DELV Semantic subtest. One child did not meet the vocabulary criteria but was allowed to participate. This child earned a raw score of 1 on the Comprehensive Receptive and Expressive Vocabulary Test–Third Edition (Wallace & Hammill, 2013), which corresponded to the 13th percentile. A raw score of 0 on the test corresponded to the 7th percentile. Thus, there was a gap in the scale such that the child could not score at the 10th percentile.

Race, ethnicity, parent marital status, and parent education of the children generally matched the demographics of the recruitment area (i.e., eastern Kansas). Participants had the following characteristics: 63% White–non-Hispanic, 19% White–Hispanic, 11% Black/African American–non-Hispanic, 4% White–unknown ethnicity, and 4% unknown race and ethnicity. In terms of parent characteristics, 70% of parents were married, 19% were single, and 11% were divorced. In terms of the mothers' education, 37% had partial college, 30% were college graduates, 22% were high school graduates, 4% had partial high school, 4% had graduate degrees, and 4% were unknown. In terms of the fathers' education, 37% were not reported (mostly from the single or divorced families), 22% were high school graduates, 22% had partial college, 7% were college graduates, 4% completed junior high school, 4% had partial high school, and 4% had graduate degrees.

Children were randomly assigned to one of four treatment intensity conditions (described below). Supplemental Tables S1a and S1b show that participants in different conditions were relatively similar.

Treatment

The treatment was based on Justice et al.'s (2005) clinical trial of interactive book reading with kindergarten children from low-income homes. Results of the prior study showed a large effect (d = 1.34) of interactive book reading (with elaboration) compared with a no-treatment control condition for the children with low vocabulary. Likewise, 77% of children made meaningful gains, operationalized as a minimum gain of 4 points on a pretreatment–posttreatment definition test of the treated words.

Treatment Materials

Materials were taken from Justice et al. (2005), who selected 10 commercially available and kindergarten-appropriate

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Table 1. Percentile scores for participants on standardized clinical tests.

Test	*M*	*SD*	Range	At or below 10th percentile (%)
RIAS Nonverbal IQ	55	25	23.0–99.0	0
CELF Core Language	3	3	0.1–10.0	100
Vocabulary: DELV Semantic	9	8	0.1–25.0	74
Vocabulary: CELF Word Classes	24	19	1.0–75.0	37
Vocabulary: CREVT Expressive	26	16	1.0–63.0	19
CELF Concepts and Following Directions[a]	7	7	0.1–25.0	85
CELF Word Structure[a]	11	8	0.4–25.0	52
CELF Recalling Sentences[a]	5	6	0.1–25.0	93
CELF Formulating Sentences[a]	8	7	0.1–25.0	78
CELF Understanding Spoken Paragraphs	7	8	0.1–25.0	81
CTOPP Nonword Repetition	22	21	1.0–75.0	44
CTOPP Phonological Memory	15	17	1.0–75.0	52
CTOPP Phonological Awareness	8	8	1.0–30.0	74
GFTA	30	21	1.0–67.0	19

Note. RIAS = Reynolds Intellectual Assessment Scale; CELF = Clinical Evaluation of Language Fundamentals–Fourth Edition; DELV = Diagnostic Evaluation of Language Variation; CREVT = Comprehensive Receptive and Expressive Vocabulary Test–Third Edition; CTOPP = Comprehensive Test of Phonological Processing–Second Edition; GFTA = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000).
[a]Scores on this subtest contributed to the CELF Core Language omnibus score.

books that contained "colorful illustrations that helped narrate the story" and "vocabulary words in the text that were unlikely to be known by the children" (p. 21). Six potentially unknown words were identified for each book using common criteria (Beck, McKeown, & Kucan, 2002). The target words identified were a mix of nouns (n = 16), verbs (n = 25), and adjectives (n = 19). The morphological form of the word was based on the form used in the book text. For example, *heaved* was presented in the book text as past tense, and thus *heaved* was used consistently throughout treatment. For the current study, the 10 books (i.e., 60 total target words) were divided into two sets of books that were matched on word type. One set contained nine nouns, 12 verbs, and nine adjectives, whereas the other set contained seven nouns, 13 verbs, and 10 adjectives. The two sets of books and words were randomized to treatment (five books, 30 treated words) or no-treatment control (five books, 30 untreated control words) for each child. Although this was not a study of treatment efficacy, including untreated control words for each child provides a benchmark for the effect of repeated testing of words over time. Limited growth is expected in untreated control words. The list of books and words is available in the KU ScholarWorks archive (see the Summary of the Stimuli file), and more detailed item-level data are available in Supplemental Table S2.

Treatment Form

The elaboration condition from the prior study (Justice et al., 2005) served as the treatment form. *Treatment form* refers to the "typical activity within which teaching episodes are delivered" (Warren et al., 2007, p. 71). In the prior study, children heard the target word in the book, a definition of the target word, and use of the target word in a supportive context sentence. A fourth element—a synonym for the target word—was added to increase elaboration. The book text, definition, supportive context sentence, and synonym for each word are shown in the KU ScholarWorks archive (see the Summary of Stimuli file).

Given the need to test higher intensities of the treatment than that used by Justice et al. (2005), exposures for the current treatment were distributed across pre–book reading, book reading, and post–book reading activities within a treatment session. An example of the highest intensity exposure is shown in Table 2 for the target word *decided*. As shown in Table 2, during the pre–book reading activity, children are shown a printed color picture intended to illustrate the target word and are told the synonym and definition of the word. Once all six target words have been previewed, the book is read to the child. As shown in Table 2, the examiner departs from the text after a target word has been read and provides the synonym for the word. Once the entire book has been read, the post–book reading activity is initiated. As shown in Table 2, a second, different printed color picture intended to illustrate the target word is shown and the child hears the supportive context sentence as well as a definition. This activity is completed for all six target words. Note that the form of the word—in this case, past tense (i.e., *decided*)—is used consistently throughout the treatment so that the child does not have to recognize consistency across changing forms (e.g., *decide*, *deciding*, *decided*). An additional example and further details are provided in Voelmle and Storkel (2015). Administration of the treatment was supported by printed scripts, which are available in the KU ScholarWorks archive (see the Treatment and Naming Scripts file).

Treatment was provided by research assistants and occurred in a one-on-one format in a quiet area at the child's school, after-school program, home, or other agreed-upon location (e.g., small room at a local library). Each treatment session focused on two books. All activities were completed for one book, and then the same types of

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Table 2. Sample treatment form for high-intensity treatment for the target word *decided*.

Activity	Verbal input	Visual input
Pre–book reading	*Decided* is like chose. [synonym] *Decided* means to make up one's mind. [definition]	 ©dreamstime.com/skypixel Reprinted with permission.
Book reading	"Then William and Mom *decided* to scare the bear that lived under the stairs." (Cooper, 1993, p. 21) [text from book] *Decided* is like chose. [synonym]	Storybook picture
Post–book reading	The boy *decided* to play with the yellow cup. [context sentence] *Decided* means to make up one's mind. [definition]	 ©istock.com/peterbooth Reprinted with permission.

activities were completed for the second book. Each session lasted approximately 20 to 30 min.

Treatment Intensity

The primary independent variable was treatment intensity, operationally defined as the cumulative number of exposures to a target word at the end of treatment. Treatment intensity is a function of the number of exposures to a target word in a book (i.e., dose) and the number of repeated readings of a book (i.e., dose frequency). As a reference, the prior study (Justice et al., 2005) provided three exposures to each target word in a book (i.e., dose = 3). Each book was read four times (i.e., dose frequency = 4). Thus, at the end of treatment, children had heard each target word 12 times. The current study used this intensity (12 cumulative exposures), twice this intensity (24 exposures), three times this intensity (36 exposures), and four times this intensity (48 exposures). These choices were based on the prior studies indicating that children with SLI need two to three times as many exposures as their peers to learn a new word (Gray, 2003; Rice et al., 1994). Table 3 lists the dose and dose frequency for the treated words in each intensity condition. Table 3 also shows that untreated control words were tested but were never taught. The KU ScholarWorks archive contains the treatment schedules (see the Treatment Schedules file) showing which books were read during each session.

Table 4 shows a comparison of the exposure for one treated word (*marsh*) across the lowest and highest intensity conditions. For intensity 12, during pre–book reading, a child hears the word one time and is provided with a synonym and a definition. During book reading, the child again hears the word one time while the book is read. During post–book reading, the child hears the word one time in a supportive context sentence. Thus, for the lowest intensity, the child hears the word three times during a session (i.e., dose = 3) and these same activities are repeated on four different days (i.e., dose frequency = 4). Turning to the illustration of intensity 48 in Table 4, during pre–book reading, a child hears the word two times when provided with a synonym and a definition. During book reading, the child hears the word another two times via reading the book and a reminder of the synonym. During post–book reading, the child hears the word a final two times via a supportive context sentence and a reminder of the definition. Thus, for the highest intensity, the child hears the word six times during a session (i.e., dose = 6), and these same activities are

Table 3. Treatment intensity conditions.

	Treatment (five books, 30 words)		Untreated control (five books, 30 words)		Treatment time (two to three sessions/week)	
Intensity	Dose	Dose frequency	Dose	Dose frequency	Sessions	Weeks
12	3	4	0	0	10	4–5
24	4	6	0	0	15	5–8
36	6	6	0	0	15	5–8
48	6	8	0	0	20	7–10

Note. Children were randomized to intensities in blocks such that four children were enrolled in a block and each child was randomized to a different intensity. For a given child, half the books and words were randomly assigned to the treatment condition and the other half were assigned to an untreated control condition. Dose is the number of exposures to a target word in a book; dose frequency is the number of repeated readings of a book across time.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Table 4. Illustration of minimum and maximum intensity for the word *marsh*.

Variable	Intensity 12	Intensity 48
Session activities (i.e., dose)		
Pre–book reading	Synonym definition: "Let's listen and look for these words in our book. [turn to marsh page] *Marsh* is like a swamp. It means a low, wet land, often thick with tall grasses."	Synonym definition: "Let's listen and look for these words in our book. [turn to marsh page] *Marsh* is like a swamp. *Marsh* means a low, wet land, often thick with tall grasses."
Book reading	Text: "They came down to a *marsh* where they saw a muskrat spring-cleaning his house."	Text synonym: "They came down to a *marsh* where they saw a muskrat spring-cleaning his house. *Marsh* is like a swamp."
Post–book reading	Context sentence: "Let's think about the words in our book. [show different picture of marsh] Ducks and beavers live in a *marsh* because they like the water."	Context sentence definition: "Let's think about the words in our book. [show different picture of marsh] Ducks and beavers live in a *marsh* because they like the water. *Marsh* means a low, wet land, often thick with tall grasses."
Number of repetitions of session activities (i.e., dose frequency)	4	8
Intensity summary	Dose 3 × dose frequency 4 = 12 cumulative exposures	Dose 6 × dose frequency 8 = 48 cumulative exposures

Note. Dose is the number of exposures to a target word in a book; dose frequency is the number of repeated readings of a book across time.

repeated on eight different days (i.e., dose frequency = 8). Note that in comparing the lowest and highest intensities, the form of the treatment is exactly the same—that is, all children hear at least one synonym, definition, book text, and supportive context sentence during a session. Higher intensities are achieved by simply repeating these treatment forms.

Treatment Fidelity

Treatment fidelity was checked for 20% of sessions. An observer watched the video of selected sessions and used a checklist to tally that each target word was administered and that the intended treatment form was used (e.g., text or script read correctly). Two scores were derived. The first score, derived by dividing the total number of exposures administered by the intended number of exposures, was 99.92%. The second score, derived by dividing the total number of correct treatment forms administered by the intended number of treatment forms, was 99.67%.

Outcome Measure

Research assistants who collected and scored the primary outcome measure were blind to the participant's assigned treatment intensity. The secondary outcome measure was collected by the research assistant who administered the treatment, but the research assistant who scored that measure was blind to the participant's assigned treatment intensity. Study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at the University of Kansas Medical Center (for more information, see Harris et al., 2009).

Primary Outcome Measure: Definition Task

The primary outcome measure was a definition task, similar to that used by Justice et al. (2005), because the goal was to find an intensity of this treatment that leads to similarly positive outcomes in children with SLI. Thus, using the same task as Justice et al. provided a clear benchmark for success. The definition task was administered prior to and immediately following treatment. The task was administered across two sessions, with 15 treated and 15 untreated control words being tested in each session. Each session began with three practice words (*bed*, *ball*, *candy*), which were words that children with SLI were likely to know. Then, treated (n = 15) and untreated (n = 15) control words were intermixed with each other and with easy known words (n = 10; e.g., *cat*, *dirty*, *sleep*). Each word was introduced with the prompt "Tell me what [word] means." Responses were audio recorded and transcribed.

Definitions were scored following the procedures of McGregor, Oleson, Bahnsen, and Duff (2013). Potential judges (i.e., principal investigator, project coordinator, five graduate research assistants) met to create an agreed-upon scoring rubric for each word by consulting dictionaries to identify critical elements for a complete and accurate definition (e.g., *flashing* = "light" and "transient"). The rubric awarded 0 points for an incorrect or absent definition (e.g., *flashing* = "flashlight"), 1 point for clearly appropriate use of the word in a sentence (e.g., *flashing* = "You're flashing your camera on") or for a vague definition (e.g., *flashing* = "change"), 2 points for a conventional definition that contains at least one critical element but lacks other critical elements (e.g., *flashing* = "lights go on" [critical element = "light"]), and 3 points for a complete and accurate definition including all critical elements (e.g., *flashing* = "bright light on and off" [critical elements = "light" and "transient"). The scoring rubrics are available in the KU ScholarWorks archive (see the Definition Scoring Rubric file). Two judges independently scored each response using the rubric. Their independent scores were then compared, and disagreements were resolved by consensus. In the event that consensus could not be reached, a third judge was consulted. These

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

types of disagreements were not formally tracked but occurred rarely.

For comparison to Justice et al. (2005), words with scores of 2 or 3 (i.e., an at least minimally accurate definition) were counted as correct, and words with scores of 0 or 1 were counted as incorrect. Table 5 shows the percentage of responses scored as 0, 1, 2, or 3 for treated and untreated words pretreatment and posttreatment. As shown in Table 5, scores of 2 tended to predominate the correct category, and scores of 0 tended to predominate the incorrect category. The final score for the definition task was computed by counting the number of words with at least minimally accurate definitions (i.e., score of 2 or 3). Overall, learning was relatively low, with an average of 3.73 (SD = 3.45, range = 0–14) treatment words correct posttreatment.

Secondary Outcome Measure: Naming Task

A definition task is heavily reliant on learning the meaning of a word. Thus, a naming task was administered during treatment to tap learning of the pronunciation of a word. The naming task was administered at four predetermined points, including the last treatment session. The naming task tested the words that were the focus of treatment for the session as well as a paired set of untreated control words. Children were shown the pre–book reading picture used in treatment and were given a prompt specific to the picture and the target word. For example, for the target word *ruffle*, children were shown a picture of a bird and asked "What does the bird do to his feathers?" Naming prompts for each target word are shown in the KU ScholarWorks archive (see the Treatment and Naming Scripts file). Responses were transcribed and scored as correct if the child named the target word (e.g., said "ruffle" for the target word *ruffle*) or incorrect if the child failed to name the target word (e.g., said "bird" for the target word *ruffle*) or failed to provide any response. Changes in grammatical form (e.g., "ruffled" or "ruffling" for the target word *ruffle*) and common misarticulations (e.g., "wuffle" for the target word *ruffle*) were ignored (i.e., responses with these differences were scored as correct). The naming score was computed by counting the number of words that were named correctly at the last test. Overall, children on average named 6.07 (SD = 4.08, range = 2–18) treatment words correctly at the last test.

Table 5. Percentage (%) of definition responses receiving a score of 0, 1, 2, or 3 at pretreatment and posttreatment for untreated control words and treated words.

Variable	0	1	2	3
Pretreatment[a]				
Untreated control words	96	3	1	0
Treated words	97	2	1	0
Posttreatment[b]				
Untreated control words	95	3	2	0
Treated words	79	8	9	4

Note. Scoring was as follows: 0 points for an incorrect or absent definition, 1 point for clearly appropriate use of the word in a sentence or for a vague definition, 2 points for a conventional definition that lacks one or more critical elements, and 3 points for a complete and accurate definition including all critical elements. For analysis, scores of 0 and 1 were aggregated (i.e., incorrect) and scores of 2 and 3 were aggregated (i.e., correct).

[a]Percentage out of 810 responses (27 children × 30 words).
[b]Percentage out of 780 responses (26 children × 30 words).

Treatment Intensity: Escalation Design Procedures

Children were enrolled in the study in blocks of four children. Within a block, each child was randomly assigned to one of the four intensities (12, 24, 36, or 48 exposures). In most cases, two blocks were run in parallel, and data were examined to determine whether changes were needed (e.g., drop or add intensities). As has been done in other escalation designs (Hunsberger et al., 2005), data analysis and decision making occurred block by block by plotting the percentage of children in each intensity who responded to the treatment. Responding to the treatment was operationally defined as a posttest score of 5 or greater for the treated words. This cut-off was selected because it was similar to that of Justice et al. (2005) in a parallel analysis (which used a cut-off of 4 or higher) and it was outside the range of scores observed for untreated control words (M = 1, SD = 0.9, range = 0–4). A pattern was considered to have been established when it had been replicated across at least three blocks. The block-by-block data are shown in Supplemental Figure S1. Intensity 12 was dropped after the fifth block. All data collection stopped after the seventh block. Thus, five children completed intensity 12, and seven children each completed intensities 24, 36, and 48.

Results

Treatment Intensity

The first goal of this study was to identify an adequate or promising intensity of interactive book reading for children with SLI. Three analyses were conducted to address this goal. The first analysis examined the percentage of children responding to the treatment on the basis of the definition task. The second analysis examined the number of words learned on the basis of the definition task. The third analysis examined the percentage of children responding to treatment on the basis of the naming task.

Definition Task: Percentage of Children Responding to Treatment

Figure 1 reports the percentage of children responding to treatment on the basis of the definition task (i.e., posttreatment score of 5 or more for treated words) for each intensity. Note that these are the same data that were used to make decisions in the escalation design procedures and that the block-by-block data are shown in Supplemental Figure S1. As shown in Figure 1, 0% of children in intensity 12 responded to the treatment. Recall that this is the intensity that was effective in promoting word learning by typically developing children from low-income homes (Justice et al.,

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Figure 1. Percentage of children responding to the treatment on the basis of a posttreatment definition score of 5 or higher for treated words in block 7 (the last block) for each treatment intensity condition (12, 24, 36, and 48). The trend line illustrates the polynomial trend also depicted by the regression equation. Posttreatment treated words are shown in shaded bars.

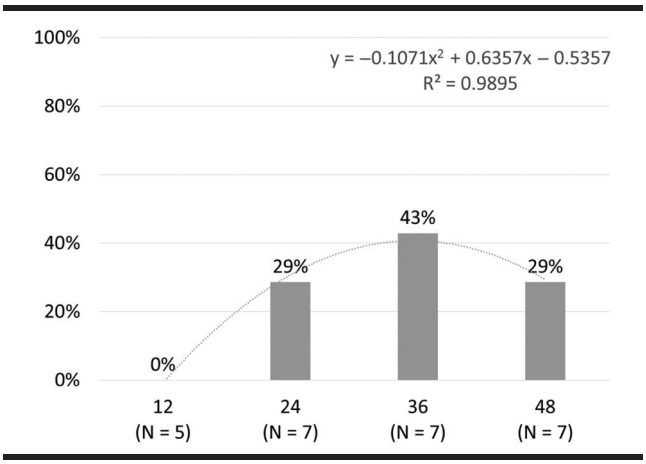
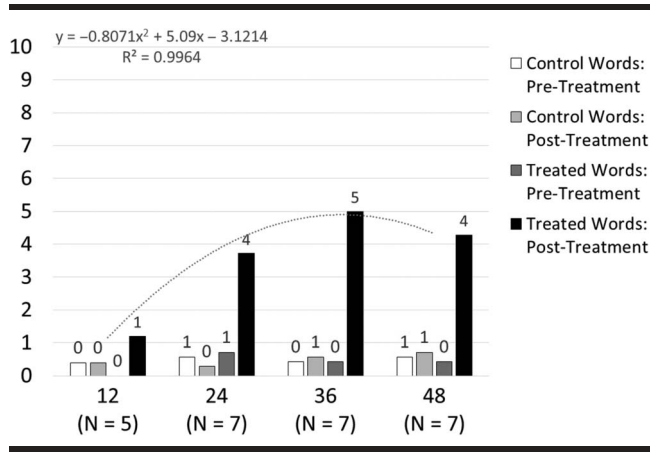


Figure 2. The average number of words with correct definitions in block 7 (the last block) for each treatment intensity (12, 24, 36, and 48). Open bars = pretreatment control words; light shaded bars = posttreatment control words; medium shaded bars = pretreatment treated words; dark shaded bars = posttreatment treated words. The trend line illustrates the polynomial trend for posttreatment treated words, also depicted by the regression equation.



2005). This intensity did not appear to be sufficient to promote word learning by children with SLI. As intensity increased from 12 to 24 exposures, the percentage of children responding to the treatment increased. When intensity increased from 24 to 36 exposures, even more children responded to treatment. At 36 exposures, 43% of children with SLI responded to the treatment. In contrast, as intensity increased from 36 to 48 exposures, the percentage of children responding to treatment decreased. This pattern is a desirable pattern in an escalation design because it indicates that a plateau has been reached where further increases in intensity are unlikely to result in more children responding to treatment. Taken together, the adequate intensity of the four intensities tested is 36 exposures because this yielded the highest percentage of children responding to the treatment, although treatment response was modest (43%) and lower than that of Justice et al. (2005; 77%).

Definition Task: Number of Words Defined Correctly

Although the magnitude of the treatment response has not typically been examined in escalation studies, it is possible that children could vary widely in the number of words learned beyond the minimal threshold used to define a treatment response. Thus, the actual number of words learned (which was based on a correct definition) was examined to complement the previous analysis of treatment response. The block-by-block data are shown in Supplemental Figure S2. Figure 2 shows the number of treated and untreated words defined correctly pretreatment and posttreatment. As shown in Figure 2, the average number of correct control words was low (i.e., < 1) across intensities at pretreatment (open bars; $M = 0.50$, $SD = 0.76$, range = 0–3) and posttreatment (light shaded bars; $M = 0.50$, $SD = 0.95$, range = 0–4). This indicates that children with SLI did not know the untreated control words prior to treatment and did not learn the control words from environmental exposure or from repeated testing during this study. In addition, the average number of correct treated words at pretreatment (medium shaded bars) was low across all intensities ($M = 0.42$, $SD = 0.76$, range = 0–3). This verifies that the treated words were unknown prior to treatment.

Turning to the average number of treated words with correct definitions posttreatment (dark shaded bars), children in intensity 12 on average learned one treated word ($SD = 2$, range = 0–3), which is relatively similar to performance on untreated control words. This corroborates the prior analysis indicating that 12 exposures was not sufficient to support word learning by children with SLI. As intensity increased from 12 to 24 exposures, the average number of treated words learned also increased. Children in intensity 24 on average learned four treated words ($SD = 1$, range = 1–5), which appears to be better than performance on untreated control words. Number of words learned continued to increase as intensity increased from 24 to 36 exposures. Children in intensity 36 on average learned five treated words ($SD = 5$, range = 0–14). In contrast, no further increase in number of words learned was observed as intensity increased beyond 36 exposures. Children in intensity 48 on average learned four treated words ($SD = 4$, range = 0–11). This analysis converges on the conclusion that 36 exposures was the adequate intensity out of the four intensities tested, although the number of treated words defined correctly posttreatment was modest even in this condition (i.e., five out of 30 words learned).

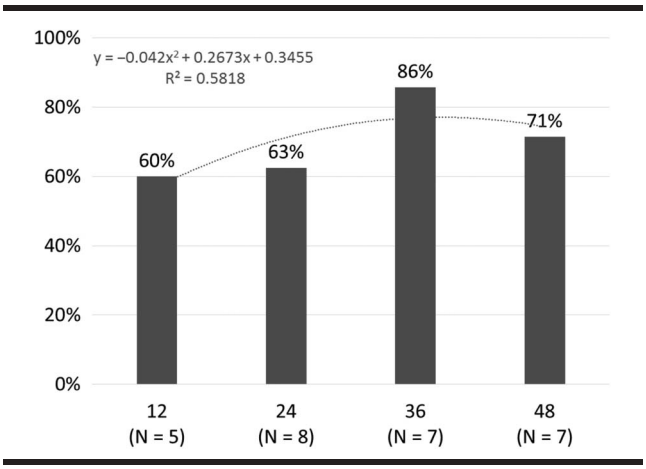
Naming Task: Percentage of Children Responding to Treatment

The final analysis used data from the naming task to determine whether the intensity that was adequate for learning of semantics (as measured by the definition task)

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

also was adequate for learning phonology (as measured by the naming task). This analysis examined the percentage of children responding to the treatment on the basis of the final naming test. Responding to the treatment was operationally defined as a score of 4 or greater for the treated words. This cut-off was selected because it was outside the range of performance observed for untreated control words (M = 1, SD = 0.7, range = 0–3). The block-by-block data are shown in Supplemental Figure S3. Note that this analysis ontains an extra child in intensity 24 compared with the prior analyses of definitions. This is the child who did not complete a posttreatment definition test but did complete treatment, including all naming tests. As shown in Figure 3, 60% of children in intensity 12 responded to the treatment. This is a more favorable picture of the benefit of intensity 12 than shown by the definition data. Although 12 exposures was not sufficient to produce the deep semantic learning needed to provide an adequate definition, 12 exposures did produce word learning when a less difficult measure of learning (i.e., naming) was used. However, greater learning was still observed with higher intensities. As intensity increased from 12 to 24 exposures, the percentage of children who responded to treatment increased slightly. It is important to note that a more marked increase in treatment response was observed as intensity increased from 24 to 36 exposures. In particular, 86% of children in intensity 36 responded to the treatment. No added benefit was observed as intensity increased beyond 36 exposures. The percentage of children responding to treatment decreased as exposures increased from 36 to 48. Thus, data from the naming test (which emphasizes phonology) converge with the findings from the definition task (which emphasizes semantics), indicating that 36 exposures is the adequate intensity. Here, the extent of benefit of the treatment (i.e., 86% of children responding to treatment) was more robust and compelling than observed for the definition data.

Figure 3. Percentage of children responding to the treatment on the basis of naming data (i.e., posttreatment naming score of 4 or higher for treated words) in block 7 (the last block) for each treatment intensity (12, 24, 36, and 48). The trend line illustrates the polynomial trend also depicted by the regression equation.



Variability in Response to Treatment

As shown in the intensity analysis, response to treatment varied across children (i.e., response to treatment is never 100% in any intensity). To better understand this variability, data from children in intensities 24, 36, and 48 were examined. Intensity 12 was excluded because 0% of children in intensity 12 responded to treatment on the basis of posttreatment definition scores for treated words. Thus, data from these children cannot help us understand the characteristics of children who respond to treatment and those who do not. In total, 21 children with posttreatment definition scores who received 24, 36, or 48 exposures were analyzed.

Pretreatment Characteristics Associated With Posttreatment Definition Scores

Correlation between pretreatment test scores and the posttreatment definition score for treated words was examined. Also, the score for treated words on the first naming test during treatment was included to determine whether early treatment performance was related to posttreatment scores. Supplemental Table S3a shows the correlations, with the first column being the primary interest. Here, the DELV Semantic score, $r(21) = .52$, $p < .05$, $r^2 = .27$, the Comprehensive Test of Phonological Processing–Second Edition (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 2013) Nonword Repetition score, $r(21) = .44$, $p < .05$, $r^2 = .20$, and the CTOPP Phonological Awareness subtest score, $r(21) = .48$, $p < .05$, $r^2 = .23$, correlated significantly with the number of treated words accurately defined posttreatment. Children with higher vocabulary scores, higher Nonword Repetition scores, or higher Phonological Awareness scores defined more treated words correctly posttreatment than did children with lower vocabulary, lower Nonword Repetition, or lower Phonological Awareness scores. Supplemental Table S3b examines the association between demographic characteristics (i.e., gender, race, ethnicity, parent marital status, parent education) and treatment response. No statistically significant effects were identified.

Classification Accuracy on the Basis of Pretreatment Characteristics

The three measures that were significantly correlated with posttreatment definition scores (i.e., DELV Semantic score, CTOPP Nonword Repetition score, and CTOPP Phonological Awareness score) were analyzed as a classification function. Classification functions typically are used to evaluate diagnostic measures. Here, the diagnostic status of participants is known (i.e., SLI vs. typical language), and then scores on a new diagnostic measure are used to predict the child's language status on the basis of an optimized cut-off score. The alignment between the known and predicted disorder status is then examined for overall accuracy, sensitivity, specificity, positive likelihood ratios, and negative likelihood ratios, with each of these measures having established criteria for interpretation. Table 6 provides a review of the computation and interpretation of these measures as applied to treatment response. As shown in Table 6,

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Table 6. Classification table and measures for analysis.

Pretreatment test score	Treatment outcome: Did not respond to treatment (–)	Treatment outcome: Did respond to treatment (+)
At or above the criterion score (predicted to respond to treatment; +)	False positive	True positive
Below the criterion score (predicted to not respond to treatment; –)	True negative	False negative

Note. Treatment outcome criteria = five or more treated words with correct definitions. *Accuracy* = (true positive + true negative)/total. *Sensitivity* = true positive/(true positive + false negative); probability that the score on the pretreatment measure predicts a child will respond to the treatment when the child actually did respond to treatment (> 80% is desirable). *Specificity* = true negative/(true negative + false positive); probability that the score on the pretreatment measure predicts a child will not respond to the treatment when the child actually does not respond to treatment (> 80% is desirable). *Positive likelihood ratio* = true positive/false positive; indicates how likely a child is to respond to treatment as opposed to not respond to treatment given a pretreatment score at or above the cut-off (≥ 3.0 is suggestive, ≥ 10.0 is informative).

Negative likelihood ratio = false negative/true negative; indicates how likely a child is to respond to treatment rather than not respond to treatment given a pretreatment score below the cut-off (≤ 0.30 is suggestive, ≤ 0.10 is informative). See https://www.medcalc.org/calc/diagnostic_test.php for a program to calculate these measures.

whether a child responded to our treatment is known using our cut-off of five or more treated words correctly defined posttreatment. Receiver operating characteristic curves are used to identify cut-off scores on the pretreatment measures to maximize accuracy in predicting which children will respond to treatment. The alignment between the actual treatment response and the predicted treatment response can then be evaluated.

Of the 21 children available for analysis, 14 were classified as not responding to treatment (i.e., four or fewer treated words defined correctly posttreatment), and seven were classified as responding to treatment (i.e., five or more words defined correctly posttreatment). Table 7 shows the selected cut-off score for each pretreatment measure (i.e., DELV Semantic score, CTOPP Nonword Repetition score, and CTOPP Phonological Awareness score) in the first column. These cut-off scores for standardized tests ranged from the sixth to the ninth percentiles.

In Table 7, the three pretreatment measures are sorted on the basis of their classification accuracy. Desirable values for accuracy, sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio are marked. As shown in Table 7, sensitivity and negative likelihood ratios were good for all three measures. In contrast, specificity and positive likelihood ratios were good only for the CTOPP Phonological Awareness subtest. Focusing on the top left panel of Figure 4, CTOPP Phonological Awareness scores are plotted against the number of treated words defined correctly posttreatment. The cut-off score for the CTOPP is shown by a horizontal line, and the cut-off definition score for a treatment response is shown by a vertical line. The CTOPP Phonological Awareness score sorts children relatively accurately into those who responded to the treatment (filled circles) and those who did not (unfilled circles). The majority of children who scored at or above the CTOPP Phonological Awareness cut-off score did respond to treatment (i.e., true positive), and the majority of children who scored below the CTOPP Phonological Awareness cut-off score did not respond to treatment (i.e., true negative). Very few children fall into the false positive or false negative quadrants in Figure 4. Overall, pretreatment CTOPP Phonological Awareness scores predicted treatment response relatively accurately.

In contrast, the other two measures did not sort children as accurately into those who responded to treatment and those who did not. This is shown in the other two panels of Figure 4. For the seven children who responded to treatment, the majority of these children scored at or above the cut-off score on the DELV or Nonword Repetition (true positive), and very few children who responded to treatment scored below the cut-off score on the DELV or Nonword Repetition (false negative). This indicates good sensitivity. Focusing now on the children who scored below the cut-off score on the DELV or Nonword Repetition, the

Table 7. Classification measures for significant pretreatment measures.

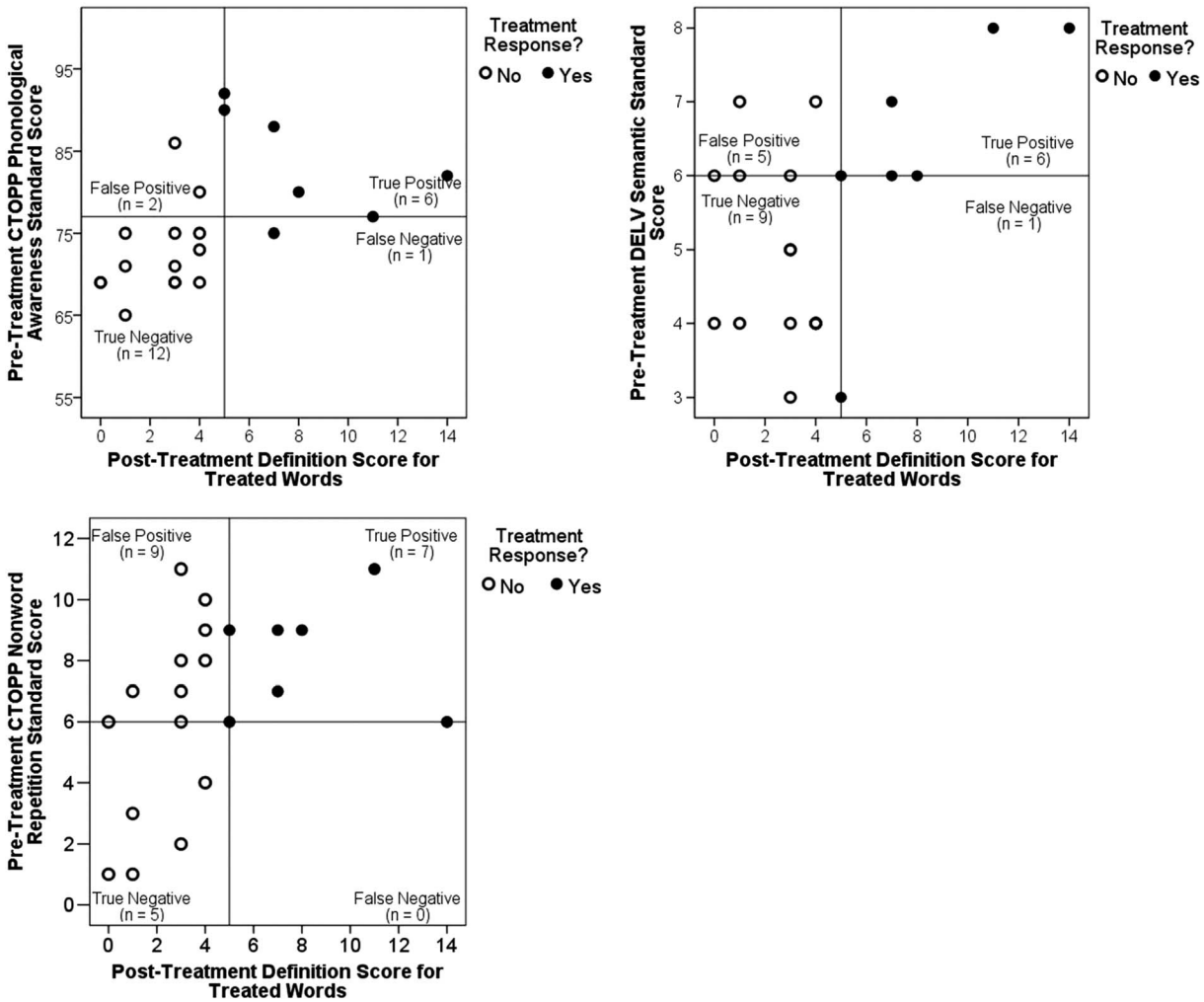
Variable	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive likelihood ratio	Negative likelihood ratio
CTOPP Phonological Awareness (cut-off = SS 77, PR 6)	86[a]	86[a]	86[a]	6.00[a]	0.17[a]
Vocabulary: DELV Semantic (cut-off = SS 6, PR 9)	71	86[a]	64	2.40	0.22[a]
CTOPP Nonword Repetition (cut-off = SS 6, PR 9)	57	100[a]	36	1.56	0.00[a]

Note. *Cut-off* refers to the criterion score used to predict the presence versus absence of a response to treatment. CTOPP = Comprehensive Test of Phonological Processing–Second Edition; DELV = Diagnostic Evaluation of Language Variation; SS = standard score; PR = percentile rank. Scores at or above the cut-off predict that the child will respond to treatment, whereas scores below the cut-off predict that the child will not respond to treatment.

[a]Desirable value.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Figure 4. Posttreatment definition scores for treated words plotted by pretreatment Comprehensive Test of Phonological Processing–Second Edition (CTOPP) Phonological Awareness standard scores (top left), Diagnostic Evaluation of Language Variation (DELV) Semantic standard scores (top right), and CTOPP Nonword Repetition standard scores (bottom left). The vertical line indicates the criteria for a treatment response on the basis of the posttreatment definition score, which also is shown by the color of the circles (unfilled = no treatment response; filled = treatment response). The horizontal line indicates the cut-off score on the pretreatment test for predicting a positive or negative response to treatment.



majority of those children did not respond to treatment (true negative), and very few did respond to treatment (false negative), indicating a desirable negative likelihood ratio. Thus, a score below the cut-off on the DELV or Nonword Repetition accurately ruled out a response to treatment. However, the poorer specificity and positive likelihood ratio indicate that a score at or above the cut-off on the DELV or Nonword Repetition was less informative. Focusing on the 14 children who did not respond to treatment (unshaded circles), an almost equal number scored at or above the cut-off (false positive) as scored below the cut-off (true negative) on the DELV or Nonword Repetition, indicating poor specificity. Focusing on the children who scored at or above the cut-off score on the DELV or Nonword Repetition, an almost equal number of children responded to treatment (true positive) as did not respond to treatment (false positive), indicating poor positive likelihood ratio. Taken together, a score below the cut-off on the DELV or Nonword Repetition indicates that a child is relatively unlikely to respond to treatment, whereas a score at or above the cut-off on the DELV or Nonword Repetition indicates that a response to treatment could occur, but there is uncertainty in this prediction.

Discussion

The current study had two main goals: (a) to identify an adequate intensity of interactive book reading to support new word learning by children with SLI and (b) to examine variability in treatment response across children with SLI.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Results showed that response to treatment and number of words learned improved as intensity increased up to 36 exposures and then plateaued as intensity increased beyond 36 exposures. Thus, 36 exposures was deemed the adequate intensity for this version of interactive book reading. However, there was variability in treatment response. Children with lower Phonological Awareness, vocabulary, and/or Nonword Repetition scores were less likely to respond positively to the treatment. In contrast, children with higher Phonological Awareness, vocabulary, and/or Nonword Repetition scores were more likely to respond to the treatment, but this was difficult to predict reliably. These findings have implications for further development of this treatment as well as for future preliminary clinical trials of language interventions.

Development of Interactive Book Reading

There is clear evidence that 36 exposures is the adequate intensity (of the four intensities tested). Thus, future research with this treatment for children with SLI should incorporate 36 exposures to the treated words. However, gains were modest, with a small number of new words learned (i.e., an average of five new words defined correctly) and with few children responding to treatment (i.e., 43% compared with 77% in Justice et al., 2005). Some could argue that these modest gains are attributable to the difficulty of the outcome measure used. This, undoubtedly, is true. In fact, a slightly more favorable picture of the treatment emerges in this study when naming data—rather than definition data—were examined. Rather than changing the outcome measure, further enhancement of the treatment is recommended. Children with SLI, like typically developing children, need rich word knowledge to support necessary language functions (e.g., reading comprehension) to set the foundation for better life outcomes (e.g., greater academic and vocational success). Thus, rather than lowering the bar on an acceptable treatment response, the treatment needs to be improved so that greater success is achieved.

Several possible avenues may enhance treatment outcomes. First, the current treatment is entirely receptive. That is, children are provided with high-intensity input but are never required to respond to that input. Prior research on children and adults with typical language has suggested that testing enhances learning over and above gains made simply through study or input (Eisenkraemer, Jaeger, & Stein, 2013; Roediger & Butler, 2011; Rowland, 2014). For example, Karpicke and Roediger (2008) showed that repeated testing increased final recall by 4 *SD* compared with repeated study. In that same study, 80 extra study trials had no effect on retention, whereas 80 extra test trials led to a 150% improvement in retention. It is important to note that the testing effect has been observed in children (Bouwmeester & Verkoeijen, 2011; Fritz, Morris, Nolan, & Singleton, 2007; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Marsh, Fazio, & Goswick, 2012; Rohrer, Taylor, & Sholar, 2010). In addition, children with SLI are known to have difficulty retrieving words (Lahey & Edwards, 1996; McGregor, 1997, 2014; McGregor & Leonard, 1989); therefore, dividing the 36 exposures across receptive input and expressive retrieval practice via testing may strengthen retrieval, thus augmenting word learning outcomes. This hypothesis warrants empirical testing.

Second, predictors of a positive treatment response included elements of phonology (i.e., phonological awareness, nonword repetition) as well as semantics (i.e., DELV Semantic subtest). This makes intuitive sense because learning a word requires learning the sound form (i.e., phonology) and the meaning (i.e., semantics). It is unclear in this small sample whether subgroups of nonresponding children had difficulty with only one or with both of these elements. A larger study would be useful in understanding these individual differences because there would be enough power to statistically disentangle the contribution of phonological and semantic measures to treatment outcomes. At the same time, it seems wise to better support and track phonological and semantic learning during treatment. To accomplish this, both phonological (e.g., naming) and semantic (e.g., defining) measures could be incorporated into the retrieval practice described previously. This phonological and semantic retrieval practice could enhance learning but also would track learning of phonology and semantics during treatment to better identify how each contributes to treatment outcomes.

Third, there are multiple ways to accumulate 36 exposures by the end of treatment. The current study used a balanced approach in which children heard the words six times during each book reading session and then each book was read on six different occasions (i.e., $6 \times 6 = 36$). It is possible that a different approach to achieving 36 exposures might be more beneficial. For example, research with adults with developmental language impairments has indicated that their primary difficulty may be with extracting word form and meaning during training—referred to as *encoding* (McGregor, Licandro, et al., 2013). This suggests that it could be better to increase the number of times a child hears a word in a book reading session so that he or she has ample exposure (and practice) to form an initial representation of the word before the end of a treatment session. Thus, achieving 36 exposures through more exposures during a session (e.g., nine exposures) but fewer repeated readings of the same book (e.g., four book readings) could enhance treatment outcomes. However, the alternative hypothesis also is supported by the literature. That is, one of the most robust findings in cognitive psychology is that distributed training, where learning from input is interspersed with gaps in training (e.g., weekly training), leads to more robust long-term learning than massed practice, where learning from input is done all at once (i.e., one training session; Philips, Kopec, & Carew, 2013). Moreover, the finding is applicable to children with SLI, who learn more new words when the same number of exposures is distributed across four training days rather than massed in one training day (Riches, Tomasello, & Conti-Ramsden, 2005). The current treatment does use a distributed exposure (i.e., repeated readings of books), but the question is

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

whether more distributed practice would enhance learning. For example, would children with SLI learn more words if they received fewer exposures during a book reading session (e.g., four exposures) but read the books more times (e.g., nine readings of the same book)? The best regimen for achieving 36 exposures remains to be discovered.

Last, it is possible that interactive book reading ultimately may not be effective, even when the treatment approach is enhanced further, for children with lower phonological awareness (i.e., less than the sixth percentile on CTOPP), vocabulary (i.e., less than the ninth percentile on DELV), or nonword repetition (i.e., less than the ninth percentile on CTOPP) scores. These specific measures may be tapping elements of online processing that are foundational to learning through interactive book reading. The phonological awareness measure taps the ability to manipulate sounds, and the nonword repetition measure taps the ability to hold novel sound sequences in memory. The vocabulary measure (i.e., DELV Semantic subtest) differs from many commercially available vocabulary measures in that half of the items focus on fast mapping. That is, both known and novel words are presented in a context before the child is asked questions about the words. This difference in the vocabulary measure may explain why vocabulary score in the current study positively correlated with treatment outcome, whereas vocabulary score on the Peabody Picture Vocabulary Test–Third Edition (a measure of receptive vocabulary knowledge; Dunn & Dunn, 1997) negatively correlated with treatment outcome in the prior study (Justice et al., 2005). However, note that cut-off scores also varied across studies (i.e., ninth percentile for current study; 16th percentile in Justice et al., 2005). All of the actions tapped by these measures presumably are necessary for learning new words from interactive book reading. That is, during interactive book reading, the child must identify that a new word was heard, hold the novel sound sequence in working memory while extracting the meaning from the visual and auditory input, and create an initial mental representation of both the word form and the meaning. As a consequence, a child with weak online processing may not be able to accomplish some or all of these steps, limiting word learning during interactive book reading. A less naturalistic treatment that minimizes online processing could be more appropriate for children with low phonological awareness, vocabulary, or nonword repetition scores.

Future Clinical Research

The intensity required for a word learning treatment likely will vary depending on the form of the treatment, the number of words targeted in a treatment session, the age of the target population, the language and cognitive abilities of the target population, and various other factors. This suggests that alternative versions of this treatment that manipulate or vary any of these factors may require a different intensity of the treatment. As noted previously, identifying the adequate intensity of a treatment for a particular population is a crucial first step in clinical research. The current study demonstrates that an escalation design can be applied successfully to a behavioral treatment to identify relatively quickly an adequate intensity for a new treatment or a new variation on an existing treatment. As a first application of this design to language treatment, our approach was somewhat conservative. For example, we continued to enroll children in the lowest intensity condition even though there was little evidence to support its effectiveness. Further development of the escalation design for behavioral interventions would assist in facilitating future clinical research. In particular, it would be helpful to have clearer procedures, similar to those developed for nontoxic drug trials (Hunsberger et al., 2005), so that data could be collected from even fewer participants while still successfully identifying the adequate intensity.

Likewise, the current study demonstrates the logic of applying a classification approach to understanding variation in treatment response. This approach has potential value to clinical practice in that the question of which treatment to use for a given client can be informed by having data to make predictions concerning the likelihood of a treatment response. Current approaches to examining variability in treatment response tend not to translate the data into a metric that can be applied to clinical decision making, but a classification approach may be able to accomplish this, providing more useful information to practitioners.

Conclusions

This study tested four intensities of interactive book reading for kindergarten children with SLI and determined that the adequate intensity for word learning is 36 exposures to the treated words. It is notable that children with lower phonological awareness, vocabulary, and/or nonword repetition scores were unlikely to respond to the treatment. In contrast, children with higher phonological awareness, vocabulary, and/or nonword repetition scores were somewhat likely to respond to the treatment, but this prediction was not definitive. Taken together, further development of this treatment is needed to increase the number of words learned as well as the percentage of children with SLI benefitting from the treatment. This study demonstrates the applicability of an escalation design for identifying an adequate treatment intensity of a language intervention and shows the utility of a classification approach to understanding potential predictors of treatment response.

Acknowledgments

This project was supported by National Institutes of Health Grants DC012824, DC05803, HD02528, and NCATS-UL1TR000001. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. We thank the staff of the Participant Recruitment and Management Core of the Center for Biobehavioral Neurosciences of Communication Disorders (supported by DC05803) for assistance with recruitment of schools and children; the staff of the Research Design and Analysis Unit of the Lifespan Institute (supported by HD02528) for assistance with study design; Frontiers: The Heartland Institute for Clinical and Translational Research

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

for REDCap support (supported by NCATS-UL1TR000001); and the staff of the Word and Sound Learning Lab (supported by DC012824) for their contributions to stimulus creation, participant recruitment, data collection, data processing, and reliability calculations. Portions of the data were presented previously at the 2015 Kansas Speech-Language-Hearing Association Convention, the 2015 Annual Convention of the American Speech-Language-Hearing Association, and the 2015 Symposium for Research in Child Language Disorders. Clinical application of the methods and findings was described previously in Voelmle and Storkel (2015) and Komesidou and Storkel (2015).

References

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49,* 941–954.

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47,* 407–420.

American Speech-Language-Hearing Association. (1997). Guidelines for screening for hearing impairment—School-age children, 5 through 18 years. *ASHA, 4,* IV-74ff–IV-74hh.

American Speech-Language-Hearing Association. (2011). *Supply and demand resource list for school-based speech-language pathologists*. Retrieved from http://www.asha.org/uploadedFiles/2011-School-SLP-Supply-Demand.pdf

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life*. New York, NY: 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,* 294–304.

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language, 65,* 32–41. doi:10.1016/j.jml.2011.02.005

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157.

Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39,* S110–S137.

Cooper, H. F. (1993). *The bear under the stairs*. London, United Kingdom: Doubleday.

Crain-Thoreson, C., & Dale, P. S. (1999). Enhancing linguistic performance: Parents and teachers as book reading partners for children with language delays. *Topics in Early Childhood Special Education, 19,* 28–39.

Dale, P. S., Crain-Thoreson, C., Notari-Syverson, A., & Cole, K. (1996). Parent-child book reading as an intervention technique for young children with language delays. *Topics in Early Childhood Special Education, 16,* 213–235.

Dollaghan, C. A. (1987). Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders, 52,* 218–222.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test—Third edition.* Circle Pines, MN: AGS.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paideia, 23,* 397–406.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24,* 174–187.

Fey, M. E., & Finestack, L. H. (2009). Research and development in child language intervention: A five-phase model. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513–531). New York, NY: Routledge.

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology, 60,* 991–1004. doi:10.1080/17470210600823595

Gathercole, S. E. (1993). Word learning in language-impaired children. *Child Language Teaching and Therapy, 9,* 187–199.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition.* Circles Pines, MN: American Guidance Service.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28,* 135–142. doi:10.1002/Acp.2956

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition, 3,* 177–182.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research, 46,* 56–67. doi:10.1044/1092-4388(2003/005)

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47,* 1117–1132. doi:10.1044/1092-4388(2004/083)

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research, 48,* 1452–1467. doi:10.1044/1092-4388(2005/101)

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196–206.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42,* 377–381.

Horton-Ikard, R., & Weismer, S. E. (2007). A preliminary examination of vocabulary and word learning in African American toddlers from middle and low socioeconomic status homes. *American Journal of Speech-Language Pathology, 16,* 381–392.

Hunsberger, S., Rubinstein, L. V., Dancey, J., & Korn, E. L. (2005). Dose escalation trial designs based on a molecularly targeted endpoint. *Statistics in Medicine, 24,* 2171–2181. doi:10.1002/sim.2102

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,* 17–32.

Karpicke, J. D., & Roediger, H. L., III. (2008, February 5). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Komesidou, R., & Storkel, H. L. (2015). Learning and remembering new words: Clinical illustrations from children with specific language impairment. *Perspectives on Language Learning and Education, 22,* 138–146.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech and Hearing Research, 39,* 1081–1098.

Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory, 20,* 899–906. doi:10.1080/09658211.2012.708757

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,* 300–335. doi:10.3102/0034654310377087

McGregor, K. K. (1997). The nature of word-finding errors of preschoolers with and without word-finding deficits. *Journal of Speech, Language, and Hearing Research, 40,* 1232–1244.

McGregor, K. K. (2014). What a difference a day makes: Change in memory for newly learned word forms over twenty-four hours. *Journal of Speech, Language, and Hearing Research, 57,* 1842–1850. doi:10.1044/2014_JSLHR-L-13-0273

McGregor, K. K., & Leonard, L. B. (1989). Facilitating word-finding skills of language-impaired children. *Journal of Speech and Hearing Disorders, 54,* 141–147.

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research, 56,* 1845–1856. doi:10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48,* 307–319. doi:10.1111/1460-6984.12008

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,* 979–1007.

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education & Development, 19,* 7–26.

Nash, M., & Donaldson, M. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48,* 426–438.

Penno, J. F., Wilkinson, I. A., & 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,* 23–33.

Philips, G. T., Kopec, A. M., & Carew, T. J. (2013). Pattern and predictability in memory formation: From molecular mechanisms to clinical relevance. *Neurobiology of Learning and Memory, 105,* 117–124. doi:10.1016/j.nlm.2013.05.003

Pile, E. J. S., Girolametto, L., Johnson, C. J., Chen, X., & Cleave, P. J. (2010). Shared book reading intervention for children with language impairment: Using parents-as-aides in language intervention. *Canadian Journal of Speech-Language Pathology and Audiology, 34,* 96–109.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test*. Lutz, FL: Psychological Assessment Resources.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55,* 33–42.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37,* 106–121.

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48,* 1397–1411.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86,* 54–64.

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37,* 401–411.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi:10.1016/j.tics.2010.09.003

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 233–239. doi:10.1037/A0017678

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463. doi:10.1037/a0037559

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). San Diego, CA: Singular.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. Bloomington, MN: Pearson.

Seymour, H. N., Roeper, T. W., de Villers, J., & de Villers, P. A. (2005). *Diagnostic Evaluation of Language Variation–Norm-Referenced*. San Antonio, TX: The Psychological Corporation.

Steele, S. C., & Mills, M. T. (2011). Vocabulary intervention for school-age children with language impairment: A review of evidence and good practice. *Child Language Teaching and Therapy, 27,* 354–370.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 40,* 1245–1260.

Voelmle, K., & Storkel, H. L. (2015). Teaching new words to children with specific language impairment using interactive book reading. *Perspectives on Language Learning and Education, 22,* 131–137.

Wagner, R. K., Torgesen, J., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition*. Austin, TX: Pro-Ed.

Wallace, G., & Hammill, D. D. (2013). *Comprehensive Receptive and Expressive Vocabulary Test–Third Edition*. Austin, TX: Pro-Ed.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13,* 70–77.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, M. C., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24,* 552–559.

Whitehurst, G. J., Fischel, J. E., Lonigan, C. J., Valdez-Menchaca, M. C., Arnold, D. S., & Smith, M. (1991). Treatment of early expressive language delay: If, when, and how. *Topics in Language Disorders, 11,* 55–68.

Downloaded From: http://lshss.pubs.asha.org/ by a Univ Of Newcastle User on 01/01/2017
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx