
AN EVALUATION OF AN EXPLICIT READ ALOUD INTERVENTION TAUGHT IN WHOLE-CLASSROOM FORMATS IN FIRST GRADE

ABSTRACT

This study describes an evaluation of a read aloud intervention to improve comprehension and vocabulary of first-grade students. Twelve teachers were randomly assigned to an intervention or comparison condition. The study lasted 19 weeks, and the intervention focused on the systematic use of narrative and expository texts and dialogic interactions between teachers and students delivered in whole-classroom formats. Read aloud intervention lessons included before-, during-, and after-reading components and explicit instruction targeted comprehension and vocabulary knowledge. Teachers in the comparison condition implemented the same amount of read aloud instruction, focusing on strategies they believed would help their students with comprehension and vocabulary. On some, but not all, outcome measures, intervention students at low risk and high risk for language difficulties outperformed comparable students in the comparison group. Implications are discussed.

Scott K. Baker
Lana Edwards Santoro
UNIVERSITY OF OREGON

David J. Chard
SOUTHERN METHODIST
UNIVERSITY

Hank Fien
UNIVERSITY OF OREGON

Yonghan Park
GANGNEUNG-WONJU
NATIONAL UNIVERSITY

Janet Otterstedt
UNIVERSITY OF OREGON

THIS article presents the results of a read aloud intervention in grade 1 designed to make read aloud lessons systematic and explicit, while maintaining the enjoyable nature of the experience that teachers and students expect. The intervention was delivered in whole-classroom formats with all students in first-grade classrooms, and was intended to be particularly effective for students at risk for language or literacy difficulties.

Read Aloud Intervention Research

Reading aloud to students in preschool and the early grades is ubiquitous in American classrooms, and considerable research has been conducted on read aloud practices (Karweit & Wasik, 1996; Morrow & Brittain, 2003; Van Kleeck, Stahl, & Bauer, 2003; Varelas & Pappas, 2006; Wilkinson & Silliman, 2000). Although relatively few studies have addressed the impact of read aloud practices on child outcomes, enough research has been conducted to begin to synthesize intervention impact (e.g., Karweit & Wasik, 1996; Mol, Bus, & de Jong, 2009; National Early Literacy Panel [NELP], 2008; Swanson et al., 2011).

Recently, the NELP (2008) examined the impact of shared-reading interventions on outcomes for children from birth to age 5 in home- and center-based settings. The NELP examined 19 experiments or quasi-experiments published between 1985 and 2003 and found a moderate impact on oral language (effect size = 0.57) and print awareness (effect size = 0.50). None of the studies included listening comprehension as an outcome.

Using more rigorous study-selection criteria, the What Works Clearinghouse (WWC; Institute of Education Sciences, 2007), reviewed eight read aloud intervention studies (four not included in the NELP review) conducted with children age 3 to 5. Three of the eight focused on interactive shared storybook reading with preschool children, defined as a set of instructional practices where parents or teachers read aloud to children individually or in small or large groups. Before, during, or after the read aloud, the adult would facilitate interactive discussions with children about the text. Outcomes were rated as potentially positive for early reading and writing (mean effect size = 0.70), mixed for oral language (mean effect size = 0.08), and there were no discernible effects for print knowledge (mean effect size = -0.10).¹ Listening comprehension was not included as an outcome in any of these studies.

Five studies focused on dialogic reading, defined by the WWC as the use of a specified prompting or cueing system to promote student discourse and comprehension during read aloud. Dialogic reading also includes the use of role-playing and group discussion after reading where the child might play the role of the storyteller with adult support. Although dialogic practices are frequently distinguished from similar practices such as interactive storybook reading, the overlap is considerable, and the essential purpose—to get children to participate verbally and actively in the read aloud experience with an adult—is the same. Across these studies, dialogic reading was used individually with students or in small groups. Overall, dialogic reading had a positive effect on oral language (effect size = 0.50) and no discernible effect on phonological processing (effect size = 0.22). Listening comprehension was not included as an outcome in these studies.

At the K–3 level, the WWC (Institute of Education Sciences, 2007) also reviewed read aloud interventions, but only selected commercially available programs for review, not interventions of the sort that were reviewed in the early childhood synthesis. One study met evidence standards for research design elements and was examined for effectiveness. In this study, Phillips, Norris, Mason, and Kerr (1990) investigated the impact of read aloud in the context of little books, texts with high-frequency words, simple sentences, and thematic topics. The intervention was delivered to kindergarten children in home and school settings and had a potentially

positive effect on general reading achievement (effect size = 0.31; Institute of Education Sciences, 2007). Listening comprehension was not measured.

In the most recent synthesis, Swanson et al. (2011) selected a broad range of studies published between 1984 and 2008 using treatment-comparison designs, multiple treatment designs, and single-group or single-subject designs. Relevant to the current study, Swanson et al. only analyzed studies in which teachers or researchers delivered the intervention (i.e., not parents) and studies that focused on students at risk for academic difficulties. This synthesis is also unique in the broad range of outcome measures analyzed. Of the 29 studies reviewed, read aloud interventions had a positive effect on language outcomes (mean effect size = 0.29), phonological awareness (mean effect size = 0.78), print concepts (mean effect size = 0.86), comprehension (mean effect size = 0.70), and vocabulary (mean effect size = 1.02).

Only four of the elementary school studies, however, included a measure of comprehension (Swanson et al., 2011). In addition, only two of the 21 treatment-comparison studies—both preschool studies—met three design issues that substantively strengthen study quality: the use of random assignment to condition, the inclusion of fidelity of implementation procedures, and the use of standardized dependent measures (Raudenbush, 2005; Shadish, 2002; U.S. Department of Education, 2003). The current study was implemented in first grade and incorporated these research design features.

Theoretical Rationale for Development of the Intervention

The current intervention integrated two instructional design features: (a) lesson content and (b) how lesson content was organized and taught. Development of lesson content was guided by Perfetti's model of reading comprehension (Perfetti, 1999; Perfetti, Landi, & Oakhill, 2005) and the report by the National Reading Panel (2000) on evidenced-based principles of vocabulary and text comprehension instruction. Regarding content delivery, we used several key principles of explicit instruction (Archer & Hughes, 2011; Engelmann & Carnine, 1982; Kame'enui & Simmons, 1990) to organize instruction in whole-classroom formats.

Intervention Focus

Although the focus of the intervention was on listening comprehension, we targeted comprehension skills also relevant for reading comprehension (Snow, 2002). The fact that listening comprehension and reading comprehension are related throughout development supports this focus, as does the observation that word identification is the primary variable that limits how closely reading comprehension skill can approach listening comprehension (Perfetti et al., 2005). That is, when students have strong word-reading skills, reading comprehension and listening comprehension should be highly aligned, suggesting that underlying strategies should support both reading and listening comprehension.

Our intention was to teach students to use the same sorts of strategies used to comprehend text when their teachers read to them in first grade that they could use to read text on their own, a situation that would become increasingly important as the texts students read independently became more complex in terms of structure, content, and vocabulary. Perfetti (1999) provided a conceptualization of compre-

hension development that can be used to frame both reading and listening comprehension interventions.

Perfetti et al. (2005) suggested there are two major classes of processing events in reading comprehension: word identification (i.e., reading words) and the “engagement in language processing mechanisms that assemble these words into messages” (p. 229). In the current read aloud intervention, teachers read the texts and taught students strategies to understand the “messages” in the words of the texts. Perfetti (1999) suggested that comprehension occurs when individuals build a good mental representation of the text. Good comprehenders do this by ensuring that the coherence of the text remains robust during text reading, and individuals who have a high “standard of coherence” (i.e., strive for a deep level of understanding) comprehend texts better than individuals who have a more lax standard.

The focus in the current read aloud intervention was on how teachers could help students develop strong mental representations of the texts being read and, over the course of the intervention, help students develop higher standards of coherence for text comprehension. Perfetti et al. (2005) identified a number of factors that influence how well students are able to construct a strong representation of the text. General background knowledge about the topic has an important influence, as does general linguistic knowledge (e.g., syntactic and morphologic knowledge). These factors influence how well readers are able to apply specific comprehension processes, including making text-based inferences, monitoring comprehension, and being sensitive to text structure, to construct mental representations of the text.

Teaching Comprehension Explicitly

Regarding intervention research on teaching comprehension strategies to students, the National Reading Panel (2000) identified seven specific strategies that offered a firm scientific basis for concluding that they improve comprehension. These strategies include monitoring comprehension and summarizing texts, representing texts structurally and graphically, asking and generating questions, and working cooperatively with others. Six of the seven approaches were active components in the read aloud intervention. Table 1 identifies these seven components and briefly describes how they were incorporated into the read aloud intervention.

Explicit instruction was the underlying foundation for how the intervention was delivered by teachers (Coyne, Kame'enui, & Carnine, 2007). To our knowledge, explicit instruction using these core principles, or others typically associated with explicit instruction (Archer & Hughes, 2011), have not been used in evaluations of read aloud interventions (Institute of Education Sciences, 2007; NELP, 2008; Swanson et al., 2011). In previous work on read alouds, an explicit instruction component has been added before or after text reading (Beck & McKeown, 2001, 2007; McKeown & Beck, 2003; Whitehurst et al., 1994; Zevenbergen, Whitehurst, & Zevenbergen, 2003), or read aloud interactions used structured, text-focused discourse guided by the teacher (e.g., Beck & McKeown, 2001). We designed the current intervention to be teacher driven and highly explicit throughout the read aloud lessons.

Compelling evidence indicates that explicit instruction has a positive impact on a range of student academic outcomes, particularly for students who are at risk for academic difficulties (Gersten, 1998; Swanson & Hoskyn, 1998; Vaughn, Gersten, & Chard, 2000). The read aloud intervention incorporated four principles of explicit

Table 1. Evidence-Based Components (National Reading Panel, 2000) of Comprehension Instruction Used in the Read Aloud Intervention

Component	Explanation
1. Comprehension monitoring	Teachers explained and demonstrated the importance of understanding what the text was about. During reading, they paused in strategic places to address difficult words in context, to make inferences, and to summarize text.
2. Cooperative learning ^a	Students were taught a specific approach for working with a partner (i.e., a book buddy). In each lesson, students worked together on text retells and other comprehension activities.
3. Graphic and semantic organizers ^a	Teachers used specific visual organizers for narrative texts and information texts. The narrative text organizer focused on story grammar elements and the information text organizer focused the K-W-L format and three strategic questions associated with learning about animals.
4. Story structure ^a	Students were taught the difference between narrative texts and information texts. Each unit focused on one information text and one narrative text and each lesson began with a discussion of ways to determine text type.
5. Question answering ^a	Teachers asked strategically placed questions throughout the lesson. Teachers also learned generic questioning approaches they could use to clarify student responses, elicit more elaborate responses, and extend discussions.
6. Question generation ^a	Teachers taught students basic questions they should ask (a) about narrative texts and expository texts, (b) to figure out word meanings when listening to texts, and (c) to predict what a text might be about before reading it and what might come next during reading.
7. Summarization ^a	Students were taught a systematic process for retelling/summarizing narrative texts and information texts.

^aDenotes components that were used explicitly and consistently in read aloud lessons.

instruction in the lesson design and associated professional development with teachers. First, lessons were organized around a model, lead, test approach (Archer & Hughes, 2011). That is, teachers used think aloud techniques (Kucan & Beck, 1997) to provide step-by-step demonstrations of the comprehension strategies and decision-making processes they wanted students to use to comprehend texts. Then, teachers led students through similar routines as the teacher read the text with students. The cycle concluded with students practicing comprehension strategies independently, with teacher support and feedback.

Second, read aloud lesson content was sequenced to become more complex over time, to build strategically on previous skills and strategies learned, and to require greater responsibility on the part of students to understand text (i.e., to develop coherent representations of text). For example, narrative texts with more complex story structures were used in later lessons, and students provided increasingly elaborate summarizations of texts over time.

Third, teachers and students engaged in frequent interactions about texts. In focus groups to prepare for the development of the intervention, we learned that teachers wanted read alouds to be highly interactive, but our classroom observations indicated that they frequently were not (Fisher, Flood, Lapp, & Frey, 2004). Read alouds might be used, for example, as “down time” after lunch or recess, and many children appeared passive as the story was read. Or, if read alouds were interactive, a handful

of children responded frequently, and many children did not respond at all. We built into the lesson structure frequent interactions between teachers and students and worked with teachers on how to interact with all students, particularly at-risk students.

Fourth, teachers provided extensive feedback to students. They affirmed students' responses and addressed aspects of students' responses they wanted to highlight and extend. Extending responses provided additional teacher models of complete responses and demonstrations of increased standards of coherence, and continued the comprehension discussion. When students responded incorrectly, teachers applied correction procedures. When the correct answer could be found in the text, teachers would draw students' attention to where in the text the correct answer could be found (a response to a literal question, for example). When the answer required a more complex inference, teachers would think aloud to demonstrate the process used to arrive at a justifiable answer.

Research Questions

Research questions addressed the impact of the whole-classroom read aloud intervention on student comprehension and vocabulary knowledge. We expected the read aloud intervention to be effective with the full range of students in first-grade classrooms, but most effective for students who were at risk for language or literacy difficulties. Specifically, we predicted that students at risk for language difficulties, or at risk for both language and literacy difficulties, would derive the greatest benefit from the intervention. We expected students with literacy difficulties, but without associated language difficulties, to derive the next greatest benefit. Finally, we expected students who were not at risk for language or literacy difficulties to also benefit from the intervention, but not to the same extent as at-risk students.

We measured intervention impact on two types of listening comprehension outcomes: (a) a receptive comprehension measure and (b) two expressive, text-based retell measures. One retell measure targeted narrative text and one targeted expository text. We also measured intervention impact on student vocabulary knowledge. We tested three categories of words. Category 1 words were taught in the intervention condition but may or may not have been taught in the comparison condition. Category 2 words were exposed or taught in both conditions, and Category 3 words were not taught in either condition.

We hypothesized that students in the intervention condition would outperform students on the vocabulary measure overall, and on Category 1 and Category 2 words specifically, because of greater word exposure (Category 1 words) and the more explicit instructional approach (Category 2 words). We addressed Category 1 words specifically because the whole-classroom format of the intervention has not been thoroughly investigated in previous research, and we wanted to test directly the assumption that exposure differences would lead to better outcomes in the intervention condition (Biemiller, 2001; Coyne, McCoach, & Kapp, 2007; Penno, Wilkinson, & Moore, 2002; Wasik & Bond, 2001). Finally, because previous research on read alouds has not found that vocabulary outcomes generalize to untaught words (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004), we did not have a specific hypothesis regarding Category 3 words.

Method

Participant Schools

We conducted a randomized control trial to assess the impact of the read aloud intervention on student comprehension and vocabulary. Three school districts in the Pacific Northwest were contacted to determine potential interest among elementary schools. A meeting was convened for interested principals and teachers. Twelve schools, with one first-grade teacher per school, agreed to participate. These 12 school-teacher units were randomly assigned to the intervention or comparison condition. School-level information is summarized in Table 2. Data from the National Center for Education Statistics indicate that intervention and comparison schools were comparable statistically on all demographic characteristics.

Teacher Participants

One first-grade classroom teacher per school provided the read aloud instruction for the study. Table 2 provides information about these teachers. Teachers were White and female, except for one White male teacher. Intervention and comparison teachers were similar in terms of age and teaching experience.

Student Participants

All 225 students in the 12 classrooms participated in the study (see Table 2). The mean number of students in intervention and comparison classrooms was 20 and 18, respectively. At pretest, students were administered two screening measures to determine their risk status for language and literacy difficulties. The Semantics Composite subtests of the Test of Language Development—Primary (TOLD, 3rd ed.) were used to identify students at risk for language difficulties. Students who scored

Table 2. Description of the Sample by Condition

	Comparison	Intervention	All
Number of students (<i>n</i>)	103	122	225
Number of teachers/schools (<i>N</i>) ^a	6	6	12
Student characteristics:			
Low risk (<i>n</i> /%)	52 (50.5)	62 (50.8)	114 (50.7)
Language risk only (<i>n</i> /%)	28 (27.2)	30 (24.6)	58 (25.8)
Literacy risk only (<i>n</i> /%)	12 (11.7)	13 (10.7)	25 (11.1)
Combined risk (<i>n</i> /%)	11 (10.7)	17 (13.9)	28 (12.4)
Teacher characteristics:			
Female (<i>N</i>)	6	5	11
White (<i>N</i>)	6	6	12
Age 35 or higher (<i>N</i>)	5	5	10
Holding master's degree (<i>N</i>)	2	3	5
Average years of teaching	12 years	15 years	13 years
School characteristics:			
Average % of FRL ^b	54	56	55
Average % of females	48	49	49
Average % of minorities	27	20	24

^a Because only one class was sampled from each participating school, the numbers of participating schools and teachers were the same.

^b FRL = students receiving free or reduced-price lunch.

below the 25th percentile on the composite were considered at risk for language difficulties. The 25th percentile was used based on score distributions of the study sample student, and because previous research has used the 25th percentile for selecting at-risk students (e.g., Simmons et al., 2007). The Dynamic Indicators of Basic Early Literacy Skills (DIBELS, 6th ed.) was used to identify students at risk for literacy difficulties. Following guidelines (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002), students with an instructional recommendation of "intensive" were considered at high risk for literacy difficulties.

These two screening measures were used to create four student subgroups: (a) at risk language, low risk literacy; (b) at risk literacy, low risk language; (c) at risk language and literacy; and (d) low risk language and literacy. Table 2 presents descriptive information about students by subgroup. Differences by condition were not significant in the ratios of at-risk students or on any pretest outcomes ($p > .10$).

The Read Aloud Intervention

We relied on four features to structure how the read aloud intervention was developed for explicit instruction. First, books that first-grade teachers commonly read to their students were used for the read alouds. Selection criteria included the book's topic, student interest, length, cost, availability, representation of diversity, text coherence, and the text's alignment with content area standards, particularly science. Texts with different levels of coherence relations (Graesser, McNamara, & Louwerse, 2003) were intended to help students develop robust comprehension strategies.

Second, narrative and expository texts were integrated systematically throughout the intervention. Within each text type, consistent routines were used to support comprehension development. With narrative texts, students were taught a story element framework to predict who or what the story was about before reading and to identify key elements of stories during reading. This framework also facilitated retells after reading. Expository texts focused on different animals and types of animals, and students were taught three important features about any animal: what they look like, what they eat, and where they live. Students were taught to listen specifically for this content during the read aloud lesson.

Third, teachers were provided a set of prescriptive lessons for the intervention. Each lesson included a long-form lesson plan and short-form lesson plan. Long-form lesson plans were designed for teacher use during lesson preparation and planning. The long-form lesson plans detailed all of the lesson content, provided sample instructional language, and outlined text excerpts to highlight during the lesson. Short-form lesson plans were designed for use during implementation and consisted of a double-sided, single-page outline of the essential lesson content and suggested text excerpts for discussions.

Fourth, dialogic interactions during read aloud lessons occurred between teachers and students, and among students. For example, students learned a structured way to work in dyads on prescribed comprehension tasks, such as summarization. The dialogue objective was that teachers closely monitor student verbalizations about text, to extend student understanding through verbal interactions and to immediately address comprehension breakdowns and misconceptions that occurred when trying to understand complex text. Verbal interactions, then, were intended to help

students build better mental representations of text and monitor comprehension. Due to the highly interactive nature of teacher-student discussions, the dynamic components of the lessons were not prescribed. However, the professional development that teachers participated in emphasized, and provided feedback with practice in, how to promote verbal interactions through the use of question-asking prompts and follow-up responses.

Structure of intervention content. The read aloud intervention was implemented over 19 weeks. Early lessons emphasized teacher demonstrations of comprehension tasks using think-alouds, explicit demonstrations, and concise explanations. As lessons progressed, students assumed greater responsibility for accomplishing comprehension tasks independently with teacher feedback.

Each 2-week unit included six or seven lessons, each lasting 30 minutes. Three lessons focused on expository text and three or four lessons (depending on the week) focused on narrative text. One expository text and one narrative text were used during each unit. Expository lessons were taught before narrative lessons to help establish background knowledge. Both intervention and comparison teachers were asked to conduct read alouds for at least 8 days in each 2-week period. Intervention teachers used the extra days to complete lessons or for related activities (e.g., doing writing-based retell activities, reviewing previously read texts).

Before, during, and after text reading. Each lesson included a before, during, and after read aloud section. Before reading, identifying the book type (narrative or expository) established the purpose for reading and preparation. Critical vocabulary was also taught. During reading, lesson content centered on story grammar elements for narrative texts and K-W-L components for expository texts. Similar to McKeown and Beck (2003), teachers read sections of narrative text prior to showing pictures. With expository text, pictures or diagrams were shown while reading to help illustrate complex content. Questioning strategies were used to prompt discussions, and these techniques emphasized higher-level thinking skills such as main idea identification and drawing inferences. Regular, structured opportunities were provided for students to discuss the text with one another. These discussions occurred in student dyads and always included a discussion prompt or focus question as an organizer.

During reading, the bulk of explicit instruction focused on higher-order comprehension skills such as inference making and comprehension monitoring. For example, teachers demonstrated for students how to draw inferences when maintaining coherence called for this. Students would then practice making inferences with the teacher or with a partner. In terms of comprehension monitoring, teachers demonstrated how to summarize a text and would draw attention to how summarization could help them remember to really understand what the text is about (Kintsch & Rawson, 2005; Perfetti et al., 2005). Instruction on new or review vocabulary words also occurred as teachers read and discussed the text.

After reading, lesson content focused on summarizing and retelling practice, and vocabulary review. Teachers modeled retelling using standard formats (Baker, Gersten, & Scanlon, 2002), and students used these formats to practice retelling with the teacher or partner. For summarizing narrative texts, the visual prompt sheet included icons representing who the story was about, what happened first, what happened next, and what happened at the end. The K-W-L chart provided the format for expository book retells. This part of the lesson was also used for systematic review of key vocabulary (Santoro, Baker, Chard, & Howard, 2007; Santoro, Chard, Howard,

& Baker, 2008). Curriculum lessons and materials, with the exception of the read aloud books, are available by request from the authors.

Comparison Condition

Teachers in the comparison condition engaged in read aloud activities at least 4 days per week during the study. Comparison teachers were asked to use read aloud procedures they normally used, particularly procedures they felt would promote student comprehension. Three specific requests were also made for the purpose of the study: (a) comparison teachers selected a narrative or expository text that would serve as the focus of the read aloud for that day, (b) comparison teachers engaged in read aloud activities for approximately 30 minutes per day, and (c) at weeks 7 and 17, comparison teachers used a book we selected for the read aloud on 2 consecutive days. Intervention teachers used this same book, also on 2 consecutive days. In using common texts, our objective was to better document read aloud instruction and fidelity across conditions. We also selected words from these texts to assess student vocabulary.

Screening Measures for Language and Literacy

Test of Oral Language Development—Primary: Third Edition (TOLD-P-3; Newcomer & Hammill, 1997). The TOLD-P-3 is a standardized, norm-referenced measure of expressive language. Reliability is reported in the manual as .95 (Kuder-Richardson 20 coefficient, Newcomer & Hammill, 1997). Three subtests of the TOLD-P-3 were individually administered at pretest to identify students at risk for language difficulties: Picture Vocabulary, Relational Vocabulary, and Oral Vocabulary. Scores on these three subtests were combined to produce a semantics composite quotient, which was used for analysis.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Kaminski & Good, 1996). We administered DIBELS at pretest to identify students at risk for literacy difficulties. DIBELS are individually administered fluency measures, and there are three first-grade subtests. Each subtest took one minute to complete. Letter Naming Fluency (LNF) measured knowledge of letter names; Phonemic Segmentation Fluency (PSF) measured student skill in identifying individual sounds in spoken words; Nonsense Word Fluency (NWF) measured phonological decoding (Good, Baker, & Peyton, 2009). All three measures are very reliable (reported between .85 and .95) (Good et al., 2004). Also, the measures demonstrate content validity and criterion-related and predictive validity (Fien et al., 2010; Good et al., 2004, 2009) and construct validity (Fien et al., 2008).

Measures of Listening Comprehension

Gates-MacGinitie Test of Reading Comprehension, Listening Comprehension Subtest (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). The Gates-MacGinitie Listening Comprehension subtest was administered to all students at pretest and posttest to evaluate receptive comprehension. On this measure, the examiner read a short story to students, repeated a short segment from the story, and then prompted students to select one of three pictures that went with that part of the story (MacGinitie et al., 2000, p. 96). Reliability is reported as .81 for the fall of first grade (Kuder-

Richardson 20 coefficient; MacGinitie et al., 2000). The average correlation of the Listening Comprehension subtest with three other reading subsets administered concurrently in the fall of first grade is reported as .55 (MacGinitie et al., 2000). Predictive validity of the Gates-MacGinitie for the fall and spring of first grade is reported as .74 (MacGinitie et al., 2000). We double scored 20% of test protocols and achieved 100% accuracy.

Narrative Retells, Strong Narrative Assessment Procedure (SNAP; Strong, 1998). The SNAP is a standardized measure of listening comprehension and was individually administered to all students at pretest and posttest. On the SNAP, students listened to a tape-recorded story while following along in a wordless picture book via an auditory signal. At the end of the story, the examiner removed the book and the student retold the story in her own words, which we audio taped for scoring and analysis. We analyzed the retells according to the manual and following procedures developed by Morrow (1985). The inclusion of specific story elements and plot episodes were counted and provided an estimate of the student's comprehension. Two trained raters coded 20% of the protocols, and interrater reliability for total score was .85. Validity for retells administered in this manner is supported by evidence for evaluating developmental trends and improvements in narrative comprehension for students in kindergarten through second grade (Dougherty Stahl, 2009; Geva & Olson, 1983; Paris & Paris, 2003).

Expository retells. We adapted SNAP procedures to assess student comprehension of expository text at pretest and posttest. Students listened to an audiotape of a text about killer whales while following along and looking at pictures. At the conclusion, students retold everything they could about the text they heard. A student's score was the number of correct concepts used in the retell. Two raters coded 20% of the protocols and interrater reliability was .98. In a study exploring the expository retell skills of first-grade students, Moss (1997) found that first graders were able to construct fairly complete and cohesive retellings.

The narrative and expository retells were transcribed according to standard transcription procedures used with the Systematic Analysis of Language Transcript software (Miller & Chapman, 1993). Coder training consisted of scoring retells first as a group, then independently, with frequent discussions regarding scoring differences throughout. After training, each coder scored a narrative and expository retell independently, and all coders achieved agreement of at least .80 on each before coding independently. Interrater agreement was determined by counting the number of line-by-line agreements and line-by-line disagreements, then dividing by the total number of agreements and disagreements. Coders, blind to condition, then scored the student transcripts for story elements, plot episodes, expository concepts, and examiner prompts.

Measure of Vocabulary Knowledge: Depth of Vocabulary Knowledge (DOK)

The National Reading Panel (2001) suggested that vocabulary growth that is sensitive to intervention effects can be measured best through researcher-developed measures. We developed the DOK measure following procedures used in previous research (Eller, Pappas, & Brown, 1988) and individually assessed students at pretest and posttest. Each DOK assessment consisted of 16 words: three academic language words (e.g., main character, index, storybook) and 13 words selected from the texts

used in classroom read alouds. The 13 text words were sampled from a pool of 63 words, 34 words from the intervention narrative texts and 29 words from the intervention expository texts. These words were categorized according to read aloud instruction. Category 1 words were taught explicitly or targeted for discussion in the intervention condition, and exposure for students in the comparison condition may have been minimal or none. Category 2 words were exposed and/or taught in both conditions when the same read aloud texts were used during weeks 7 and 17 of the study. Category 3 words also were selected from the books used during weeks 7 and 17 but were not highlighted for classroom instruction.

On the DOK assessment, students verbally told the meaning of each word and used each word in a sentence. Responses were scored using a modified version of Eller's vocabulary scoring criteria (Eller et al., 1988). Each word was given a score for definition (0–2 points) and use (0–2 points). Interrater agreement based on total score was .95, and internal consistency, as measured by Cronbach's alpha, was .80 and .87 at pretest and posttest, respectively.

Implementation of the Read Aloud Intervention

Training intervention teachers. Prior to implementation, intervention teachers participated in a full-day training that included the rationale for the intervention and an overview of the lessons. Selected lessons were modeled by trainers, and video clips were used to discuss critical features of instruction and implementation. Teachers also practiced teaching the read aloud lessons with feedback from trainers.

Teacher training continued during the first 2 weeks of implementation. A staff member observed each teacher during narrative and expository text instruction and provided feedback regarding intervention components and delivery. Also, a half-day follow-up training session was provided at mid-implementation (week 9) to review lesson formats and introduce new lessons for the second half of the intervention.

Implementation fidelity. Two observations were conducted in each intervention classroom to measure fidelity of implementation. The observation instrument was based primarily on components taken from the short form of the lesson plan teachers used to guide instruction. Teachers were aware of the purpose of the observation. We did not provide them with a copy of the instrument, but teachers were aware that the components of the instrument mirrored the instructional objectives on the read aloud lesson plan. Observers were trained to use the fidelity instrument and achieved reliability of at least .85 during training before conducting classroom observations. Critical intervention components were rated on a 0–1 scale according to the presence or absence of each component, and on a 0–2 scale based on quality of implementation. For the intervention, teachers were required to implement specific components (e.g., teaching specific vocabulary terms) and suggestions were also made regarding implementation activities, depending on the flow of the lesson. On required components, overall fidelity was .84.

Comparison classrooms were also observed. None of the comparison teachers were observed using any of the intervention books or materials (except when asked, during weeks 7 and 17). This is not surprising given that intervention and comparison teachers were in different schools. Time spent on read aloud activities was measured in both conditions. Teachers in intervention and comparison classrooms spent an average of 27.7 and 29.2 minutes on read aloud activities, respectively.

Analytic Strategy

In this analysis, students were nested within classrooms and classrooms were nested within schools. Because participation was limited to one classroom per school, the hierarchical linear models (HLM) (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) were constructed at two levels, student and classroom. Intervention condition was included as a classroom-level predictor. We limited the number of classroom-level predictors to one (i.e., intervention condition) because of the small sample size (Raudenbush & Bryk, 2002). At the student level, initial language and literacy risk status were included as predictors to explore the effect of risk status and their interactions with the intervention on student outcomes. We used low risk (i.e., students who were not at risk for language or literacy difficulties) as a reference category in our analytic models, and thus the relative contributions of the identified risks (language risk, literacy risk, and combined risk) should be interpreted in relation to the reference category. We applied multilevel ANCOVA models to our analyses assuming random intercepts but fixed slopes (Raudenbush & Bryk, 2002). Preliminary analyses revealed that the random components of the slopes were not statistically significant.

For the analyses, we examined two different models for each outcome: a main-effect only model and an interaction-effect model. The main-effect only model assumed that the intervention effects would be consistent across different risk-status groups, and the interaction-effect model assumed that intervention effects would be different for students depending on risk status. We compared the fits of these contrasting models, and the results will be discussed based on the model with better fit. The specific models are as follows: (a) main-effect only model: $Y_{ij} = \gamma_{00} + \gamma_{01} \times (\text{intervention}) + \gamma_{10} \times (\text{language risk}) + \gamma_{20} \times (\text{literacy risk}) + \gamma_{30} \times (\text{combined risk}) + u_{0j} + r_{ij}$; and (b) interaction-effect model: $Y_{ij} = \gamma_{00} + \gamma_{01} \times (\text{intervention}) + \gamma_{10} \times (\text{language risk}) + \gamma_{11} \times (\text{language risk} \times \text{intervention}) + \gamma_{20} \times (\text{literacy risk}) + \gamma_{21} \times \gamma_{30} \times (\text{combined risk}) + \gamma_{31} \times (\text{combined risk} \times \text{intervention}) + u_{0j} + r_{ij}$.

Standardized effect sizes were calculated based on clustered randomized design (Hedges, 2007; Spybrook, 2008). The formula for calculating effect sizes was $d = \gamma / \sqrt{\tau + \sigma^2}$, where γ is the adjusted group mean difference, τ is the classroom-level variance, and σ^2 is the student-level variance.

Missing Data

Some students in the final analytic sample ($n = 225$) did not complete one or more of the assessments at pretest or posttest, primarily due to absence on assessment dates. This missing information was minimal: less than 7% of students had a missing pretest or posttest score on any of the assessments. To analyze the potential impact of missing data, students with missing data on any test were compared to students without missing data on that test, in relation to all other available test scores. Test scores between students with and without missing data were not significantly different across the intervention and comparison conditions, confirming that data were missing at random.

Because data were missing at random, we used the listwise deletion for analysis on each of the four posttests. That is, the final analysis only included students with

Table 3. Descriptive Statistics of Student Performance by Intervention Condition and by Risk Status

	Pretest Scores						Posttest Scores					
	Comparison		Intervention		All		Comparison		Intervention		All	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Listening comprehension (<i>n</i> = 213):												
Low risk	14.46	(4.18)	14.71	(3.18)	14.60	(3.66)	17.76	(2.08)	17.36	(2.13)	17.54	(2.11)
Language risk only	10.92	(3.88)	11.13	(4.85)	11.04	(4.41)	14.42	(3.75)	15.30	(3.81)	14.91	(3.77)
Literacy risk only	10.55	(2.30)	15.69	(2.36)	13.33	(3.47)	15.91	(3.05)	18.54	(1.27)	17.33	(2.58)
Combined risk	9.82	(3.12)	8.40	(2.41)	9.00	(2.77)	12.27	(3.88)	13.00	(2.27)	12.69	(3.00)
Total	12.59	(4.26)	13.09	(4.26)	12.87	(4.26)	16.08	(3.47)	16.40	(3.08)	16.26	(3.26)
Narrative retell (<i>n</i> = 214):												
Low risk	8.62	(5.83)	10.14	(6.39)	9.44	(6.16)	13.70	(6.38)	16.71	(5.30)	15.33	(5.99)
Language risk only	8.16	(5.90)	6.47	(5.92)	7.24	(5.92)	12.24	(4.46)	12.53	(7.14)	12.40	(6.02)
Literacy risk only	8.33	(5.09)	7.46	(6.24)	7.88	(5.61)	10.83	(4.95)	13.85	(7.32)	12.40	(6.36)
Combined risk	3.11	(3.18)	6.38	(4.56)	5.20	(4.35)	6.78	(5.40)	11.63	(5.33)	9.88	(5.75)
Total	7.95	(5.71)	8.40	(6.22)	8.20	(5.99)	12.31	(5.96)	14.64	(6.35)	13.60	(6.27)
Expository retell (<i>n</i> = 208):												
Low risk	2.73	(2.44)	2.61	(2.17)	2.66	(2.28)	4.85	(2.94)	5.85	(2.50)	5.40	(2.74)
Language risk only	1.52	(1.38)	1.38	(1.40)	1.44	(1.38)	4.04	(2.87)	3.93	(3.38)	3.98	(3.13)
Literacy risk only	1.67	(.98)	3.31	(2.63)	2.52	(2.14)	3.42	(2.23)	5.54	(2.44)	4.52	(2.54)
Combined risk	1.11	(.93)	1.40	(1.35)	1.29	(1.20)	3.78	(3.19)	3.47	(1.96)	3.58	(2.43)
Total	2.13	(2.03)	2.22	(2.06)	2.18	(2.05)	4.36	(2.87)	5.03	(2.83)	4.73	(2.86)
Vocabulary (<i>n</i> = 215):												
Low risk	13.52	(7.92)	13.05	(8.33)	13.27	(8.11)	24.44	(10.62)	35.07	(10.86)	30.15	(11.95)
Language risk only	6.44	(4.87)	6.87	(5.95)	6.67	(5.44)	17.08	(8.17)	23.03	(11.92)	20.33	(10.72)
Literacy risk only	6.33	(4.74)	9.77	(6.47)	8.12	(5.85)	19.08	(6.86)	31.08	(11.88)	25.32	(11.38)
Combined risk	3.18	(2.56)	4.75	(3.02)	4.11	(2.90)	11.45	(7.03)	17.31	(8.31)	14.93	(8.21)
Total	9.67	(7.56)	9.97	(7.71)	9.83	(7.63)	20.45	(10.22)	29.11	(12.80)	25.16	(12.45)

Note.—Statistics are based on the scores of students who completed both pretest and posttest of respective measures.

pretest and posttest scores on the specific assessment. Consequently, the number of students in the analysis varied slightly depending on measure, as shown in Table 3.

Results

We examined the impact of the read aloud intervention on four student outcomes: listening comprehension, narrative retell, expository retell, and vocabulary. The listening comprehension measure assessed receptive language, and the other two retell comprehension measures and the vocabulary measure assessed expressive language. Table 3 summarizes the descriptive statistics of student outcomes in pretest and posttest by measure, intervention condition, and risk status.

Before conducting the analysis of intervention effects, we first checked the comparability of the students in intervention and control conditions by comparing their pretest scores across different risk-status groups. This was done by analyzing the interaction-effect models with pretest scores as outcomes. As shown in Table 4, there were no statistically significant differences in students' pretest scores between inter-

Table 4. Examination of Pretest Difference between Intervention and Comparison Groups

Fixed Effect	Listening Comprehension			Narrative Retell			Expository Retell			Vocabulary		
	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t
Intercept, γ_{00}	14.25	.59	24.28***	8.62	.81	10.62***	2.73	.28	9.79***	13.57	1.07	12.71***
× Intervention, γ_{01}	.70	.81	.87	1.52	1.10	1.38	-.10	.37	-.26	-.56	1.47	-.38
Language risk, γ_{10}	-3.47	.85	-4.07***	-.46	1.40	-.33	-1.21	.49	-2.47*	-7.21	1.64	-4.40***
× Intervention, γ_{11}	-.57	1.17	-.49	-3.21	1.90	-1.69	-.04	.65	-.07	1.29	2.22	.58
Literacy risk, γ_{20}	-3.59	1.15	-3.12**	-.29	1.84	-.16	-1.06	.62	-1.71	-6.79	2.13	-3.18**
× Intervention, γ_{21}	4.33	1.58	2.74**	-2.39	2.55	-.94	1.74	.86	2.03*	3.37	2.94	1.15
Combined risk, γ_{30}	-4.15	1.19	-3.49**	-5.50	2.08	-2.65**	-1.62	.70	-2.31*	-10.35	2.21	-4.69***
× Intervention, γ_{31}	-2.56	1.55	-1.65	1.74	2.63	.66	.39	.89	.43	2.14	2.91	.74
Random Effect	Variance	df	χ^2	Variance	df	χ^2	Variance	df	χ^2	Variance	df	χ^2
Level 2, u_{0j}	.61	10	22.71*	.01	10	8.70	.00	10	10.91	1.63	10	20.26*
Level 1, r_{ij}	12.51			32.85			3.71			42.89		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

vention and comparison conditions except for the students with literacy risk only, for whom the intervention condition had higher pretest scores than the control condition on two measures—listening comprehension and expository retell (see γ_{21} for these two outcomes in Table 4). This is also apparent in the mean pretest scores in Table 3. Therefore, we interpret our results about intervention effects with caution by considering these preexisting differences for the students with literacy risk.

In the next section, we present findings on read aloud intervention effects by comparing student performance by condition on the four primary posttest measures. Then, we provide a more detailed analysis of performance on the vocabulary measure, to examine intervention effects by varying degrees of word exposure during instruction.

Impact of the Read Aloud Intervention

As described earlier, we fit both the main-effect only model and the interaction-effect model for each outcome. In the main-effect only models, the intervention effects (γ_{01}) were assumed to be consistent across all risk-status groups, while the intervention effects were assumed to be different depending on risk-status group in the interaction-effect models. The results of the interaction-effect models revealed that no statistically significant interaction terms were found on narrative retell, expository retell, or vocabulary. On listening comprehension, a significant interaction was found between literacy risk and intervention condition ($\gamma_{21} = 3.09, t = 2.60, p = .01$). This significant interaction, however, should not be taken too seriously in consideration of the preexisting interaction between literacy risk and intervention condition on listening comprehension (i.e., pretest difference between intervention and comparison conditions for students with literacy risk only; see Table 5). Further analysis that controlled for pretest scores also confirmed that there were no statistically significant interactions between risk status and intervention condition on any outcome after preexisting differences were taken into account.

We compared the model fits between the main-effect only and interaction-effect models, drawing on their deviances and Bayesian information criterion (BIC) indices for each outcome. Deviance tests revealed that the interaction-effect models were not significantly better than the simpler main-effect models on any of the four outcome measures: listening comprehension [$\chi^2(3) = 6.97, p = .07$], narrative retell [$\chi^2(3) = 3.79, p = .28$], expository retell [$\chi^2(3) = 2.90, p > .50$], and vocabulary [$\chi^2(3) = 3.67, p = .30$]. BIC indices also favored the main-effect only models over the interaction-effects model on all outcomes: listening comprehension (1,061.85 for the main-effect only model vs. 1,070.97 for the interaction-effect model), narrative retell (1,466.05 vs. 1,478.51), expository retell (1,127.37 vs. 1,140.49), and vocabulary (1,704.17 vs. 1,716.73). In sum, multiple sources of evidence support the conclusion that the main-effect only models fit the data better than the interaction-effect models on all four outcomes. This indicates that the read aloud intervention effects were rather consistent across different risk-status groups. Thus, the results below are discussed focusing on the main-effect only models presented in Table 5.

The results of the main-effects models revealed that the read aloud intervention had significant effects on students' narrative retell and vocabulary outcomes. On the narrative retell measure, students in the intervention condition performed better at posttest than students in the comparison condition by 2.42 points on average ($\gamma_{01} =$

Table 5. Examination of Intervention Effects on Posttest Outcomes (Main-Effect Only Models)

Fixed Effect	Listening Comprehension			Narrative Retell			Expository Retell			Vocabulary		
	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t
Intercept, γ_{00}	17.33	.47	37.18***	14.22	.69	20.62***	5.09	.33	15.33***	25.12	1.45	17.32***
× Intervention, γ_{01}	.44	.60	.73	2.42	.78	3.11*	.78	.38	2.06	9.35	1.79	5.23***
Language risk, γ_{10}	−2.90	.45	−6.47***	−3.01	.94	−3.21**	−1.34	.45	−2.97**	−10.12	1.61	−6.27***
Literacy risk, γ_{20}	−.20	.60	−.34	−3.08	1.28	−2.40*	−.98	.62	−1.58	−4.77	2.19	−2.18*
Combined risk, γ_{30}	−4.67	.59	−7.94***	−6.12	1.23	−4.99***	−2.22	.60	−3.71***	−16.12	2.13	−7.59***
Random Effect	Variance	df	χ^2	Variance	df	χ^2	Variance	df	χ^2	Variance	df	χ^2
Level 2, u_{0j}	.68	10	33.79***	.00	10	6.97	.00	10	4.23	4.33	10	21.88*
Level 1, r_{ij}	6.84			33.70			7.82			96.40		
Model Fit	Deviance = 1,024.32 (N of parameters = 7) BIC = 1061.85			Deviance = 1,428.14 (N of parameters = 7) BIC = 1,466.05			Deviance = 1,089.52 (N of parameters = 7) BIC = 1,127.37			Deviance = 1,666.28 (N of parameters = 7) BIC = 1,704.17		
Standardized effect size	.16			.42			.28			.93		

* $p < .05$.

** $p < .01$.

*** $p < .001$.

2.42, $t = 3.11$, $p = .01$) across different risk-status groups. The standardized effect size was moderate (0.42), equal to a difference of approximately 16 percentile points. Independent of intervention condition, at-risk students scored lower on the narrative retell measure at posttest than low-risk students. Compared to low-risk students, narrative retell posttest scores were lower for students with language risk only ($\gamma_{10} = -3.01$, $t = -3.21$, $p = .002$), students with literacy risk only ($\gamma_{20} = -3.08$, $t = -2.40$, $p = .02$), and students with combined language and literacy risk ($\gamma_{30} = -6.12$, $t = -4.99$, $p < .001$).

Similarly, the intervention effect was statistically and substantially significant on the vocabulary outcome. Students in the intervention condition demonstrated higher vocabulary scores than students in the comparison condition, on average by 9.35 points across different risk-status groups ($\gamma_{01} = 9.35$, $t = 5.23$, $p < .001$). The standardized effect size was high (0.93), equaling a difference of approximately 32 percentile points. Independent of intervention condition, in relation to low-risk students, vocabulary scores were significantly lower for students with language risk only ($\gamma_{10} = -10.12$, $t = -6.27$, $p < .001$), students with literacy risk only ($\gamma_{20} = -4.77$, $t = -2.18$, $p = .03$), and students with combined language and literacy risk ($\gamma_{30} = -16.12$, $t = -7.59$, $p < .001$).

The read aloud intervention did not have a statistically significant effect on student listening comprehension ($\gamma_{01} = 0.44$, $t = 0.73$, $p = .48$) or expository retell ($\gamma_{01} = 0.78$, $t = 2.06$, $p = .07$), although the trend on the expository retell measure approached significance.

Follow-up Analyses of the Vocabulary Outcomes by Word Type

Additional analyses were conducted to explore vocabulary outcomes based on three categories of word types: words taught only in the intervention condition (Category 1), words exposed in both conditions and taught explicitly in the intervention condition but not necessarily taught explicitly in the comparison condition (Category 2), and words exposed in both conditions but not necessarily taught explicitly in either condition (Category 3).

As shown in Table 6, we examined the read aloud intervention effects on the vocabulary outcomes by each of three word categories. For this purpose, we fit the main-effect only models on the basis of our previous findings. Results were consistent with our predictions. Intervention effects (γ_{01}) were statistically significant for Category 1 words ($\gamma_{01} = 3.85$, $t = 5.60$, $p < .001$) and Category 2 words ($\gamma_{01} = 2.19$, $t = 3.66$, $p = .005$), but not for Category 3 words ($\gamma_{01} = 0.36$, $t = 1.25$, $p = .24$). The standardized effect sizes were 1.05 for Category 1 words, 0.65 for Category 2 words, and 0.17 for Category 3 words.

Discussion

Read aloud practices are used extensively in preschool and school settings, but there is substantial variability in how teachers conduct these practices, and no consistent format is used by the majority of teachers (Fisher et al., 2004). Furthermore, little experimental research has been conducted to determine whether certain practices benefit students more than others. In a recent synthesis, for example, Swanson et al. (2011) reported that 29 read aloud intervention studies had been implemented in

Table 6. Examination of Intervention Effects in Vocabulary Subscales

Fixed Effect	(Maximum Expected Effect) Category 1 Words			(Medium Expected Effect) Category 2 Words			(Minimum Expected Effect) Category 3 Words		
	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t
Intercept, γ_{00}	7.93	.55	14.40***	8.66	.48	17.87***	2.91	.25	11.64***
× Intervention, γ_{01}	3.85	.69	5.60***	2.19	.60	3.66**	.36	.28	1.25
Language risk, γ_{10}	-3.20	.59	-5.46***	-2.45	.54	-4.55***	-1.18	.34	-3.49**
Literacy risk, γ_{20}	-1.28	.80	-1.61	-1.65	.73	-2.25*	-.26	.46	-.56
Combined risk, γ_{30}	-5.37	.77	-6.95***	-4.78	.71	-6.75***	-1.63	.44	-3.68**
Random Effect	Variance	df	χ^2	Variance	df	χ^2	Variance	df	χ^2
Level 2, u_{0j}	.72	10	24.78**	.49	10	22.11*	.00	10	10.17
Level 1, r_{ij}	12.68			10.71			4.40		
Standardized effect size		1.05			.65			.17	

* $p < .05$.

** $p < .01$.

*** $p < .001$.

preschool or school settings, and 8 used random assignment to condition, 4 in preschool settings, and 4 in elementary school settings (K–3). Only three of these studies investigated outcomes targeting higher-order skills such as comprehension and vocabulary, and these studies were conducted in preschool settings.

The purpose of the current study was to investigate the impact of a read aloud intervention in first grade. Our goal was to develop a read aloud intervention that preserved the pleasurable dimension of read aloud practice while simultaneously constructing systematic opportunities for teachers to explicitly instruct students in ways to better understand the texts. The content of this instruction was comprehension strategies and vocabulary knowledge. The intervention was organized around the concept of explicit instruction where teachers assisted students in developing strong representations of text (Perfetti et al., 2005) by using approaches to comprehension instruction supported by the National Reading Panel (2000).

An important aspect of this study is that the intervention was tested under the kinds of conditions in which read aloud practices are commonly used in the early elementary grades (Duke & Martin, 2008; Fisher et al., 2004; Morrow & Brittain, 2003). Three issues are relevant. First, all children in first-grade classrooms participated. Thus, the format used for delivery was whole-classroom instruction. Second, the length of time for the intervention was about 30 minutes per day, 4 days per week. This is about the amount of time teachers indicate they have available for read aloud practices (Fisher et al., 2004; Good et al., 2009). Third, read aloud lessons were text based, and specific narrative and expository texts were juxtaposed systematically to improve comprehension.

By constructing a whole-classroom approach, our objective was to develop an intervention that would address the broad range of experiences, knowledge, and skills students typically bring to each first-grade classroom. We wanted the intervention to be particularly sensitive to the needs of students at risk for academic difficulties, but also to be effective with students who were not at risk (i.e., children who were at low risk for difficulties). Two types of risk categories were identified, students at risk for language difficulties and students at risk for literacy difficulties (Catts, Fey, Tomblin, & Zhang, 2002; Gough & Tunmer, 1986; Rumelhart, 1977; Stanovich, 1980). We hypothesized that the intervention would have the greatest impact on students at risk for language difficulties, followed by students at risk for literacy difficulties, and finally students who were not at risk for academic difficulties (Catts, Fey, Zhang, & Tomblin, 1999).

Study Outcomes by Risk Categories

Two types of HLM models were fit to the outcome data, main-effect only models and interaction-effect models. Based on our differential effects prediction, we expected that the interaction-effects model would fit the data better. However, the main-effects only models fit the data better, indicating that intervention impact was consistent across the four risk groups. Overall, the intervention had a statistically significant impact on the narrative retell measure and the vocabulary measure. The intervention did not have an effect on the information retell measure or on the listening comprehension measure. In the following section, the effect sizes of our intervention are discussed after controlling for pretest scores as summarized in Table 7.

Table 7. Standardized Effect Sizes of the Read Aloud Intervention after Controlling for Pretest Scores

	Listening Comprehension	Narrative Retell	Expository Retell	Vocabulary
Overall main effect:				
Effect size for all students	.08	.42	.25	1.02
Effect by risk status:				
Effect size for low-risk students	-.22	.46	.41	1.33
Effect size for language-risk only students	.34	.17	-.02	.64
Effect size for literacy-risk only students	.29	.62	.55	1.11
Effect for students with combined risk	.60	.68	-.17	.56

Expressive comprehension outcomes. On the narrative retell measure, the significant impact produced an effect size of 0.42, which is equal to a difference of approximately 16 percentile points between intervention and comparison students. Effect sizes based on the four risk groups ranged from 0.17 (language risk students) to 0.68 (both language and literacy risk). On the information retell measure, the overall effect size was 0.25, and although not statistically significant, the trend favored the intervention condition. Not detecting a significant information retell effect may reflect an underpowered study.

There are two plausible reasons why the intervention may have had a stronger impact on the narrative retell measure than on the information retell measure. First, the way impact was measured may have been more sensitive to narrative retellings. We used the SNAP for the narrative retells, and adapted this procedure for the expository retellings. Our adaptation may not have accurately captured student knowledge of the information text. On the surface, it might appear that a retelling assessment measure would be more sensitive to narrative texts than information texts. However, we attempted to align the focus of instruction (i.e., teaching structured routines for remembering content) with how retellings were measured (the number of correct concepts), and previous research has found that first graders are capable of constructing cohesive retellings of information content (Leung, 2008; Moss, 1997; Smolkin & Donovan, 2001; Varelas & Pappas, 2006).

Second, the intervention may simply have had a stronger effect on narrative comprehension than expository comprehension. Although we believe that both narrative and expository lessons were balanced in terms of teacher-directed instruction and scaffolds to support student learning, it was much more difficult to select appropriate information texts than narrative texts. In almost all cases, the information texts needed to be sequenced and organized for instruction in substantially nonlinear ways. In other words, whereas teachers could read the narrative texts through from beginning to end following the lesson plans, teachers did not read the information texts this way. Rather, they had to skip from one part of the text to another depending on the information being sought, and frequently not all of the text was read to students because not all of it was particularly relevant.

Receptive comprehension outcomes. We used the Gates Listening Comprehension subtest as the receptive measure of listening comprehension. The difference between intervention and comparison conditions was not significant and the effect size ($d = 0.08$) was close to zero, indicating no trend one way or the other. The Gates was the most distal outcome measure we administered, and two factors may have

influenced the nonsignificant findings. First, the stimulus and response format used in test administration was not a common feature of the intervention. On the Gates, students were read an item and selected one of three responses that represented the best answer. Although this strict multiple-choice format was not used as a regular part of the interactions between teachers and students in the intervention, part of the training teachers received was to provide selection-type responses for students as a way to make student response formats easier and to increase engagement of all students. However, anecdotal information about how this was actually carried out during intervention lessons suggests that teachers used rather simple yes/no formats for these questions (e.g., “Did the main character want the reward?”) rather than more sophisticated approaches involving two or more actual selections (e.g., “Did the main character want the reward *for herself or to help her friend?*”).

The Gates also included stimulus test content that students may not have been exposed to frequently during the intervention. On the Gates, students listened to the examiner describe an action such as a boy walking a dog along the river. Students saw three pictures and chose the one that best described the examiner’s statement. In other words, the stimulus was brief and represented a discrete action more than a story. In the intervention, teachers worked with students on comprehending specific sentences and isolated phrases. However, the bulk of intervention activities addressed larger chunks of information contained in stories and information texts.

Vocabulary outcomes. We assessed vocabulary outcomes overall and in relation to three word types that differed in terms of exposure during instruction. Overall, the intervention had a large impact on vocabulary outcomes ($d = 1.02$), translating to a 35 percentile difference between intervention and comparison students. Effect sizes for the four risk groups ranged from 0.56 (both language and literacy risk) to 1.33 (low risk). Outcomes based on word exposure during instruction fell out in the predicted manner. Thus, we confirmed that a relatively low intensity intervention could have a strong impact ($d = +1.05$) on vocabulary outcomes when teachers taught words explicitly (Category 1 words) in the context of typical read aloud lessons. On words taught explicitly or exposed in both intervention and comparison groups, the intervention had a moderate impact on word learning outcomes. A reasonable conclusion is that the greater degree of explicit instruction in word meanings used in the intervention condition was responsible for the moderate impact ($d = 0.65$) on Category 2 words. On words identified as important but not targeted explicitly for direct teaching, the intervention had a nonsignificant implementation and the effect size was small ($d = 0.17$).

This study confirms that explicit vocabulary instruction has a benefit on outcomes and that this instruction can occur in the context of integrated read aloud routines conducted in whole-classroom settings. The positive outcome on vocabulary confirms previous research on read aloud impacts (Swanson et al., 2011). Of course, we were hopeful that the intervention would have a positive impact on Category 3 words, which were words presented to children in both conditions and which teachers determined whether to teach directly.

We wanted teachers in the intervention condition to explicitly teach many of these Category 3 words to students even though the intervention procedures did not require them to do so. Either this did not happen or teachers did teach these words to students but did not use instructional procedures that were strong enough to make a difference. Anecdotal information based on classroom observations would suggest

that teachers did embellish read aloud lessons in appropriate ways—improvising based on student responses. However, teachers may not have improvised in a way that resulted in robust vocabulary instruction with Category 3 words.

Effective vocabulary instruction is difficult to do (Biemiller, 2001, 2004), especially when teachers rely on spontaneous teaching approaches to achieve deep understanding of word meanings. What frequently occurs, and what we observed anecdotally, is that teachers regularly paused during text reading, acknowledged a potentially difficult word, gave a quick definition of the word, and continued to read. This process may help with text comprehension, but for students who do not know the meaning of the word, this approach may be insufficient for building meaning beyond a strictly associative level of knowledge (Coyne, McCoach, et al., 2007; Coyne et al., 2004; Stahl & Nagy, 2006). More intentional instruction may be necessary for building deeper meaning, and for most teachers this may take deliberate planning before the lesson. More extensive training on teacher-developed vocabulary instruction integrated within a comprehensive read aloud professional development approach might be needed for more robust vocabulary instruction in the context of words not specifically identified for explicit teaching.

Implications

We offer several implications based on the findings and previous research. At the broadest level, we believe this study supports the value of conceptualizing read aloud practices for specific academic purposes. Currently, substantial use of read aloud practices occurs in the early elementary grades (and in preschool), but there is a lack of structure in how read aloud practices are implemented. Prior to the time students are independently reading texts that are complex in terms of language and structure, we suggest that read aloud practices can have a positive impact on higher-order skills including comprehension and vocabulary (Dickinson, McCabe, & Anastasopoulos, 2003; Leung, 2008; Smolkin & Donovan, 2001). Further, we suggest that Perfetti's framework for comprehension development, organized around the creation of vivid mental representations of texts, is well suited to read aloud environments. That is, teachers have abundant opportunities to help students deepen their standards for text coherence through read aloud instruction, and this focus in preschool and the early grades could be beneficial for students as they transition to reading complex texts independently.

In terms of structure, we think there is value in the idea of organizing read aloud lessons to include before-, during-, and after-reading components. However, we clearly acknowledge that in a 30-minute lesson, it is a significant challenge to fit these parts in consistently. In the current study, we found that components of the after-reading part of the lesson were sometimes missed in order to complete the lesson in 30 minutes. The primary problem with this is that the after-reading portion of the lesson may provide the best opportunity students have to process and practice essential content. The solution is either to increase the amount of time teachers have for read aloud lessons or spend less time on the before- and during-reading segments. There are challenges and drawbacks with either solution, but the before, during, and after organization is worth further research.

In terms of delivering read aloud lessons, we suggest that the study supports explicit instruction generally, but the level of prescription we provided has important

caveats. We suggest that the seven principles of comprehension instruction the National Reading Panel (2000) indicated were supported scientifically and offer a useful way of organizing read aloud instruction. We believe we were able to integrate six of these principles on a regular basis in the intervention. Basically, these principles address (a) how teachers instruct students in using comprehension strategies, (b) how teachers and students can work together to construct meaning from texts, and (c) how students can practice using comprehension strategies to develop strong representations of texts.

We are not convinced that the level of prescription we used in the intervention condition is highly feasible or perhaps necessary for achieving significant outcomes. Book availability can change over time, and the lack of flexibility in terms of book selection will be problematic for many teachers. Because the prescriptive nature of the lessons was so closely tied to the specific text being used, providing more flexibility in terms of book selection means, by definition, that lessons for books not in the read aloud set will have to be developed in some other manner. Perhaps one solution is a combination of preselected books with detailed lesson plans and books teachers select and design lessons for following structured lesson development guidelines. We suggest there are several important aspects of the prescriptive nature of the lessons that should be preserved.

First, we suggest that the systematic integration of narrative and information texts is a useful approach. This integration increases the amount of expository text used in the early grades, where expository text exposure has been neglected (Duke, 2000; Jeong, Gaffney, & Choi, 2010); where it has been used, expository text has been shown to have some benefits for learners (Duke & Kays, 1998). Second, we think the majority of teachers would provide better read aloud experiences for their students if they used developed read aloud lessons or developed high-quality read aloud lessons on their own prior to instruction. Thinking aloud to demonstrate comprehension monitoring or teaching vocabulary explicitly for deep understanding is very difficult to do effectively unless teachers have reviewed or planned this instruction before the lesson. Although many aspects of read aloud lessons cannot be planned before the lesson, including the myriad interactions that occur dynamically between teachers and students, there are important instructional goals that can be met when teachers prepare lessons beforehand.

Limitations

The most important limitation in interpreting the results is the relatively small sample size. Although several hundred students participated, the unit of intervention and analysis was the classroom ($n = 12$). The small sample size made it difficult to detect small effects, which may have been relevant with the expository retell measure and the Category 3 vocabulary words.

Another limitation is that beyond the measure of fidelity, we were not able to collect observation data that could have provided deeper insights into the nature of the read aloud practices used in comparison classrooms. Although we know comparison teachers were not using intervention materials, and we know that comparison teachers were engaging in read aloud practices for the same amount of time that intervention teachers were, we also suspect that many comparison teachers were using instructional approaches and techniques that were highly engaging and effec-

tive with their students. Without more detailed observations, however, we are not able to describe these practices with high specificity. Future research should address the active ingredients of read aloud instructional practices and the associations between specific practices and student outcomes.

Note

1. The WWC uses six levels to rate intervention impact. The two positive ratings are Positive Effects (strong) and Potentially Positive Effects (not as strong but no overriding contradictory evidence). On the negative end are Negative Effects (strong) and Potentially Negative Effects. In the middle are Mixed Effects (evidence of inconsistent effects) and No Discernible Effects (no affirmative evidence of effects).

References

- Archer, A. L., & Hughes, C. A. (2011). Exploring the foundations of explicit instruction. In A. L. Archer & C. A. Hughes (Eds.), *Explicit instruction: Effective and efficient teaching* (pp. 1–22). New York: Guilford.
- Baker, S. K., Gersten, R. M., & Scanlon, D. (2002). Procedural facilitators and cognitive strategies: Tools for unraveling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities Research and Practice*, 17, 65–77.
- Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *Reading Teacher*, 55, 10–20.
- Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, 107, 251–271.
- Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *American Educator*, 25, 24–28, 47.
- Biemiller, A. (2004). Teaching vocabulary in the primary grades: Vocabulary instruction needed. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 28–40). New York: Guilford.
- 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.
- Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, 3, 331–361. doi:10.1207/s1532799xssr0304_2
- Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2007). *Effective teaching strategies that accommodate diverse learners* (3rd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction with embedded instruction and incidental exposure. *Learning Disabilities Quarterly*, 30, 70–88.
- Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, 12, 145–162.
- Dickinson, D. K., McCabe, A., & Anastopoulos, L. (2003). A framework for examining book reading in early childhood classrooms. In S. A. Stahl, A. Van Kleeck, & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 91–108). Mahwah, NJ: Erlbaum.
- Dougherty Stahl, K. A. (2009). Synthesized comprehension instruction in primary classrooms: A story of successes and challenges. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 25, 334–355.
- Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, 35, 202–224. doi:10.1598/rrq.35.2.1
- Duke, N. K., & Kays, J. (1998). "Can I say 'Once upon a time'?: Kindergarten children developing knowledge of information book language. *Early Childhood Research Quarterly*, 13, 295–318.

- Duke, N. K., & Martin, N. M. (2008). Comprehension instruction in action: The elementary classroom. In C. C. Block, L. M. Morrow, & S. R. Parris (Eds.), *Comprehension instruction: Research-based best practices* (pp. 241–257). New York: Guilford.
- Eller, R. G., Pappas, C. C., & Brown, E. (1988). The lexical development of kindergarten learning from written context. *Journal of Reading Behavior*, *20*(1), 5–23.
- Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington.
- Fien, H., Baker, S. K., Smolkowski, K., Smith, J. M., Kame'enui, E. J., & Thomas Beck, C. (2008). Using nonsense word fluency to predict reading proficiency in K–2 for English learners and native English speakers. *School Psychology Review*, *37*, 391–408.
- Fien, H., Park, Y., Baker, S. K., Smith, J. L. M., Stoolmiller, M., & Kame'enui, E. J. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review*, *39*, 631–653.
- Fisher, D., Flood, J., Lapp, D., & Frey, N. (2004). Interactive read-alouds: Is there a common set of implementation practices? *Reading Teacher*, *58*, 8–17.
- Gersten, R. M. (1998). Recent advances in instructional research for students with learning disabilities: An overview. *Learning Disabilities Research and Practice*, *13*, 162–170.
- Geva, E., & Olson, D. (1983). Children's story-retelling. *First Language*, *4*(11), 85–109. doi:10.1177/014272378300401102
- Good, R., Baker, S. K., & Peyton, J. (2009). Making sense of nonsense word fluency: Determining adequate progress in early first grade reading. *Reading & Writing Quarterly*, *25*, 33–56.
- Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., . . . Flindt, N. (2004). *Technical adequacy and decision making utility of DIBELS*. (Technical Report No. 7). Eugene: University of Oregon.
- Good, R. H., Simmons, D. C., Kame'enui, E. J., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene: University of Oregon.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education (RASE)*, *7*(1), 6–10.
- Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York: Guilford.
- Hedges, L. V. (2007). Effect sizes in cluster randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341–370.
- Institute of Education Sciences. (2007). What Works Clearinghouse. Accessed September 10, 2007, from <http://ies.ed.gov/ncee/wwc/>
- Jeong, J., Gaffney, J. S., & Choi, J.-O. (2010). Availability and use of informational texts in second-, third-, and fourth-grade classrooms. *Research in the Teaching of English*, *44*(4), 435–456.
- Kame'enui, E. J., & Simmons, D. (1990). *Designing instructional strategies: The prevention of academic learning problems*. Columbus, OH: Merrill.
- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, *25*, 215–227.
- Karweit, N. L., & Wasik, B. A. (1996). The effects of story reading programs on literacy and language development of disadvantaged preschoolers. *Journal of Education for Students Placed at Risk*, *1*, 319–348.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In C. Hulme & M. J. Snowling (Eds.), *The science of reading: A handbook* (pp. 209–226). Malden, MA: Blackwell.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*, 271–299.
- 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. doi:10.1080/02702710801964090
- MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2000). *Gates-MacGinitie reading tests*. Itasca, IL: Riverside.

- McKeown, M. G., & Beck, I. L. (2003). Taking advantage of read-alouds to help children make sense of decontextualized language. In A. Van Kleeck, S. Stahl, & E. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 159–176). Mahwah, NJ: Erlbaum.
- Miller, J. F., & Chapman, R. S. (1993). *SALT: Systematic Analysis of Language Transcripts*. Madison, WI: Language Analysis Laboratory.
- 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. doi:10.3102/0034654309332561
- Morrow, L. M. (1985). Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity. *Elementary School Journal*, *85*, 647–661.
- Morrow, L. M., & Brittain, R. (2003). The nature of storybook reading in the elementary school: Current practices. In A. Van Kleeck, S. Stahl, & E. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 140–158). Mahwah, NJ: Erlbaum.
- Moss, B. (1997). A qualitative assessment of first graders' retelling of expository text. *Reading Research and Instruction*, *37*, 1–13.
- National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00–4769). Washington, DC: National Institute of Child Health and Human Development. Retrieved from <http://www.nichd.nih.gov/publications/nrp/smallbook.htm>
- National Reading Panel. (2001). *Put reading first: The research building blocks for teaching children to read. Kindergarten through grade 3*. Washington, DC: National Institute for Literacy.
- Newcomer, P. L., & Hammill, D. D. (1997). *Test of language development: Primary* (3rd ed.). Austin, TX: Pro-Ed.
- Paris, A., & Paris, S. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*, 36–76. doi:10.1598/RRQ.38.1.3
- 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*, 23–33. doi:10.1037/0022-0663.94.1.23
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Haggort (Eds.), *The neurocognition of language* (pp. 167–208): Oxford: Oxford University Press.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford: Blackwell.
- Phillips, L., Norris, S., Mason, J. M., & Kerr, B. (1990). *Effect of early literacy intervention on kindergarten achievement* (Technical Report No. 520). Champaign: Center for the Study of Reading, University of Illinois at Urbana-Champaign.
- Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, *34*(5), 25–31. doi:10.3102/0013189X034005025
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 573–603). Hillsdale, NJ: Erlbaum.
- Santoro, L. E., Baker, S. K., Chard, D. J., & Howard, L. (2007). The comprehension conversation: Using purposeful discussion during read-alouds to promote student comprehension and vocabulary. In B. Taylor & J. E. Ysseldyke (Eds.), *Effective instruction for struggling readers, K–6* (pp. 206–263). New York: Teachers College Press.
- Santoro, L. E., Chard, D. J., Howard, L., & Baker, S. K. (2008). Making the very most of classroom read alouds to promote comprehension and vocabulary. *Reading Teacher*, *61*, 396–408.
- Shadish, W. R. (2002). Revisiting field experimentation: Field notes for the future. *Psychological Methods*, *7*, 3–18. doi:10.1037/1082-989X.7.1.3
- Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., . . . Kaufman, N. (2007). Attributes of effective and efficient kindergarten reading intervention: An

- examination of instructional time and design of instructional specificity. *Journal of Learning Disabilities*, **40**, 331–347.
- Smolkin, L. B., & Donovan, C. A. (2001). The contexts of comprehension: The information book read aloud, comprehension acquisition, and comprehension instruction in a first-grade classroom. *Elementary School Journal*, **102**, 97–122.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Snow, C. E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Spybrook, J. (2008). Power, sample size, and design. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 273–312). Charlotte, NC: Information Age.
- Stahl, S. A., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, **16**(1), 32–71.
- Strong, C. J. (1998). *The Strong Narrative Assessment Procedure (SNAP)*. Eau Claire, WI: Thinking Publications.
- Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., . . . Tackett, K. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities*, **44**, 258–275. doi:10.1177/0022219410378444
- Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, **68**, 277–321.
- U.S. Department of Education. (2003). *Random assignment in program evaluation and intervention research: Questions and answers*. Retrieved from <http://www.ed.gov/rschstat/eval/resources/randomqa.pdf>
- Van Kleeck, A., Stahl, S. A., & Bauer, E. B. (2003). *On reading books to children: Parents and teachers*. Mahwah, NJ: Erlbaum.
- Varelas, M., & Pappas, C. C. (2006). Intertextuality in read-alouds of integrated science-literacy units in urban primary classrooms: Opportunities for the development of thought and language. *Cognition and Instruction*, **24**, 211–259.
- Vaughn, S., Gersten, R. M., & Chard, D. J. (2000). A search for the underlying message in the intervention research on learning disabilities. *Exceptional Children*, **67**, 99–114.
- Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, **93**, 243–250.
- 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 homes for children from low-income families. *Developmental Psychology*, **30**, 679–689.
- Wilkinson, L. C., & Silliman, E. R. (2000). Classroom language and literacy learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 337–360). Mahwah, NJ: Erlbaum.
- Zevenbergen, A. A., Whitehurst, G. J., & Zevenbergen, J. A. (2003). Effects of a shared-reading intervention on the inclusion of evaluative devices in narratives of children from low-income families. *Journal of Applied Developmental Psychology*, **24**, 1–15. doi:10.1016/s0193-3973(03)00021-2