LSHSS

Research Article

Classroom-Based Narrative and Vocabulary Instruction: Results of an Early-Stage, Nonrandomized Comparison Study

Sandra Laing Gillam,[a] Abbie Olszewski,[b] Jamison Fargo,[a] and Ronald B. Gillam[a]

Purpose: This nonrandomized feasibility study was designed to provide a preliminary assessment of the impact of a narrative and vocabulary instruction program provided by a speech-language pathologist (SLP) in a regular classroom setting.
Method: Forty-three children attending 2 first-grade classrooms participated in the study. Children in each classroom were divided into high- and low-risk subgroups on the basis of their performance on a narrative test. Narrative and vocabulary instruction was provided by an SLP in 1 classroom for three 30-min periods per week for 6 weeks.
Results: The children in the experimental classroom made clinically significant improvements on narrative and vocabulary measures; children in the comparison classroom did not. Within the experimental classroom, children in the high-risk subgroup demonstrated greater gains in narration and fewer gains in vocabulary than children in the low-risk subgroup. There were no subgroup differences in the comparison classroom.
Conclusion: These preliminary results provide early evidence of the feasibility of implementing a narrative instruction program in a classroom setting. Children at a high risk for language difficulties appeared to profit more from the narrative instruction than from the embedded vocabulary instruction. More extensive research on this instructional program is warranted.

Key Words: school-based services, intervention, at risk

One approach to improving language skills in elementary-aged students is to have speech-language pathologists (SLPs) implement interventions in regular classroom settings (Justice, 2006; Nippold, 2011). Unfortunately, school-based SLPs who want to implement evidence-based practices have very little evidence to draw on, because in only a handful of studies have researchers assessed the effects of SLP interventions in regular classrooms (see Cirrin et al., 2010). In the few intervention studies that have been conducted in the classroom, researchers have focused on vocabulary and phonological awareness intervention (Hadley, Simmerman, Long, & Luna, 2000; Throneburg, Calvert, Sturm, Paramboukas, & Paul, 2000; Wilcox, Kouri, & Caswell, 1991). These researchers examined the gains made by children with identified speech and language problems as well as children who were developing typically or were at risk for academic failure.[1] However, we could find no studies in which SLPs provided classroom instruction for discourse-level targets such as narration.

The unique knowledge that SLPs possess about the underlying oral language comprehension and production skills children need to become proficient in oral language may be used to supplement curricular instruction provided by regular classroom teachers in support of the Common Core State Standards (CCSS; National Governors Association and Council of Chief State School Officers, 2011). The CCSS contain goals that target aspects of narrative comprehension, narrative production, and vocabulary knowledge (Squires, Gillam, & Reutzel, 2013). The purpose of this study was to evaluate the impact of a narrative intervention program with embedded vocabulary instruction on the performance of all the children (low risk and high risk) in a regular classroom setting. An SLP was the primary provider of the

[a]Utah State University, Logan
[b]University of Nevada School of Medicine, Las Vegas
Correspondence to Sandra Laing Gillam: sandi.gillam@usu.edu

Editor: Marilyn Nippold
Associate Editor: LaVae Hoffman
Received January 25, 2013
Revision received June 18, 2013
Accepted March 4, 2014
DOI: 10.1044/2014_LSHSS-13-0008

[1]Because children who are at risk for language impairment are also at risk for academic failure, we will use the general term *at risk* to refer to both groups.

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

instruction, but the classroom teacher remained in the classroom during all sessions; facilitated participation by assisting children during the lessons; consulted with the SLP on methods, procedures, and lesson plans; and was asked to follow the teaching procedures throughout the day.

Teaching Narrative Structure

Explicit instruction on simple story elements (character, setting, initiating event, attempts, complications, consequences) and complex episodic structures provides children with a framework upon which to learn and practice oral language skills (Hayward, Gillam, & Lien, 2007; Lynch et al., 2008). The predictability afforded by story structure and the causal connections that exist among events, actions, motivations, and consequences contained in narratives should provide children with focused experiences related to listening to and using specific cognitive and linguistic forms that support the language of the curriculum. Instruction within the narrative context should make the language-learning task less demanding, more meaningful, and more authentic (Tomasello, 2003).

Teaching story elements such as setting, initiating event, internal response, attempt, and consequence has been shown to improve oral language production and reading comprehension (Davies, Shanks, & Davies, 2004; Reutzel, 1985). For example, Hayward and Schnieder (2000) examined the effects of story grammar instruction on the oral narration skills of 13 school-age children (ages 4–6) with language disorders. Children received instruction in small group settings (two to three children per group) for long (12 intervention sessions) or short (eight intervention sessions) durations. Children participated in activities that required them to identify story grammar elements, determine when elements were missing in stories, and organize stories into chronological order. Color-coded cue cards were used that contained line drawings with the words (a) *when,* (b) *what + what doing,* (c) *what happens 1-2-3,* and (d) *who + feelings + action* written on them. More than half of the children who participated in the instruction demonstrated large improvements in the structural complexity and content of their oral narratives.

In a study of older children, Nathanson, Crank, Saywitz, and Ruegg (2007) asked whether teaching story structure would retrospectively improve recall of a history lesson that had been taught 2 weeks earlier. Two weeks after participating in a history lesson, 39 children with learning disabilities in first through fifth grades were randomly assigned to a treatment group that received instruction in story structure (narrative elaboration treatment; NET) or to a control group. NET involved teaching students to associate cue cards with aspects of character and setting details as well as character actions, behaviors, and feelings. Students in NET used the cue cards to organize retellings of video vignettes and received feedback on the accuracy of their recounts. Children in the control condition participated in similar activities but did not learn to use the cue cards. Students who participated in NET remembered significantly more information from the history lesson than children in the control group. The results of this study suggest that teaching story structure using visual graphic organizers may be useful for improving the content and quality of information children with learning disabilities report, even after a significant time delay between exposure to information and oral reporting of the information.

More recently, Westerveld and Gillon (2008) used visual graphic organizers to teach story structure to 10 children between the ages of 7;11 and 9;2 (years;months) with language and reading disabilities. They were interested in determining whether story structure knowledge would be associated with gains in general oral language skills, narrative comprehension, and narrative production. Instruction focused on setting, character, problem, goal/plan, attempts, resolution, and conclusions. After being taught story maps, children were asked to identify and discuss story elements and to retell stories. Results indicated that the children's ability to answer questions about story elements improved, but their ability to include more story elements in their oral narratives did not.

The children's failure to include more story elements may have been related to how the intervention program was implemented. Most of the children's instructional time was spent answering questions and identifying story elements rather than incorporating and using all of the elements together to create and tell stories. Children may need more frequent and intensive practice in constructing and generating oral narratives that contain all of the story elements if they are to use them consistently in their spontaneously produced oral narratives.

A similar approach was taken by Swanson, Fey, Mills, and Hood (2005), who provided children with multiple opportunities to practice using story elements they had been taught. Ten 7–9-year-old children with specific language impairment (SLI) received an intervention in which story elements (setting, character, problem, resolution, complication, and ending) were taught in the context of authentic stories. Stories were read to children in their entirety and section by section. In this way, children were exposed to all of the story elements and asked to focus on each one individually. Eighty percent of children demonstrated clinically significant improvements in their oral narratives after 50-min individual sessions provided three times weekly over the course of 6 weeks.

Narrative Contexts for Teaching Vocabulary

Instruction in story comprehension with an explicit focus on improving vocabulary has been associated with improvements in reading comprehension over time (Clarke, Snowling, Truelove, & Hulme, 2010; Hulme & Snowling, 2011; McKeown & Beck, 2006; Snowling & Hulme, 2012). For example, Biemiller and Boote (2006) conducted two experiments in which they investigated approaches to teaching vocabulary to elementary-school-age children attending kindergarten (n = 43) and first (n = 37) and second (n = 32) grades. In Study 1, researchers assessed vocabulary outcomes after children listened to stories multiple times with or

without explanations of word meanings. The meanings of four to six words were explained each day. Children in second grade made gains in vocabulary knowledge after hearing stories twice, whereas children in kindergarten and first grade demonstrated substantially more benefit only after hearing stories four times. Vocabulary gains increased by 10% when word explanations were provided during repeated readings, as compared to conditions in which children engaged in repeated reading without word explanations. Multiple exposures to stories accompanied by word explanations resulted in increased vocabulary knowledge for children in kindergarten and first and second grades.

In Study 2, the same children listened to repeated stories and received embedded vocabulary instruction. As in Study 1, teachers read stories to children and asked them to answer comprehension questions afterward. This time, one or two word meanings that were critical to understanding the content of the story were taught before the story was read. After the selected vocabulary words had been taught, teachers held review sessions and invited children to provide definitions for the new words they learned. Findings revealed that children demonstrated gains in vocabulary knowledge at posttest that were 41% above pretest, with first graders demonstrating the greatest benefit. These results suggest that children learned new words that were embedded in narrative contexts when they received instruction before the story was read, when stories containing the words were repeated often, and when children were given opportunities to discuss the words and their meanings.

Results of the previous studies suggest that narrative instruction should include explicit teaching of story elements that are represented by graphic organizers or visual cues together with demonstrations of the relationships between complete stories and their component parts. Narrative-based vocabulary instruction appears to be most successful when vocabulary targets are embedded in stories that are read repeatedly, when word meanings are explained explicitly within the narrative context, and when children use the target vocabulary during story retelling and discussions about the text. We developed a narrative intervention program that incorporated these critical components of instruction and designed a preliminary study to assess its effects.

According to Fey and Finestack (2009), feasibility studies assessing the outcomes of an intervention should be conducted to determine whether larger-scale and more costly randomized clinical trials are warranted. We conducted a feasibility study in which we compared the narrative and vocabulary outcomes of children who attended a first-grade class that received narrative instruction from an SLP to the outcomes of children who attended another first-grade class that followed the standard, district-approved curriculum (with no specific instruction on narration). We assessed the effects of the narrative instruction for children who were at risk (the high-risk subgroup) and those who were not at risk (the low-risk subgroup) in each classroom. Children were classified as high risk when they demonstrated standard scores on a measure of narrative proficiency of 90 (25th percentile) or lower. The research questions were as follows:

1. Do high-risk and low-risk students who received whole-class narrative instruction from an SLP perform differently on narrative measures than high-risk and low-risk students who did not receive this instruction?
2. Do high-risk and low-risk students who received whole-class narrative instruction from an SLP perform differently on vocabulary measures than high-risk and low-risk students who did not receive this instruction?

Method

Participants

The participants were 43 students in two classrooms in a Title 1 elementary school located in northern Utah. There were 11 boys and 10 girls (21 total children) in the experimental classroom and nine boys and 13 girls (22 total children) in the comparison classroom. Each classroom had seven bilingual children (three boys, four girls). Three children in the comparison classroom moved away before we completed pretesting, leaving 19 total children who participated in pre- and posttesting. Children in the experimental classroom ranged in age from 6;6 to 7;4 (M_{age} = 7;0). Children in the comparison classroom ranged in age from 6;7 to 7;4 (M_{age} = 7;0). The majority of the students (86%) were from underrepresented ethnic groups, and 75% qualified for free or reduced-cost lunch. Two children in each classroom had Individual Education Plans (IEPs) and were receiving speech therapy for delayed articulation skills.

The Test of Narrative Language (TNL; Gillam & Pearson, 2004), a standardized measure of narrative comprehension and production for children ages 5–12, was administered to all of the children in both classrooms. The TNL yields narrative comprehension (NC) and oral narration (ON) scaled scores and a total narrative language index score (TNLAI). As a group, children in the experimental classroom obtained a mean TNLAI standard score of 87.00 (SD = 18.26), whereas the children in the comparison classroom obtained a mean standard score of 86.42 (SD = 15.34). The TNLAI scores for the experimental and comparison classrooms were highly similar at pretest (p = .915, Cohen's d = 0.03), with 97.7% overlap in the distributions.

Consistent with other studies of academic risk in children (e.g., Frisk et al., 2009; Vaughn, Linan-Thompson, & Hickman, 2003), we set the risk cut-point at a standard score of 90 (equivalent to the 25th percentile) on the TNLAI. The 25th percentile corresponds to a Z value of –0.68, which is commonly used as the cut-point for standardized language screening (Law, Boyle, Harris, Harkness, & Nye, 2000). In the experimental classroom, there were 10 children in the low-risk subgroup (TNL performance at or above the 25th percentile) and 11 children in the high-risk subgroup (TNL performance below the 25th percentile). In the comparison classroom, there were seven children in the low-risk subgroup and 12 children in the high-risk subgroup. The means and standard deviations for the pretest TNL scores for the high- and low-risk subgroups in the experimental and comparison classrooms are shown in Table 1.

Table 1. Narrative and vocabulary scores at pretest for high- and low-risk children in the experimental and comparison classrooms.

Measure	Experimental *M* (*SD*)	Comparison *M* (*SD*)	*p*
Narrative			
TNL			
Total	87.00 (18.26)	86.42 (15.34)	.915
High risk	72.82 (9.96)	77.75 (10.27)	
Low risk	101.50 (13.58)	101.29 (10.23)	
MISL			
Total	5.60 (3.33)	6.74 (3.36)	.294
High risk	4.36 (3.08)	5.67 (2.64)	
Low risk	7.11 (3.06)	8.57 (3.87)	
Vocabulary			
Total	24.55 (7.72)	24.32 (8.86)	.930
High risk	21.60 (8.18)	19.92 (5.95)	
Low risk	27.50 (6.28)	31.86 (8.09)	

Note. Vocabulary raw scores were obtained from a criterion-referenced vocabulary probe and rubric. None of the comparisons yielded statistically significant differences. TNL = Test of Narrative Language; MISL = Monitoring Indicators of Scholarly Language.

Assessment Procedure

Children were assessed before and after a classroom narrative intervention was delivered to the experimental classroom. Pretesting began in January and posttesting was conducted in March. Research assistants who were blind to the purpose of the investigation and to group assignment administered, transcribed, and scored all of the outcome measures. All assessments were digitally recorded and transcribed according to Systematic Analysis of Language Transcription (SALT) conventions (Miller & Chapman, 2004) using Sony digital voice editors. The TNL and outcome measures (spontaneous narratives and vocabulary responses) were transcribed verbatim, with the inclusion of both child and examiner utterances when applicable. Research assistants segmented the transcripts into communication units (C-units; Loban, 1976) that consisted of an independent main clause and any phrases or clause(s) subordinate to it. Each transcript was checked by a second research assistant for word spelling, mazing, morpheme segmentation, and utterance segmentation. All transcription disagreements were resolved as the two transcribers listened to the digital recording together for a third time. As a check on the accuracy of the original transcription and coding process, 30% of the transcripts were retranscribed. Percentage of agreement between primary and secondary transcribers was 98% for C-unit segmentation and 95% for the identification of mazes. The interrater reliability for the outcome measures is reported in the following sections.

Outcome Measures

Narrative probe and rubric. Children were asked to tell a story after looking at a single picture that did not contain an obvious initiating event. Three single-scene choices were available at pre- and posttesting and were presented in a randomized, counterbalanced fashion. In one scene, children were holding hands in a circle; in another, children were running around in the rain; and, in the last scene, a monkey was riding on the back of a dog. The spontaneous narratives children produced in response to single-scene prompts were scored using a progress monitoring tool, the Monitoring Indicators of Scholarly Language (MISL; Gillam & Gillam, 2013). The MISL was designed to measure the complexity of macrostructure elements (character, setting, initiating event, internal response, plan, attempt, and consequence) and microstructure elements (coordinated and subordinated conjunctions, adverbs, metacognitive verbs, and elaborated noun phrases) in self-generated stories. The MISL yielded two subscale scores, Macrostructure (range = 0–21) and Microstructure (range = 0–15), and one total score that was a combination of the subscale scores (range = 0–36).

Initiating events, actions, and consequences were linked together in the scoring system because they must be interrelated to form a coherent episode. For example, for initiating events, a score of 0 was awarded when there was nothing stated in the story to motivate a character into action. A score of 2 was awarded if the initiating event clearly motivated a character's actions. For example, if the child said, "The spaceship landed. And the children ran over to see it," the initiating event would be coded as 2 because the stated action of running to see the spaceship was tied directly to the spaceship having landed. Compare this to the utterances "The spaceship landed. And the children ate their lunch." In this example, there was an event (spaceship landing) that may have led to an action. However, there is no clear relationship between the proposed initiating event (spaceship landing) and the stated action (ate lunch). In this example, the statement "The spaceship landed" would receive a score of 1. Macrostructure definitions, examples, and scoring criteria are shown in Table 2.

The Microstructure subscale of the MISL consisted of five literate language features including coordinating conjunctions and subordinating conjunctions, adverbs, metacognitive verbs, and elaborated noun phrases. Definitions for the literate language structures, examples, and scoring criteria are shown in Table 3.

Two research assistants trained to score the MISL independently rated de-identified transcripts of the single-scene narratives that were elicited from the children. The raters scored 10 stories and met to check their scoring agreement. Any scoring discrepancies that remained were resolved by the first author. Then, the research assistants coded 10 more stories and met again to check agreement. This procedure of coding 10 stories, meeting to resolve discrepancies, and oversight by the first author was incorporated to control for drift in scoring and was conducted until all of the stories had been rated. Interrater reliability values, calculated after every 10 stories, were scored independently and were combined to form interrater reliability scores for each macrostructure and microstructure element. Interrater reliability was 100% for character and coordinating conjunctions; 96% for setting, initiating event, action, and subordinating conjunctions; 98% for internal response and plan; 92% for consequence and metacognitive verbs; and 90% for elaborated

Table 2. Macrostructure subscale: Definitions, examples, and scoring criteria.

Story element	Definition	Example(s)	0	1	2	3
Character	Agent who performs actions	*The **girl** ran.* ***Sally** ate.*	None, ambiguous pronoun	1, nonspecific pronoun	1, formal name	>1, formal name
Setting	Time, place	*They in Central **park**.*	None	Nonspecific	1 specific	>1 specific
Initiating event (IE)	Event motivates characters into action	*The spaceship landed.*	None	Event stated, does not motivate action	1 event stated, motivates action	>1 event stated, motivates action
Internal response	Feelings characters have about the IE	*The dog out. The girls were **afraid** it would bite them.*	None	Feeling words, not related to IE	1 feeling word, related to IE	>1 feeling word, related to IE
Plan	Thoughts of characters related to the IE with intent to act	*They **decided** to try to hide the treasure.*	None	Use of cognitive state verb, not tied to IE	Use of cognitive state verb, clearly tied to IE	>1 use of cognitive state verb, clearly tied to IE
Attempt	Action taken by characters in response to the IE, including complicating actions	*The children **climbed** into the tree to escape.* <u>Complication:</u> *Sara realized **she had forgotten her purse.***	None	Use of action verb(s) not tied to IE	Use of action verb tied to IE	The inclusion of a complicating action
Consequence	End result of characters actions in relation to the IE	*John **made it to school**.*	None	Outcome of action linked to another action, no IE	1 outcome of action related to IE	2 or more outcomes, related to IE

Table 3. Microstructure subscale: Literate language structures, definitions, examples, and scoring criteria.

Literate language structure	Definition	Example(s)	0	1	2	3
Coordinating conjunctions	Coordinates nouns, verbs, or clauses	*FANBOYS (for, and, nor, but, or, yet, so)*	None	1 exemplar	2 different exemplars	>2 different exemplars
Subordinating conjunctions	Joins a subordinate clause to a main clause	***Before** you go, get your coat.* *That is the park **where** she loved to play.*	None	1 exemplar	2 different exemplars	>2 different exemplars
Mental/linguistic verbs (metacognitive verbs)	Verbs that refer to passive states of cognition, thinking, knowing, or perception, about the world	*She **knows** what she needs to do.* *He **decided** to try to swim across the river.*	None	1 exemplar	2 different exemplars	>2 different exemplars
Adverbs	Adjectives used to modify the verb	*She walked **slowly**.*	None	1 exemplar	2 different exemplars	>2 different exemplars
Elaborated noun phrases	Modifiers + noun, includes articles, possessives, determiners, *wh*-words, adjectives	***A** dog ate my homework.* ***The big, black** dog ate my homework.*	None	1 exemplar	2 different exemplars	>2 different exemplars

noun phrases. Interrater reliability for total scores was 97% for Macrostructure, 93% for Microstructure, and 95% for the total MISL score.

Means and standard deviations for pretest scores on the MISL are shown in Table 1. Independent *t* tests revealed no significant differences between the experimental and comparison groups on their MISL scores ($p = .294$).

Vocabulary probe and rubric. A criterion-referenced vocabulary probe and rubric was used to measure students' understanding of words related to story grammar, literacy knowledge, feelings, verbs, and adjectives (Supplemental Appendix A). Children were asked to provide a definition for the vocabulary word with no prompting or contextual cues. The examiner said, "Tell me what _____ means." If children did not respond or said, "I don't know," the prompt was given again saying, "The word is _____. Can you tell me what it means?" Consistent with previous studies investigating approaches to improving vocabulary in young children, each response was awarded a score depending on its accuracy (Zipoli, Coyne, & McCoach, 2011). A score of 0 was awarded when a child produced a definition that was wrong, replied with "I don't know," or did not respond. A score of 1 was awarded when a child's response described the word using a function or attribute but was not a synonym or complete definition. A score of 1 was also given when children provided a sentence using the word in a way that clearly demonstrated appropriate conceptual knowledge related to its meaning, for example, "The illustrator was drawing." A score of 2 was given for responses that contained accurate function or attribute information in an utterance that resembled a formal definition, for example, "The illustrator is the person who drew the drawings."

Means and standard deviations for vocabulary scores at pretest are shown in Table 1. Independent *t* tests revealed no significant differences between children in the experimental and comparison classrooms for their knowledge and/or ability to define the vocabulary words that were to be taught in the narrative program at pretest ($p = .930$).

The two research assistants who originally scored the vocabulary responses achieved point-by-point interrater reliability of 95% on scoring. A third research assistant independently scored 20% of the protocols and achieved 97% interrater reliability.

General Intervention Procedure

The narrative language instruction was provided to all children in the experimental classroom for a total of 30 min, three times per week, over the course of 6 weeks. While the SLP was providing instruction in the experimental classroom, an undergraduate student in speech-language pathology assisted the teacher in the comparison classroom. This procedure was implemented to maintain the same adult-to-student ratio in the comparison and experimental classrooms during reading and listening instruction time (see Hadley et al., 2000). The undergraduate student did not provide additional instruction in narration but was asked to assist students in sounding out words or in summarizing a story, for example.

Classroom Curricula

The teachers in the experimental and comparison classrooms implemented their classroom instruction in a similar fashion using the same daily schedule, shown in Table 4. The narrative intervention was provided in the experimental classroom from 10:25 a.m. to 10:55 a.m., whereas the comparison class received their typical instruction in reading and listening comprehension at that time. Otherwise, instruction in both classrooms remained the same throughout the day. On the days that the SLP did not provide instruction in the experimental classroom, the teacher conducted the same reading and listening comprehension activities that were used in the comparison classroom.

The reading and listening comprehension instruction included activities designed to facilitate comprehension before, during, and after reading. Lessons used to facilitate comprehension before reading included (a) activation of background knowledge prior to oral assisted reading, (b) asking students to predict what stories would be about from their titles, and (c) instruction in the use of graphic organizers to include critical aspects of stories such as beginning, middle, and end. Teachers in both classrooms used a variety of procedures to facilitate comprehension during reading including (a) modeling, (b) guided and independent practice, and (c) think-alouds to explain how to use visualization and story maps to recall information that was read. The two teachers also engaged students in character analysis discussions and encouraged them to reread to monitor their comprehension. After reading and/or discussing stories, students were engaged in activities that required them to recall key ideas contained in the text, including writing, retelling, summarizing, dramatizing, and answering questions.

Vocabulary instruction also occurred during reading and listening comprehension time in both classrooms. This involved discussions about three to four Tier 2 words each day. *Tier 2 words* may be described as those words that have "general utility" but are not specific to a certain topic or subject (Beck & McKeown, 2007, p. 195). For example, the teacher may explain the word *search* as a synonym for the

Table 4. Daily classroom schedule.

Time	Subject
8:10–8:50	Opening/fluency
8:50–9:50	Power hour (guided reading and individual group work)
9:50–10:05	Morning recess
10:05–10:25	Writing
10:25–10:55	Comprehension (reading and listening)
10:58–11:38	Lunch and recess
11:40–12:20	Math
12:20–1:00	Computer/PE/library/music
1:00–1:20	Math (continued)
1:20–1:50	Word work
1:50–2:25	Units/cleanup
2:30	Dismissal

Note. PE = physical education.

concept *to look for*. The teachers introduced vocabulary using student-friendly definitions and examples to promote learning of the new words.

Narrative Instruction Program

The basic teaching tasks and contexts used in the narrative intervention program included story modeling, story retelling, story generation and comprehension instruction. The narrative program consisted of three phases. In Phase I, Teaching Story Grammar Elements, students heard and told stories that contained simple episodes (e.g., initiating event, attempt, consequence). In Phase II, Elaboration: Making Stories Sparkle, students participated in lessons designed to encourage the use of more complex narratives by including complicating actions into their stories. They also practiced incorporating dialogue, coordinated and subordinated conjunctions, adverbs, adjectives, and metacognitive verbs in their narratives. In Phase III, Independent Storytelling, children took part in multiple lessons to foster their ability to create and tell complex and elaborated stories on their own. Each phase ended with a literature unit prepared to allow the children practice incorporating the skills they had learned over the course of successive sessions.

The narrative approach incorporated aspects of embedded instruction to teach vocabulary (Coyne, McCoach, & Kapp, 2007). During story modeling the SLP who conducted the intervention sessions told stories using wordless picture books. Six wordless books created for this project were presented over a period of 6 weeks. The books in Phase I contained eight to 10 pictures illustrating stories about going camping, losing an umbrella, or looking for friends. The books used in Phase II were 15 pictures in length and contained more complicating actions. These stories addressed topics such as lost pets, hungry mice, and problems getting a Christmas tree home from the woods. No wordless books were used in Phase III.

The content in the wordless books was used to teach and highlight story grammar elements (SGEs; e.g., character, setting, initiating event, internal response, plan, attempt, consequence), literate language structures such as causal connections (because, since, so that, in order to), and the target vocabulary. Some of the vocabulary was specific to the SGEs (e.g., *character, setting*) and some was specific to the stories children were being exposed to (e.g., *alley cat, distracted*).

Children in the experimental class were encouraged to retell stories told by the SLP as well as those created and told by other children. Games, symbolic icons, and graphic organizers were used to maintain group engagement and to support memory during the course of natural narrative discourse. The organization of the classroom was such that children were seated at desks arranged in groups of four to six. When retelling or telling stories, these natural groups were used and the children in each group took turns telling their stories to the entire class. Children in other groupings were engaged in marking bingo cards containing story elements being used by children telling stories. The SLP and classroom teacher walked around the room during this time to facilitate the engagement of the other students.

As the instruction progressed, children were exposed to more elaborate and complex stories and were expected to become more independent and proficient in their storytelling. In Phases II and III, children worked in small groups to create stories that corresponded to single-scene pictures or to verbal prompts. For example, early in instruction, children were asked to develop a story after seeing a picture of a boat going over a waterfall. Later, as children became more proficient, they were presented with pictures that did not contain obvious initiating events, such as a picture of a smiling child sitting in the snow.

The children were asked to draw their stories using a process similar to pictography, as proposed by Ukrainetz (1998). In our intervention, children drafted their stories onto storyboards that contained story grammar icons. Storyboards were graphic organizers that included all of the SGEs in sequential order in separate squares marked with visual icons to cue students as to what should be represented therein (see Supplemental Appendix B). Students' drafts were represented in the form of stick-figure drawings that allowed for quick representation of concepts and ideas. Students used their self-made graphics to support oral storytelling. At the end of each phase of instruction, children were encouraged to tell stories without the use of their storyboards, although they were cued using icons when SGEs were omitted.

Story comprehension was targeted during instruction in drafting and storytelling, and relied on the use of the SGE icons and graphic organizers. Lessons were designed so that children practiced answering questions related to each of the SGEs after listening to or telling stories. When children responded incorrectly or did not respond to questions, they were given visual support through story pictures, icons, and graphic organizers.

The classroom teacher was told that she was free to use the terminology and procedures we were implementing in the narrative instruction throughout the day, but she was not given specific instructions about how to do so. She was asked not to share information and terminology related to the narrative program with the teacher of the comparison classroom. The classroom teacher reported that she encouraged students to take their storyboards home to write stories with parents and caregivers and to bring them back to share in the general education classroom during reading and listening activities. However, we did not monitor the classroom outside the time we were engaged in our instruction and cannot report on her use of information from the program in her daily activities.

Embedded Vocabulary Instruction

Target vocabulary was introduced and defined during discussions of the wordless storybooks. Approximately eight vocabulary words were introduced each week. The target vocabulary related to book concepts (e.g., *author, illustrator, dialogue*), story elements (e.g., *setting, initiating event*), internal responses (e.g., *frustrated, bored*), adverbs (e.g., *quickly, frantically*), verbs (e.g., *discover, stroll*), adjectives (e.g., *sneaky,*

beautiful), and words that were specific to understanding the content of the wordless books (e.g., *alley cat*).

Children were encouraged to use the target vocabulary in discussions about the stories. The SLP facilitated the use of causal connections (e.g., *because, so*) through contingent facilitation devices such as modeling, imitation, and growth-relevant recasts (see Supplemental Appendix C). In many cases, the SLP highlighted symbolic icons while explaining the meaning of words. For example, to teach the word *icon,* the SLP told children a story about two dogs that went camping and were chased by a bear. During the storytelling, the SLP showed children various pictures representing the story elements (see Supplemental Appendix D for examples of icons used) and told a portion of the story. For example, "There were two dogs named Bob and Max [holding up the character icon]. Bob and Max were walking in a forest [holding up the setting icon] during the day [again, holding up the setting icon]." After reading the story in its entirety and including all of the day's icons and vocabulary, the SLP conducted a comprehension review. If children experienced difficulty with a concept, the SLP conducted a mini-lesson using examples of iconic symbols that children might encounter such as a stop sign, a women's restroom sign, and so forth, and revisited the icons representing the story elements (e.g., character, setting, feelings) to ensure students understood the relationships.

Children were encouraged to repeat the target words when they were presented in hopes of reinforcing their phonological representations as in Brackenbury and Pye (2005) and Zipoli et al. (2011). Students were asked to repeat or paraphrase definitions provided by teachers, or to create their own version of definitions. To encourage accuracy, when they asked students to define words, teachers said, "Tell me what the word means but don't say the word." This was done to discourage students from defining words by repeating the word as they were wont to do; for example, in response to the teacher asking "What is character?" the student answers, "Character."

Treatment Fidelity

An intervention observation checklist (IOC) was designed to monitor treatment fidelity for aspects of teaching narrative structure, the use of language facilitation devices, and explicit teaching of vocabulary words. The IOC contained specific headings related to each lesson plan. As an example, see the following excerpts from the IOC related to teaching children to use elaboration in their stories:

- Review bingo cards with children.
- Identify all story elements present in story.
- Facilitate a discussion with the children surrounding the story elements.
- If children answer any of the story comprehension questions inaccurately, review the story elements.
- Review critical differences between the two hamster stories.
- Alter the nature of your discussion to needs of the children (when children exhibit difficulty, rephrase, restate, or provide more examples to illustrate the concept).
- Show icons during story comparisons to talk about how stories differ with respect to the SGEs.

Each of the bulleted points on the IOC was associated with instructions that were included in the curriculum manual for the program. For example, in order to check off the box associated with reviewing the bingo cards with children, the SLP must have (a) mentioned all of the story elements by name, (b) pointed to them on a bingo card shown to the class, and (c) ensured that children had placed a chip on the appropriate square. Each session was observed by one member of the research team who completed an IOC that was specific to the lesson being taught. If fidelity fell below 85% for any lesson, the SLP and the research staff met immediately after the lesson to discuss areas of concern. There was only one session during which the SLP failed to meet the minimum fidelity requirement. The information in that lesson was reviewed in the following session. A separate member of the research team observed 20% of the instructional lessons along with another research team member who was present in the classroom. Interrater reliability for implementation of lessons was calculated point by point and was 90% or greater for all of the lessons that were judged.

Data Analysis

The purpose of the study was to assess the impact of a narrative language program with embedded vocabulary instruction delivered to children in a regular education classroom. The narrative dependent variable was the MISL score obtained from the story-generation task using a single-scene prompt. The vocabulary dependent variable was the score obtained on the researcher-designed vocabulary probe. Because this study was the first to explore these procedures in a classroom setting, we used a more limited sampling design, with only two classrooms representing intervention and comparison groups. Such a design likely violated the statistical assumption of independence of observations because students were nested within classrooms. The sample of one experimental classroom and one control classroom precluded the use of statistical methods for clustered data (e.g., mixed-effects models or generalized estimating equations) due to the inclusion of only two clusters ($N = 2$). Thus, the primary approach to the analysis of our outcomes was an examination of the descriptive information derived from the outcomes measured before (pretest) and after (posttest) intervention. Basic descriptive statistics, including the 95% confidence interval (CI) around the mean for each outcome, were computed for the classroom data as a whole as well as for the low- and high-risk subgroups in each classroom, and then evaluated for clinically meaningful differences (performance outside the upper bound of the 95% CI).

We also assessed outcomes by examining effect sizes. Using the formula suggested by Cumming (2012) for

within-subjects variables, we used the average standard deviation of the pre- and posttest measures as the standardizer for the Cohen's *d* effect sizes. According to standard interpretative guidelines, a Cohen's *d* value of 0.8 is considered to be a large effect, 0.5 is considered to be a medium-sized effect, and 0.2 is considered to be a small effect (Cohen, 1988). However, in a systematic review of the effect sizes reported across 176 studies of whole-class instructional activities, the mean effect size was 0.18, with a standard deviation of 0.41 (Lipsey et al., 2012). Small and large effects reflect ±1 *SD* from the medium effect. Thus, for whole-class instruction, a medium-sized effect would be 0.18 and a large effect would be 0.59.

Results

Narrative Outcomes

Table 5 presents the means and standard deviations for the total MISL scores at pre- and posttest, the pretest 95% confidence intervals, and Cohen's *d* values for children in the experimental and comparison classrooms. Recall that we used three single-scene pictures as the story prompts. To test the possibility that some single scenes might be associated with higher scores on the MISL than others, we compared the total MISL scores at pretest using a repeated-measures analysis of variance (ANOVA), with scene (1, 2, 3) as the within-subjects factor and classroom (experimental, comparison) as the between-subjects factor. Results indicated that the scenes yielded similar total scores, $F(1, 2) = 1.12$, $p = .339$, partial eta-squared = .07, regardless of group $F(1, 2) = 0.58$, $p = .454$, partial eta-squared = .02, and there was no significant interaction between group and scene number, $F(1, 2) = 1.17$, p = .324, partial eta-squared = .07. Therefore, MISL scores from the three scenes were collapsed in all subsequent analyses.

Experimental classroom. As can be seen in Table 4, the posttest total MISL score for the children in the experimental classroom ($M = 8.79$) fell outside of the upper limit of their pretest 95% CI (7.40) and was associated with a large effect size ($d = 0.82$). The finding of posttest performance outside the 95% CI combined with a large effect size suggests that the children in the experimental classroom made clinically significant improvements in oral narration after receiving narrative instruction.

We wondered whether the high-risk and low-risk subgroups within the experimental classroom benefitted similarly from the instruction. The pretest MISL score of the low-risk group (7.11) was outside the upper limit of the high-risk group's pretest 95% CI, with a moderately large Cohen's *d* of 0.78, suggesting that that the two groups had clinically different narrative skills before intervention. After the narrative instruction, both subgroups in the experimental classroom improved, but only the performance of the high-risk group reached clinical significance. The mean posttest MISL score of the children in the high-risk group ($M = 8.80$) exceeded the upper limit of their pretest 95% CI (6.71) and was associated with a large, within-group effect size ($d = 1.0$). However, this was not true for the children in the low-risk group. Their posttest MISL mean of 8.78 did not exceed the upper boundary of the pretest 95% CI (9.13). However, there was some degree of improvement from pre- to posttest as demonstrated by a within-subject effect size of 0.53, which was high in relation to the effect sizes from studies of whole-class instructional programs (Lipsey et al., 2012). The children in the high-risk group started out with much lower

Table 5. Pretest and posttest means and standard deviations, 95% confidence intervals (CIs) at pretest, and Cohen's *d* values for the MISL scores from the story-generation task and the scores on the vocabulary rubric for children in the experimental and comparison classrooms (total and high- and low-risk groups).

Measure	Pretest *M* (*SD*)	Pretest 95% CI	Posttest *M* (*SD*)	Cohen's *d*
MISL scores				
Experimental				
Total	5.89 (3.10)	[4.39, 7.40]	8.79 (4.00)*	0.82
High-risk	4.80 (2.86)	[2.89, 6.71]	8.80 (5.11)*	1.0
Low-risk	7.11 (3.06)	[5.09, 9.13]	8.78 (2.59)	0.59
Comparison				
Total	7.31 (3.36)	[5.67, 8.95]	7.94 (2.77)	0.21
High-risk	6.00 (2.79)	[4.09, 7.91]	6.80 (1.93)	0.34
Low-risk	9.50 (3.27)	[7.03, 11.97]	9.83 (3.76)	0.09
Vocabulary scores				
Experimental				
Total	24.55 (7.72)	[20.73, 28.37]	36.10 (14.85)*	1.02
High-risk	21.60 (8.18)	[16.93, 26.27]	28.80 (13.60)*	0.66
Low-risk	27.50 (6.28)	[22.84, 32.17]	45.40 (9.40)*	2.28
Comparison				
Total	24.94 (9.18)	[20.80, 29.08]	22.88 (9.31)	–0.22
High-risk	20.10 (6.54)	[15.44, 24.77]	18.60 (6.07)	–0.24
Low-risk	31.86 (8.09)	[26.28, 37.43]	29.00 (10.07)	–0.31

*Posttest mean is outside the upper bound of the 95% CI of the pretest mean.

MISL scores than the children in the low-risk group. But, they experienced a clinically significant benefit from the narrative instructional program, which resulted in very similar MISL posttest scores for the high-risk (M = 8.80) and low-risk (M = 8.78) subgroups. The standardized mean difference (Cohen's d) scores for the high-risk and low-risk groups dropped to –0.005 at posttest, indicating that the two subgroups performed at essentially the same level after instruction.

Comparison classroom. Looking at the comparison class as a whole, the mean posttest MISL score (M = 7.94) did not exceed the upper limit of the pretest 95% CI (8.95) and was associated with a small effect size (d = 0.21). It does not appear that the standard curriculum resulted in clinically significant changes in children's storytelling skills.

The high- and low-risk subgroups in the comparison classroom had clinically different narrative production skills before intervention. As was the case for the experimental classroom, the pretest mean MISL score of the low-risk group in the comparison classroom (9.50) was outside the upper limit of the pretest 95% CI of the high-risk group (7.91). There was a large subgroup difference in narrative skills as indicated by a Cohen's d value of 1.15. However, it did not appear that the standard curriculum had a differential effect on the narrative production skills of the high-risk and low-risk students. The posttest performance of the children in the high-risk group (M = 6.80) did not exceed the upper limit of the pretest 95% CI (7.91) and was associated with a small effect size (d = 0.34). Similarly, the posttest performance of the children in the low-risk group (M = 9.83) did not exceed the upper limit of the pretest 95% CI and was associated with a negligible effect size (d = 0.09). Furthermore, the differences between the high-risk and low-risk children's MISL scores at posttest were very similar to their differences at pretest, with a posttest standardized difference (Cohen's d) value of 1.01. Unlike the experimental classroom, where the high-risk children caught up to the low-risk children after intervention, the low-risk children in the comparison classroom retained their narrative advantage over the high-risk children across the duration of the study.

Classroom comparisons. The narrative skills of the children in the experimental classroom improved much more than the narrative skills of the children in the comparison classroom. The extent of the intervention effect can be examined by calculating the difference between the posttest scores for the children in the experimental and comparison classrooms in relation to their pretest scores. In this manner, it is possible to proportion the intervention effect against the comparison classroom pre–post change (see Lipsey et al., 2012, for an explanation of this descriptive approach). The posttest difference between the experimental and comparison classrooms for the composite MISL score was .85 (8.79 – 7.94). The pretest and posttest difference for the composite MISL score for the comparison classroom was .63 (7.94 – 7.31). Therefore, the instruction resulted in pre–post improvement in the MISL score that was 1.35 times (.85 / .63) greater than the improvement for the comparison classroom. The effect sizes tell a similar story. Looking at the effect sizes for the two classrooms as a whole, the pre- to posttest effect size for the MISL scores obtained by children in the experimental classroom (d = 0.82) was more than 3 times larger than the effect size for the children in the comparison classroom (d = 0.21).

Comparing the subgroups in each of the classrooms, the children in the high-risk subgroup in the experimental classroom had more than twice the gain in narrative production (d = 1.0) than the high-risk subgroup in the comparison classroom (d = 0.34). The findings were even more compelling for children in the low-risk subgroup in the experimental classroom, whose effect size was six times greater (d = 0.59) than the effect size observed for the children in the low-risk subgroup in the comparison classroom (d = 0.09). These analyses suggest that the narrative instruction program provided to an entire class of first graders by an SLP is very promising for improving the complexity of children's stories.

Vocabulary Outcomes

Experimental classroom. Recall that a criterion-referenced vocabulary probe was used to measure students' understanding of words related to story grammar, literacy knowledge, feelings, verbs, and adjectives. The vocabulary skills of the children in the experimental classroom improved after intervention. The mean posttest performance for the vocabulary score for all the children in the experimental classroom (M = 36.10) fell outside of the upper limit of the pretest 95% CI (28.37) and was associated with large effect size (d = 1.02). Taken together, the posttest performance outside the 95% CI and the large effect size suggest clinically significant improvement in the ability to define words related to narration after receiving the intervention.

It appears that the embedded vocabulary instruction differentially affected the children in the high-risk and low-risk subgroups within the classroom. The posttest performance of the children in the high-risk group (M = 28.80) exceeded the upper limit of the pretest 95% CI (26.27) and was associated with a moderate effect size (d = 0.66). The vocabulary gains were much greater for children in the low-risk group, whose posttest performance (M = 45.40) well exceeded the upper limit of the pretest CI (32.17) and was associated with a very large effect size (d = 2.28). There were relatively large differences in the vocabulary abilities of the high-risk and the low-risk subgroups before receiving the instruction (d = 0.81), and the subgroup differences increased after instruction (d = 1.43) due to the disproportional benefits of instruction for the low-risk subgroup.

Comparison classroom. The vocabulary scores for the children in the comparison classroom did not improve from pretest to posttest. The mean posttest vocabulary score (M = 22.88) did not exceed the upper limit of the pretest CI (29.08) and was associated with a small and negative effect size (d = –0.22). The standard curriculum did not affect the children's ability to define words related to narration.

Examining the subgroups, the children in the high- and low-risk subgroups had clinically significant differences on their vocabulary skills before and after intervention (d = 1.59

pretest and 1.26 posttest). The posttest performance of the children in the high-risk group (M = 18.60) did not exceed the upper limit of the pretest 95% CI (24.77) and was associated with a small, negative effect size (d = –0.24). Similarly, the posttest performance of the children in the low-risk group (M = 29.00) did not exceed the upper limit of the pretest CI and was associated with a small, negative effect size (d = –0.31).

Classroom comparisons. The vocabulary skills of the children in the experimental classroom improved much more than the vocabulary skills of the children in the comparison classroom. Once again, we examined the extent of the intervention effect by computing the difference between the posttest scores for the children in the experimental and comparison classrooms in relation to the pretest to posttest change for the comparison classroom. The difference between the pretest and posttest vocabulary scores of the children in the comparison classroom was –2.06 (22.88 – 24.94) whereas the posttest difference between the experimental and comparison classrooms was 3.22 (26.10 – 22.88). The pre- to posttest effect sizes for the vocabulary scores obtained by children in the experimental classroom were more than 10 times larger than the effect sizes for the children in the comparison group.

Inspection of the effect sizes for the subgroups revealed that children in the high-risk subgroup in the experimental classroom had three times more gain (d = 0.66) than children in the high-risk subgroup in the comparison classroom (d = –0.24). The findings were most compelling for children in the low-risk subgroup in the experimental classroom, whose effect size was more than 10 times greater (d = 2.28) than the effect size observed for the children in the low-risk subgroup in the comparison classroom (d = –0.31).

MISL comparisons. The MISL measured macrostructure and microstructure elements of the children's stories. We wondered whether some aspects of the children's stories were more likely to change after instruction than others. We noticed that the posttest scores for most of the macrostructure elements measured on the MISL increased from pretest to posttest, with the exception of internal response. A frequency count of the character scores from the MISL revealed that most children in both classrooms referred to characters using determiners and pronouns such as *that girl, this boy,* or *the dog,* earning a MISL score of 1. However, eight children in the experimental classroom included two or more proper names for their characters as compared to only one participant who did so in the comparison classroom. Our findings for the use of setting were similar to those for character. Prior to instruction, the majority of children in both classrooms earned scores of 0 for setting; after instruction, 13 children in the experimental classroom as compared to eight in the comparison classroom included specific information about where or when the story took place.

There were 16 children in the experimental classroom who included an overt plan (e.g., "he decided to run") that was clearly related to an initiating event (e.g., a bear coming down the path) and a resulting consequence (e.g., "he got away") as compared to seven children who did so in the comparison classroom. More than half of the children in the experimental classroom used words that referred to a character's motivation and implied a causal relationship among initiating events, plans, and actions. Finally, 43% of the children in the experimental classroom, as compared to 20% in the comparison classroom, incorporated a complicating action into their stories.

Discussion

Narration

Following intervention, the narratives produced by children in the experimental classroom were observed to be more complex as measured by the MISL progress-monitoring tool. Children in the comparison classroom, whether they were identified as low or high risk, were not observed to make quantifiable gains in the complexity of their stories. When we compared gains across the two classrooms, we observed that the pre- to posttest effect sizes for the scores on the MISL were three times larger for children in the experimental as compared to the comparison classroom.

Within the experimental classroom, children who were designated as high risk experienced the most benefit from the narrative instruction. When we compared the high-risk groups across the two classrooms, we found the effect sizes for children in the experimental class were more than twice those of children in the comparison classroom (d = 1.0 vs. d = 0.34). Children who were designated as low risk in the experimental classroom evidenced effect sizes that were as much as six times greater than those of children at low risk in the comparison classroom. An important result of the intervention was that the high-risk children in the experimental classroom caught up to the low-risk children, in terms of the complexity of their stories, after intervention.

One aspect of the narrative instruction that may have differed from the traditional classroom was the focus on teaching children to specify causal links among initiating events, internal responses, plans, attempts, and consequences. Inclusion of causal links among story elements was measured using a scoring system that was specifically designed to detect their use. The MISL scoring system was weighted to reflect whether children achieved this standard by anchoring the scores at 2. For example, when a child provided an initiating event that was clearly linked to a plan, attempt, or consequence, the use of the element was awarded a score of 2. Similarly, when internal responses, plans, attempts, or consequences were distinctly related to the stated initiating event, each was awarded a score of 2. It was possible for a child to score 2 each for initiating event, attempt, and consequence and 1 for internal response if the word used to reflect how the character was feeling could not be directly linked to the initiating event. Thus, the scoring system was well suited for capturing the kinds of incremental changes that occurred in children's narratives as a result of the instruction provided.

The following is an example of a story told by one of the participants in the experimental classroom after instruction.

> Once upon a time there was a family that wanted to go to their grandma's house and play with their cousins. When they were driving there, they ran out of gas. Then, we had a mile left on their gas, and their grandma's house was two miles away. So they found a gas station and they filled it with gas. Then, they got to play with their cousins.

Consider the story told by a child in the comparison classroom while looking at the same scene:

> One time those guys were outside playing. They asked their mom. Their mom said yes, so they were playing. When they were playing they saw some chipmunks singing. That's the end.

As is shown in these examples, the story told by the child in the comparison classroom is largely descriptive in nature. However, the story told by the child in the experimental classroom contains notable causal relationships between the ideas (e.g., wanted to go, drove there, almost ran out of gas, got gas) and incorporated clear links between the initiating event (e.g., wanting to play with cousins) and the consequence (e.g., playing with cousins).

Another important goal of the narrative instruction implemented in the experimental classroom was to teach children to include information about how their characters were feeling, what they were planning, and why. This emphasis required children to make inferences about internal states, motivations, feelings, and plans. Research has shown that story knowledge is facilitated when instruction includes a focus on internal states and motivations of characters (Dunning, 1992). Young elementary-school-age children are typically not proficient in generating inferences or recalling information from stories about character motivation (Shannon, Kameenui, & Baumann, 1988). Dunning (1992) showed that instruction targeting children's skill in generating inferences about why characters decided to take particular actions was associated with greater gains in comprehension than instruction that focused on discussion of actions without talking about internal states. It is possible that instruction in the comparison classroom focused more on specific anchor standards for reading comprehension in first grade that place emphasis on recalling key details and explicit, factual information contained in stories than on motivational states of characters.

Finally, the narrative instruction placed substantial importance on story complexity. Toward that end, children were asked to include complicating actions into their stories. At posttesting, more children in the experimental classroom were observed to include complicating actions than were children in the comparison classroom. Consider the following example of an interaction between the SLP and children in the experimental classroom as they developed a story containing a complicating action together:

> **Students:** Mini and Bubba did not have a Christmas tree. So, they decided to go into the woods and find one. They found a beautiful Christmas tree. They took it home and decorated it for Christmas.
> **SLP:** Can you think of something that might get in Mini and Bubba's way when they try to take the tree home? Can you think of a complication?
> **Students:** It's too heavy for them.
> **SLP:** What will they do now?
> **Students:** Go home and get the car.
> **SLP:** OK. Tell the story again with the complication.
> **Students:** Mini and Bubba did not have a Christmas tree. So, they decided to go into the woods and find one. They found a beautiful Christmas tree but it was too heavy. So, they decided to go home and get the car. They came back with the car and put the tree on top. They drove home with the tree and decorated it for Christmas.

It did not appear that microstructure elements were as amenable to instruction as macrostructure elements. The only microstructure element with posttest scores that exceeded the upper limit of the pretest CI for children in the experimental group was elaborated noun phrases. The children in the experimental classroom were observed to provide more detail in their stories by using elaborated noun phrases such as *a big white thing* or *a blue shirt*. It is possible that children did not have the cognitive resources to allocate to both macrostructure and microstructure within the short time span of the instruction (Colozzo, Gillam, Wood, Schnell, & Johnston, 2011). Future studies may extend instruction to determine if microstructure elements improve more significantly after children have had sufficient time to internalize the macrostructure elements they were learning.

Vocabulary

The vocabulary instruction that was embedded within the narrative intervention yielded vocabulary gains that far exceeded those of the children in the comparison classroom. Vocabulary was taught by placing words into contextually relevant sentences and passages, by highlighting the words explicitly, and by reviewing them repeatedly. This approach has been shown to be instructionally time efficient and effective for teaching words to young children (Nelson & Stage, 2007; Zipoli et al., 2011). Although this was beneficial for the experimental classroom as a whole, the children designated as low risk demonstrated greater benefit than the high-risk children. Children who were performing at the lower end of the high-risk group experienced the least growth, suggesting that our approach to teaching vocabulary to the entire classroom was not sufficient to meet their learning needs.

This finding is not altogether surprising given the research that shows that children who are demonstrating difficulty learning words in authentic contexts or embedded instruction may require more intensive, explicit instruction to learn (Beck & McKeown, 2007). This instruction may be provided in a number of contexts including small-group or one-on-one instruction. Growth related to knowledge about words related to story and feelings was notable; however, our instruction had little to no additional impact on children's use of adjectives. Initially, this was a surprising finding,

because we observed that children in the experimental classroom used more elaborated noun phrases than children in the comparison classroom. Our rubric awarded credit for elaboration when children used an article or determiner (*a, the*) preceding a noun, although more points were awarded for phrases such as *the big dog* that included an adjective as well. We noticed that in some cases when children were encouraged to include adjectives, they would leave out important macrostructure elements such as setting or initiating event. Therefore, we were careful to notice when our instruction was overtaxing them in this way. This may in part explain why children in the experimental classroom were not using many adjectives; however, it does not explain why this was the case in the comparison classroom.

Limitations

Given the relatively small scope of this investigation (one experimental classroom and one comparison classroom) it is not possible to determine the extent to which the results were due to the particular intervention that was used or to the curriculum–intervention combination. We have no way of knowing whether the effects would have been different had our study examined possible differences between classrooms using different curricula, for example. Our vocabulary results may be due in part to the fact that children in the comparison classroom were not specifically taught the words that were taught to students in the experimental classroom. We were primarily interested in determining whether the words that were taught could be learned when they were embedded in the stories used to teach story structure. The fact that the children in the comparison group did not demonstrate similar vocabulary gains suggests that the approach to embedded vocabulary instruction employed in the experimental classroom has the potential to result in robust vocabulary learning.

It is possible, although unlikely, that our findings reflected general effects of a better teacher in the experimental classroom than in the comparison classroom. The likelihood of this explanation is reduced somewhat by the fact that the students in both classrooms performed similarly on all of our pretest measures. Because the pretesting was done in January, the students in the study had been in the two classrooms for one full semester before the study started. If the teacher in the experimental classroom was better than the teacher in the comparison classroom, the children in the experimental classroom should have performed better than the children in the comparison classroom on our pretest measures. Nonetheless, additional research in which intervention is provided in many classrooms is needed to rule out potential curriculum–intervention interaction effects and classroom teacher effects.

A final limitation of this study relates to the type of children who were included and how our results relate to services provided to children with diagnosed language impairments. The children in this study attended a Title 1 school in a low-income area. Only two children in each class had been diagnosed with speech-language impairments. In all four cases, the speech therapy the children received primarily focused on articulation skills. It would appear that few children in this study had language impairments. However, in a large-scale epidemiological study conducted across the state of Iowa, Tomblin and colleagues (1997) found that only 29% of the children who were identified as having a specific language impairment by their evidence-based criteria had actually been diagnosed as such by school personnel. It is likely that a number of the children in the high-risk subgroups in our study may have had undiagnosed language impairments. The results of this pilot study suggest that such children would be likely to profit from the type of whole-class intervention that the SLP provided.

Conclusion

The purpose of this feasibility study was to examine the outcomes of a narrative program that included embedded vocabulary instruction to first-graders who attended a regular classroom in a Title 1 school. The students in the experimental and comparison classrooms performed very similarly on our measures of narration and vocabulary before the intervention, and quite differently after intervention. Children in the experimental classroom demonstrated clinically significant improvement on measures of narrative complexity and vocabulary knowledge after receiving the whole-class narrative instruction. The narrative and vocabulary skills of the children in the comparison classroom did not change. Further, there were differential effects of the instruction for high-risk and low-risk children. With respect to narration, the instruction appeared to be more beneficial for children at a high-risk for language impairment than children at a low-risk. However, the vocabulary results suggest just the opposite: low-risk children tended to benefit more than high-risk children.

This study contrasted delivery and nondelivery of a narrative intervention. It did not contrast delivery of the narrative intervention by an SLP with delivery by a non-SLP and, therefore, cannot address the value of delivery by an SLP. However, such an investigation might make an interesting next step in this line of research and could help to define the frequently asserted "value-added" component of SLPs in schools.

The teachers in the experimental and comparison classrooms followed the same daily curriculum and met weekly to plan instruction. The primary difference between the experimental and comparison classrooms, as far as narrative and vocabulary skills were concerned, was the narrative instruction provided in the experimental classroom. The gains made by children in the experimental classroom may have occurred because the program placed greater emphasis upon the creation and generation of stories that contained complete and complex episodes than did the regular classroom curriculum, resulting in more proficiency in oral narrative skills. A larger-scale trial is necessary to confirm this hypothesis. However, for the present, this study provides preliminary evidence that a narrative instruction program implemented in a regular classroom setting yielded

narrative and vocabulary improvements for both high-risk and low-risk children.

Acknowledgments

Work reported in this article was supported in part with funds from the Institute for Educational Sciences awarded to Sandra Laing Gillam and Ronald B. Gillam and to Utah State University, Logan, Utah (R324A100063). The authors would like to thank Jason Anthony, Anthony Wheeler, Natalie Nelson, Julise Nelson, Brittney Lamb, Amanda Frandsen, Beki Ballum, Dave Long, and the students, faculty, and staff at Bridger Elementary and the Logan City School District who made this work possible.

References

Beck, I., & McKeown, M. (2007). Different ways for different goals, but keep your eye on the higher verbal goals. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 182–204). New York, NY: Guilford Press.

Biemiller, A., & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology, 98,* 44–62. doi:10.1037/0022-0663.98.1.44

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments: Issues for clinical assessment. *Language, Speech, and Hearing Services in Schools, 36,* 5–16. doi:10.1044/0161-1461 (2005/002)

Cirrin, F., Schooling, T., Nelson, N., Diehl, S., Flynn, P., Staskowski, M., … Adamczyk, D. F. (2010). Evidence-based systematic review: Effects of different service delivery modes on communication outcomes for elementary school-age children. *Language, Speech, and Hearing Services in Schools, 41,* 233–264. doi:10.1044/0161-1461(2009/08-0128)

Clarke, P., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomized controlled trial. *Psychological Science, 21,* 1106–1116. doi:10.1177/0956797610375449

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 54,* 1609–1627. doi:10.1044/1092-4388(2011/10-0247)

Coyne, M., McCoach, B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30,* 74–88.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Davies, P., Shanks, B., & Davies, K. (2004). Improving narrative skills in young children with delayed language development [References]. *Educational Review, 56,* 271–286. doi:10.1080/00131910420020118 1

Dunning, D. B. (1992). *Instructional questions that clarify story character's feelings and motivations: Their effect on students' narrative comprehension* (Technical Report No. 563). Champaign: University of Illinois Center for the Study of Reading. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/17842/ctrstreadtechrepv01992i00563_opt.pdf?sequence=1

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: Psychology Press.

Frisk, V., Montgomery, L., Boychyn, E., Young, R., vanRyn, E., McLachlan, D., & Neufeld, J. (2009). Why screening Canadian preschoolers for language delays is more difficult than it should be. *Infants and Young Children, 22,* 290–308. doi:10.1097/IYC.0b013e3181bc4db6

Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Gillam, S., & Gillam, R. (2013). *Monitoring Indicators of Scholarly Language (MISL)*. Logan: Utah State University.

Hadley, P., Simmerman, A., Long, M., & Luna, M. (2000). Facilitating language development for inner-city children: Experimental evaluation of a collaborative, classroom-based intervention. *Language, Speech, and Hearing Services in Schools, 31,* 280–295. doi:10.1044/0161-1461.3103.280

Hayward, D. V., Gillam, R. B., & Lien, P. (2007). Retelling a script-based story: Do children with and without language impairments focus on script and story elements? *American Journal of Speech-Language Pathology, 16,* 235–245. doi:10.1044/1058-0360(2007/028)

Hayward, D., & Schneider, P. (2000). Effectiveness of teaching story grammar knowledge to pre-school children with language impairment. An exploratory study. *Child Language Teaching and Therapy, 16,* 255–284. doi:10.1177/026565900001600303

Hulme, C., & Snowling, M. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science, 20,* 139–142. doi:10.1177/0963721411408673

Justice, L. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37,* 284–297. doi:10.1044/0161-1461(2006/033)

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine and Child Neurology, 42,* 190–200.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., … Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncser/

Loban, W. (1976). *Language development: Kindergarten through grade twelve* (Research Report No. 18). Urbana, IL: National Council of Teachers of English.

Lynch, J.S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology, 29,* 327–365. doi:10.1080/02702710802165416

McKeown, M., & Beck, I. (2006). Issues in the advancement of vocabulary instruction: Response to Stahl and Fairbanks meta-analysis. In K. A. Dougherty-Stahl & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 262–271). New York, NY: Guilford Press.

Miller, J., & Chapman, R. (2004). *Systematic analysis of language transcripts* [Computer software]. Madison: University of Wisconsin—Madison, Waisman Center, Language Analysis Laboratory.

Nathanson, R., Crank, J., Saywitz, K., & Ruegg, E. (2007). Enhancing the oral narratives of children with learning disabilities. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 23,* 315–331. doi:10.1080/10573560701501610

National Governors Association and Council of Chief State School Officers. (2011, October 27). *Common Core State Standards Initiative*. Retrieved from www.corestandards.org

Nelson, R., & Stage, S. (2007). Fostering the development of vocabulary knowledge and reading comprehension through contextually based multiple meaning vocabulary instruction. *Education and Treatment of Children, 30,* 1–22. doi:10.1353/etc.2007.0003

Nippold, M. (2011). Language intervention in the classroom: What it looks like. *Language, Speech, and Hearing Services in Schools, 42,* 393–394. doi:10.1044/0161-1461(2011/ed-04)

Reutzel, D. R. (1985). Story maps improve comprehension. *The Reading Teacher, 38,* 400–404.

Shannon, P., Kameenui, E., & Baumann, J. (1988). An investigation of children's ability to comprehend character motives. *American Educational Research Journal, 25,* 441–462. doi:10.3102/00028312025003441

Snowling, M., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language and Communication Disorders, 47,* 27–34. doi:10.1111/j.1460-6984.2011.00081.x

Squires, K., Gillam, S., & Reutzel, R. (2013). Characteristics of children who struggle with reading: Teachers and speech-language pathologists collaborate to support young learners. *The Early Childhood Education Journal, 41,* 401–411.

Swanson, L., Fey, M., Mills, C., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology, 14,* 131–141. doi:10.1044/1058-360(2005/014)

Throneburg, R., Calvert, L., Sturm, J., Paramboukas, A., & Paul, P. (2000). A comparison of service delivery models: Effects on curricular vocabulary skills in the school setting. *American Journal of Speech-Language Pathology, 9,* 10–20. doi:10.1044/1058-0360.0901.10

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten. *Journal of Speech, Language, and Hearing Research, 40,* 1245–1260. doi:10.1044/jslhr.4006.1245

Ukrainetz, T. (1998). Stickwriting stories: A quick and easy narrative representation strategy. *Language, Speech, and Hearing Services in Schools, 2,* 197–206.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to treatment as a means for identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Westerveld, M., & Gillon, G. (2008). Oral narrative intervention for children with mixed reading disability. *Child Language Teaching and Therapy, 24,* 31–54. doi:10.1177/0265659007084567

Wilcox, J., Kouri, T., & Caswell, S. (1991). Early language intervention: A comparison of classroom and individual treatment. *American Journal of Speech-Language Pathology, 1,* 49–62.

Zipoli, R., Coyne, M., & McCoach, B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education, 32,* 131–143. doi:10.1177/0741932510361262