

# Large group narrative intervention in Head Start preschools: Implications for response to intervention

Journal of Early Childhood Research  
2015, Vol. 13(2) 196–217  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1476718X13515419  
ecr.sagepub.com  


**Trina D Spencer**

Northern Arizona University, USA

**Douglas B Petersen**

University of Wyoming, USA

**Timothy A Slocum**

Utah State University, USA

**Melissa M Allen**

University of Wyoming, USA

## Abstract

This study investigated the effect of a large group narrative intervention on diverse preschoolers' narrative language skills with aims to explore questions of treatment efficacy and differential response to intervention. A quasi-experimental, pretest/posttest comparison group research design was employed with 71 preschool children. Classrooms were randomly assigned to treatment and comparison conditions. Intervention consisted of explicit teaching of narrative structure via repeated story retell practice, illustrations and icons, and peer mediation. Children's narrative language and comprehension were assessed at Pretest, Posttest, and 4 weeks after treatment. Statistically significant differences between treatment and comparison groups were found on retell and story comprehension measures. A priori classification criteria resulted in 28 percent of the participants identified as Minimal Responders on the story retell measure and 19 percent as Minimal Responders on the story comprehension measure. Children who were dual-language learners did not have a different pattern of response than monolingual English speakers. Low-intensity narrative intervention delivered to a large group of children was efficacious and can serve as a targeted language intervention for use within preschool classrooms. A culturally and linguistically appropriate, dynamic approach to assessment identified children for whom intensified intervention would be recommended.

---

## Corresponding author:

Trina D Spencer, Institute for Human Development, Northern Arizona University, PO Box 5630, Flagstaff, AZ 86011, USA.

Email: tds95@nau.edu

**Keywords**

dual-language learners, dynamic assessment, narratives, response to intervention

The ability to understand and use complex language contributes appreciably to academic success (Chaney, 1998; Speece et al., 1999; Walker et al., 1994), but children who are from low socioeconomic backgrounds or who are dual-language learners (DLLs) are at particular risk for inadequate language growth (August et al., 2005; Bowey, 1995; Chaney, 1994; Fazio et al., 1996; Hart and Risley, 1995; Pearson, 2002; Snow et al., 1998; Whitehurst, 1997) and overidentification in special education programs (Artiles et al., 2002; Aud et al., 2010; Fletcher and Navarrete, 2003). It is imperative to accurately identify those who are struggling with language acquisition and to provide appropriate, differentiated instruction. The early, accurate identification of children at risk for language-learning difficulty can lead to appropriate, specialized intervention.

**Response to intervention procedures**

Response to intervention (RtI) is a framework for identifying children with emerging difficulties and leads to the subsequent provision of differentiated instruction according to individual children's needs. RtI integrates valid assessment and evidence-based instruction throughout the implementation of the model (National Professional Development Center on Inclusion (NPDCI), 2012). In most RtI models, all children receive instruction at the classroom level (Tier 1) and then through systematic assessment of their response to instruction are given additional, specialized intervention with increasing levels of intensity (Tier 2 and/or Tier 3) according to their needs. The principles of RtI place the focus on children's learning responses by measuring growth from one assessment point to another. Thus, early identification and allocation of educational supports within RtI models do not rely on single-time static assessment, which is particularly biased for children who are culturally and linguistically diverse, or who come from low socioeconomic backgrounds (Demsky et al., 1998; Laing and Kamhi, 2003; McCauley and Swisher, 1984; Rodekohr and Haynes, 2001; Scheffner-Hammer et al., 2002; Spaulding et al., 2006; Valencia and Rankin, 1985; Valencia and Suzuki, 2001). RtI models designed to address the learning needs of all children reflect advancements in early education for a group of children that is becoming more and more diverse. In 2008, 31 percent of the children attending Head Start programs were identified as DLLs (Yandian, 2009). That percentage is likely to increase in the coming years.

RtI requires two main components: evidence-based instruction that is powerful enough to effect changes in key domains over a brief period of time and reliable and valid assessment procedures that are sensitive to those changes. Research on the relationship between RtI and student achievement outcomes with school-age children has yielded positive results. A meta-analysis by Burns et al. (2005) indicated that RtI procedures resulted in large, positive effect sizes for students, and that through the RtI process, less than 2 percent of students were placed into special education. Van Der Heyden et al. (2007) examined RTI procedures implemented across several schools. Results of their study indicated that RTI reduced the number of students evaluated for special education services and greatly reduced overrepresentation of culturally and linguistically diverse children in special education.

Many of the RtI components are found within recommended practices for early childhood education (The Division for Early Childhood of the council for Exceptional Children (DEC), National Association for the Education of Young Children (NAEYC), and National Head Start Association (NHSA) (2012)), although there is no explicit legislation specifying the use of RtI

with this population. RtI models have been proposed within the early childhood system (Buysee and Peisner-Feinberg, 2010; Coleman et al., 2009; Fox et al., 2010; Jackson et al., 2009); however, actual full implementation is at a low level (i.e. 4%) in Head Start programs (Linan et al., 2010). Lack of tiered evidence-based curricula is one specific challenge to implementation of the RtI model within early childhood education (Greenwood et al., 2011).

## **Narrative language intervention**

Narrative language curricula and interventions hold particular promise for application to the RtI process for early childhood education. Narratives are often used to share information and assess learning in the classroom (Petersen, Gillam, and Gillam, 2008; Mehta et al., 2005). Narration, because of its required use of complex language, serves as a bridge between oral and written language (Scott, 1995; Westby, 1985). Early narrative proficiency, specifically knowledge of story structure (or story grammar), is one of the best predictors of decoding, reading comprehension, and written language (Wellman et al., 2011). Furthermore, narration includes a set of discrete, constrained skills that can be taught and assessed in a relatively brief amount of time at an early age. Presumably, explicit teaching of story grammar and other important narrative-related language features, along with subsequent measurement of the response to that teaching, could be a means to identify and help children who are struggling with language acquisition across tiers of RtI.

Interventions designed to improve narrative skills have been investigated with individuals and small groups of children (Petersen, 2011; Petersen et al., 2010; Petersen et al., 2014). Yet for preschool children, research is limited. Tyler and Sandoval (1994) implemented a narrative intervention with preschool children with language impairments. During individual sessions, participants practiced retelling stories, while a clinician used expansion and recasting techniques to support retellings. Hayward and Schneider (2000) delivered narrative intervention in small groups of two or three children with language delays. Their intervention consisted of a variety of components such as narrative practice, visual supports, systematically introducing targeted features, receptive identification of missing components, and prompting from a clinician. These investigations provide evidence that narrative intervention improved preschoolers' narrative skills, although they were primarily conducted with young children with language impairment at intensity levels aligned with Tier 3 of an RtI framework. In contrast, Spencer and Slocum (2010) examined the effects of a narrative intervention on retelling and personal experience generation skills of preschoolers with limited language skills attending Head Start. Intervention was delivered in a small group format and included individual and group narrative practice, pictures and icons as prompts, verbal prompting from the instructor, and receptive identification games for active responding as intervention components. Children made substantial gains in their retell and personal narrative productions indicating that a variety of children, not just children with language impairment, benefit from a Tier 2 narrative intervention when provided in a small group context.

There are no investigations of whole group narrative intervention to improve preschoolers' oral narration. However, one study featured a general preschool program designed to target book reading and abstract language with children from low socioeconomic status (SES) homes (Feagans and Farran, 1994). The preschool instruction provided some advantages in narrative understanding and retelling at the beginning of kindergarten, although these advantages diminished over time. Another program, story telling and retelling (STaR), has been evaluated as part of Success for All for preschool children (Karweit, 1989). The purpose of STaR is to enhance children's print awareness and oral language skills and to teach listening and reading comprehension strategies. Although STaR was not evaluated separately, children receiving Success for All have made substantial reading gains compared to a control group in several research studies (Success For All Foundation, 2007).

There are several potential benefits of embedding story-telling experiences into preschool classrooms. First, all children, not just those who have significant language impairments, can benefit from narrative intervention. Kindergarten readiness assessments (e.g. HELP for Preschoolers; Vort Corporation, 2004) and core curricula (Petersen, 2011) commonly include oral language, story structure, and listening comprehension objectives. Second, teaching in a large group is more efficient and cost-effective than teaching one child or a few children at a time. Finally, children who may require more intense language intervention can be identified by assessing their response to narrative intervention provided to the whole class, consistent with RtI principles. To our knowledge, no research to date has investigated the feasibility of analyzing response to narrative intervention in a large group to identify children in need of more intensive services.

Early childhood educators are in great need of a narrative language intervention that can be implemented in a whole-class arrangement. Given their interest in tiered language intervention, Spencer and Slocum (2010) asked teachers about the feasibility of using small group narrative intervention procedures for a large group in a classroom. All teachers agreed that the procedures were (a) appropriate for young children, (b) enjoyable for the children, and (c) could be adapted for use in a classroom. Given the heavy emphasis on evidence-based practice in education, early childhood educators need to know whether a less expensive, less intensive, whole-class intervention can improve preschoolers' narrative language production and comprehension.

## Current study

This study represents a systematic extension of individual (e.g. Petersen et al., 2010; Petersen et al., 2014) and small group (e.g. Spencer and Slocum, 2010) narrative intervention to be delivered to a whole class of Head Start preschoolers. Materials and teaching procedures were adapted to maintain high rates of active responding while explicitly targeting story grammar elements within a narrative retell genre. There are three distinct purposes of this study. First, we conducted an efficacy study to investigate the effect of narrative intervention, delivered in a large group arrangement (i.e. whole class), on diverse preschoolers' narrative language skills. Second, we conducted responsiveness analyses to identify children for whom the low-intensity intervention was not sufficient. Third, we examined the extent to which DLLs benefited from the instruction compared to monolingual English speakers. The specific research questions were as follows:

1. Do children who receive a large group, narrative intervention differ significantly from children who do not receive narrative intervention on measures of story retell, story comprehension, and story generation?
2. Do children who receive a large group, narrative intervention respond differently to intervention?
3. Do English-speaking children and DLLs who receive a large group, narrative intervention respond differently to intervention?

## Method

### *Participants and setting*

Seventy-one Head Start preschoolers participated in this study. Participants were drawn from four preschool classrooms located in a single Head Start building. The program was located in a Western US city with a population of approximately 50,000. Hispanic (10.2%) and Asian (3.7%) groups constituted the largest minority groups in this predominately White (82.6%) community. Two of

**Table 1.** Participant characteristics (N = 71).

	Treatment group (n = 36)	Comparison group (n = 35)
Age in months, <i>M</i> ( <i>SD</i> )	57.53 (3.23)	57.71 (4.12)
Gender, <i>n</i> (%)		
Male	19 (53)	17 (49)
Female	17 (47)	18 (51)
Language, <i>n</i> (%)		
English only	13 (36)	13 (37)
Mostly English	2 (6)	4 (11)
Half English/half other	9 (25)	9 (26)
Mostly other	11 (31)	5 (14)
Other only	1 (3)	4 (11)

SD: standard deviation.

the four teachers had worked at Head Start for 6 years, whereas the other two were first-year teachers. All four teachers used the Creative Curriculum in their classrooms and had similar daily schedules that included large group activities, rotating centers (e.g. writing and literacy, blocks, dramatic play), and self-selected play. Teachers attended weekly professional development and planning sessions to ensure proper implementation of the curriculum. Each classroom consisted of 17–20 children, 1 lead teacher, and 1 teaching assistant. Intervention sessions took place in the classroom where large group instruction typically took place (e.g. on a rug). Assessment sessions occurred in small observation rooms between classrooms.

During the consent process, parents were asked about the amount of English and other languages their children use. Parents were asked to select one of five language options: (a) English only, (b) Mostly English, (c) Half English and half other, (d) Mostly other, and (e) Other only. Because approximately half of the students' parents spoke Spanish, a Spanish version of the parent permission form was developed and a Spanish-speaking researcher (D.B.P.) spoke to these families directly about the amount of English and Spanish the children spoke. Although many of the children came from English- and Spanish-speaking homes, two families spoke Karrin, one family spoke Napali, and one family spoke Mandarin Chinese. These parents completed an English permission form and were asked about their child's language use in English.

The four classrooms were divided into two groups by matching them, as close as possible, on the number of students with limited English language skills and teachers' experience. Table 1 displays participants' gender, age, and English use for treatment and comparison groups.

## Research design

A quasi-experimental, pretest/posttest comparison group research design was employed for this study. After matching the groups, they were randomly assigned to treatment and comparison conditions. In both groups, participants' narrative language and comprehension were assessed prior to implementing narrative instruction with the treatment group (Pretest), immediately following the completion of the 3-week (12 sessions) treatment phase (Posttest), and then again following a 4-week period of no treatment (Follow-up). During the treatment phase, T.D.S. delivered the standardized narrative intervention to both treatment classes in place of the regular large group activity. The typical large group activity that was temporarily replaced by the narrative intervention involved the teacher reading storybooks aloud and asking the children factual and inferential questions

about the story. Teachers of the comparison classrooms continued to read to their classes during this time in the schedule and were unaware of the narrative intervention procedures being taught to the treatment group. Storybook reading for the comparison group and the Story Champs instruction for the treatment group lasted between 15 and 20 minutes.

### *Instrumentation*

The Narrative Language Measures: Preschool (NLM:P; Petersen and Spencer, 2012) was used to assess narrative production and comprehension. The NLM:P includes three subtests with standardized administration procedures. The Test of Narrative Retell (TNR) involves an examiner reading a short story to a child and asking the child to retell the story. In the Test of Story Comprehension (TSC), the examiner reads a short story to a child and then asks five factual questions and one inferential question about the story (e.g. “Who was this story about?” and “How do you think he/she felt at the end of the story?”). The Test of Personal Generation (TPG) is administered while playing and conversing with a child in the most natural manner possible. The examiner tells a brief story about him or herself and asks the child whether something like that has ever happened to him or her and provides an opportunity for the child to tell his or her own personal story.

Forty parallel stories are used in the administration of the NLM:P so that the measures can be repeated. For example, NLM stories have a consistent structure (e.g. five main story grammar elements), length (67–70 words), and language complexity (e.g. subordination, causal and temporal ties, dialogue, adverbs, and adjectives), which is based on the narrative development of 5-year-old children (Hughes et al., 1997). In addition, stories reflect experiences that are common among preschoolers such as getting hurt, playing a game, or quarreling with a sibling.

The NLM:P was initially developed for outcome measurement in narrative intervention studies because there were no adequate instruments appropriate for this age group. Its primary purpose is to facilitate the standardized elicitation and analysis of children’s narratives so that treatment effects can be detected. Nonstandardized narrative analysis has been employed in hundreds of studies and is well established as an acceptable language assessment approach (Petersen, Gillam and Gillam, 2008). Although the NLM:P is a newly developed, researcher-made instrument, its reliability and validity are promising. As evidence of construct validity, the retell measure significantly correlates with norm-referenced, narrative language measures, including the Renfrew Bus Story (Cowley and Glasgow, 1994;  $r = .88$ ) and the Index of Narrative Complexity (Petersen, Gillam and Gillam, 2008;  $r = .93$ ). The NLM:P’s alternate form reliability ( $r = .77$ ,  $p < .001$ ) is also adequate (Petersen and Spencer, 2012).

### *Story champs instruction*

**Materials.** Story Champs, a multi-tiered narrative language program, was used for narrative instruction and intervention (e.g. Spencer & Slocum, 2010; Petersen et al., 2014); however, only the large group procedures were examined in this study. Story Champs’ materials consist of 12 personal-themed stories with corresponding pictures, story grammar icons, and champ checks. Story Champs stories have the same structural features as the NLM:P stimulus stories. The five pictures correspond to five major narrative structural elements: character, problem, feeling, action, and ending. The pictures are simple black and white drawings with a few colored features. Pictures were printed on  $8\frac{1}{2} \times 11$ -inch cardstock and laminated. Story grammar icons are symbols representing each major story grammar element (i.e. character, problem, feeling, action, and consequence) printed on  $4 \times 4$ -inch cards. Champ checks are  $5 \times 7$ -inch laminated cardstock boards with flaps (corresponding to each story grammar element) that open to reveal a picture of a trophy.

**Table 2.** Five main parts of stories with a sample story, and a description of story gestures.

	Character	Problem	Feeling	Action	Ending
Sample story	One day, John was riding his bike down the street because it was his favorite thing to do.	He rode his bike and he accidentally crashed into a rock. John cut his knee.	He was sad because the cut hurt.	Then John ran home and said to his mom, "I need a band-aid." She put a band-aid on his cut.	After he got a band-aid, John's knee felt better.
Story gestures	Open hand on top of head	Thumbs down	Pointer finger near eye and motion downward like a tear falling	Closed right hand at left shoulder slowly moving across body and opening in a motion emulating a star moving across the sky	Thumbs up

*Arrangements.* Prior to the treatment phase, teachers helped arrange the children in their class into pairs for the partner retell step (see step 4 below). Children were paired together if they got along well, were not likely to be disruptive together, and if one child was capable of helping the other one. For example, pairs of children were commonly arranged so that children with stronger English language skills were paired with children with weaker English language skills. Partners were not adjusted throughout the treatment phase unless children were absent. On occasion, an adult served as a child's partner.

*Instructor.* Because this was an early efficacy study, T.D.S. and developer of Story Champs served as the instructor. T.D.S. is an experienced preschool teacher, school psychologist, and early childhood special educator. Although the classroom teacher and teaching assistant did not assist during the large group activities, they helped during the partner retell step (see step 4 below).

*Prompting and corrections.* During each step of Story Champs in which vocal responses were expected of the children, the instructor provided prompting as needed. In general, more direct prompts (e.g. "David got a shot. Now you say, David got a shot.") were used during early sessions and less direct prompts (e.g. "What was David's problem? Tell me what happened in this picture.") were used during later sessions. However, a strict prompt hierarchy was not followed. Because the children had very different language skills, the instructor relied on clinical judgment to adjust prompting. Prompting was focused on helping children identify and retell the structural elements of the story, and specific language complexity features were not emphasized. Even so, those features were modeled every time the instructor told or retold the stories. On occasion when a group response was expected and few children responded, the instructor said, "I need to hear all of your voices. Let's try it again." If a child or the group gave an incorrect response, the instructor did not say it was wrong. She simply modeled the correct response and said, "Let's try it again." When moving around the room helping partners perform their independent champ checks responsibilities, the instructor was more likely to use questions to help children come to the correct answer than a model correction.

**Procedures.** During the treatment phase, 12 sessions of Story Champs were delivered to both classes in the treatment group and each session lasted approximately 15–20 minutes (view video clip at [http://www.youtube.com/watch?v=u9\\_2f3DRDfY](http://www.youtube.com/watch?v=u9_2f3DRDfY)). Story Champs sessions occurred 4 days a week (Monday through Thursday) for 3 weeks. Stories used during instruction were randomly selected without replacement for each of the two treatment classes separately. In other words, the two treatment classes did not receive the stories in the same order and every story was unfamiliar to the students. Before beginning each Story Champs session, the instructor selected a story and attached the five pictures to a large canvas so all the children could see them.

**Step 1: model.** In the first step, with the participants sitting on a large rug in a circle, the instructor modeled the story by reading it word for word. The instructor paused at each main structural element to attach the corresponding story grammar icon and to name it (e.g. “Hannah is the character”). At the end of step 1, the instructor asked the group to name the five parts of the story by saying, “Everybody, what are the five main parts of a story? Get ready.” The teacher pointed to each icon in sequence as the group named the five main parts (i.e. “Character, problem, feeling, action, and ending”).

**Step 2: story gestures.** In the second step, the instructor told the story again and prompted the children to play story gestures. In story gestures, children made a gesture when the instructor told parts of the story. For example, as the instructor told about the problem, the children made a thumbs down sign (see Table 2 for a description of story gestures). They made gestures for each main part as the instructor told the story.

**Step 3: team retell.** In the third step, the instructor asked the group to retell the story. She said, “Everybody, think about what happens in this picture. Who/What is the *story grammar element*? Just think about it and I’m going to call on someone.” After a brief pause so all the children had time to formulate a silent response, the instructor asked a child to provide that part of the story. Prompts were provided, if necessary. The instructor restated the target response at least twice and asked for the class to say it with her. For example, “That’s right. David had to get a shot. Everyone say it with me, “David had to get a shot.” That procedure was repeated until individual children told all five main parts in order. At the end of step 3, the instructor summarized the entire story to provide a cohesive model.

**Step 4: partner retell.** In the fourth step, students were paired with a partner to retell the whole story individually. Pairs of children were ushered to separate locations in the classroom for them to work together. One partner, who was to be the champ checker (i.e. the tutor) first, was given the champ checks. Champ checks were designed to hang loosely around one child’s necks so that the story grammar icons were visible to help the other child retell the story. As the storyteller retold the story, the champ checker listened carefully for each part of the story. When the storyteller told each part, the champ checker opened a corresponding flap. If necessary, the champ checker helped the storyteller or told the partner what part(s) of the story they forgot to tell. When the storyteller finished the story, the roles switch. The storyteller became the champ checker and the champ checker became the storyteller. Children’s roles were systematically alternated and those assignments were distributed just before each session. This was important so the same child did not complete the storyteller role first every time. During the partner retell step, the instructor, classroom teacher, and teaching assistant walked around to praise children for being good “checkers” and for retelling the story, or to provide help as needed. In the first two Story Champs sessions, the instructor and an assistant demonstrated the duties of the champ checker and the storyteller.

*Step 5: champ ceremony.* After both children completed the storyteller role, their champ checker flaps were open to reveal a trophy. The instructor called for the children to return to the rug and asked each pair to hold up their trophy. The adults and children clapped and cheered for the story champs (i.e. all the children).

### *Fidelity of implementation*

During half of the Story Champs sessions, an undergraduate research assistant observed T.D.S. delivering the intervention and monitored fidelity of implementation. For each of the sessions observed, she completed a procedural checklist with 14 items reflecting the key procedures described above, documenting the extent to which the instructor implemented each step accurately. Fidelity monitoring indicated the instructor implemented the procedures as intended in all the observed sessions, resulting in 100 percent fidelity of implementation.

### *Social validity*

Once the treatment group had completed 12 sessions of Story Champs, the teachers were asked to respond to a social validity teacher questionnaire. Following the completion of the study, T.D.S. delivered the same Story Champs intervention to the comparison classrooms. Following those Story Champs sessions, the teachers were asked to respond to the same social validity teacher questionnaire. It is important to note that teachers were never directly trained in the procedures, but they were present during every Story Champs session. The questionnaire included six items (e.g. Children enjoy Story Champs, My student's language improved as a result of Story Champs, I am interested in using Story Champs with future students) with options to rate each statement on a 1–5 Likert scale where 5 was *Strongly Agree* and 1 was *Strongly Disagree*. All four teachers rated every item with a 5, indicating maximum feasibility and impact.

### *Data collection*

*Administration.* A team of six research assistants who were blind to group assignment administered the NLM:P. The research assistants received 2 hours of training that included discussion, video demonstrations, and practice. Prior to administering the NLM:P for research purposes, each research assistant administered a practice TNR, TSC, and TPG with 100 percent accuracy. T.D.S. provided the training and used a procedural fidelity checklist to score their administration accuracy.

At each data collection point (Pretest, Posttest, and Follow-up), children were given three opportunities to complete the NLM:P tasks within one session, which lasted approximately 5–10 minutes depending on the subtest. For instance, an examiner brought an individual child to a quiet area just outside the classroom to administer one of the subtests. In a TNR session, the examiner read a story and asked the child to retell the story. Regardless of whether the child retold the story or not, the examiner read another (pre-selected) story and asked for the child to retell it. This repeated until the child had heard three different stories and was given an opportunity to retell each one of them. This procedure was used for all three subtests. Additionally, each subtest (i.e. TNR, TSC, and TPG) was administered on different days so that no child was asked to complete more than one assessment session per day. However, the TNR was always administered to children first, followed by the TSC, and then the TPG.

Stories used during administration were randomly selected without replacement for each student. Participants never heard the same story twice and each participant heard different, but

equivalent, stories in a different order and for different subtests. This was important so that changes in narrative language could be attributed to learning and not to the variability between stimulus stories or familiarity with the stories.

Given that the NLM:P is a new, researcher-made instrument, we randomly selected 30 percent of the audio-recorded assessment sessions across Pretest, Posttest, and Follow-up time points for fidelity evaluation by an independent research assistant. The research assistant listened to each audio session and completed the procedural fidelity checklist used during training. Fidelity of test administration was 91 percent (64%–100%) for the TNR, 93 percent (67%–100%) for the TPG, and 98 percent (80%–100%) for the TSC.

**Scoring.** During the administration of the TSC, examiners wrote participants' responses to questions on a score sheet in real time. Immediately following the TSC administration session, the examiner reviewed the participant's responses and scored them according to a 0–2 rating scale of completeness and clarity of responses, where 2 would be most complete. A total of 12 points was possible on the TSC.

For narrative retells and personal generations, examiners recorded participants' responses using digital voice recorders. A second set of research assistants, who were also blind to group assignment, transcribed recordings and scored them according to the NLM:P scoring guides. Initially, transcription, communication-unit segmentation, and NLM:P scoring procedures were taught to the entire group of research assistants sequentially by D.B.P. using preschool stories from children who were not participants. Following group training, D.B.P. and a graduate student in speech-language pathology individually trained each research assistant and provided extensive scoring practice using seven different preschoolers' retells and personal generations until 100 percent accuracy was documented across both narrative genres.

Retell and personal story narratives were scored for story completeness, also called story grammar, and language complexity. The story grammar subscale includes five categories that correspond to the main parts of stories: character, problem, feeling, action, and ending. Retell and personal story responses were examined for the inclusion and completeness of each category using a 0–3 Likert-type scale, where 3 points is the most complete and 0 means that component was not included in the narrative. A bonus point was given if a narrative response included a complete episode (i.e. problem, action, and ending). The language complexity subscale includes six language features that reflect language development of typical 5-year-old children and correspond to features embedded in the NLM:P stimulus stories. The language complexity features include opening, causal markers, temporal markers, dialogue, adverbs, and adjectives. Retell and personal story narrative responses were examined for inclusion and scored according to frequency. For example, if a child's response included two causal markers and one adjective, the language complexity score would be 3. The NLM:P stimulus stories were constructed to have a score of 24 total points (16 for story grammar and 8 for language complexity) and because they were used to elicit narrative responses, 24 is considered the total points possible.

### *Scoring reliability*

To document the extent to which scoring was consistent, 10 percent of children's TNR and TPG transcribed narratives from Pretest, Posttest, and Follow-up time points were randomly selected to be scored by an independent scorer. Research assistants were the primary coders and D.B.P. served as the independent scorer. Point-by-point scoring agreement from transcribed stories was 96 percent (83%–100%) for retells and 94 percent (75%–100%) for personal stories. Because children's responses to comprehension questions were recorded in real time and scored immediately

following the assessment session, the research assistants who administered the TSC also scored them. T.D.S. served as an independent scorer for 30 percent of the TSCs, selected randomly from Pretest, Posttest, and Follow-up time points. Point-by-point agreement for the story comprehension subtest was 91 percent (72–100).

### Data analysis

At each data collection point (Pretest, Posttest, and Follow-up), children were given three opportunities to complete each of the NLM:P tasks (TNR, TSC, and TPG). The main reason for providing multiple opportunities was to increase the likelihood of capturing a valid response from participants. Preschool children are not generally accustomed to individual testing situations and can be easily distracted or confused. To reduce the impact of fleeting attention and novelty of testing on the results, we selected each child's highest scored retell, highest scored set of comprehension responses, and highest scored personal story to use for analysis.

Of the 71 children enrolled, 3 children (2 from the treatment group and 1 from the comparison group) were unavailable for Follow-up testing. For each of these children, we substituted their groups' means at Follow-up for their missing scores.

## Results

This study examined the narrative language skills of preschool children attending a Head Start program, of which four classrooms were randomly assigned to either a treatment condition or a control condition. For the pretest measures, an analysis of variance (ANOVA) was conducted, with an alpha level of .05. A two-step procedure was used for analyses that compared effects of intervention. First, preliminary analyses were completed to determine whether it was appropriate to complete a one-way analysis of covariance (ANCOVA) due to the assumption of homogeneity of regression slopes and assumption of homogeneity of variance. The interaction between the group and a pretest measure was used to examine the homogeneity of slopes' assumption, while Levene's test was completed to evaluate the assumption that the population variances for the two groups were equal. Second, based on the results of the preliminary analyses, either ANCOVA or simple main effects tests were conducted. Follow-up tests were completed to determine pairwise comparisons among the adjusted means using the least significant difference (LSD) procedure, with an alpha level of .05. Effect sizes were calculated as standardized mean difference ( $d$ ) for pairwise comparisons. Values greater than 0.20 were considered small, greater than 0.50 were considered medium, and greater than 0.80 were considered large (Cohen, 1977).

### Story retell (TNR results)

**Pretest.** ANOVA was conducted and indicated that there were no significant differences between the treatment and control conditions,  $F(1, 69) = 0.24, p = .629$  (see Table 3).

**Posttest.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story retell, and the dependent variable, posttest story retell, did not differ significantly as a function of group,  $F(1, 67) = 0.39, p = .534$ . Levene's test indicated that the error variance of the dependent variable, posttest story retell, did not differ significantly as a function of group,  $F(1, 69) = 0.31, p = .581$ . ANCOVA was conducted since the preliminary analyses were not significant. The ANCOVA was significant,  $F(1, 68) = 4.11$ , mean squared error ( $MSE$ ) = 129.69,

**Table 3.** Means for narrative language measures for treatment and control conditions at pretest, posttest, and follow-up.

Testing	Treatment ( <i>n</i> = 36)	Control ( <i>n</i> = 35)
<b>Story retell</b>		
Pretest	5.19 ( <i>SD</i> = 6.33)	5.94 ( <i>SD</i> = 6.65)
Posttest*	11.00 ( <i>SE</i> = 0.94)	8.29 ( <i>SE</i> = 0.95)
Follow-up*	11.51 ( <i>SE</i> = 0.96)	8.62 ( <i>SE</i> = 0.97)
<b>Story comprehension</b>		
Pretest	4.89 ( <i>SD</i> = 4.39)	4.97 ( <i>SD</i> = 3.75)
Posttest*	6.92 ( <i>SE</i> = 0.45)	5.42 ( <i>SE</i> = 0.46)
Follow-up		
Low**	5.53 ( <i>SE</i> = 0.52)	1.45 ( <i>SE</i> = 0.58)
Medium**	7.80 ( <i>SE</i> = 0.38)	4.99 ( <i>SE</i> = 0.39)
High	10.07 ( <i>SE</i> = 0.53)	8.54 ( <i>SE</i> = 0.57)
<b>Story generation</b>		
Pretest	5.56 ( <i>SD</i> = 5.14)	7.41 ( <i>SD</i> = 5.55)
Posttest	6.62 ( <i>SE</i> = 0.90)	5.82 ( <i>SE</i> = 0.91)
Follow-up	7.87 ( <i>SE</i> = 0.78)	7.42 ( <i>SE</i> = 0.80)

*SD*: standard deviation; *SE*: standard error; ANCOVA: analysis of covariance.

Pretest results were reported as means and *SD*s while posttest and follow-up results were reported as adjusted means and *SE*s. ANCOVA was completed for all comparisons if the homogeneity assumptions were met. They were met for all comparisons except story comprehension at follow-up, so the simple main effect tests were conducted.

\*  $p < .05$ ; \*\* $p < .017$ .

$p = .046$ ,  $d = 0.49$ . Based on the LSD procedure, the treatment group had a significantly larger adjusted mean than the adjusted mean for the control group (see Table 3).

**Follow-up.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story retell, and the dependent variable, follow-up story retell, did not differ significantly as a function of group,  $F(1, 67) = 0.01$ ,  $p = .912$ . Levene's test indicated that the error variance of the dependent variable, follow-up story retell, did not differ significantly as a function of group,  $F(1, 69) < .01$ ,  $p = .980$ . ANCOVA was conducted since the preliminary analyses were not significant. The ANCOVA was significant,  $F(1, 68) = 4.45$ ,  $MSE = 146.93$ ,  $p = .039$ ,  $d = 0.51$ . Based on the LSD procedure, the treatment group had a significantly larger adjusted mean than the adjusted mean for the control group (see Table 3).

### Story comprehension (TSC results)

**Pretest.** ANOVA was conducted and indicated that there were no significant differences between the treatment and control conditions,  $F(1, 69) = 0.01$ ,  $p = .932$  (see Table 3).

**Posttest.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story comprehension, and the dependent variable, posttest story comprehension, did not differ significantly as a function of group,  $F(1, 67) = 0.14$ ,  $p = .714$ . Levene's test indicated that the error variance of the dependent variable, posttest story comprehension, did not differ significantly as a function of group,  $F(1, 69) = .06$ ,  $p = .808$ . ANCOVA was conducted since the preliminary analyses were not significant. The ANCOVA was significant,  $F(1, 68) = 5.41$ ,  $MSE = 39.88$ ,  $p = .023$ ,  $d = 0.56$ . Based on the LSD procedure, the

treatment group had a significantly larger adjusted mean than the adjusted mean for the control group (see Table 3).

**Follow-up.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story comprehension, and the dependent variable, follow-up story comprehension, differed significantly as a function of group,  $F(1, 67) = 5.20, p = .026$ . Levene's test indicated that the error variance of the dependent variable, follow-up story comprehension, did not differ significantly as a function of group,  $F(1, 69) = .739, p = .39$ .

Based on the results of the significant interaction effect, simple main effects test were conducted that allow for heterogeneity of slopes. Differences among groups were assessed at low (1 standard deviation (*SD*) below the mean = 0.87), medium (mean = 4.93), and high (1 *SD* above the mean = 8.99) values on the covariate. A *p* value of .017 (.05/3) was required for significance for each of these tests. If any one simple main effect was significant, then pairwise comparisons were evaluated at the same level of .017 using the LSD procedure. The simple main effects test was significant for a low pretest story comprehension score,  $F(1, 67) = 27.33, p < .0001, d = 1.26$ , and for a medium value on the covariate,  $F(1, 67) = 26.55, p < .0001, d = 1.24$ . In contrast, the simple main effects test was not significant for a high value on the covariate,  $F(1, 67) = 3.90, p = .052, d = 0.48$ . The treatment group produced significantly greater follow-up story comprehension scores than the control group for both low and medium pretest story comprehension scores (see Table 3).

### Story generation (TPG results)

**Pretest.** ANOVA was conducted and indicated that there were no significant differences between the treatment and control conditions,  $F(1, 69) = 2.13, p = .149$  (see Table 3).

**Posttest.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story generation, and the dependent variable, posttest story generation, did not differ significantly as a function of group,  $F(1, 67) = 24.51, p = .358$ . Levene's test indicated that the error variance of the dependent variable, posttest story generation, did not differ significantly as a function of group,  $F(1, 69) = 1.12, p = .294$ . ANCOVA was conducted since the preliminary analyses were not significant. The ANCOVA was not significant,  $F(1, 68) = 0.38, MSE = 10.85, p = .539, d = 0.15$  (see Table 3).

**Follow-up.** The evaluation of the homogeneity of slopes assumption indicated that the relationship between the covariate, pretest story generation, and the dependent variable, follow-up story generation, did not differ significantly as a function of group,  $F(1, 67) = 0.003, p = .955$ . Levene's test indicated that the error variance of the dependent variable, follow-up story generation, did not differ significantly as a function of group,  $F(1, 69) = 0.89, p = .769$ . ANCOVA was conducted since the preliminary analyses were not significant. The ANCOVA was not significant,  $F(1, 68) = 0.16, MSE = 3.53, p = .688, d = 0.10$  (see Table 3).

### Responsiveness for the total treatment group

Responsiveness was calculated for the children in the treatment group ( $n = 36$ ) using a criterion raw best score at posttest and a criterion gain score from pretest to posttest for the story retell and story comprehension measures. The criteria were set at 0.5 *SD* below the mean, based on the work by McMaster et al. (2005). Four groups of children were identified based on the level and gain criteria: Minimal Responders (MR), children who did not meet either criterion; Gainers (G), children who met the gain criterion but not the level criterion; Levelers (L), children who met the

**Table 4.** Responsiveness for the story retell and story comprehension tasks at posttest for the treatment group.

Responsiveness	Number of children	Mean raw score (SD)	Mean raw gain (SD)
<b>Story retell</b>			
Minimal Responder	10	0.50 (1.27) <sup>a,b,c</sup>	0.20 (0.42) <sup>c</sup>
Gainer	2	3.00 (0.00) <sup>d,e</sup>	2.50 (0.71) <sup>e</sup>
Leveler	3	17.67 (3.21)	-0.33 (0.58) <sup>f</sup>
Responder	21	15.33 (3.90)	8.86 (5.04)
<b>Story comprehension</b>			
Minimal Responder	7	0.00 (0.00) <sup>a,b,c</sup>	0.00 (0.00) <sup>a,c</sup>
Gainer	4	3.50 (1.00) <sup>d,e</sup>	3.00 (1.15) <sup>d</sup>
Leveler	8	9.13 (2.47)	-0.63 (0.74) <sup>f</sup>
Responder	17	9.47 (2.45)	3.82 (2.86)

SD: standard deviation.

<sup>a</sup>Minimal Responder and gainer comparison significant at  $p < .05$ .

<sup>b</sup>Minimal Responder and Leveler comparison significant at  $p < .05$ .

<sup>c</sup>Minimal Responder and responder comparison significant at  $p < .05$ .

<sup>d</sup>Gainer and Leveler comparison significant at  $p < .05$ .

<sup>e</sup>Gainer and responder comparison significant at  $p < .05$ .

<sup>f</sup>Leveler and responder comparison significant at  $p < .05$ .

level criterion at pretest and had no room to gain and Responders (R), children who met both the criteria. A one-way ANOVA was conducted to evaluate the relationship between responsiveness type and level and the relationship between responsiveness type and gain. Follow-up tests were conducted to evaluate pairwise differences among the means. Due to the small sample size, post hoc comparisons were conducted using the Games–Howell test, which does not assume equal variances among the groups.

**Story retell.** For the posttest story retell level criterion, scores of 7 and above were considered responsive. For the story retell gain (i.e. pretest to posttest) criterion, scores of 2 or higher were considered responsive. The R group had 21 participants (58.3%), the L group had 3 participants (8.3%), the G group had 2 participants (5.5%), and the MR group had 10 participants (27.7%). The ANOVA was significant for the level,  $F(3, 32) = 55.1, p < .001$ , and gain,  $F(3, 32) = 13.3, p < .001$ , criteria. The R, L, and G groups had a significantly greater level than the MR group ( $p < .001; p = .023; p = .001$ ), while the R and L groups had a significantly greater level than the G group ( $p < .001; p = .039$ ). There was no significant difference between the R and L groups ( $p = .70$ ). As regards to gain, the R group had significantly greater gains than the L, G, and MR groups ( $p < .001$  for all). There were no other significant differences for gain ( $p = .22; p = .55; p = .11$ ; see Table 4).

**Story comprehension.** For the posttest story comprehension level criterion, scores of 5 and above were considered responsive. For the story comprehension gain (i.e. pretest to posttest) criterion, scores of 1 or higher were considered responsive. The R group had 17 participants (47.2%), the L group had 8 participants (22.2%), the G group had 4 participants (11.1%), and the MR group had 7 participants (19.4%). The ANOVA was significant for the level,  $F(3, 32) = 39.9, p < .001$ , and gain,  $F(3, 32) = 11.1, p < .001$ , criteria. The R, L, and G groups had a significantly greater level than the MR group ( $p < .001; p < .001; p = .018$ ), while the R and L groups had a significantly

**Table 5.** Percentage of children identified by responsiveness type at posttest for the treatment condition by language.

Language	Minimal Responder (%)	Gainer (%)	Leveler (%)	Responder (%)
<b>Story retell</b>				
English ( <i>n</i> = 16)	18.8	6.3	12.5	63.5
DLL ( <i>n</i> = 20)	35.0	5.0	5.0	55.0
Total ( <i>n</i> = 36)	27.8	5.6	8.3	58.3
<b>Story comprehension</b>				
English ( <i>n</i> = 16)	12.5	6.3	37.5	43.8
DLL ( <i>n</i> = 20)	25.0	15.0	10.0	50.0
Total ( <i>n</i> = 36)	19.4	11.1	22.2	47.2

DLL: dual-language learner.

greater level than the G group ( $p < .001$ ;  $p = .001$ ). There was no significant difference between the R and L groups ( $p = .99$ ). As regards to gain, the R group had significantly greater gains than the L and MR groups ( $p < .001$ ;  $p < .001$ ), while the G group had significantly greater gains than the L and MR groups ( $p = .041$ ;  $p = .013$ ). There were no significant differences between the L and MR groups ( $p = .17$ ) or the R and G groups ( $p = .80$ ; see Table 4).

### Responsiveness by language

A two-way contingency table analysis was conducted to evaluate whether language learning (i.e. English only or DLL) impacted type of responsiveness (see Table 5). For this analysis, children classified as DLL spoke a language other than English at least half of the time (see Table 1). We included children who fell into the Half English/Half other category as DLLs because during interviews with parents, all children were determined to be sequential bilinguals and not simultaneous bilinguals. Language and responsiveness type were not found to be significantly related for the story retell task at posttest, Pearson  $\chi^2(3, N = 36) = 1.56, p = .67$ , or the story comprehension task at posttest, Pearson  $\chi^2(3, N = 36) = 4.43, p = .22$ . Follow-up pairwise comparisons were not completed due to the lack of significance.

## Discussion

### Intervention efficacy

The primary purpose of this study was to test the efficacy of a large group narrative intervention on preschoolers' narrative language skills. As this was one of the first examinations of a whole-class narrative intervention, the results provide early support for its efficacy. Statistically significant differences between the experimental and control groups were found for the story retell and story comprehension measures at Posttest and Follow-up; however, the intervention did not affect personal story generations. This is not surprising considering that the intervention was implemented at a low dose in a whole-class arrangement and did not directly teach or provide practice telling personal stories.

The observed improvements in story retell and story comprehension are quite meaningful. Preschoolers who have adequate narrative comprehension are able to benefit from teacher-led storybook reading and other classroom-based literacy instruction (McCabe and Rollins, 1994).

Early narrative comprehension is associated with later reading comprehension (Catts et al., 2002; Dickinson and McCabe, 2001; Griffin et al., 2004) and most early learning guidelines include objectives related to retelling the beginning, middle, and end of a story.

While narrative retelling and answering questions about stories are important, improved personal stories offer children additional social and language benefits (Boudreau, 2008; Johnston, 2008). Personal stories may be imperative for children with language-learning difficulties, whose language limitations interfere in social competence and acceptance (Botting and Conti-Ramsden, 2000; Hart et al., 2004). Adequate personal story-telling skills can lead to increased peer attention and more opportunities to practice narrative language (McCabe and Marshall, 2006).

Given the social importance of personal stories for young children, it seems necessary to further support the development of this narrative genre. However, a whole-class intervention may not be the most appropriate context for teaching personal stories. The personal nature of the Story Champs stories presents opportunities for children to volunteer a story, but children who volunteer in a large group will likely be the children with sufficient narrative skills. Within a whole-class arrangement, it is difficult to ensure every child has an opportunity to tell a personal story while providing high-quality teacher support. For that reason, it is likely that the large group Story Champs intervention should maintain a focus on teaching story grammar through a retell genre only. When transfer of story grammar improvements from retell to personal stories is desired, then something more or different is needed.

Fortunately, adding something more or different aligns with preschool RtI models in which varying levels of intervention intensity are available (Coleman et al., 2009). Large group intervention should not be the end all-be all approach, but it is certainly an economical place to begin intervention. Typically, as a group size decreases, intensity and the ability to tailor an intervention to individual children increase (Fuchs and Fuchs, 2006). This is evident in Spencer and Slocum (2010) small group narrative intervention, in which the transfer of story grammar from retell to personal stories was systematically engineered. Their participants received intervention in groups of four, and a third of the intervention steps addressed personal stories. Petersen et al. (2014) further increased the intensity of narrative intervention and explicitly focused the intervention toward personal stories of children with autism. This was accomplished by decreasing the group size to one and increasing the duration of intervention sessions. When these three studies are viewed collectively, they represent multiple tiers of narrative intervention using the same personal-themed stories, illustrations, colored icons, and carefully scaffolded teaching procedures (i.e. Story Champs). If provided sequentially (large group, small group, individual), the system naturally increases its intensity, explicitness, individualization, and focus on personal stories.

### *Responsiveness analysis*

We first tested the efficacy of the large group, low-intensity narrative intervention, and considering only 58.3 percent (retell measure) and 47.2 percent (story comprehension measure) of the treatment group responded favorably to our intervention, the statistically significant results are quite remarkable. While this level of response may seem low for a test of intervention efficacy, it is important to understand the differences in populations that preschools serve compared to elementary schools. Public prekindergarten and Head Start preschools restrict enrollment to children with documented need (e.g. poverty); elementary schools do not. Understandably, preschools are filled with children who are at risk due to low income and cultural and linguistic differences (Kaiser et al., 2002; Yandian, 2009). Logically, there will be a higher proportion of preschool children who demonstrate language-learning limitations than in K-3 schools. Therefore, it is reasonable to expect a low-intensity language intervention to be inadequate for a large number of

at-risk children and that further intensified interventions will be needed. In fact, 27.7 percent did not respond to the intervention and 5.6 percent more made moderate gains but never reached an adequate performance level as measured by the story retell measure.

A critical component of any RtI model is accurate and early identification of children for whom initial, low-intensity instruction/intervention is insufficient. Our a priori criterion for both level and gain, set at .5 *SDs* below the mean, resulted in statistically significant classifications as expected between the four resultant categories across the retell and story comprehension measures, with few exceptions. These findings suggest that the four classifications, R, L, G, and MRs, divided children according to meaningful selection criteria.

The four classifications based on responsiveness can inform early childhood educators about the level of instruction/intervention that children may need. For example, after a brief intervention the L and R may no longer require explicit narrative intervention or could continue in a low-intensity large group arrangement. The G group made some progress and it would be appropriate to continue with the large group intervention or slightly increase the intensity of their intervention (e.g. small group intervention). The MRs, the children with the greatest needs, may already be showing signs that they are in need of a significantly different intensity for instruction and be assigned to a small group or individual intervention. If this conceptual model had been applied based on the story retell measure, which reflects expressive and receptive language skills (Petersen and Spencer, 2012), then the large group intervention would be sufficient for 64.6 percent of the children and 33.4 percent (G and MRs) could have been assigned to Tier 2 or 3 interventions.

Current practice in RtI involves the use of seasonal (fall, winter, and spring) screening procedures to help identify children who require supplemental intervention. Although performance level and gain can be obtained from seasonal assessment, there are noteworthy limitations to this approach in preschool. First, most children are in preschool for only the year prior to their entrance into kindergarten. This means that by the time gain scores from fall to winter can be collected, children only have a few months left in preschool to receive the intensified intervention they need. Oral language deficits typically do not remediate in such a short period. Second, if children's static performance in fall is compared to a criterion performance level, the assessment is likely to over-identify children based on cultural and language differences and limited time in an enriched instructional environment. For instance, if in this study we had identified children for Tier 2 intervention using the static pretest measures, all our participants except the L would have qualified. Based on the retell measure, we would have qualified 91.7 percent of the treatment group for Tier 2 language intervention. Although we believe that 91.7 percent of the children needed narrative language instruction, they did not all need an intensified small group or individualized version of the intervention. In fact, more half of them benefited sufficiently from the low-intensity whole-class narrative intervention. Had we not conducted the responsiveness analysis, we would not have understood this phenomenon.

### *Responsiveness by language*

Our third research question is specifically related to the demographics of preschoolers today. Questions of efficacy and responsiveness are important, but if an intervention is not effective for diverse populations then it is severely limited. Although there is a heightened interest in examining the effect of bilingual instruction on English and Spanish language development (e.g. Barnett et al., 2007), in our study, all adult-delivered procedures were in English. On occasion, we observed some children assisting their partners during the peer-tutoring step (i.e. champ checks) in Spanish. For this English narrative intervention, we found that there were no statistically significant differences between responders for English-speaking children and DLLs. Essentially, responders came

from both language groups. As expected, however, there were more L from the English-only group and there were more MRs from the DLL group. Based on the retell subtest, 35 percent of the DLLs did not make progress in the low-intensity whole-class English narrative intervention. Not surprising, the majority of these children were those who spoke mostly or only their first language. It should be noted that 18.8 percent of the English-only participants were classified as MRs based on the retell measure. Our anecdotal observations suggest that these children were either extremely shy or had developmental delays inhibiting their ability to learn in a large group arrangement. These results and observations lead us to believe that our process of early identification accurately detected children who need strategic, intense language intervention.

### *Limitations and future directions*

Despite the favorable outcomes and interesting secondary findings, there are a number of limitations worth mentioning. First, the study design was weakened by our inability to randomly assign children individually to conditions. This early efficacy study should be followed by replications using stronger experimental designs with true random assignment. Second, T.D.S., who is an expert teacher and clinician, delivered the intervention. Highly trained interventionists may be acceptable for initial examinations of efficacy, but Head Start teachers should deliver the intervention to document effectiveness of the intervention. The program's external validity depends on its feasibility in typical Head Start classrooms with typical Head Start resources. A third limitation involves the manner in which we categorized children as DLLs. Given the resources available for this study, we were unable to conduct comprehensive language proficiency assessments to more accurately describe our participants. Because we relied on parent interviews, our language information was susceptible to reporter bias. Future research should correct this limitation by supplementing parent interviews with direct assessments of language proficiency. Fourth, the NLM:P used to document narrative outcomes is a new, researcher-developed language instrument and has not been fully validated. The NLM:P stories were initially designed for use as outcome measurement tool (Spencer and Slocum, 2010) and has been shown to be sensitive to intervention effects. Nonetheless, only preliminary evidence of the NLM:P's technical adequacy has been documented (Petersen and Spencer, 2012). Finally, this study has broad implications for early childhood RtI models; however, this study did not actually take place within an RtI system. The investigators have drawn extensively from their theoretical foundation, clinical experience, and previous research to extrapolate meaningful RtI-relevant interpretations. At this point, narrative assessment and tiered narrative interventions have been investigated in a patchwork manner. Future research needs to tie it altogether and examine how it functions as a whole within an early childhood RtI context.

### *Conclusions and implications*

Young children receiving need-based preschool education are diverse with many risk factors that impair language development (Kaiser et al., 2002; Yandian, 2009). Depressed language skills are likely the most difficult to remediate (Hart and Risley, 1995). Narrative language is academically and socially relevant and serves as a critical outcome for young children. Traditional assessment practices have not overcome obstacles to valid assessment and proper early identification of culturally and linguistically diverse children. Based on our findings, we conclude that a low-intensity, whole-class narrative intervention improved story retelling and story comprehension skills of English-speaking children and DLLs. Importantly, our dynamic approach to early identification yielded a manageable number of children for whom Tier 2 or Tier 3 language intervention would be recommended.

The large group Story Champs intervention is simple, brief, fun, and effective. As RtI models increase in practice among early childhood educators, such interventions will be useful to teachers interested in improving oral language skills of their diverse students. As evidenced in this study, the low-intensity narrative intervention was appropriate for native English speakers as well as DLLs. Considered together with previous narrative intervention studies, this intervention may function as one element of a promising tiered narrative intervention program, addressing the diverse language needs of preschoolers.

Tiered intervention is one component of an effective RtI system; valid assessment is the other. The expansion of RtI to early childhood education has been plagued with many assessment challenges. Individualized testing is not commonplace for preschoolers from middle-class backgrounds and is often even more foreign to those who come from impoverished language environments, resulting in biased test scores. Dynamic assessment approaches can reduce cultural and linguistic bias (Peña, 2000). When the majority of the children tested have limited English language skills due to poverty or cultural and linguistic differences, a less biased assessment approach is critical to ensure early and accurate identification children.

For schools, accurate identification can result in more precise allocation of interventions (and resources) to those children who truly need them. Early identification that leads to intensive and preventative language intervention may reduce the number of children entering kindergarten without the language foundations necessary for reading. From a practical perspective, the identification procedures implemented in this study took approximately 4 weeks (i.e. Pretest to Posttest). Although this study employed 12 sessions of Story Champs before retesting, fewer sessions may be needed to identify responsiveness. Spencer and Slocum (2010) found that all but one participant made substantial gains within three sessions. Their intervention was more intense than what was provided in this study; however, RtI via dynamic approaches may not necessitate 12 sessions. If a test-teach-test format was completed shortly after children begin preschool, identification could be accomplished months before winter screening occurs.

The implications of our responsiveness results are far-reaching and believed to be the most substantial contribution of this study. Early childhood RtI models promise to maximize educational outcomes for children. However, innovations such as dynamic approaches to assessment will be necessary to overcome the challenges inherent in early childhood education.

## Acknowledgements

The authors would like to thank the undergraduate and graduate students at Utah State University and the University of Wyoming who collected and analyzed the hundreds of narratives for this study.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

- Artiles AJ, Harry B, Reschly DJ, et al. (2002) Over-identification of students of color in special education: a critical overview. *Multicultural Perspectives* 4: 3–10.
- Aud S, Fox M and KewalRamani A (2010) *Status and trends in the education of racial and ethnic groups*. US Department of Education, National Center for Education Statistics, NCEES 2010-015, July. Washington, DC: US Government Printing Office.
- August D, Carlo M, Dressler C, et al. (2005) The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice* 20: 50–57.
- Barnett WS, Yarosz DJ, Thomas J, et al. (2007) Two-way and monolingual English immersion in preschool education: an experimental comparison. *Early Childhood Research Quarterly* 22: 277–293.

- Botting N and Conti-Ramsden G (2000) Social and behavioral difficulties in children with language impairment. *Child Language Teaching & Therapy* 16: 105–120.
- Boudreau D (2008) Narrative abilities: advances in research and implications for clinical practice. *Topics in Language Disorders* 28: 99–114.
- Bowey JA (1995) Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology* 87: 476–487.
- Burns MK, Appleton JJ and Stehouwer JD (2005) Meta-analytic review of responsiveness-to-intervention research: examining Field-Based and Research Implemented Models. *Journal of Psychoeducational Assessment* 23(4): 381–394.
- Buysee V and Peisner-Feinberg E (2010) Recognition & response: Response to intervention for pre-kindergarten. *Young Exceptional Children* 13: 2–13.
- Catts HW, Fey ME, Tomblin JB, et al. (2002) A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research* 45: 1142–1157.
- Chaney C (1994) Language development, metalinguistic awareness, and emergent literacy skills of 3-year-old children in relation to social class. *Applied Psycholinguistics* 15: 371–394.
- Chaney C (1998) Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics* 19: 433–446.
- Cohen J (1977) *Statistical Power Analysis for the Behavioral Sciences*. Rev. ed. New York: Academic Press.
- Coleman MR, Roth FP and West T (2009) *Roadmap to Pre-K RTI: Applying Response to Intervention*. New York: National Center for Learning Disabilities, Inc. Available at: <http://www.rtinetwork.org/images/roadmaptoprekrti.pdf>
- Council for Exceptional Children (CEC) (2007) CEC's position on response to intervention (RTI): the unique role of special education and special educators. *Council for Exceptional Children, Arlington, VA*, October.
- Cowley J and Glasgow C (1994) *The Renfrew bus story*. Centreville, DE: The Centreville School.
- Densky YI, Mittenberg W, Quintar B, et al. (1998) Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale-Revised. *Assessment* 5: 115–121.
- Dickinson DK and McCabe A (2001) Bringing it all together: the multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice* 16: 186–202.
- Fazio BB, Naremore RC and Connell P (1996) Tracking children at risk for specific language impairment: a 3-year longitudinal study. *Journal of Speech and Hearing Research* 39(3): 611–624.
- Feagans L and Farran DC (1994) The effects of day care intervention in the preschool years on the narrative skills of poverty children in kindergarten. *International Journal of Behavioral Development* 17: 503–523.
- Fletcher TV and Navarrete LA (2003) Learning disabilities or difference: a critical look at issues associated with the misidentification and placement of Hispanic students in special education programs. *Rural Special Education Quarterly* 22: 37–46.
- Fox L, Carta JJ, Strain P, et al. (2010) Response to intervention and the Pyramid model. *Infants and Young Children* 23: 3–13.
- Fuchs D and Fuchs LS (2006) Introduction to response to intervention: what, why, and how valid is it? *Reading Research Quarterly* 41(1): 93–99.
- Greenwood CR, Bradfield T, Kaminski R, et al. (2011) The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children* 43: 1–22.
- Griffin TM, Hemphill L, Camp L, et al. (2004) Oral discourse in the preschool years and later literacy skills. *First Language* 24: 123–147.
- Hart B and Risley TR (1995) *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H Brookes Publishing Company.
- Hart K, Fujiki M, Brinton B, et al. (2004) The relationship between social behavior and severity of language impairment. *Journal of Speech, Language, and Hearing Research* 47: 647–662.
- Hayward D and Schneider P (2000) Effectiveness of teaching story grammar knowledge to pre-school children with language impairment: an exploratory study. *Child Language Teaching & Therapy* 16(3): 255–284.
- Hughes D, McGillivray L and Schmidek M (1997) *Guide to Narrative Language: Procedures for Assessment*. Eau Claire, WI: Thinking Publications.

- Jackson S, Pretti-Frontczak K, Harjusola-Webb S, et al. (2009) Response to intervention: implications for early childhood professions. *Language, Speech and Hearing Services in Schools* 40: 424–434.
- Johnston JR (2008) Narratives twenty-five years later. *Topics in Language Disorders* 28: 93–98.
- Kaiser AP, Cai X, Hancock TB, et al. (2002) Teacher-reported behavior problems and language delays in boys and girls enrolled in Head Start. *Behavioral Disorders* 28(1): 23–39.
- Karweit NL (1989) Preschool programs for students at risk of school failure. In: Slavin RE, Karweit NL and Madden NA (eds) *Effective Programs for Students at Risk*. Boston, MA: Allyn & Bacon, pp. 75–102.
- Laing S and Kamhi A (2003) Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech and Hearing Services in Schools* 34: 44–55.
- Linas MW, Carta JJ and Greenwood CR (2010) Taking a snapshot of early childhood response to intervention across the United States: 2009 and 2010. In: *Head Start research conference*, Washington, DC. Available at: <http://www.crtiec.org/aboutcrtiec/documents/crg2010snapshotPoster3by65size.pdf>
- McCabe A and Rollins PR (1994) Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology* 3: 45–56.
- McCabe PC and Marshall DJ (2006) Measuring the social competence of preschool children with specific language impairment: correspondence among informant ratings and behavioral observations. *Topics in Early Childhood Special Education* 26: 234–246.
- McCauley RJ and Swisher L (1984) Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders* 49: 34–42.
- McMaster KL, Fuchs D, Fuchs LS, et al. (2005) Responding to nonresponders: an experimental field trial of identification and intervention methods. *Exceptional Children* 71: 445–463.
- Mehta PD, Foorman BR, Branum-Martin L, et al. (2005) Literacy as a unidimensional multilevel construct: validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading* 9(2): 85–116.
- National Professional Development Center on Inclusion (NPDCI) (2012) *Response to intervention (RTI) in early childhood: building consensus on the defining features*. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute, National Professional Development Center on Inclusion. Available at: <http://npdci.fpg.unc.edu>
- Pearson BZ (2002) Narrative competence among monolingual and bilingual school children in Miami. In: Oller DK and Eilers RE (eds) *Language and Literacy in Bilingual Children*. Clevedon: Multilingual Matters Ltd, pp. 135–174.
- Peña E (2000) Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly* 21(2): 87–97.
- Petersen DB (2011) A systematic review of narrative-based language intervention with children who have language impairment. *Communication Disorders Quarterly*, 32(4): 207–220.
- Petersen, DB, Brown CL, Ukrainetz TA, et al. (2014). Systematic individualized narrative intervention on the personal narratives of children with autism. *Language, Speech, and Hearing Services in Schools* 45: 67–86.
- Petersen DB, Gillam SL and Gillam, RB (2008). Emerging procedures in narrative assessment: The index of narrative complexity. *Topics in Language Disorders*, 28: 115–130.
- Petersen, DB, Gillam SL, Spencer, TD, et al (2010). The effects of literate narrative intervention on children with neurologically based language impairments: An early stage study. *Journal of Speech, Language, and Hearing Research*, 53, 961–981.
- Petersen, DB and Spencer, TD (2012). The Narrative Language Measures: Tools for language screening, progress monitoring, and intervention planning. *Perspectives on Language Learning and Education*, 19(4), 119–129.
- Rodekohr RK and Haynes WO (2001) Differentiating dialect from disorder: a comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders* 34: 255–272.
- Scheffner-Hammer C, Pennock-Roman M, Rzasa S, et al. (2002) An analysis of the test of language development—primary for item bias. *American Journal of Speech-Language Pathology* 11: 274–284.
- Scott CM (1995) A discourse approach to syntax teaching. In: Tibbits DF (ed.) *Language Intervention: Beyond the Primary Grades*. Austin, TX: Pro-Ed, pp. 435–463.

- Snow CE, Burns MS and Griffin P (eds) (1998) *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Spaulding TJ, Plante E and Farinella KA (2006) Eligibility criteria for language impairment: is the low end of normal always appropriate? *Language, Speech and Hearing Services in Schools* 37: 61–72.
- Speece DL, Roth FP, Cooper DH, et al. (1999) The relevance of oral language skills to early literacy: a multivariate analysis. *Applied Psycholinguistics* 20: 167–190.
- Spencer TD and Slocum, TA (2010). The effect of a narrative intervention on story retelling and personal story generation skills of preschoolers with risk factors and narrative language delays. *Journal of Early Intervention*, 32(3), 178–199.
- Success for All Foundation (2007) Independent reviews of success for all. Available at: <http://www.successforall.org/SuccessForAll/media/PDFs/IndependReviews.pdf>
- The Division for Early Childhood of the council for Exceptional Children (DEC), National Association for the Education of Young Children (NAEYC), and National Head Start Association (NHSA) (2012). Frameworks for response to intervention in early childhood: Description and implications. Retrieved from [http://www.dec-sped.org/About\\_DEC/Whats\\_New?id=191](http://www.dec-sped.org/About_DEC/Whats_New?id=191)
- Tyler AA and Sandoval KT (1994) Preschoolers with phonological and language disorders: treating different linguistic domains. *Language, Speech and Hearing Services in Schools* 25: 215–234.
- VanDerHeyden AM, Witt JC and Gilbertson DA (2007). Multiyear evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, 45: 225–256.
- Valencia R and Rankin RJ (1985) Evidence of content bias on the McCarthy scales with Mexican American children: implications for test translation and nonbiased assessment. *Journal of Educational Psychology* 77: 197–207.
- Valencia R and Suzuki L (2001) *Intelligence Testing and Minority Students: Foundations, Performance Factors, and Assessment Issues*. Thousand Oaks, CA: SAGE.
- Vort Corporation (2004) *Hawaii Early Learning Profile*. Palo Alto, CA: Vort Corporation.
- Walker D, Greenwood C, Hart B, et al. (1994) Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development* 65: 606–621.
- Wellman RL, Lewis BA, Freebairn LA, et al. (2011) Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech and Hearing Services in Schools* 42: 561–579.
- Westby C (1985) Learning to talk-talking to learn: oral-literate language differences. In: Simon CS (ed.) *Communication Skills and Classroom Success*. San Diego, CA: College Hill Press, pp. 181–218.
- Whitehurst GJ (1997) Language processes in context: language learning in children reared in poverty. In: Adamson LB and Ronski MA (eds) *Research on Communication and Language Disorders: Contribution to Theories of Language Development*. Baltimore, MD: Paul H Brookes Publishing Company, pp. 233–266.
- Yandian S (2009) *How to Welcome and Support Head Start Dual language Learners When They Get to Elementary School*. Washington, DC: US Department of Health and Human Services, Office of Head Start, Administration for Children and Families. Available at: [http://www.ncela.gwu.edu/files/webinars/8/yandian\\_ballantyne\\_mar09.pdf](http://www.ncela.gwu.edu/files/webinars/8/yandian_ballantyne_mar09.pdf)