Dynamic Assessment of School-Age Children's Narrative Ability:

An Experimental Investigation of Classification Accuracy

Elizabeth D. Peña, **Ronald B. Gillam**, **Melynn Malek**, **Roxanna Ruiz-Felter**, **Maria Resendiz**, **Christine Fiestas**, and **Tracy Sabel**
Ron Gillam is now at Utah State University, Logan, UT; Tracy Sabel is now with Infant Parent Program, Texas Early Childhood Intervention, Austin, TX; Melynn Malek is now at DeTar Hospital, Victoria, TX; Roxanna Ruiz-Felter is now at Richardson Independent School District, Richardson, TX.

Abstract

Two experiments examined reliability and classification accuracy of a narration-based dynamic assessment task.

Purpose—The first experiment evaluated whether parallel results were obtained from stories created in response to 2 different wordless picture books. If so, the tasks and measures would be appropriate for assessing pretest and posttest change within a dynamic assessment format. The second experiment evaluated the extent to which children with language impairments performed differently than typically developing controls on dynamic assessment of narrative language.

Method—In the first experiment, 58 1st- and 2nd-grade children told 2 stories about wordless picture books. Stories were rated on macrostructural and microstructural aspects of language form and content, and the ratings were subjected to reliability analyses. In the second experiment, 71 children participated in dynamic assessment. There were 3 phases: a pretest phase, in which children created a story that corresponded to 1 of the wordless picture books from Experiment 1; a teaching phase, in which children attended 2 short mediation sessions that focused on storytelling ability; and a posttest phase, in which children created a story that corresponded to a second wordless picture book from Experiment 1. Analyses compared the pretest and posttest stories that were told by 2 groups of children who received mediated learning (typical and language impaired groups) and a no-treatment control group of typically developing children from Experiment 1.

Results—The results of the first experiment indicated that the narrative measures applied to stories about 2 different wordless picture books had good internal consistency. In Experiment 2, typically developing children who received mediated learning demonstrated a greater amount of pretest to posttest change than children in the language impaired and control groups. Classification analysis indicated better specificity and sensitivity values for measures of response to intervention (modifiability) and posttest storytelling than for measures of pretest storytelling. Observation of modifiability was the single best indicator of language impairment. Posttest measures and modifiability together yielded no misclassifications.

Conclusion—The first experiment supported the use of 2 wordless picture books as stimulus materials for collecting narratives before and after mediation within a dynamic assessment paradigm. The second experiment supported the use of dynamic assessment for accurately identifying language impairments in school-age children.

Contact author: Elizabeth D. Peña, University of Texas at Austin, 2504-A Whiteside Avenue, Room 7-214, Austin, TX 78712. E-mail: lizp@mail.utexas.edu.

Keywords

assessment; bias; multicultural; narratives; children

Traditionally, speech-language pathologists diagnose children with language impairments (LI) by comparing their performance on standardized tests with the performance of their same-age peers (Tomblin, Records, & Zhang, 1996). There is growing dissatisfaction with single-time assessment because of the potential for measurement error (Bracken, 1988; McCauley & Swisher, 1984a, 1984b; Plante & Vance, 1994) as well as the interfering effects of potential cultural bias (Demsky, Mittenberg, Quintar, Katell, & Golden, 1998; Rodekohr & Haynes, 2001; Scheffner-Hammer, Pennock-Roman, Rzasa, & Tomblin, 2002; Valencia & Rankin, 1985; Valencia & Suzuki, 2001). Recently, Spaulding, Plante, and Farinella (2006) reviewed 43 commercially available tests of language that are currently in use. Of the 9 tests that provided classification accuracy statistics, only 5 classified children with 80%s or better accuracy. Classification accuracy of children from diverse cultural and linguistic backgrounds is likely to be even less accurate (Demsky et al., 1998; Peña & Quinn, 1997; Scheffner-Hammer et al., 2002; Valencia & Suzuki, 2001). When a test's content and/or the conventions for obtaining responses are inconsistent with a child's culture or experience, the interpretation of obtained test scores is uncertain (Sternberg, 2002; Sternberg & Grigorenko, 2002a, 2002b; Sternberg et al., 2002).

Peformance on nonword repetition tasks has been proposed as a diagnostic marker of LI (Bishop, North, & Donlan, 1996; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000). An advantage of using a nonword repetition task is that it appears to be culturally and linguistically unbiased because it is processing dependent rather than language dependent. Investigators have documented that children from various cultural backgrounds performed similarly on nonword repetition measures (Campbell, Dollaghan, Needleman, & Janosky, 1997; Ellis Weismer et al., 2000). A number of investigators have suggested that nonword repetition is a potential clinical marker of language impairment (Baddeley, Gathercole, & Papagno, 1998; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990). Other researchers have found that nonword repetition tasks yield accurate identification rates only when combined with language-based tasks (Conti Ramsden, 2003; Ellis Weismer et al., 2000). While performance on nonword repetition can be used to inform diagnostic decisions, results from such tasks are limited with respect to determining what to treat and what clinical methods to use.

Dynamic assessment provides information that is clinically useful, and it is one promising solution to the problems inherent in the cultural and linguistic bias of standardized tests (Gutierrez-Clellen & Peña, 2001; Peña, 2000; Peña, Iglesias, & Lidz, 2001). Combining assessment and teaching processes within a single assessment procedure has the potential to reveal observable and consistent differences between children with and without LI, regardless of their cultural and linguistic diversity (Gillam, Peña, & Miller, 1999; Tzuriel, 2000).

Dynamic assessment differs from traditional, static assessment in three important ways (Lidz, 1991,1996;Sternberg & Grigorenko, 2002b). First, examiners and children interact extensively during the assessment process. Rather than merely administering tests, examiners in dynamic assessment teach the cognitive-linguistic strategies that children need to perform optimally on a given task. Secondly, a focus of dynamic assessment is on the observation of learning processes and strategies during the teaching phase. Examiners look for evidence of emerging skills and strategies as they watch children attempt to learn a new skill (Haywood & Tzuriel, 1992;Haywood & Wingenfeld, 1992). Finally, dynamic assessment measures more than the demonstration of a skill at one point in time. Pretest-to-posttest comparisons of performance

and examination of emerging learning strategies during mediation sessions can reveal children's latent capacities for change (Bain & Olswang, 1995;Lidz & Peña, 1996;Olswang & Bain, 1996;Sternberg, 2002).

Current approaches to dynamic assessment have been influenced by Vygotskian sociocultural theory and Feuerstein's theory of mediated learning experiences (MLE; Lidz, 1996; Tzuriel, 2000). Vygotsky (1986) believed that cognitive and linguistic development occur as a function of symbolic mediation. Specifically, Vygotsky posed developmental mechanisms in which natural psychological processes such as memory, perception, concept formation, and attention are altered through contexts in which parents, teachers, or more competent peers attempt to teach children something new (Vygotsky, 1986). Altered psychological processes drive the development of language, which becomes a symbolic tool that regulates learning (Kozulin, 2002). A Vygotskian approach to dynamic assessment would focus on comparisons between preteaching and postteaching performance. The idea is that competent learners should perform a task better after instruction than before instruction as a result of altered psychological processes.

Feuerstein's (1979;Feuerstein, Miller, Rand, & Jensen, 1981) MLE theory extends Vygotskian theory to the formal educational process. Feuerstein defined MLE as an active departure from typical development. In MLE, learning is directed by teachers and parents who intentionally focus the child's acquired symbolic tools in ways that efficiently reorganize independent learning. One significant difference from Vygotskian approaches to dynamic assessment is that Feuerstein's approach focuses on children's behavior during MLE rather than on pretest-to-posttest change. The extent of the child's improvement after mediation is not measured against his or her own age peers. Rather, examiners observe how children learn to use and apply psychological tools in a learning situation. The critical measure is the assessment of changes in cognitive strategies (Kozulin, 2002), not changes in task performance.

The dynamic assessment protocol that is the subject of our investigations combines both approaches in documenting children's response to intervention. Examiners measure change from pretest to posttest, and they document changes in children's cognitive strategies (referred to as child responsivity) within the MLE sessions. Our application of dynamic assessment is based on Lidz's (1991,1997;Lidz & Thomas, 1987) application of cognitive dynamic assessment. Here, dynamic assessment uses a test-teach-retest method that allows clinicians to gain information about how children learn and use cognitive strategies in language learning. Interactive approaches such as dynamic assessment potentially provide additional information about language learning processes that can complement traditional assessment practices. But to recommend the use of dynamic assessment for the purpose of diagnosis, the reliability of the pre- and posttest measures and the degree to which dynamic assessment differentiates between children with and without language impairment, regardless of racial/ethnic background, should be tested.

Research on the dynamic assessment of language has demonstrated its potential for differentiating between language differences and true language impairment in children from nonmainstream backgrounds. Peña, Quinn, and Iglesias (1992) compared the pre- and posttest performance of African American and Latino American children with and without language impairment. The Expressive One-Word Picture Vocabulary Test (Gardner, 1979) was used as the pre- to posttest measure of vocabulary. Children received two 30-min MLEs in small groups. While there were no significant differences between the two groups of children at pretest, those with typical development earned significantly higher posttest scores than the children with LI. Observations of modifiability during MLE (operationalized as examiner effort, child modifiability, and extent of transfer) also significantly differentiated the two groups.

In a follow-up study, Peña, Iglesias, and Lidz (2001) replicated their earlier findings and added a no-treatment control group to assess possible carryover effects. Children with typical development who received MLE improved significantly more than children with LI and children in the no-treatment control group. Ukrainetz, Harpell, Walsh, and Coyle (2000) used a similar design to examine Native American children's response to MLE that focused on categorization. Stronger language learners improved more than weaker language learners at posttest and demonstrated higher modifiability during MLE.

Bain and Olswang (1995) examined predictive, construct, and concurrent validity of a dynamic assessment procedure that used a hierarchy of cues to predict response to language intervention. In their study, children were able to produce more two-word combinations in the supported cuing condition, providing evidence of construct validity. Children who were responsive to the cuing also were most responsive to intervention, providing evidence of predictive validity. Concurrent validity with language sample analysis yielded mixed results.

The application of dynamic assessment to narrative language has promise as a less (culturally and experientially) biased assessment tool than standardized tests because it provides information about the child's thought processes, emerging skills, and learning potential. Narrative assessment has high content validity because narratives are routinely part of the discourse that occurs at home and at school. For example, at home, parents often ask children to tell personal stories about experiences they have had, and children listen to and read narratives as part of their language arts instruction at school (Geist & Aldridge, 2002; Jordan, Snow, & Porche, 2000; Riding & Tite, 1985). Narrative intervention is often incorporated into treatment plans because of its functional nature and its relationship to academic demands (Davies, Shanks, & Davies, 2004; Gillam, McFadden, & van Kleeck, 1995; Hayward & Schneider, 2000; McCartney et al., 2004; McFadden, 1998; Peterson, Jesso, & McCabe, 1999; Schoenbrodt, Kerins, & Gesell, 2003; Stiegler & Hoffman, 2001; Ukrainetz, 1998). While current publications provide anecdotal evidence and guidelines for the dynamic assessment of narratives (Gillam & McFadden, 1994; Gillam et al., 1999; Gutierrez-Clellen, Peña, & Quinn, 1995; Gutierrez-Clellen & Quinn, 1993; Iglesias, 1985; Iglesias & Gutierrez-Clellen, 1988; L. Miller, Gillam, & Peña, 2001; Peña & Gillam, 2000), no experimental studies have examined the clinical efficacy of using dynamic assessment of narratives to classify language impairment in culturally and linguistically diverse children. The current investigation explores the utility of dynamic assessment for use in identification of LI.

L. Miller et al. (2001) published a manual describing their application of dynamic assessment to narratives. They provided evidence of face validity based on the literature on narrative development. Further, they applied Lidz's (1987, 1991, 2002) description of mediated learning to their intervention framework. This assessment procedure is not a normed test, so the authors did not provide evidence of test stability, interitem correlation, or classification accuracy.

The present investigation consisted of two experiments. The first experiment was a preliminary study to evaluate the reliability of the narrative measures. To explore the diagnostic utility of dynamic assessment, it was first necessary to demonstrate that the two stories used in the pre- and posttest phases of dynamic assessment yielded equivalent estimates of children's narrative performance without an intervening mediation. Children told stories in counterbalanced orders to evaluate alternative-forms and internal-consistency estimates of reliability. The second experiment focused on the application of dynamic assessment of narratives with children from diverse racial/ethnic backgrounds. First, children with and without LI were compared within a pretest-posttest, control group design. Next, we compared the classification accuracy of static and dynamic assessments of narrative performance by racial/ethnic group as well as for the group overall.

Experiment 1

Method

Participants—A sample of 58 first- and second-grade children from central Texas participated in this study. Children were from African American (38%), European American (34%), and Latino American (28%) backgrounds, as reported by parents. The groups were balanced for grade (48% first graders; 52% second graders), but the gender distribution favored girls (64%) over boys (36%). Table 1 summarizes the ethnicity, gender, and grade data for each group (see Order 1 and Order 2).

Children took home permission forms that contained general information about the study. Parents were asked to give permission for participation as well as permission for school record review. Children in the typically achieving groups met at least three of the following criteria:

1. Teachers indicated no concerns regarding children's expressive language, receptive language, and/or speech.

2. Parents indicated (via questionnaire) no concerns regarding children's expressive language, receptive language, and/or speech.

3. Classroom observation using Patterson and Gillam's (1995) classroom observations of peer interaction indicated fewer than 15% syntactic, semantic, and/or pragmatic errors in a 10-min observation during play or group activities.

4. Children scored within one standard deviation of the mean on the Test of Language Development—Primary, Third Edition; TOLD-P: 3; Newcomer & Hammill, 1997) or the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999).

A research assistant not involved in other aspects of the study was responsible for group assignment into the first and the second experiments. Children with LI were assigned to Experiment 2, as described below. Children with typical development were assigned to one of three typically achieving groups stratified by grade and ethnicity during the accrual phase of the study. Two of the three groups constituted the population for the current study. One group of typically developing children (Order 1) received the story, *Two Friends* (L. Miller, 2000b) first, followed by the story, *Bird and His Ring* (L. Miller, 2000a). The second group of typically developing children (Order 2) received *Bird and His Ring* first, followed by *Two Friends*. The third group constituted the experimental group in Experiment 2. Examiners were unaware of the children's language history, ability, and group assignment.

Procedure

Data collection: Children in both groups produced stories that corresponded to the two wordless picture books. The stories were collected 4 to 6 weeks apart, according to procedures published in *Dynamic Assessment and Intervention* (L. Miller et al., 2001). In keeping with Berman and Slobin (1994), participants were presented with a wordless picture book and were instructed to think of a story to go with the pictures. Children could look at the pictures as long as they wished. When children indicated that they were ready, they told their story while looking at the pictures. Stories were audiotaped using a Marantz audio recorder and a lavalier microphone.

Transcription and coding of narrative samples: Audiotapes were transcribed and analyzed according to Systematic Analysis of Language Transcripts (SALT; J. Miller & Chapman, 2002) procedures. During transcription, participants' stories were segmented into C-units (i.e., each main clause and its subordinating clauses; Loban, 1976; J. Miller & Chapman, 1983). Mazes (e.g., nonlinguistic vocalizations, repetitions, false starts, and abandoned utterances)

were marked with parentheses and were excluded from the word count. Graduate student research assistants who were unaware of group assignment rated the stories for 10 aspects of narrative language that yielded three category scores—Story Components (Setting: Time and Place, Character Information, Causal Relationships, and Temporal Order of Events), Story Ideas and Language (Complexity of Ideas, Knowledge of Dialogue, Complexity of Vocabulary, Grammatical Complexity, and Creativity), and Episode Structure (combinations of various story grammar elements)—using the dynamic assessment of narratives protocol described by L. Miller et al. (2001). All the ratings, except Episode Structure, were based on a Likert 5-point scale ranging from *none stated/given* to *well specified/detailed*. The Episode Structure scale used a 7-point scale that ranged from *none* to *multiple episodes*. Three of the items were adapted slightly from the original in cases where the criteria had overlapping descriptions (these adaptations are presented in Appendix A). The three category scores were summed to yield a total story score ranging from 10 to 52.

In addition to the narrative ratings presented above, productivity measures including mean length of utterance (MLU) of words per C-unit, total number of words, number of different words, number of clauses, and number of clauses per C-unit were calculated. These calculations were generated by SALT.

Reliability: Interrater reliability for the narrative ratings was calculated on 20 randomly selected stories (10 from each book). Each rater scored the transcripts independently. The Pearson product-moment correlation for two independent ratings was .93. Pearson product-moment correlations were .93 for *Two Friends* and .94 for *Bird and His Ring*.

Interrater reliability for number of C-units, number of clauses, and number of different words was calculated on 14 of the stories (7 from each book). An independent rater trained in the transcription methods randomly selected the audiotapes, listened to the tapes, transcribed the samples, segmented utterances by C-units, and identified and tallied the total number of clauses. Pearson product-moment correlations for the two independent calculations were number of C-units = .93, number of clauses = .93, number of different words = .98.

Results

The goal of the first experiment was to determine whether the two books yielded comparable measures of children's narrative performance without intervening mediation sessions. Total story scores (sum of Story Components, Story Ideas and Language, and Episode Structure) and productivity measures (MLU for words, total number of words, number of different words, number of clauses, number of C-units, and number of clauses per C-unit) for the two books were compared. Parallel-forms reliability and coefficient alpha were calculated for each of the two books for the total story scores. Next, possible differences for ethnicity and gender were explored. Means and standard deviations by book, time, gender, and ethnicity are displayed in Table 2. Results indicated that stories produced for the two books yielded equivalent total story scores and productivity measures with similar score variances (indexed by standard deviation).

Parallel-forms reliability using Pearson correlations for the narrative scores across the two books was .88. Coefficient alpha, which yields information about interitem stability, was .824 for the *Two Friends* stories and .800 for the *Bird and His Ring* stories. These values are considered to be good to very good (DeVellis, 1991).

Repeated measures analysis of variance (ANOVA) was used to evaluate effects of time. Children's total story scores and productivity measures were compared at Time 1 and Time 2. A repeated measures ANOVA for total story scores, with time as the within-subjects factor, revealed a significant main effect for time, $F(1, 57) = 19.239$, $p < .001$, $\eta_p^2 = .252$. This is a small-to-moderate effect. On average, children's total story scores were 3.328 points higher

on the second story than the first story. For productivity, a repeated measures ANOVA was conducted with time and measure as the within-subjects factors. There was a significant main effect for time, $F(1, 57) = 4.455$, $p = .039$, $\eta_p^2 = .072$. This was a very small effect size. Table 3 contains the means for Time 1 and Time 2 for total story score and productivity measures.

Ethnicity and gender differences were evaluated with independent repeated measures mixed ANOVA. We report F values that are based on between and within solutions to mixed ANOVA because sphericity assumptions were met in all cases. Of interest were main effects or interactions with ethnicity or with gender on total narrative scores. For tests of potential ethnicity differences, book (*Two Friends* and *Bird and His Ring*) was the within-subjects factor and ethnicity (African American, European American, and Latino American) was the between-subjects factor. Main effects for ethnicity, $F(2, 55) = 0.985$, $p = .2380$, $\eta_p^2 = .035$, and book, $F(1, 55) = 2.454$, $p = .123$, $\eta_p^2 = .043$, were not significant, and there were no significant interactions between book and ethnicity. Children with different racial/ethnic backgrounds scored similarly on the stories (African American = 28.75, European American = 27.10, and Latino American = 25.63). For tests of potential gender effects, book (*Two Friends* and *Bird and His Ring*) was the within-subjects factor, and gender (male and female) was the between-subjects factor. Main effects for gender, $F(1, 56) = .568$, $p = .454$, $\eta_p^2 = .010$, and book, $F(1, 56) = 2.845$, $p = .097$, $\eta_p^2 = .048$, were not significant, and there were no significant interactions between book and gender. Total story scores were similar for boys ($M = 26.33$) and girls ($M = 27.88$).

Two repeated measures mixed ANOVAs were computed to explore possible ethnicity and gender differences on the productivity measures. For tests of potential ethnicity effects, book (*Two Friends* and *Bird and His Ring*) was the within-subjects factor, and ethnicity (African American, European American, and Latino American) was the between-subjects factor. The productivity measures (number of words, number of different words, number of C-units, MLU for words, number of clauses, and clauses per C-unit) were the dependent variables. There were no significant differences for ethnicity, $F(2, 55) = .731$, $p = .486$, $\eta_p^2 = .026$, or book, $F(1, 55) = 1.503$, $p = .225$, $\eta_p^2 = .027$, and no significant interactions between ethnicity and book. Children with different racial/ethnic backgrounds scored similarly on the productivity measures (see Table 2).

To test potential gender differences, the within-subjects factor in the mixed measures ANOVA was book (*Two Friends* and *Bird and His Ring*); gender (male and female) was the between-subjects factor. The dependent variables were number of words, number of different words, number of C-units, MLU for words, number of clauses, and clauses per C-unit. There were no significant differences for gender, $F(1, 56) = .388$, $p = .536$, $\eta_p^2 = .006$, or book, $F(1, 56) = 1.304$, $p = .258$, $\eta_p^2 = .023$, and no significant interactions between gender and book.

Discussion

The results of this study indicated that the wordless picture books, *Two Friends* and *Bird and His Ring*, elicited narratives that were equivalent with respect to total story scores (Story Components, Story Ideas and Language, and Episode Structure) and productivity (MLU, number of words, number of different words, number of clauses, number of C-units, and number of clauses per C-unit). These results provide evidence of parallel-forms reliability (Allen & Yen, 1979). Furthermore, these stories yielded similar narrative and productivity scores for males and females and for children from various racial/ethnic groups. Children increased their total story scores and productivity slightly from the first to the second stories. Therefore, evaluation of change over time using these materials requires a correction to make direct comparisons of pre- to posttest performance. Table 3 provides data for each measure that can be used to correct pretest scores allowing direct comparisons.

Use of wordless picture books has several advantages for dynamic assessment of narratives and for formative assessment of narrative ability during treatment. But many books that are frequently used to elicit children's narratives were not designed for the purpose of comparing change in performance over time. One advantage of using alternative, parallel books is reduction in text exposure. Children respond to materials that are similar but different for each administration, decreasing the likelihood that they learned the story from the initial exposure. The use of alternative books also maintains children's interest. Anecdotally, no child complained of having already told the story when asked to tell a story from the second book. The stability of the story scores and productivity supports extension to examining pre- to posttest change in response to dynamic assessment.

Experiment 2

This experiment had two main purposes. First, it was designed to compare pre- to posttest dynamic assessment changes in story creation by children with and without LI. Story changes by children in the two treatment groups were also compared to story changes by children in a no-treatment, control group (from Experiment 1, Order 1). Second, the experiment was designed to explore the diagnostic utility of the dynamic assessment procedure for classifying children as LI. One goal in this second analysis was to examine the potential bias of the measures used at pretest. Valencia and Suzuki (2001) have argued that the type of bias analysis that is performed should be directly related to how a measure will be used. Because diagnostic decisions are a hallmark of test use in speech-language pathology, an important bias analysis involves determining whether misclassifications vary as a function of race or ethnicity. Overall, we were interested in investigating which kinds of measures (story scores, productivity measures at pretest and posttest, and modifiability observations obtained during MLE) best differentiated children with typical development from children with LI.

Method

Participants—A total of 71 children from first and second grade in Central Texas and Southern California participated in this experiment. Children were assigned to one of three groups. One group included 27 children with typical development (TD group). The second group included 14 children with language impairment (LI group). There was also a no-treatment, control group (CON group) that consisted of the 30 children from the first experiment who received Order 1 (*Two Friends* then *Bird and His Ring*). Data for this experiment were collected simultaneously with the first experiment.

While we have no specific socioeconomic status (SES) information on individual children, we do have such data from the different schools from which the children were drawn. Children were recruited from seven schools in four school districts (three in Texas, one in California). All participant schools had enrollment of at least 50% economically disadvantaged children or children who received free or reduced-cost lunches. The mean (collapsed across the two types of identifiers) was 68.54% and ranged from 54.70% to 89.50%. This broad range of SES was evident for both the California and Texas schools.

Children were identified with typical language development using the same criteria as described in Experiment 1. The LI group was identified on the basis of standardized tests, history of LI, teacher or parent concern, and classroom observation. Children placed in the LI group met at least two of the following criteria:

1. Diagnosis of a language disorder by a certified speech-language pathologist.

2. Teacher or parent concern regarding their language expression or comprehension at school or at home.

3. Performance less than or equal to 1.25 standard deviations below the mean on the TOLD-P:3 (Newcomer & Hammill, 1997) or the CASL (Carrow-Woolfolk, 1999).

Overall, children were balanced for grade (48% first grade and 52% second grade) and ethnicity (35% African American, 32% European American, 30% Latino American, 3% did not report). There were fewer girls in the LI and TD groups but more girls in the CON group. There were proportionally more European American boys in the TD group and proportionally more African American girls in the CON group. Table 1 provides frequency counts for gender, ethnicity, and grade.

To ensure that the different proportions by gender and ethnicity did not affect the results of the study, we conducted preliminary ANOVA and repeated measures mixed ANOVA comparing differences for gender and ethnicity at pretest. As in Experiment 1, F values reflect between and within solutions to mixed ANOVA because all sphericity assumptions were met. Children in the TD and CON groups were included in this analysis. Independent analyses were conducted to reduce Type 2 error. These analyses included narrative score or productivity measures as the within-subjects factors and gender or ethnicity as the between-subjects factors. In keeping with results from Experiment 1, findings indicated no significant main effects for any of the gender and ethnicity analyses nor any interactions between type of score, gender, or ethnicity. For total story score, there were no significant effects for gender, $F(1, 55) = 2.802$, $p = .100$, $\eta_p^2 = .048$, or ethnicity, $F(3, 53) = 0.655$, $p = .584$, $\eta_p^2 = .036$. Similarly, there were no significant effects for gender, $F(3, 54) = 0.592$, $p = .992$, $\eta_p^2 = .031$, or ethnicity, $F(3, 53)$ $p = .624$, $\eta_p^2 = .031$, for the narrative productivity measures. Subsequent analyses were collapsed across gender and ethnicity variables.

Procedure

Design: A pretest-posttest, control group design was used to evaluate children's response to MLE. All children were told a story based on *Two Friends* for the pretest and *Bird and His Ring* for the posttest. Pretest and posttest stories were rated as described in Experiment 1, according to the scoring criteria in Appendix A. All examiners were unaware of group assignment.

MLE: Children in the TD and LI groups received two individual 30-min MLE sessions focusing on narrative skills and strategies. These sessions were conducted by speech-language pathologists or graduate students in communication sciences and disorders who were unaware of language ability. The MLE sessions were video-taped using a Sony Hi-8 recorder and audiotaped using a Marantz tape recorder for reliability purposes.

The general goal of the MLE sessions was to increase the length and complexity of children's stories. Examiners and children reviewed and discussed the child's version of the pretest story (*Two Friends*) during Session 1. First, the examiner read the child's story aloud. The child and the examiner discussed story components (setting: time and place, character information, and temporal order of events) and episode structure using examples from the child's story. In Session 2, examiners led children through the process of creating a story that corresponded with the wordless picture book, *One Frog Too Many* (Mayer & Mayer, 1975). The MLE scripts were designed to teach story components (e.g. setting, character information, temporal order of events, and causal relationships) and episode structure (e.g. initiating event, internal response, plan, attempt, consequence, and reaction/ending; see Appendix B for complete scripts for each MLE session). Scripting out the intervention standardized the clinician-child interactions. Further, the scripts were designed to provide comprehensive information about stories. The clinicians used puppets and pictures of backgrounds (e.g. mountains, forest) to demonstrate how to tell a complete story. The scripts were written to incorporate the five mediation strategies of intention to teach, mediation of meaning, mediation of transcendence,

mediation of planning, and mediation of transfer (Lidz, 1991). The sessions were somewhat flexible as the clinicians were encouraged to respond to each child's individual needs.

To begin each MLE session, the clinician explained the goal of the activity (intention to teach) and the purpose of the activity (mediation of meaning):

> Today we're going to talk about telling complete stories. When people tell stories they include a number of parts. They tell what the problem is, what the characters did, how they solve the problem, and how they feel about that. As you tell the story, let's talk about the characters, where the story takes place, and when it takes place.

The clinician continued the introduction, relating the storytelling activity to the children's home and school activities (mediation of transcendence):

> It's important to be able to tell good stories because you tell stories to your friends all the time, and you read and write stories in school. So learning to tell complete stories helps you do better in school.

The clinician then assisted the children with including the story components (setting, character information, causal relationships, and temporal order of events) and episode structure (initiating event, internal response, plan, attempt, consequence, and reaction/ending) into their stories using the books, puppets, and pictures of the background (mediation of planning):

> Stories need to tell us when and where something happened because that helps us understand the world the character lives in. So, what do we need to think about [when and where or setting]? [Use background sheet to illustrate setting, and then compare with book]. [Refer to p. 1 in *Two Friends*] How does this story start? [Pause, wait for response.] Where do you think they are? [Pause, wait for response.] What time do you think it is? [Pause, wait for response]. So, to say where and when, you could say ... [pause, let them fill in, if they don't, give example "one morning the dog and cat stood by the river"]. That tells us when and where.

To conclude the activity, the clinician reviewed the story components and episode structure and discussed changes observed in the children's ability to produce a complete story. Ways to use and practice their storytelling skills were summarized (mediation of transfer):

> Wow, that was good! So in this story you remembered to include [list what they included]. So always remember to talk about the setting, character information, order, and why things happen. What is the setting? [Let child fill in, assist him or her.] What is character information? [Let child fill in, assist her or him.] The order is what? [Let child fill in, assist him or her.] What are the reasons things happened? [Let child fill in, assist her or him.] This is important because it tells about the world the characters live in, the order of the story, and the reasons the characters did what they did.

Modifiability scores were derived for children who participated in the MLE sessions (based on Peña, 2000; Peña et al., 2001). At the conclusion of the second MLE session, clinicians made judgments of examiner effort, indicating how much effort and support were required during the MLE session (based on a 5-point Likert scale). High examiner effort received a score of 1, while low examiner efforts received a score of 5, according to criteria described in Peña et al. (2001). Clinicians also rated child responsivity on a 5-point Likert scale after the second MLE session. This rating indicated the child's level of responsiveness to learning during the two MLE sessions. High child responsivity was scored as a 5, while low child responsivity was scored as a 1. Responsivity and examiner effort scores at the conclusion of the second session were added together to yield a total modifiability score (range = 2-10).

Clinician training: Six graduate student clinicians conducted the mediation sessions. The clinicians were trained how to use the MLE script and how to rate child responsivity and

examiner effort. Training was completed through the use of videotaped examples of previous clinician's MLE sessions and MLE practice sessions by the clinicians themselves. The videotaped examples and practice sessions were critiqued by using the Mediated Learning Experience Rating Scale (Lidz, 1991). Further experience and practice were provided using the scripts and materials (e.g. books, background, and puppets) during group training sessions until the clinicians were familiar and confident with the procedure. During the study, the clinicians on the research team attended weekly meetings to discuss any questions, concerns, or difficulties in scoring or mediation that had occurred the previous week.

Fidelity of treatment: Six videotaped MLE sessions were randomly selected and evaluated to document the consistency of the implementation of MLE. The Mediated Learning Experience Rating Scale (Lidz, 1991) was used to rate the mediator's inclusion of each component of MLE (e.g., intentionality, transcendence, meaning, self-evaluation, transfer, and competence). Each component was rated on a 4-point scale (ranging from 0 to 3), for a possible total of 18. A rating of 3 indicates that the mediator included a statement of principle or a general rule. A rating of 2 indicates that the examiner consistently used the component of mediation and provided explanations and examples. A rating of 1 indicates that mediation was in evidence but was not elaborated for the child. A score of 0 indicates that the component was not used during the session. The mean total MLE score for the sessions was 17.25 indicating that all components of MLE were present consistently and that a general rule was provided frequently.

Results

This study compared the pretest and posttest performance of children in the LI, TD, and CON groups. A mixed within and between-participants design was used to compare the performance of the LI and TD children on the narrative tasks. Two separate repeated measures ANOVAs were computed that compared the three groups (LI, TD, and CON) on the total story scores and productivity measures derived from the two stories. For this set of analyses, the between-subjects factor was group (LI, TD, and CON), and the within-subjects factor was time (pretest and posttest). Discriminant function analyses were computed to explore the classification accuracy of children with and without language impairment on the basis of the scores derived from the dynamic assessment protocol. Descriptive statistics for all the dependent variables were calculated and are displayed in Table 4.

Narrative Ratings—The first set of analyses examined total story scores. First, total pretest and posttest narrative scores were compared. Next, narrative category scores (i.e., Story Components, Story Ideas and Language, and Episode Structure) were compared using repeated measures ANOVA to further explore the contribution of each of these scores to the total story score differences. For these analyses, time (pretest and posttest) was the within-subjects variable, and group (TD, LI, and CON) was the between-subjects variable. Post hoc analyses were conducted for significant effects using Scheffé's test for multiple comparisons.

For total story score, repeated measures ANOVA with time (pretest and posttest) as the within-subjects variable and group (TD, LI, and CON) as the between-subjects variable yielded significant main effects for time, $F(1, 68) = 32.942$, $p < .001$, $\eta_p^2 = .326$, and group, $F(2, 68) = 12.384$, $p < .001$, $\eta_p^2 = .267$, and a significant Time × Group interaction, $F(2, 68) = 4.418$, $p = .016$, $\eta_p^2 = .115$. These effects were small to moderate in size. Post hoc analyses using Scheffé's test for multiple comparisons demonstrated that pretest-posttest change was greater for the TD group than it was for the CON group (mean difference = 4.25, $p < .01$). Note that the gain for the TD group was more than 1 standard deviation above their pretest score.

The next three analyses examined the effects of MLE on three category scores that constituted the total story score. We were specifically interested in whether MLE sessions that targeted story components and episode structure resulted in changes in those specific areas. As before, repeated measures ANOVA with time (pretest and posttest) as the within-subjects variable and group (TD, LI, and CON) as the between-subjects variable were conducted for each of the narrative category scores (Story Components, Story Ideas and Language, and Episode Structure). Table 5 displays the ANOVA results for these follow-up analyses. Scheffé's comparisons for each of the narrative category scores are presented in Tables 6 and 7.

Scheffé's comparisons for mean pretest-to-posttest gain (see Table 6) suggest that children in the TD group made the most improvements on Story Components (mean difference = 3.15) in contrast to children in the CON group (mean difference = 0.57) and the LI group (mean difference = 1.36). The differential gains observed in the total score are explained mainly by change in performance on story components. While there were differences between the LI and TD groups (mean difference = 1.79) and for TD and CON groups (mean difference = 2.58), there were no differential effects of MLE on Story Ideas or Episode Structure (see Table 7).

Productivity Measures—One question of interest was whether the MLE sessions affected the productivity measures similarly across the three groups. A repeated measures mixed ANOVA was computed with productivity measures (MLU per C-unit, number of words, number of different words, number of clauses, and clauses per C-unit) as the dependent variables, time (pretest and posttest) as the within-subjects factor, and group (LI, TD, CON) as the between-subjects factor. Results for the repeated measures mixed ANOVA revealed significant main effects for time, $F(1, 68) = 13.189$, $p = .001$, $\eta_p^2 = .162$, and group, $F(2, 68) = 8.278$, $p = .001$, $\eta_p^2 = .196$, and a significant Time × Group × Measure interaction, $F(10, 340) = 2.225$, $p = .016$, $\eta_p^2 = .061$. Generally, children from all groups scored slightly higher at posttest compared with their pretest scores. Further, children in the LI group demonstrated significantly lower pretest-to-posttest change in comparison with children in the TD group ($p < .001$) and the CON group ($p = .013$; see Tables 6 and 7).

Scheffé's tests for multiple comparisons were used to explore differential performance of the three groups for each of the productivity measures. Generally, performance patterns indicated that the TD group showed a greater pre- to posttest gain than the LI or CON groups. Table 6 contains the mean group pretest-to-posttest change, and Table 7 displays the mean gain differences between groups.

Overall, children in the TD group demonstrated higher scores than the children in the LI group on the amount of talk (e.g., words, C-units, clauses). They also demonstrated higher scores than the CON group with respect to proportion of talk (e.g., MLU for words, clauses per C-unit). TD children made their stories more complex by increasing story length as well as by increasing the amount of information included in an utterance. In comparison, the gains the LI group made after MLE were similar to those the CON group made with no intervention. Comparing the amount of change, children in the TD group generally demonstrated gains of 1/2 to above 1 standard deviation based on their pretest scores, while children in the LI and CON groups made more modest gains (see Table 7).

Classification Analysis—An important question is whether the results of dynamic assessment can help clinicians diagnose language ability with greater precision and reliability. We examined the classification accuracy of each measure at pretest independently and in combination to yield the smallest number of measures that provided the best classification accuracy and conducted a bias analysis. For modifiability and the posttest measures, we again combined different sets of predictor variables in an iterative manner to yield the smallest number of measures that provided the best classification accuracy. Pretest measures were

considered to be static measures, because they were derived before the MLE sessions. Modifiability and posttest measures were considered to be dynamic measures because they represented performance during or after MLE.

Classification accuracy of pretest measures: The pretest measures generally had low rates of correct classification (see Table 8). A cutoff score of 1 standard deviation below the TD group mean was used to classify the groups. We selected this cutoff score based on inspection of the differences in the TD and LI means. This cutoff score is consistent with other reports of performance differences between children with and without LI (Spaulding et al., 2006;Tomblin, Records, & Zhang, 1996). On average, sensitivity was 26% with a range from 0% (number of C-units) to 71% (total score and Story components). Average specificity was 88%, with a range from 74% (total score, Story Components, and Episode Structure) to 100% (number of C-units). Generally, high specificity came at a cost of low sensitivity with number of C-units being the most extreme example (100% specificity and 0% sensitivity). Discriminant function analysis demonstrated that for the pretest measures, a combination of Story Components, Story Ideas and Language, and Episode Structure together yielded a specificity of 70.4% and sensitivity of 78.6%. These are marginally acceptable classification rates.

Likelihood ratios use specificity and sensitivity calculations to determine the increased odds of having the condition under study (e.g., LI) when test results are positive (Table 8). Positive likelihood ratios ranged from 0.00 to 5.18 for the pretest measures, with a mean of 1.58. The positive likelihood ratio is multiplied by the base rate, in the present case, 7% incidence of language impairment in the population (Tomblin et al., 1997), to yield a positive predictive value. The positive predictive value is an estimate of the accuracy of the diagnosis of LI if a child scores below the cut score. Positive predictive values for the pretest measures ranged from 0% to 28%, with a mean of 10%. The negative likelihood ratio is multiplied by (1 - the base rate) to yield a negative predictive value. The negative predictive value is an estimate of the accuracy of concluding that a child does not have impairment if he or she scores above the cutoff score. The negative predictive values ranged from 93% to 97%. Together, these results indicate that the narrative measures at pretest were not sensitive to LI while accuracy of negative findings of LI was high.

Recall that there were no racial/ethnic group differences on the total story scores, three category scores, and productivity measures at pretest. But there was a high rate of misclassification. In this analysis, we wanted to know whether misclassification of individual children on the pretest measures was related to racial/ethnic group (Table 9). We began by calculating the means and standard deviations for each measure at pretest for typical children who told the *Two Friends* story first (CON and TD groups) and children in the LI group. Generally, the LI group scored between 0.31 (number of C-units) to 1.18 (total story score) standard deviations (*Z* scores) below the TD mean across the 10 measures. These findings are consistent with other work comparing TD and LI groups on language sample measures (Dunn, Flax, Sliwinski, & Aram, 1996;Hewitt, Hammer, Yont, & Tomblin, 2005;Norbury & Bishop, 2003) and on standardized test instruments (Peña, Spaulding, & Plante, 2006;Spaulding et al., 2006). A cutoff score of 1 standard deviation below the TD mean is therefore well within the LI range. Based on this cutoff score, we calculated the percentage of typical children from each racial/ethnic group who scored at or above the cutoff score (correctly classified cases) and those who scored below the cutoff score (misclassified cases). On 7 of the 10 pretest measures, 90% or more of the European American children scored within the normal range (-1 standard deviation from the mean and above). Eighty percent or more of the European American children scored within the normal range on all but 1 of the 10 measures.

These pretest results differ for the African American and Latino American children. An unacceptably high proportion of African American children in the TD and CON groups scored

more than -1 standard deviation from the mean. There were only 3 out of 10 pretest measures for which 80% of the African American children scored within the normal range. Results for Latino American children were similar. There were only 2 of 10 pretest measures in which 90% or more of the Latino American children scored within the normal range. Eighty percent or more of the Latino American children scored within the normal range on 7 of the 10 pretest measures. In practical terms, these results suggest classification bias at pretest. On a case-by-case basis, African American and Latino American children were much more likely than European American children to fall below the normal range on the pretest measures of narrative language.

Classification analysis of posttest and dynamic measures: At posttest, the 10 measures demonstrated higher sensitivity values ($M = 64\%$) in comparison with the pretest measures ($M = 26\%$; see Table 8). Overall, posttest specificity ($M = 83\%$) was slightly lower than at pretest ($M = 88\%$). Discriminant analysis demonstrated that the combination of Story Components and Episode Structure yielded the best classification (specificity = 85.2%, sensitivity = 78.6%). This classification rate is in the "fair" range (Plante & Vance, 1994). The positive likelihood ratios ranged from 2.27 and 5.82 with a mean of 3.88. The current findings indicate better positive and negative prediction values for the posttest results. The positive predictive values for the posttest ranged from 15% to 30% with a mean of 22%. The posttest sensitivity was just over two times higher than the pretest sensitivity. As expected, the negative predictive values (how likely you are to be normal if you score above the cutoff score) were high for both pretest and posttest measures.

By itself, the modifiability score was the most accurate measure, with 93% sensitivity and 82% specificity. Children with typical development had higher modifiability scores ($M = 7.93$, $SD = 1.492$) than children with LI ($M = 3.40$, $SD = 1.24$). The positive likelihood ratio for the modifiability score was 5.17. Moving the cutoff score to the mean of the centroids (5.66) increased sensitivity and specificity to 93% and 96%, respectively. The positive likelihood ratio for this cutoff score was 23.25, with a positive predictive value of 64%.

We combined the modifiability score with the posttest scores to see if these combinations led to correct classification. As before, classification values were calculated in an iterative fashion to obtain the fewest variables that yielded the best classification rates. Results indicate that the combination of modifiability, number of different words, total number of words, and the Story Components score yielded 100% correct classification. Analysis of narratives produced after mediated learning with clinician judgments of response to instruction resulted in excellent classification with no bias.

Discussion

The results of the current study are consistent with those of earlier studies that examined the effects of MLE on word learning (Peña, Iglesias, & Lidz, 2001; Peña, Quinn, & Iglesias, 1992). As in these previous studies, children with and without LI performed differently on measures of language performance after mediation. In the current study, children told more complete and complex stories after they received two mediation sessions that focused on general narration skills related to episode structure and complexity, character development, dialogue, and ways to express temporal and causal relationships between events. More important, the benefits of mediation differed for children with and without LI. Typical children had greater pre- to posttest gains and earned higher modifiability ratings during MLE in comparison with children with LI.

The pretest measures of narration did not yield accurate classification of children with and without LI. Depending on the pretest measure of narration, either the false-positive rate or the false-negative rate was unacceptably high. The posttest measures of narration that occurred

after two mediation sessions were much more accurate and stable. Narrative analysis appears to be a better assessment tool after intervention than it is before intervention.

Similar to results reported by Sternberg et al. (2002), correlations between pretest and posttest (total narrative score) were higher for children in the control group ($r = .759$) than children in the MLE group ($r = .483$). For children who received MLE, the pretest score predicted posttest scores only fairly well. However, the pretest and posttest scores remained fairly stable for those children who did not receive MLE. These results illustrate the notion that language performance in low-performing children with typical development can be changed with even short-term instruction. A diagnosis of LI based on single-time use of static test measures may not be accurate for all children, particularly those whose experiences on tasks such as storytelling vary from mainstream expectations. Such children may perform below the normal range even though they have normal language learning abilities (i.e., language difference). Dynamic assessment follow-up with children who perform below expectations can help clinicians differentiate between children's language difference and LI.

The strongest and best predictor of language ability was the clinician's rating of modifiability, which was a clinical judgment of the extent of the examiner's teaching effort and the child's responsiveness to instruction. A goal of dynamic assessment is to examine response to instruction by focusing on use of cognitive tools as an indication of underlying ability (Jensen & Feuerstein, 1987). It is likely that underlying cognitive skills such as attention to task, working memory, problem solving, and flexibility were evident during one-to-one teaching and were interpreted as child responsivity (Lidz, 1991; Peña, 2000). This perspective is consistent with research that implicates information processing deficits as markers of LI (Bishop, 1992; Gillam, Cowan, & Marler, 1998; Hoffman & Gillam, 2004).

Dynamic assessment is somewhat time-consuming in comparison with some other types of assessments. But if information gained from dynamic assessment helps clinicians make reliable diagnostic decisions that accurately differentiate between language differences and LI, the practical advantages would clearly outweigh the disadvantages. As noted by Laing and Kamhi (2003), the process of dynamic assessment provides practical information about real-time language comprehension and use in a functional learning context during the teaching phase of dynamic assessment (MLE). Clinicians can observe how children problem solve, respond to feedback, and persist in attending to the tasks. Such clinical information is useful for developing intervention strategies. The incorporation of narrative assessment within the dynamic assessment paradigm provides a great deal of clinically relevant information from an investment of about 1 hr. Very often, standardized tests yielding only a single score (or profile of scores) have little clinical utility from a similar investment of time.

Clinicians are often faced with making diagnostic decisions about children whose low scores on standardized tests may reflect cultural, experiential, and/or linguistic differences. Because a static approach to assessment evaluates performance at one point in time, low performance due to language differences, fatigue, or other factors can be misinterpreted as LI. Additionally, information about a child's thought processes, emerging skills, or learning potential cannot be inferred from such one-time evaluations (Olswang & Bain, 1996). Dynamic assessment, with its emphasis on the learning process, provides a reliable means for differentiating between language difference and LI. Overall, the results of this study support the theoretical constructs of dynamic assessment and show that dynamic assessment of narratives provides a clinically applicable, culturally fair way of testing the language of children who are suspected of LI.

Appendix A. Adapted scoring criteria for stories

Item	Score	Scoring rule
Complexity of Ideas	5	(Same as in manual)
	4	The midpoint between *some complexity* and *complex (abstract)* if a child produced one additional detail beyond what was shown on the page
	3	*Some complexity* would be marked if a child produced one or two nonliteral ideas in his or her story
	2	The midpoint between *simple (concrete)* and *some complexity* would be marked if a child used a listing or description of each panel
	1	*Simple (concrete):* Children told a story that did not contain any literal ideas
Grammatical Complexity	5	*Complex sentences:* Indicates two or more examples of complex sentences
	4	(Same as in manual)
	3	(Same as in manual)
	2	(Same as in manual)
	1	(Same as in manual)
Creativity	5	(Same as in manual)
	4	The midpoint between *uninteresting* and *somewhat captivating* indicates production of at least 1 creative element (e.g., humor, irony, suspense, metaphors, and surprises)
	3	*Somewhat captivating* indicates production of 2-3 creative elements
	2	The midpoint between *somewhat captivating* and *interesting and captivating* indicates production of 4-5 creative elements
	1	*Interesting and captivating* indicates production of more than 5 creative elements

Note. See original for additional scoring criteria. From *Dynamic Assessment and Intervention: Improving Children's Narrative Skills*, by L. Miller, R. B. Gillam, and E. D. Peña, 2001, Austin, TX: Pro-Ed. Copyright 2001 by Pro-Ed. Adapted with permission.

Appendix B (p. 1 of 3). MLE scripts

Mediation 1 uses *Two Friends*; Mediation 2 uses *One Frog Too Many*.

Mediation 1

[Begin by showing the child the *Two Friends* book.] Remember when I showed you this book? You said...[read back the story they told]. Was that a good story? Why or why not? [Tell them what parts of the episode they included/excluded in a way they can understand.]

Today we're going to talk about telling complete stories. When people tell stories they include a number of parts. They tell what the problem is, what the characters did, how they solve the problem, and how they feel about that. As you tell the story, let's talk about the characters, where the story takes place, and when it takes place.

It's important to be able to tell good stories because children tell each other stories all the time, and you read and write stories in school. So, learning to tell complete stories helps you communicate better and do better in school. Now, why is it important to tell better stories? [Help child to explain that stories are important for school and for communication.]

First, let's talk about the different parts that need to be in a story. Story tellers start their stories by telling when and where something happened. That helps us understand the world the character lives in. So, what do we need to think about when we start a story [when & where or setting]?

[Refer to p. 1 in *Two Friends*] How does this story start? [pause, wait for response, help child to respond when needed] Where do you think they are? [pause, wait for response]. What time

to you think it is? [pause, wait for response]. How would you start a story in a way that tells where and when the story takes place? [Pause, let them fill in, if they don't, give an example such as "one morning the dog & cat stood by the river" that tells us when and where.]

We also need to know about the characters. Good story tellers tell listeners about who the characters are and what they're like. We also need to include what? [character information]. Let's think about the characters? What do they look like? [pause, wait for response] Do the dog and the cat have names? [pause, wait for response] You could say, Bill the dog and Sally the cat were talking about what they were going to do that day. You can also tell what they look like or think of names that describe them. For example, I could say, Triangles—and who would that describe? Yes, the cat, Triangles the cat was thinking about...[can additionally use toys or puppets to name/describe].

In stories, we also want to talk about what happened first, second, and last, and why things happened (order & causal relationships). This is important because it helps us understand the order of the story, and the reasons the characters (people) did what they did.

What would happen if you told the story backwards or out of order? [Help child state that it would be hard to know what happened when; or that it would be hard to know why it happened.] At the beginning of the story, first, they were...[help child to describe], then [turn page] [help child describe]. We use words like first, next, and then, to describe what happened and why it happened (order & causal relationships) [using puppets, let child act out the story and explain the order and causal relationships].

Let's tell a story that includes all these pieces. [Help child tell story with setting, time, place, characters, temporal order and causality] [Wow, that was good] [Example: Triangles the cat and Rex the dog were standing by the river talking. While they were talking, Rex fell asleep. So, Triangles left because she had no one to talk to.] In this story you remembered to include... [list what they included].

Always remember to talk about the setting (when & where), character information, order (temporal) and causal relationships. In the story, what is the setting? [let child fill in, assist them], what should we say about the characters? [let child fill in, assist them]; what happened first? Then what? Then what? [let child fill in, assist them] and why did the cat leave? Why did the dog look for the cat? [let child fill in, assist them]. It's important to include these things because they tell us about the world the characters live in (setting), the order of the story (order), and the reasons the characters did what they did (causal relationships).

Tell me what these are again. Character information [let child respond, assist if necessary]; setting (when & where) [let child respond, assist if necessary]; order (temporal) [let child respond, assist if necessary]; & causal relationships [let child respond, assist if necessary].

Next, we're going to talk about telling complete stories. When people tell stories they include a beginning, a middle, and an end. They tell what the problem is at the beginning and how the characters feel about the problem. For the middle, they talk about the actions the characters take to solve the problem. At the end, they talk about how the characters eventually solved the problem, and how they feel once the problem is solved.

Do you remember why stories are important [expand on what child says, e.g., It's important to be able to tell good stories because children tell each other stories all the time, and you read and write stories in school. So, learning to tell complete stories helps you communicate better and do better in school.]

Let's talk about the different parts that need to be in a story. When people tell stories, they need to know what happened to start the action in the story. This is called problem. What do we need to include? [Problem] [refer to p.1 in *Two Friends*]. How does this story start? [child answers] [turn the page].What do you think caused the problem? [let them fill in]. To include the problem, you would say...[pause, let them fill it in, if they don't give example "One morning, the cat and the dog were talking and the dog fell asleep, that tells us what started the problem] [reflect back what they said—use expansion/extension as needed] [let child act out with puppets.]

After the problem we talk about how the character feels about it. That is important because it makes the story interesting and helps us understand why they did things. What happens to the dog on this page (page 2)? Yes, he falls asleep. That's the problem. Over here (p. 3), how does the cat feel about the dog falling asleep? Ok, so what are we calling them? [then continue using the names selected]. Right, Sally is sad because Bill fell asleep. Why do you think she felt sad? [Yes, Sally feels sad because Bill didn't want to talk to her any more.] What do you think she said to Bill? [Hey Bill, wake up and talk to me.] But, did he wake up? [No, even though Sally tried to wake him up, he didn't.] So (p. 4) what does Sally do? [wait for child response]. Yes. When Bill wouldn't wake up, Sally decided to leave. You need to include how characters feel about what happened.

After we talk about how characters feel, we talk about how they try to solve the problem. We also need to include what [The attempts]. [refer to p. 8 in *Two Friends*] What does the dog do? [pause, let them fill it in, if they don't give example "He asks the animals if they have seen the cat."]

After we talk about what the character does, we need to tell how the problem was solved. What happened after the dog looked for the cat? [child responds that the dog found the cat]. That's right; what was the problem? [the cat was gone], and how did the dog solve the problem? [he looked for the cat and he found it].

After talking about how the problem was solved, story tellers can tell how the character feels about it or their reaction. How did the story end? [pause, let them fill it in, if they don't give example "The dog and the cat became friends again."] Do you think they were happy? How would you feel, why?

Stories include problems, the way people feel about them, what they do to try to solve them, what happens, and how they feel at the end. Why is this important? [This is important because it helps your friends understand your story and helps you do better in school.] I want you to tell me the story of the two Friends again [let child use puppets to tell the story if they choose].

How are you going to remember to tell a complete story with all the different parts? [discuss strategies to include specific components of story and a complete episode]

Mediation 2

Show the child the *One Frog Too Many* book. Remember the *Two Friends* story that you told? Today we're going to use this book called *One Frog Too Many* to keep talking about telling stories.

Today we're going to talk about telling stories again. Remember, last time we talked about what the problem is, how the characters feel about the problem, what the characters did, how they solve the problem, and how they feel once the problem is solved. As you tell the story, include information about the characters, where the story takes place, and when it takes place. Do you remember why it's important to tell complete stories? [It's important to be able to tell

good stories because children tell each other stories all the time, and you read and write stories in school. Learning to tell complete stories helps you communicate better and do better in school.]

First, let's talk about the different parts that need to be in a story. Stories need to tell us when and where something happened because it helps the listener understand about the world the characters live in. What do we need to think about [when & where/setting] [refer to p. 1 in *One Frog Too Many*] How does this story start? [pause, wait for response] Where do you think they are? [pause, wait for response] What time do you think it is? [pause, wait for response] To say where and when, you could say...[pause, let them fill in, if they don't, give example "one morning the boy received a present at his house" "that tells us when and where (setting)"] [use background sheet to discuss setting and compare to book]

We also need to know about the characters. Character information tells the listener about who they are and what they're like. We also need to include what? [character information] Let's think about the characters. What do they look like? Do the boy, the frog, the dog, and the turtle have names? [pause, wait for response] You could say, Pete the boy got a present. Bosco the dog, Timmy the turtle, and Benny the frog were watching him.... You can also tell what they look like or think of names that describe them. For example, I could say Greeny-and who would that describe? Yes, the frog. Greeny the frog was watching the boy [use puppets, let child name and describe them].

In stories, we also want to talk about what happened first, second, and last, and why things happened (Order). This is important because it helps the people understand the order of the story, and the reasons the characters (people) did what they did (Causal relationships).

What would happen if you told the story backwards or out of order? [help child say something like, It would be hard to know what happened when or why it happened.] At the beginning of the story, first, they were...[help child describe], then [help child describe]. We use words like first, next, and then, to describe what happened and why it happened. [encourage child to use puppets to act out the story, discussing order and causality]

Let's tell a story that includes all these pieces. [help child tell story with setting, time, place, characters, temporal order and causality] [example: One morning, Pete the boy found a present at his house. All his animal friends watched while he opened his present. First, Pete opened the present and found a baby frog. Benny the big frog was very jealous and bit the baby frog. Then Benny got into trouble because Pete was very mad that he hurt the baby frog.] [Let child use puppets to act out the story.]

Always remember to talk about the setting, time, and characters, to use temporal words like first, second and last, and also to talk about why things happen. This is important because it tells about the world the characters live in, the order of the story, and the reasons the characters did what they did.

Now we're going to talk about stories and parts of stories. Remember, a story has a beginning, middle, and end. And we're going to talk about how those parts work together.

[Show the child the *One Frog Too Many book*.] We're going to talk about telling complete stories. When people tell stories they include the beginning, middle, and end. They tell what the problem is at the beginning and how the characters feel about the problem. For the middle, they talk about the actions the characters take to solve the problem. At the end, they talk about how the characters solved the problem, and how they feel when the problem is solved.

Why is it important to tell complete stories? [It's important to be able to tell complete stories because children tell each other stories all the time, and you read and write stories in school. Learning to tell complete stories helps you communicate better and helps you do better in school.]

First, let's review the different parts that need to be in a story. When people tell stories, they need to know what happened to start the action in the story. Do you remember what this is called? [assist child in responding: This is called a problem.] So what do we need to include? [Problem/IE] [refer to p.1 in *One Frog Too Many*] How does this story start? [pause, wait for response] What do you think caused the problem (IE)? So, to include the problem you would say...[pause, let them fill it in, if they don't give example "One morning, the boy received a present. He opened it and found a new baby frog. That tells us what started the problem. [let child act out the initiating event, or examiner can demonstrate]

[Refer to p. 6] So what happens to the baby frog on this page? [pause, wait for response] Yes he was bitten. Over here how does the boy feel? [pause, wait for response] Ok, so what are we calling him? [then continue using the name selected] Right, Tim is mad. Why do you think he felt mad? Yes, Tim was mad because Ronnie bit Teeny's leg. What do you think he said to Ronnie? "Hey Ronnie, you aren't nice." And did he behave? No, he kicked the little frog off the boat. You need to include how characters feel about what happened.

After we talk about how they feel, we talk about what they do to try to solve the problem. [act out or let child act out using the puppets] What do the characters do here to solve the problem? [pause, let them fill it in, if they don't give example "The boy, the dog, the turtle, and the big frog all tried to find the little frog."] [act out or let child act out the attempts using the puppets]

After we talk about what they do, then we talk how the problem was solved. How was the problem solved? [pause, let them fill it in, if they don't give example "The boy and the animals went home and waited. The frog jumped in the window."] Right that's how the problem was solved. [use puppets to act out or let child act out describing the solution.]

After talking about how the problem was solved, we need to talk about how the character feels about it, or their reaction. How did the story end? [pause, let them fill it in, if they don't give example "The big frog and the little frog became friends.] Do you think they were happy? [pause, wait for response] How would you feel, why? [pause, wait for response] [use puppets to demonstrate and discuss reaction]

Stories include what the problem is, how the characters feel about the problem, the actions the characters take, how they eventually solve the problem, and how they feel once the problem is solved. This is important because it helps your friends understand your story and helps you do better in school.

Tell me this story remembering all these parts [child tells story—assist in helping child tell a story with a complete episode if necessary, use puppets or let child use puppets to tell the story if they wish] Tell me what the important parts of a story are [have child tell examiner in his/her own words—assist if necessary].

How are you going to remember to tell complete stories with all the parts? [discuss strategies to include specific components of story]

References

Allen, M.; Yen, W. Introduction to measurement theory. Wadsworth; Belmont, CA: 1979.

Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. Psychological Review 1998;105:158–173. [PubMed: 9450375]

Bain B, Olswang L. Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. American Journal of Speech-Language Pathology 1995;4:81–92.

Berman, RA.; Slobin, DI. Relating events in narrative: A crosslinguistic developmental study. Erlbaum; Hillsdale, NJ: 1994.

Bishop DVM. The underlying nature of specific language impairment. Journal of Child Psychology and Psychiatry 1992;33:3–66. [PubMed: 1737831]

Bishop DVM, North T, Donlan C. Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. Journal of Child Psychology and Psychiatry 1996;37:391–403. [PubMed: 8735439]

Bracken BA. Ten psychometric reasons why similar tests produce dissimilar results. Journal of School Psychology 1988;26:155–166.

Campbell T, Dollaghan C, Needleman H, Janosky J. Reducing bias in language assessment: Processing-dependent measures. Journal of Speech and Hearing Research 1997;40:519–525.

Carrow-Woolfolk, E. Comprehensive assessment of spoken language. AGS.; Circle Pines, MN: 1999.

Conti-Ramsden G. Processing and linguistic markers in young children with specific language impairment (SLI). Journal of Speech, Language, and Hearing Research 2003;46:1029–1037.

Davies P, Shanks B, Davies K. Improving narrative skills in young children with delayed language development. Educational Review 2004;56:271–286.

Demsky YI, Mittenberg W, Quintar B, Katell AD, Golden CJ. Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale—Revised. Assessment 1998;5:115–121. [PubMed: 9626387]

DeVellis, R. Scale development: Theory and applications. 26. Sage; Newbury Park, CA: 1991.

Dollaghan C, Campbell TF. Nonword repetition and child language impairment. Journal of Speech, Language, and Hearing Research 1998;41:1136–1146.

Dunn M, Flax J, Sliwinski M, Aram D. The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. Journal of Speech and Hearing Research 1996;39:643–654. [PubMed: 8783141]

Ellis Weismer S, Tomblin JB, Zhang X, Buckwalter P, Chynoweth JG, Jones M. Nonword repetition performance in school-age children with and without language impairment. Journal of Speech, Language, and Hearing Research 2000;43:865–878.

Feuerstein, R. The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. University Park Press; Baltimore: 1979.

Feuerstein R, Miller R, Rand Y, Jensen M. Can evolving techniques better measure cognitive change? Journal of Special Education 1981;15:201–219.

Gardner, MF. Expressive One-Word Picture Vocabulary Test. Academic Therapy Publications; Novato, CA: 1979.

Gathercole SE, Baddeley AD. The role of phonological memory in vocabulary acquisition: A study of young children learning new names. British Journal of Psychology 1990;81:439–454.

Geist E, Aldridge J. The developmental progression of children's oral story inventions. Journal of Instructional Psychology 2002;29:33–39.

Gillam RB, Cowan N, Marler JA. Information processing by school-age children with specific language impairment: Evidence from a modality effect paradigm. Journal of Speech, Language, and Hearing Research 1998;41:913–926.

Gillam RB, McFadden TU. Redefining assessment as a holistic discovery process. Journal of Childhood Communication Disorders 1994;16:36–40.

Gillam, RB.; McFadden, T.; van Kleeck, A. Improving the narrative abilities of children with language disorders: Whole language and language skills approaches. In: Fey, M.; Windsor, J.; Reichle, J., editors. Communication intervention for school-age children. Brookes; Baltimore: 1995. p. 145-182.

Gillam RB, Peña ED, Miller L. Dynamic assessment of narrative and expository discourse. Topics in Language Disorders 1999;20(1):33–47.

Gutierrez-Clellen VF, Peña E. Dynamic assessment of diverse children: A tutorial. Language, Speech, and Hearing Services in Schools 2001;32:212–224.

Gutierrez-Clellen VF, Peña ED, Quinn R. Accommodating cultural differences in narrative style: A multicultural perspective. Topics in Language Disorders 1995;15(4):54–67.

Gutierrez-Clellen VF, Quinn R. Assessing narratives in diverse cultural/linguistic populations: Clinical implications. Language, Speech, and Hearing Services in Schools 1993;24:2–9.

Hayward D, Schneider P. Effectiveness of teaching story grammar knowledge to preschool children with language impairment: An exploratory study. Child Language Teaching and Therapy 2000;16:255.

Haywood, HC.; Tzuriel, D. Interactive assessment. Springer-Verlag; New York: 1992.

Haywood HC, Wingenfeld SA. Interactive assessment as a research tool. Journal of Special Education 1992;26:253–268.

Hewitt LE, Hammer CS, Yont KM, Tomblin JB. Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. Journal of Communication Disorders 2005;38:197. [PubMed: 15748724]

Hoffman LM, Gillam RB. Verbal and spatial information processing constraints in children with specific language impairment. Journal of Speech, Language, and Hearing Research 2004;47:114–125.

Iglesias, A. Cultural conflict in the classroom: The communicatively different child. In: Ripich, DN.; Spinelli, FM., editors. School discourse problems. College-Hill; San Diego, CA: 1985. p. 79-96.

Iglesias, A.; Gutierrez-Clellen, V. The cultural linguistic minority student. In: Yoder, D.; Kent, R., editors. Decision making in speech-language pathology. B. C. Decker; Hamilton, Ontario, Canada: 1988. p. 119-120.

Jensen, M.; Feuerstein, R. The learning potential assessment device: From theory to practice. In: Lidz, CS., editor. Dynamic assessment: An interactional approach to evaluating learning potential. Guilford Press; New York: 1987. p. 379-402.

Jordan GE, Snow CE, Porche MV. Project ease: The effect of a family literacy project on kindergarten students' early literacy skills. Reading Research Quarterly 2000;35:524–546.

Kozulin A. Sociocultural theory and the mediated learning experience. School Psychology International 2002;23:7–35.

Laing SP, Kamhi A. Alternative assessment of language and literacy in culturally and linguistically diverse populations. Language, Speech, and Hearing Services in Schools 2003;34:44–55.

Lidz, CS. Dynamic assessment: An interactional approach to evaluating learning potential. Guilford Press; New York: 1987.

Lidz, CS. Practitioner's guide to dynamic assessment. Guilford Press; New York: 1991.

Lidz CS. Dynamic assessment and the legacy of L.S. Vygotsky. School Psychology International 1996;16:143–154.

Lidz CS. Dynamic assessment: Psychoeducational assessment with cultural sensitivity. Journal of Social Distress and the Homeless 1997;6:95–111.

Lidz CS. Mediated learning experience (MLE) as a basis for an alternative approach to assessment. School Psychology International 2002;23:68–84.

Lidz CS, Peña ED. Dynamic assessment: The model, its relevance as a non-biased approach and its application to Latino American preschool children. Language, Speech, and Hearing Services in Schools 1996;27:367–372.

Lidz, CS.; Thomas, S. The preschool learning assessment device: Extension of a static approach. In: Lidz, CS., editor. Dynamic assessment: An interactional approach to evaluating learning potential. Guilford Press; New York: 1987. p. 288-326.

Loban, W. Language development: Kindergarten through grade twelve. National Council of Teachers of English; Urbana IL: 1976.

Mayer, M.; Mayer, M. One frog too many. Penguin Books; New York: 1975.

McCartney E, Boyle J, Bannatyne S, Jessiman E, Campbell C, Kelsey C, et al. Becoming a manual occupation? The construction of a therapy manual for use with language impaired children in mainstream primaryschools. International Journal of Language and Communication Disorders 2004;39:135–148. [PubMed: 14660190]

McCauley R, Swisher L. Psychometric review of language and articulation tests for preschool children. Journal of Speech and Hearing Disorders 1984a;49:34–42. [PubMed: 6700200]

McCauley R, Swisher L. Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. Journal of Speech and Hearing Disorders 1984b;49:338–348. [PubMed: 6389982]

McFadden TU. The immediate effects of picto-graphic representation on children's narratives. Child Language Teaching and Therapy 1998;14:51–67.

Miller J, Chapman RS. Using microcomputers to advance research in language disorders. Theory Into Practice 1983;22:301–307.

Miller, J.; Chapman, RS. SALT for Windows—Research version 7.0. University of Wisconsin—Madison, Waisman Center, Language Analysis Laboratory; Madison: 2002.

Miller, L. Bird and his ring. Neon Rose Productions; Austin, TX: 2000a.

Miller, L. Two friends. Neon Rose Productions; Austin, TX: 2000b.

Miller, L.; Gillam, RB.; Peña, ED. Dynamic assessment and intervention: Improving children's narrative skills. Pro-Ed.; Austin, TX: 2001.

Newcomer, P.; Hammill, D. Test of Language Development—Primary. Third Edition. Pro-Ed.; Austin, TX: 1997.

Norbury CF, Bishop DVM. Narrative skills of children with communication impairments. International Journal of Language and Communication Disorders 2003;38:287–313. [PubMed: 12851080]

Olswang L, Bain B. Assessment information for predicting upcoming change in language production. Journal of Speech and Hearing Research 1996;39:414–423. [PubMed: 8729929]

Patterson, S.; Gillam, RB. Team collaboration in the evaluation of language in students above the primary grades. In: Tibbits, D., editor. Language intervention: Beyond the primary grades. Pro-Ed.; Austin, TX: 1995. p. 137-181.

Peña ED. Measurement of modifiability in children from culturally and linguistically diverse backgrounds. Communication Disorders Quarterly 2000;21(2):87–97.

Peña, ED.; Gillam, RB. Dynamic assessment of children referred for speech and language evaluations. In: Lidz, CS.; Elliott, J., editors. Dynamic assessment: Prevailing models and applications. 6. Elsevier Science; Oxford, England: 2000.

Peña, E.; Gillam, R.; Miller, L. Dynamic Assessment of Narratives: Revised Scripts. 2003. Retrieved from http://www.neonrose.net/dynamicassessment/script.pdf

Peña ED, Iglesias A, Lidz CS. Reducing test bias through dynamic assessment of children's word learning ability. American Journal of Speech-Language Pathology 2001;10:138–154.

Peña ED, Quinn R. Task familiarity: Effects on the test performance of Puerto Rican and African American children. Language, Speech, and Hearing Services in Schools 1997;28:323–332.

Peña ED, Quinn R, Iglesias A. The application of dynamic methods to language assessment: A nonbiased procedure. Journal of Special Education 1992;26:269–280.

Peña ED, Spaulding TJ, Plante E. The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. American Journal of Speech-Language Pathology 2006;15:247–254. [PubMed: 16896174]

Peterson C, Jesso B, McCabe A. Encouraging narratives in preschoolers: An intervention study. Journal of Child Language 1999;26:49–67. [PubMed: 10217889]

Plante E, Vance R. Selection of preschool language tests: A data-based approach. Language, Speech, and Hearing Services in Schools 1994;25:15–24.

Riding RJ, Tite HC. The use of computer graphics to facilitate story telling in young children. Educational Studies 1985;11:203–210.

Rodekohr RK, Haynes WO. Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. Journal of Communication Disorders 2001;34:255–272. [PubMed: 11409607]

Scheffner-Hammer C, Pennock-Roman M, Rzasa S, Tomblin JB. An analysis of the Test of Language Development—Primary for item bias. American Journal of Speech-Language Pathology 2002;11:274–284.

Schoenbrodt L, Kerins M, Gesell J. Using narrative language intervention as a tool to increase communicative competence in Spanish-speaking children. Language, Culture and Curriculum 2003;16:48–59.

Spaulding TJ, Plante E, Farinella KA. Eligibility criteria for language impairment: Is the low end of normal always appropriate? Language, Speech, and Hearing Services in Schools 2006;37:61–72.

Sternberg RJ. Raising the achievement of all students: Teaching for successful intelligence. Educational Psychology Review 2002;14:383–393.

Sternberg RJ, Grigorenko EL. Difference scores in the identification of children with learning disabilities: It's time to use a different method. Journal of School Psychology 2002a;40:65–83.

Sternberg, RJ.; Grigorenko, EL. Dynamic testing: The nature and measurement of learning potential. Cambridge University Press; New York: 2002b.

Sternberg RJ, Grigorenko EL, Ngorosho D, Tantufuye E, Mbise A, Nokes C, et al. Assessing intellectual potential in rural Tanzanian school children. Intelligence 2002;30:141–162.

Stiegler LN, Hoffman PR. Discourse-based intervention for word finding in children. Journal of Communication Disorders 2001;34:277–303. [PubMed: 11508896]

Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. Journal of Speech and Hearing Research 1997;40:1245–1260.

Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. Journal of Speech and Hearing Research 1996;39:1284–1294. [PubMed: 8959613]

Tzuriel D. Dynamic assessment of young children: Educational and intervention perspectives. Educational Psychology Review 2000;12:385–435.

Ukrainetz TA. Stickwriting stories: A quick and easy narrative representation strategy. Language, Speech, and Hearing Services in Schools 1998;29:197–207.

Ukrainetz TA, Harpell S, Walsh C, Coyle C. A preliminary investigation of dynamic assessment with Native American kindergartners. Language, Speech, and Hearing Services in Schools 2000;31:142–154.

Valencia R, Rankin RJ. Evidence of content bias on the McCarthy scales with Mexican American children: Implications for test translation and nonbiased assessment. Journal of Educational Psychology 1985;77:197–207.

Valencia, R.; Suzuki, L. Intelligence testing and minority students: Foundations, performance factors, and assessment issues. Sage; Thousand Oaks, CA: 2001.

Vygotsky, LS. Thought and language. MIT Press; Cambridge, MA: 1986.

Table 1

Participants in Study 1 and Study 2

Experiment	Group	Ethnicity	Grade	Sex		Group total
				F	**M**	
1 & 2	Order 1	AA	1	6	1	30
			2	3	4	
		EA	1	2	2	
			2	2	2	
		LA	1	3	1	
			2	2	2	
1	Order 2	AA	1	1	3	28
			2	2	2	
		EA	1	3	2	
			2	6	1	
		LA	1	4	0	
			2	3	3	
2	TD	AA	1	2	0	27
			2	0	3	
		EA	1	2	4	
			2	3	4	
		LA	1	2	1	
			2	1	3	
		Unknown		0	2	
2	LI	AA	1	1	3	14
			2	1	1	
		EA	1	0	1	
			2	0	1	
		LA	1	1	2	
			2	1	2	

Note. For Order 1 in Experiment 1, *Two Friends* was used as the pretest; *Bird and His Ring* was used as the posttest. Children in Order 1 were also the control group for Experiment 2. For Order 2 in Experiment 1, *Bird and His Ring* was used as the pretest; *Two Friends* was used as the posttest. F = female; M = male; TD = typical development; LI = language impairment; AA = African American; EA = European American; LA = Latino American.

Table 2

Parallel-forms reliability: Story scores and productivity measures by story, gender, and ethnicity

Book	Gender	Ethnicity		Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
Two Friends	F	AA	M	28.67	10.25	14.42	4.00	124.75	55.92	19.42	6.31	25.00	1.26
			SD	10.71	4.43	5.62	1.48	48.90	22.97	4.58	1.59	10.19	0.30
		EA	M	28.85	10.92	13.62	4.31	133.92	61.62	19.85	6.63	26.31	1.31
			SD	8.43	3.15	4.29	1.55	64.69	22.56	7.99	1.30	12.62	0.25
		LA	M	27.33	10.00	13.08	4.25	121.83	52.08	19.67	6.23	24.33	1.23
			SD	4.62	2.92	2.27	0.97	30.69	8.11	4.31	1.06	7.08	0.15
	M	AA	M	28.70	9.20	15.30	4.20	124.20	54.40	21.00	5.97	23.60	1.15
			SD	6.02	3.16	4.35	1.32	37.94	17.38	6.02	1.30	5.68	0.22
		EA	M	28.57	10.43	14.29	3.86	134.14	60.29	19.57	6.70	24.57	1.22
			SD	8.52	2.99	4.39	2.04	68.61	22.46	7.72	0.80	12.96	0.18
		LA	M	22.25	7.75	11.50	3.00	98.75	40.75	14.00	6.93	20.00	1.38
			SD	9.14	2.75	5.92	1.41	33.78	11.41	2.58	1.38	6.48	0.22
	Both	All	M	27.98	10.02	13.90	4.07	125.45	55.62	19.48	6.39	24.52	1.25
			SD	7.86	3.34	4.34	1.42	48.84	18.94	6.0	1.25	9.52	0.23
Bird and His Ring	F	AA	M	30.58	11.17	15.33	4.08	161.08	61.67	21.92	7.24	28.58	1.30
			SD	7.34	3.56	3.77	1.31	60.04	20.97	6.23	1.08	10.30	0.31
		EA	M	25.92	9.38	12.69	3.92	137.38	52.69	19.31	7.16	23.92	1.25
			SD	8.17	3.59	3.77	1.55	44.33	10.27	6.05	1.02	7.26	0.18
		LA	M	26.00	8.83	13.42	3.75	133.42	49.42	18.92	7.05	24.25	1.26
			SD	6.90	2.95	3.96	1.14	50.50	18.45	6.02	1.22	10.21	0.20
	M	AA	M	26.70	9.00	13.40	4.30	151.70	52.10	23.50	6.52	26.90	1.17
			SD	5.79	3.33	3.63	0.67	73.62	12.78	11.55	1.20	12.49	0.23
		EA	M	24.57	9.43	12.14	3.00	116.00	47.00	15.86	7.20	19.14	1.19
			SD	8.10	2.76	4.45	1.63	40.63	16.37	4.67	0.79	6.67	0.15
		LA	M	22.75	7.00	12.50	3.25	132.50	48.75	18.00	7.54	27.00	1.53
			SD	6.18	1.41	4.80	1.26	35.83	13.23	5.48	1.44	6.06	0.17
	Both	All	M	26.66	9.41	13.43	3.83	141.02	52.81	19.98	7.08	25.10	1.26
			SD	7.28	3.27	3.91	1.30	53.86	16.13	7.27	1.10	9.55	0.23

Note. F = female; M = male; AA = African American; EA = European American; LA = Latino American; total = total score; Comp = Story Components; I & L = Story Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words.

Table 3

Means for Time 1 and Time 2

		Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
Time 1	M	25.66	9.12	12.79	3.76	126.16	51.09	18.97	6.65	23.86	1.26
	SD	7.81	3.2	4.35	1.41	49.35	15.78	6.76	1.20	8.96	0.23
Time 2	M	28.98	10.31	14.53	4.14	140.31	57.34	20.50	6.82	25.76	1.25
	SD	7.00	3.33	3.70	1.30	53.60	18.81	6.48	1.25	9.98	0.22
Gain	M	3.32	1.19	1.74	0.38	14.15	6.25	1.53	0.17	1.90	-0.01

Note. Total = total score; Comp = Story Components; I & L = Story Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words.

Table 4

Descriptive statistics for all dependent variables

	Group		Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
Pretest	TD	*M*	23.56	8.30	11.48	3.78	129.33	55.70	22.04	5.81	25.59	1.16
		SD	5.39	2.38	3.52	1.34	71.49	20.58	10.45	1.1	13.63	0.21
	LI	*M*	16.79	5.93	8.14	2.71	91.21	38.93	17.86	5.06	18.64	1.04
		SD	4.96	1.44	3.28	1.27	34.4	14.34	5.11	1.27	5.84	0.16
	CON	*M*	27.43	9.80	13.77	3.87	118.73	52.33	18.80	6.28	24.03	1.27
		SD	8.60	3.44	4.89	1.46	40.45	16.14	4.87	1.32	8.10	0.26
Posttest	TD	*M*	29.74	11.44	13.85	4.44	180.48	68.48	25.74	7.17	32.74	1.29
		SD	6.33	3.21	3.27	0.97	64.18	21.77	9.86	1.09	12.36	0.17
	LI	*M*	20.36	7.29	9.93	3.14	102.64	41.86	16.64	6.16	18.50	1.11
		SD	6.45	2.7	3.73	1.23	36.52	10.71	4.45	1.59	6.06	0.26
	CON	*M*	29.37	10.37	15.00	4.00	147.47	55.67	20.77	7.11	26.43	1.27
		SD	7.01	3.43	3.68	1.23	50.79	16.38	6.03	1.26	9.11	0.25

Note. Total = Total score; Comp = Story Components; I & L = Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words; TD = typical development; LI = language impairment; CON = control.

Table 5

ANOVA results for category scores

Measure	Effect	df	F	p	η_p^2
Story Components	Time	1,68	18.806	<.001	.217
	Group	2,68	9.790	<.001	.224
	Time × Group	2,68	5.017	<.001	.129
Story Ideas and Language	Time	1,68	22.109	<.001	.245
	Group	2,68	10.976	<.001	.244
	Time × Group	2,68	0.992	.376	.028
Episode Structure	Time	1,68	5.447	.023	.074
	Group	2,68	6.208	.003	.154
	Time × Group	2,68	0.977	.359	.030

Note. ANOVA = analysis of variance.

Table 6

Scheffé's comparisons: Mean pretest-to-posttest gain by group

	Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
TD	6.19**	3.15**	2.37*	0.67	51.15**	12.78*	3.70	1.36**	7.15*	0.13*
LI	3.57	1.36	1.79	0.43	11.43	2.93	-1.21	1.10*	-0.14	0.06
CON	1.93	0.57	1.23	0.13	28.73*	3.33	1.97	0.83**	2.40	0.01

Note. Total = total score; Comp = Story Components; I & L = Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words; TD = typical development; LI = language impairment; CON = control.

*
$p < .05.$

**
$p < .01.$

Table 7

Scheffé's comparisons: Mean gain differences between groups

	Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
TD-LI	2.61	1.79[*]	0.58	0.24	39.72[**]	9.85[*]	4.92[**]	0.26	7.29[**]	0.06
TD-CON	4.25[**]	2.58[**]	1.14	0.53	22.41	9.44[*]	1.74	0.53[*]	4.75[*]	0.13[**]
LI-CON	1.64	0.79	0.55	0.30	-17.30	-0.40	-3.18	0.28	-2.54	-0.06

Note. Total = total score; Comp = Story Components; I & L = Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words; TD = typical development; LI = language impairment; CON = control.

[*] $p < .05$.

[**] $p < .01$.

Table 8

Classification accuracy of pretest, modifiability, and posttest measures

Measure	Cutoff score	Sensitivity (LI as LI)	Specificity (TD as TD)	Positive likelihood ratio	Negative likelihood ratio	Positive predictive value	Negative predictive value
Pretest							
Total	18.16	71%	74%	2.73	.39	17%	97%
Comp	6.91	71%	74%	2.73	.39	17%	97%
I & L	7.96	57%	89%	5.18	.48	28%	96%
ES	2.44	57%	74%	2.19	.58	14%	96%
No. words	57.84	14%	93%	2.00	.92	13%	93%
NDW	35.12	57%	89%	5.18	.48	28%	96%
No. C-units	11.59	0%	100%	0.00	1.00	0%	93%
MLU-W	4.71	29%	85%	1.93	.84	13%	93%
No. clauses	11.96	14%	93%	2.00	.92	13%	93%
Clauses/C-unit	0.95	36%	85%	2.40	.75	15%	95%
Modifiability	6.43	93%	82%	5.17	.09	28%	99%
Posttest							
Total	23.40	71%	82%	3.94	.35	23%	97%
Comp	8.23	71%	78%	3.22	.37	20%	97%
I & L	10.59	71%	85%	4.73	.34	26%	98%
ES	3.47	64%	89%	5.82	.40	30%	97%
No. words	116.30	71%	82%	3.94	.35	23%	97%
NDW	46.71	71%	85%	4.73	.34	26%	98%
No. C-units	15.89	50%	78%	2.27	.64	15%	95%
MLU-W	6.08	57%	85%	3.80	.51	22%	96%
No. clauses	12.44	57%	78%	2.59	.55	16%	96%
Clauses/C-unit	1.11	57%	85%	3.80	.51	22%	95%

Note. LI = learning impairment; TD = typical development; Total = total score; Comp = Story Components; I & L = Ideas and Language; ES = Episode Structure; NDW = number of different words; MLU-W = mean length of utterance in words.

Table 9

Bias analysis at pretest

Group		Total	Comp	I & L	ES	No. words	NDW	No. C-units	MLU-W	No. clauses	Clauses/C-unit
AA	-1 SD and above	84	74	79	74	79	84	89	74	79	79
	Less than -1 SD	16	26	21	26	21	16	11	26	21	21
EA	-1 SD and above	86	86	90	62	90	95	95	90	95	90
	Less than -1 SD	14	14	10	38	10	5	5	10	5	10
LA	-1 SD and above	73	60	73	87	93	87	87	80	87	93
	Less than -1 SD	27	40	27	13	7	13	13	20	13	7

Note. All values are percentages. Total = total score; Comp = Story Components; I & L = Story Ideas and Language; NDW = number of different words; MLU-W = mean length of utterance in words; AA = African American; EA = European American; LA = Latino American; -1 *SD* and above = percentage of children in the typical group who scored -1 standard deviation or above; below -1 *SD* = percentage of children who scored more than 1 standard deviation below the mean.