Properties of the Narrative Scoring Scheme Using Narrative Retells in Young School-Age Children

John Heilmann
East Carolina University, Greenville, NC

Jon F. Miller
Ann Nockerts
University of Wisconsin—Madison

Claudia Dunaway
San Diego Unified School District, CA

Purpose: To evaluate the clinical utility of the narrative scoring scheme (NSS) as an index of narrative macrostructure for young school-age children.

Method: Oral retells of a wordless picture book were elicited from 129 typically developing children, ages 5–7. A series of correlations and hierarchical regression equations were completed using microstructural measures of vocabulary and grammar to predict NSS scores.

Results: While the NSS was significantly correlated with age and each of the microstructural measures, the hierarchical regression analyses revealed a unique relationship between vocabulary and narrative macrostructure.

Conclusion: The NSS is an efficient and informative tool for documenting children’s development of narrative macrostructure. The relationship between the NSS and microstructural measures demonstrates that it is a robust measure of children’s overall oral narrative competence and a powerful tool for clinicians and researchers. The unique relationship between lexical diversity and the NSS confirmed that a special relationship exists between vocabulary and narrative organization skills in young school-age children.

Key Words: narrative, language sample analysis, story grammar, vocabulary

Analysis of oral narratives provides a rich source of data that documents children’s language use in a naturalistic context. Narrative analysis is a highly effective clinical and research tool, as it allows examiners to analyze multiple linguistic features simultaneously using a single short sample. Examiners have the opportunity to document children’s productive vocabulary and grammar using microstructural analyses as well as children’s broader text-level narrative organization skills by utilizing macrostructural analyses (see Westby, 2005).

Microstructural analyses primarily focus on children’s linguistic form and content, which are measured within individual utterances. Linguistic form is commonly assessed by analyzing children’s grammatical and syntactic abilities using mean length of utterance (e.g., Brown, 1973; Miller, 1981) and various measures of sentence complexity (e.g., Nippold, Hesketh, Duthie, & Mansfield, 2005; Schuele & Tolbert, 2001; Scott & Stokes, 1995). Measures of linguistic content are used to document children’s productive vocabulary skills, which typically calculate children’s lexical diversity using measures such as type–token ratio (Templin, 1957) and number of different words (Miller, 1987; Miller & Klee, 1995). While these measures are the most commonly cited microstructural analyses in the literature, this list is in no way exhaustive. Microstructural analyses continue to be reviewed, critiqued, and reanalyzed (see Justice et al., 2006, for a review and discussion).

Macrostructural analyses, on the other hand, examine children’s language skills beyond the utterance level and document children’s ability to relate concepts that transcend the individual utterance. Most macrostructural analyses of children’s narratives are rooted in the story grammar tradition, which proposes that all stories have a setting and episode system (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979). The setting provides background information
on the characters and their environment, while the episode system includes three main components that occur in all stories: (a) a problem (initiating event and/or internal response), (b) attempts at solving the problem, and (c) consequences/outcomes. To be a complete episode, a narrator must include all three of these key components (see Strong, 1998, for a review). These settings and episodes can be combined in an infinite number of ways to create individual stories.

**Relationship Between Microstructural and Macrostructural Measures**

Previous research has demonstrated that lexical and grammatical gains play an important role in children’s acquisition of narrative proficiency (Berman & Slobin, 1994; Bishop & Donlan, 2005). In their seminal work, Berman and Slobin (1994) documented the role of linguistic forms taking on new functions that aid in the organization of narratives. This work described the trade-offs that occurred when different microstructural features were used for different functions, including the organization of narratives. Children’s narrative organization skills were positively related to advances in use of microstructural features, including grammatical forms (e.g., verb tense, aspect, and voice), lexical forms (e.g., lexical aspect and manner/cause of verbs of motion), and lexico-grammatical features (e.g., locative particles, prepositional phrases, and connectives). This pattern of results is consistent with Slobin’s long-standing claim that “new forms first express old functions, and new functions are first expressed by old forms” (Slobin, 1973, p. 184). Bishop and Donlan (2005) observed that children’s ability to encode and retell a story was more strongly related to linguistic ability than nonverbal IQ. Children’s microstructures, including complex syntax and relating causal concepts, were more influential in event memory and story organization than their cognitive skills. These data documenting the unique relationship between microstructures and macrostructures are contrary to evidence demonstrating that children’s narrative organization skills emerge from more general cognitive capabilities, including executive function (van den Broek, 1997), and sociolinguistic processes (Eaton, Collis, & Lewis, 1999; McCabe, 1997; Peterson & McCabe, 2004; Quasthoff, 1997a, 1997b).

**Narrative Skills of Children With Language Impairment**

Given the strong relationship between microstructural and macrostructural measures, it is not surprising that children with language impairment have substantial difficulty using appropriate vocabulary and grammar when telling stories (Boudreau & Hedberg, 1999; Gillam & Johnston, 1992; Pearce, McCormack, & James, 2003; Reilly, Losh, Bellugi, & Wulfeck, 2004) in addition to substantial difficulty with text-level organization of narratives (Boudreau & Hedberg, 1999; Manhardt & Rescorla, 2002; Merritt & Liles, 1987; Pearce et al., 2003; Reilly et al., 2004). The presence of both microstructural and macrostructural deficits in children with language impairment counters theories that identify the primary cause of language impairment as deficiencies with grammatical competence (e.g., Clahsen, 1989; Gopnik & Crago, 1991; Rice, Wexler, & Cleave, 1995). Alternatively, the global narrative difficulties of children with language impairment suggest that their language deficits may be due to broader information-processing deficits, such as a reduced processing capacity (Boudreau, 2007; Colozzo, Garcia, Megan, Gillam, & Johnston, 2006). Further study of the relationship between children’s microstructural and macrostructural language skills will provide a better understanding of the nature of the impairment. Better understanding of children’s narrative skills in general, and macrostructural skills in particular, also has substantial clinical implications. Difficulty with narrative organization can have a dramatic impact on children with language impairment, as such discourse-level skills are required to effectively communicate. While vocabulary and grammar deficits limit children’s ability to produce fully competent utterances, impaired macrostructural skills affect children’s ability to generate coherent and age-appropriate extended discourse. Furthermore, narratives are a major component of the school curriculum, and children are expected to understand and use appropriate narrative form effectively. Therefore, narrative macrostructure skills must be efficiently and accurately documented, and should be considered within an extensive assessment protocol for children with language difficulties.

**Methods of Measuring Narrative Macrostructure**

While most narrative macrostructure measures are rooted in the story grammar tradition and share the same underlying principles, the coding of narrative macrostructure varies widely throughout the literature. Some narrative macrostructure coding schemes documented children’s inclusion of specific story grammar components (e.g., Strong, 1998) or identified the presence or absence of specific story grammar components for a given story (e.g., Berman, 1988; Boudreau & Hedberg, 1999; Miles & Chapman, 2002; Reilly et al., 2004). In these studies, children who produced more story grammar components and/or more advanced story grammar features were thought to have stronger narrative organization skills. The second major class of macrostructural measures used text-level judgments of children’s narrative proficiency (Applebee, 1978; Hedberg & Westby, 1993; Stein, 1988). Rather than identifying the presence or absence of specific story grammar components, these measures required holistic judgments by the examiner to rate the quality and developmental level of the narrative.

The major advantage of the simple story grammar coding procedures is that it facilitates a relatively high level of accuracy across coders. There is less room for differences across coders, as the coder is only responsible for identifying the presence or absence of specific story-related themes. The major disadvantage of simple story grammar analyses is that they are potentially limited in their ability to account for the abstract interutterance concepts and qualitative aspects of the narrative, or the story’s “sparkle” (Peterson & McCabe, 1983). McFadden and Gillam (1996) demonstrated that holistic ratings of children’s narratives capture the more refined aspects of narratives, such as charm and depth, and that these ratings were better than simple story grammar analyses for documenting differences between children with language impairment and their typically developing peers between the ages of 9;0 (years; months) and 11;7.
Additional Skills Required to Tell an Effective Story: Beyond Story Grammar

The development of narratives in children and adults has been studied extensively and has revealed additional areas of advancement beyond inclusion of story grammar features, including children’s use of literate language and cohesive devices (Bamberg & Damrad-Frye, 1991; Halliday & Hasan, 1976; Hedberg & Westby, 1993; Wigglesworth, 1997). Use of literate language occurs when children use abstract language features commonly used by teachers and found in the curriculum (Westby, 2005). Some key literate language features related to narrative competence include use of metacognitive verbs (e.g., think or know), metalinguistic verbs (e.g., say or talk), and elaborated noun phrases (e.g., the boy in the restaurant with the frog; see Nippold, 2007, and Westby, 2005, for a review). Bamberg and Damrad-Frye (1991) identified that more sophisticated and later developing narratives included abstract language such as metacognitive and metalinguistic verbs. In their analysis of productions of Frog, Where Are You? (Mayer, 1969), Bamberg and Damrad-Frye identified that these abstract language skills emerged at age 5 years, demonstrated robust development through adulthood, and were essential for relating the hierarchical relationships between events in complex narrative productions. Additional studies have identified that literate language skills were present in children’s oral narratives during the preschool years (Curenton & Justice, 2004), developed through the school years and into adolescence (Greenhalgh & Strong, 2001; Nippold, 2007; Pelligrini, Galda, Bartini, & Charak, 1998), and were used less frequently by children with language impairment (Greenhalgh & Strong, 2001).

Another high-level narrative feature that continues to develop through the school years is the cohesiveness of children’s narrative productions (Halliday & Hasan, 1976; Wigglesworth, 1997). To tell a story, narrators must effectively use cohesive devices to carry concepts across individual utterances. Three major categories of cohesive devices are (a) referential cohesion, which allows a narrator to maintain appropriate reference to the characters, objects, and locations across utterances using both noun phrases and pronouns; (b) conjunctive cohesion, which allows a narrator to sustain concepts across phrases and utterances using conjunctive words and phrases (e.g., and, but, besides, on the other hand, finally, in addition); and (c) lexical cohesion, which allows a narrator to effectively use vocabulary to link concepts across utterances. Measures of cohesiveness can be a sensitive index of language use, as children with language impairment have more difficulty with correct use of cohesive ties (Hedberg & Westby, 1993; Liles, 1985; Strong & Shaver, 1991). While cohesion is often considered a microstructural measure, we treated it as a macrostructural measure because aspects of cohesion transcend individual utterances and are necessary for producing coherent narratives.

The Narrative Scoring Scheme: A Comprehensive Measure of Narrative Proficiency

Our goal in developing the narrative scoring scheme (NSS) was to create a metric that documents the range of skills required for school-age children to effectively tell a coherent and interesting story (see Appendix). To extend beyond simple story grammar analyses, the NSS incorporates multiple aspects of the narrative process into a single scoring rubric and provides an overall impression of the child’s narrative ability. This metric combines both the basic features of the story grammar approaches as well as the higher level narrative skills that continue to develop through the school-age years. In addition to adding higher level narrative skills in the scoring scheme, the NSS uses a combination of discrete coding criteria and examiner judgment. The NSS was created to improve on the simple story grammar measures by requiring the examiners to make interutterance text-level judgments, which have been shown to be more effective than discrete coding schemes in identifying children with language impairment (McFadden & Gillam, 1996). By breaking the judgments into seven skill areas, examiners have the opportunity to reflect on each component of the narrative process and judge the child’s proficiency in that area. This combination of explicit scoring guidelines and flexibility to allow for examiner judgment reflects the hybrid nature of the NSS. The scores from the seven NSS categories are combined to provide a single composite score, which allows the examiner to generate an index of children’s overall narrative ability.

The first step in developing the NSS was establishing the key components from the story grammar literature, which included the story’s introduction, the major conflicts and resolutions (conflict resolution), and a conclusion. To document children’s use of literate language, the NSS includes categories that assess metacognitive and metalinguistic verbs. The mental states component documents children’s abilities to use metacognitive verbs (e.g., think and know) to describe the characters’ thoughts and feelings. The character development component of the NSS also documents children’s literate language skills by measuring the ability to use metalinguistic verbs (e.g., talk and say), differentiate between main and supporting characters, and talk in the first person to depict the characters in the story. The NSS evaluates two separate aspects of cohesive ties that were adapted from Halliday and Hasan (1976). The referencing component measures aspects of referential cohesion, including appropriate use of pronouns and antecedents. The cohesion component documents the conjunctive and lexical aspects of cohesion, including appropriate ordering, emphasis of critical events, and transitions between events.

Goals of the Study

Children’s performance on the NSS has been reported in studies that examined the narrative organization skills of native Spanish-speaking children who were learning English as a second language (Miller et al., 2006). The goal of the present study was to describe the NSS from a clinical perspective and to further analyze the linguistic properties of the measure in a group of children who were fluent in English. Furthermore, the literature has revealed that there is a special relationship between children’s microstructural and macrostructural language skills (Berman & Slobin, 1994; Bishop & Donlan, 2005). To better understand the linguistic properties of the NSS and to extend our understanding of the relationship...
between microstructural and macrostructural measures, a second goal was to document the relationship between children’s vocabulary, grammar, and narrative organization skills. To achieve this, we examined the relationship between measures of vocabulary, grammar, and the NSS. These analyses further clarified the key constructs that the NSS is measuring and also provided additional evidence for the role of vocabulary and grammar in the development of young school-age children’s narrative organization skills. This study addressed the following questions:

1. Are age and measures of vocabulary and grammar significantly correlated with NSS scores in narrative retells of young school-age children?
2. Are measures of vocabulary uniquely related to NSS scores in narrative retells of young school-age children?
3. Are measures of grammar uniquely related to NSS scores in narrative retells of young school-age children?

Method

Participants

A total of 129 typically developing children age 5–7 years were recruited for this study. The children were recruited from public schools in the San Diego (CA) City School and Cajon Valley School Districts. Administrators from the two districts assisted with obtaining informed consent from each of the children’s primary caregivers. The pool of potential participants was reviewed by the school’s speech-language pathologists (SLPs) and classroom teachers to identify children who qualified for the study. To participate in the study, children were required to have average scores on all summative classroom, district, and state assessments. The classroom teachers reviewed the records for each student to identify him or her as average performing. The child’s academic record was also reviewed to identify him or her as fluent English. This designation was made by the child’s respective school district and was based on a passing grade on an English proficiency test and grade-level academics. While academic data were used as inclusionary criteria for the participants, test scores and descriptions of performance were not recorded and were not available for further analysis. The majority of the participants were native English speakers. A small percentage of the children were Spanish/English bilingual and designated as “fluent English.” This designation was made by the child’s respective school district and was based on a passing grade on an English proficiency test and grade-level academics. The SLPs and teachers confirmed that each child met the inclusionary criteria and enrolled the eligible children in the study. The individual children participating in the study provided verbal assent prior to completing the protocol.

Table 1 summarizes the demographic data for the participants. Sixty-one percent of the participants were in kindergarten, 36% of the children were in first grade, and 2% were in preschool. The numbers of male and female participants were roughly equal, with slightly fewer boys than girls. The school SLPs attempted to recruit students who reflected the racial and ethnic diversity of their schools. The SLPs first identified the children’s ethnicity (Hispanic or not Hispanic) and then documented the non-Hispanic children’s race. A review of the race and ethnicity data revealed that the sample is a relatively heterogeneous group. Socioeconomic status was measured by calculating the highest number of years of education that the child’s mother completed. On average, mothers completed 14.4 years of education (SD = 2.5), with a range of 9–20 years. The majority of the mothers completed at least some college, while only eight of the children’s mothers did not complete high school.

Procedure

Each participant completed a narrative retell of the wordless picture book Frog, Where Are You? (Mayer, 1969). The purpose of collecting the narratives was to establish a narrative database reflecting typically developing children’s oral narrative skills and to further our understanding of children’s developing narrative competence. The head SLP from each district was responsible for training 18 school-based SLPs to elicit the oral narratives. The head SLP and clinicians met on three separate occasions and had the opportunity to practice the protocol with each other several times. The clinicians read the scripted instructions to the students and cued them to listen to a taped version of the story while following along with the pictures in the book. The students then retold the story using the book as an aid. Examiner prompts were limited to encouragement to begin the story and open-ended cues to continue the retell. The scripted instructions and audiotaped story script were adapted from the Strong Narrative Assessment Procedure (Strong, 1998) and were used to facilitate high fidelity among the numerous examiners completing the language sample elicitation.

Transcription and Coding

The children’s narrative productions were digitally recorded and sent to the Language Analysis Lab at the University of Wisconsin—Madison, where they were transcribed by trained research assistants who had at least 10 hr of transcription.

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>Preschool</td>
<td>3</td>
</tr>
<tr>
<td>Kindergarten</td>
<td>79</td>
</tr>
<tr>
<td>First grade</td>
<td>47</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>69</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>87</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16</td>
</tr>
<tr>
<td>Othera</td>
<td>15</td>
</tr>
<tr>
<td>African American</td>
<td>7</td>
</tr>
<tr>
<td>No data</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. Ethnicity data were collected for children who were Hispanic or Latino. Race data are provided for all children who were non-Hispanic/non-Latino.

Other races/ethnicities included Arabic (2), Chinese (3), Japanese (2), Korean (1), Filipino (5), Portuguese (1), and Samoan (1).
experience using standard coding conventions for Systematic Analysis of Language Transcripts (SALT) software (Miller & Iglesias, 2008). Utterances were segmented into communication units (C-units; Labov & Waletsky, 1967), which included a main clause and all dependent clauses. The transcripts began and ended with the child’s first and last utterance, respectively. See Miller and Iglesias (2008) for a full review of the transcription conventions.

After completing the orthographic transcription, the research assistant reviewed the transcript and completed the NSS. To score the NSS, the transcriber carefully reviewed the narrative transcript and assigned a score of 0–5 for each of the seven categories summarized in the Appendix. Categories that could not be scored received a score of zero or NA. Scores of zero were given if the child did something that precluded the examiner from scoring a section of the NSS, such as including a part of the story or refusing to complete the task. If there was an error on the part of the examiner (e.g., not following the protocol or problems with the recording), sections of the NSS that were affected were not scored, and the examiner coded the section as not applicable for analysis (i.e., NA). For all other sections, scores of 1 reflected minimal presence/immature performance, scores of 3 reflected emerging skills, and scores of 5 reflected proficient performance. Transcribers also had the opportunity to assign scores of 2 and 4 if performance was judged to be between the major anchors. To improve the accuracy of the scoring procedure, specific guidelines were provided for scores of 1, 3, and 5 (see the Appendix). These guidelines assisted the transcriber in assigning an accurate score that reflected the child’s performance in each component of the narrative process and reduced the abstractness of the narrative macrostructure concepts. The scoring across the seven categories received equal weighting because the literature revealed that each of these seven narrative aspects is necessary for telling a well-developed story. Furthermore, keeping the scoring rules straightforward and consistent facilitates simple and accurate scoring. A comprehensive training procedure is available on the SALT Web site (www.saltsoftware.com/training/handcoded/) that includes an overview of the NSS; scoring tips; a description of how to enter NSS scores into a SALT file; excerpts from samples demonstrating minimal/immature, emerging, and proficient performance across each section of the NSS; and a set of practice transcripts.

In addition to the NSS, the transcribers completed coding for the subordination index (SI; Scott & Stokes, 1995; Strong, 1998). The transcribers added a code to each C-unit that summarized the number of independent and dependent clauses. C-units that were incomplete, unintelligible, or nonverbal, or that had an error at the utterance level, were excluded from the analysis. Utterance-level errors included incorrect word order, omission of more than two words in an utterance, and utterances that simply did not make sense. Elliptical responses to examiner questions were also excluded from the SI analysis. Utterances in which the child inappropriately omitted the subject or copula were coded and received a score of zero. After each C-unit was coded, the SI was generated by dividing the total number of clauses (both main and subordinate) by the total number of C-units. After all the narratives were transcribed and coded, the transcripts were analyzed using SALT (Miller & Iglesias, 2008), which produced a rectangular data file summarizing each dependent measure for each of the transcripts. This file was formatted for statistical analysis using SPSS Version 16.0.

**Agreement**

Accuracy of the transcription and coding process was examined at three levels. Protocol accuracy was calculated by the principal investigator, who reviewed 10% of the written transcripts to identify whether the transcribers were adhering to the transcription conventions. Percentage agreement between the transcribers and principal investigator was 98% for segmentation rules, 99% for word-level codes, and 98% for coding of reduplications and reformulations (i.e., mazes). To determine transcription accuracy and coding agreement, 10% of the narratives were independently transcribed and coded for the NSS and SI by a second research assistant. Transcription accuracy was calculated by comparing the independent transcripts at the word and morpheme level (94% agreement), utterance segmentation decisions (98% agreement), placement of mazes (93% agreement), and utterance types (100% agreement).

Calculating agreement for the two coding schemes presented a greater challenge. Simple interrater agreement scores can be misleading, as small differences between coders (e.g., NSS scores of 24 and 25) are treated the same as large differences between coders (e.g., NSS scores of 16 and 28). Therefore, agreement for the NSS and SI coding was calculated using Krippendorff’s alpha, which accounted for both chance agreement and the degree of difference between transcribers (Krippendorff, 1980). Alpha values accounting for differences in ordinal data were calculated using the summed NSS and SI scores that were calculated by the two independent transcribers (α = .92 for SI; α = .79 for NSS). Krippendorff established benchmarks for alpha values, with ≥.80 reflecting adequate agreement and values between .67 and .80 reflecting acceptable agreement for exploratory research and drawing tentative conclusions. For a review of the accuracy and agreement process and a further discussion of Krippendorff’s alpha, see Heilmann et al. (2008).

**Language Sample Measures**

To test the relationship between microstructural measures and the NSS, we used the following language sample measures that repeatedly have been found to be robust and developmentally sensitive to the population and context used in the present study (i.e., typically developing 5–7-year-olds who produced oral narratives):

*Length/productivity.* Number of total words (NTW) is a measure of productivity that documents the amount of information provided in the story (Allen, Kertoy, Sherblom, & Pettit, 1994; Paul & Smith, 1993). In addition, NTW was used to assist with statistically controlling sample length, which has the potential to affect additional language sample measures.

*Vocabulary.* Number of different words (NDW) is a measure of lexical diversity that provides a robust estimate of children’s productive vocabulary (Klee, 1992; Miller, 1987; Miller & Klee, 1995) and has been widely used as an index of
Results

The NSS scores were first reviewed to identify how many child and examiner errors precluded scores to be administered. In the present study, a total of 903 scores were completed using the children’s narrative transcripts (seven categories were scored across 129 transcripts). Across the 903 sections, only six sections received scores of zero (four stories lacked an introduction, and two children omitted a conclusion), and only one of the 903 sections received a score of NA. 99% of the NSS sections were able to be scored correctly, confirming that the training and elicitation procedures facilitated a high level of child compliance and examiner fidelity during the elicitation process.

The skewness statistic was calculated for the NSS to determine if there was an unequal distribution of NSS scores across the sample (see Coolican, 2004, for a review). Skewness measures of zero indicate a perfectly normal distribution, while skewness values below –0.8 or above 0.8 have been described as “noticeably skewed” (Bourque & Clark, 1992, p. 69). The skewness statistic for the NSS was –0.5, indicating that scores were more concentrated toward the ceiling but were not noticeably skewed according to the criteria established by Bourque and Clark.

Descriptive statistics for age, the NSS, and each of the microstructures are presented in Table 2. Table 2 also summarizes the bivariate correlations between the NSS, children’s age, and each of the microstructures. All correlations were significant and were moderate in strength (Cohen, 1988).

To further explore the covariance structure between the variables and to identify unique relationships between the microstructures and the NSS, two separate hierarchical regression equations were completed. Hierarchical regressions allow examination of variance that is uniquely explained by a given variable. The first hierarchical regression equation is summarized in Table 3. Sample length was first controlled by entering NTW into Model 1. Length was controlled because measures of lexical diversity are inevitably influenced by the NTW in the sample (Malvern & Richards, 2002). That is, the more total words that a child produces, the more opportunity he or she has to produce different words.

Model 2 identified the unique relationship between vocabulary and NSS after controlling for length. Taken together, NTW and NDW were significantly correlated with NSS scores ($r = .58$). Adding NDW in Model 2 increased the explained variance from 24% to 33%, a net increase of 9%. A one-way analysis of variance was completed to determine whether the increase in explained variance was significant, and an $F^2$ statistic was calculated to estimate its effect size. According to Cohen (1988), effect sizes of 0.02, 0.15, and 0.35 are considered small, medium, and large, respectively. The 9% increase in explained variance was significant, $F(1, 124) = 16.5$, $p < .001$, $F^2 = 0.12$, demonstrating that NDW was uniquely related to the NSS beyond the effect of length.

Model 3 further controls for vocabulary (NDW) to determine whether NDW is unique to the NSS score. As expected, adding vocabulary to Model 2 resulted in a significant increase in explained variance from 33% to 35%, a net increase of 2%. The 2% increase in explained variance was significant, $F(1, 123) = 4.5$, $p = .04$, $F^2 = 0.03$, supporting the unique relationship between vocabulary and the NSS.

Model 4 presents the results of the final regression equation. As expected, adding grammatical structures to Model 3 resulted in a significant increase in explained variance from 35% to 41%, a net increase of 6%. The 6% increase in explained variance was significant, $F(1, 122) = 7.0$, $p = .01$, $F^2 = 0.05$, supporting the unique relationship between grammatical structures and the NSS.

These findings indicate that vocabulary, grammatical structure, and length all have unique contributions to the NSS score.

### Table 2. Descriptive statistics for language sample measures and correlations with the narrative scoring scheme (NSS).

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Correlation with NSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6.0</td>
<td>0.7</td>
<td>5.0–7.0</td>
<td>.30</td>
</tr>
<tr>
<td>Narrative macrostructure (NSS)</td>
<td>19.0</td>
<td>3.0</td>
<td>11.0–26.0</td>
<td>.50**</td>
</tr>
<tr>
<td>Productivity (NTW)</td>
<td>264.1</td>
<td>77.2</td>
<td>133.0–608.0</td>
<td>.50**</td>
</tr>
<tr>
<td>Vocabulary (NDW)</td>
<td>92.5</td>
<td>19.0</td>
<td>47.0–150.0</td>
<td>.58**</td>
</tr>
<tr>
<td>Grammar (MLCU)</td>
<td>7.0</td>
<td>0.9</td>
<td>4.8–9.6</td>
<td>.44**</td>
</tr>
<tr>
<td>Grammar (SI)</td>
<td>1.1</td>
<td>0.1</td>
<td>1.0–1.3</td>
<td>.35**</td>
</tr>
</tbody>
</table>

*Note. NTW = number of total words; NDW = number of different words; MLCU = mean length of C-unit; SI = subordination index.

*Significant at $p < .001$.
related to children’s NSS scores above and beyond the variability explained by NTW.

Model 3 was completed to identify if there was a unique relationship between the two grammatical measures (MLCU and SI) and NSS scores, after controlling for NTW and NDW. The measures in this final model were significantly correlated with the NSS \( r = .61 \). Adding MLCU and SI to the third regression equation added an additional 2% prediction to NSS scores, which was not significant, \( F(2, 122) = 2.9, p = .060, r^2 = 0.03 \). In this first hierarchical regression analysis, NDW uniquely predicted NSS scores after controlling for length using NTW. The two grammatical measures, however, did not add unique prediction of children’s NSS scores. To confirm that NDW was the major unique predictor of NSS scores, a second hierarchical regression equation was completed and is summarized in Table 4. Again, sample length was controlled by entering NTW in the first model. The two grammatical measures were next entered in the second equation, which resulted in a combined correlation of \( r = .57 \). Adding SI and MLCU to NTW increased the explained variance from .24 to .30, documenting that the grammatical measures explain 6% of the variance in NSS after controlling for sample length. This increase was significant, \( F(2, 123) = 6.0, p = .002, r^2 = 0.08 \). NDW was added to the third model to test the unique prediction of vocabulary on NSS scores. Adding NDW in the third model resulted in a 5% increase in explained variance between the microstructures and NSS, which was significant, \( F(1, 122) = 8.7, p = .003, r^2 = 0.06 \).

In sum, this series of hierarchical regression equations documented that children’s use of vocabulary is the major significant and unique microstructural variable in predicting their story organization skills as measured by the NSS. Children’s productive grammar, while significantly correlated with NSS scores, did not provide unique prediction of the children’s narrative macrostructure ability.

**Discussion**

Upon establishing a set of reference databases for children’s narrative retells, the Language Analysis Lab at the University of Wisconsin—Madison set out to identify a clinically useful measure of children’s narrative organization skills. The NSS was created to bring together the benefits of concretely scoring criteria combined with judgment of text-level constructs. The NSS also incorporated higher level narrative components, including cohesive markers and measures of literate language, to measure a wider range of skills than traditional story grammar analyses.

To determine whether the NSS was developmentally appropriate for the children in this study, the skewness statistic was calculated. This analysis revealed that the data were not noticeably skewed according to Borque and Clark’s (1992) criteria and that the NSS appeared to be a sensitive measure for school-age children who produced an oral retell. (We are currently completing a more thorough investigation of the developmental sensitivity of the NSS as compared to alternative methods of documenting narrative macrostructure skills.)

To document the relationship between the microstructural features of language samples and children’s performance on the NSS, a series of correlation and hierarchical regression analyses were completed. These analyses confirmed that a close relationship existed between children’s productivity, vocabulary, grammar, and narrative macrostructure skills. The correlation analyses documented that age and each of the microstructural measures (NTW, NDW, MLCU, and SI) were significantly correlated with children’s narrative organization skills. It is noteworthy that the correlation between age and NSS was the weakest observed correlation. The socio-cultural theory of narrative macrostructure development proposes that children who have more experience with stories will have greater narrative competence (e.g., Eaton et al., 1999; Stein & Glenn, 1979). The children in the present sample spanned 2 years in age. The older children in this sample likely had more experience listening to and telling stories. However, it was the children’s vocabulary and grammar skills that were most strongly related to their narrative macrostructure scores. While this study did not directly control for the amount of experience the children had with narratives, the data provide some additional evidence for the importance of children’s linguistic proficiency in predicting narrative organization skills (Berman & Slobin, 1994; Bishop & Donlan, 2005).

Two separate hierarchical regression analyses were completed to identify the unique relationships between each of the microstructures and NSS scores. After controlling for length, the unique relationship between the NSS and measures of vocabulary and grammar was calculated. The analyses showed that children’s productive vocabulary skills were the only unique predictor of narrative organization skills. Grammatical measures, on the other hand, provided no unique prediction of NSS scores. The unique importance of vocabulary in predicting narrative organization skills was a novel finding. Bishop and Donlan (2005) documented that children’s use of complex syntax and expression of causal concepts uniquely predicted children’s ability to organize their oral narratives. Bishop and Donlan examined children between 7 and 9 years of age, while the present study investigated children age 5–7 years. The children in the present study were using minimal complex syntax. As observed in Table 2, SI values averaged 1.1, showing that children produce approximately one subordinate clause every 10 utterances. The children’s use of subordination may have been influenced by the story used in this study: the children’s SI values, on average, were just slightly lower than the SI value from the story script (SI = 1.15). These low levels of subordination could explain, in part, the modest correlations and lack of a unique relationship between the grammatical measures and the NSS.

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>( r )</th>
<th>Adjusted ( r^2 )</th>
<th>( r^2 ) change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Productivity (NTW)</td>
<td>.50</td>
<td>.24*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Productivity (NTW)</td>
<td>.57</td>
<td>.30**</td>
<td>.06*</td>
</tr>
<tr>
<td></td>
<td>Grammar (MLCU, SI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Productivity (NTW)</td>
<td>.61</td>
<td>.35*</td>
<td>.05*</td>
</tr>
<tr>
<td></td>
<td>Grammar (MLCU, SI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vocabulary (NDW)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at \( p < .01 \).

**TABLE 4. Summary of hierarchical regression analysis with vocabulary measures uniquely predicting NSS scores in Model 3.**
**Relationship Between Vocabulary and the NSS**

The present study revealed that there is a special and important relationship between narrative organization and vocabulary skills that emerges prior to children becoming fully literate. This study demonstrated that the development of story schema and vocabulary acquisition is developing along a similar path. The importance of vocabulary in narrative organization skills is not surprising given the literature describing the development of narrative form. Preschool and young school-age children typically produce narratives that simply chain sequences of events in temporal order (Berman, 1988) and provide simple descriptive sequences (Stein & Glenn, 1979). It is not until the later school-age years that children hierarchically organize the events in their narrative productions (Berman, 1988) and take multiple perspectives to relate the events (Stein & Glenn, 1979). To produce these more advanced narratives, children must use complex syntax (Bamberg & Damrad-Frye, 1991; Bishop & Donlan, 2005). Before children are proficient in using complex syntax, they likely have to rely on their vocabulary skills to organize their narrative productions. Furthermore, the emerging literacy literature has documented the importance of vocabulary in the development of children’s narrative and comprehension skills. There is a well-documented relationship between children’s vocabulary skills and reading comprehension (see Scarborough, 2001). Furthermore, preschool and young school-age children acquire new vocabulary through repeated exposures to narrative form (Robbins & Ehri, 1994; Senechal & Cornell, 1993) and by receiving adult scaffolding that highlights the story’s structure (Hargrave & Senechal, 2000; Penno, Wilkinson, & Moore, 2002).

**Limitations**

The results from the hierarchical regression equations could have been affected by the sampling context. Because all of the measures were collected from a single language sample, high internal validity was achieved. While each of the language sample measures theoretically and empirically reflects its respective linguistic domain, strong intercorrelations between language sample measures have been observed (Miller & Klee, 1995). The hierarchical regression equations did afford better understanding of the covariance structure. However, vocabulary and grammatical measures acquired from alternate tasks may provide a more informative test of the relationship between vocabulary, grammar, and narrative macrostructure.

**Clinical Implications**

The NSS was created as a clinically useful comprehensive narrative macrostructure measure. To be clinically feasible, an assessment tool must be able to be completed in a short amount of time to accommodate the busy schedule of the SLP. In our lab, trained transcribers could complete the NSS using a narrative transcript in approximately 3 min. In addition to efficient scoring, the NSS was developed to facilitate accurate scoring both within and between examiners. The Krippendorff alpha analyses revealed that the NSS had lower inter-rater agreement than we would ideally observe and that the NSS alpha was not as high as the alpha for the SI. The difference in alphas between the NSS and SI was not surprising, however, as the NSS requires much greater individual judgment when compared to the relatively straightforward scoring rules for the SI. However, recall that these subjective ratings are often the most sensitive when identifying children with language impairment (e.g., McFadden & Gillam, 1996). Our goal is to continue developing new training methods and complete additional research identifying ways to increase coding accuracy for the NSS and other macrostructural analyses of children’s natural language use.

Upon scoring the NSS, clinicians are provided with a clinically useful benchmark for children’s overall narrative proficiency. The data described in this study used the composite NSS score, which was the summed score for all seven sections of the NSS. We proposed that the NSS composite score provided an index of children’s overall narrative organization skills. The NSS data described in this study are available as part of the SALT Narrative Story Retell database and can be downloaded for free of charge from the SALT Web site. SALT software, used to access the data, can compare a target child’s NSS scores with those of age-matched peers. In addition to comparing NSS scores to the SALT database, composite NSS scores can be useful for monitoring progress and documenting treatment outcomes by collecting multiple narrative samples from a child and documenting changes in NSS scores over time.

The NSS also provides examiners with the opportunity to identify specific aspects of the narrative process that are difficult for a child. Because the NSS separately judges seven aspects of the narrative process, examiners can evaluate performance on each section of the NSS to identify areas of strength and areas that require intervention. Compared with narrative macrostructure measures that make holistic text-level judgments of narrative proficiency, there is greater specificity in the NSS scoring procedure. Having a detailed narrative performance profile facilitates a more accurate description of the child’s performance and can assist in the development of treatment goals. For example, a child who performed poorly on the referencing and cohesion sections of the NSS but did well on the other sections likely has difficulty using cohesive devices. Treatment goals could include improving the child’s use of referential, lexical, and conjunctive cohesion.

Increasing our understanding of the relationships between vocabulary, grammar, and narrative macrostructure has important clinical implications for documenting functional outcomes and identifying treatment goals. One goal of language intervention programs is for the therapy tasks to generalize to functional tasks. Telling a well-formed narrative is a functional task that is important to children’s academic success (Bishop & Edmundson, 1987; Griffin, Hemphill, Camp, & Wolf, 2004; O’Neill, Pearce, & Pick, 2004). The data from the present study demonstrated that vocabulary skills were uniquely related to children’s story organization skills. Therefore, a treatment procedure that increases a child’s vocabulary skills would have broader, more functional outcomes if concurrent increases in his or her narrative macrostructure skills were documented. Similarly, we may expect that treatments addressing narrative macrostructure skills could also result in...
concurrent increases in vocabulary skills. Clearer understanding of the relationship between microstructures and macrostructures will facilitate selection of appropriate treatment goals. Understanding these relationships can aid in identifying the appropriate microstructures to address when implementing interventions that focus on linguistic macrostructures, such as narrative organization. It is important to note that the data in this study are purely correlational and that further research is needed before causal relationships between vocabulary, grammar, and narrative macrostructure can be identified.

In sum, narrative language assessment is an effective method for documenting children’s language skills. The NSS was developed by the Language Analysis Lab as a clinically useful index of children’s narrative organization skills. Given its clinical feasibility and its robust relationship with other linguistic measures, the NSS provides clinicians and researchers with an additional tool to document children’s global language skills using a functional and curriculum-oriented task.

Acknowledgments

This research was supported by National Institutes of Health Grant 5 T32 DC005459 (“Interdisciplinary Research Training in Speech-Language Disorders”) and by the Language Analysis Lab at the University of Wisconsin—Madison. This work was also made possible through close collaboration with two public school districts in San Diego County: San Diego City Schools and Cajon Valley Union Schools. The authors express appreciation to Karen Andriacchi and the research assistants from the Language Analysis Lab for their assistance with transcription and organization of the data. Portions of this article were presented at the 2006 American Speech-Language-Hearing Association Annual Convention in Miami, FL.

References


Received March 24, 2008
Revision received December 1, 2008
Accepted November 30, 2009
DOI: 10.1044/1058-0360(2009/08-0024)

Contact author: John Heilmann, East Carolina University, Department of Communication Sciences and Disorders, Health Sciences Building, Room 3310T, Greenville, NC 27858-4353. E-mail: heilmannj@ecu.edu.
## The Narrative Scoring Scheme

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Proficient</th>
<th>Emerging</th>
<th>Minimal/immature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
<td>Setting - Child states general place and provides some detail about the setting (e.g., reference to the time of the setting—daytime, bedtime, or season). - Setting elements are stated at appropriate place in story.</td>
<td>Setting - Child states general setting but provides no detail. - Description or elements of story are given intermittently through story. - Child may provide description of specific element of setting (e.g., the frog is in the jar).</td>
<td>- Child launches into story with no attempt to provide the setting.</td>
</tr>
<tr>
<td><strong>Characters</strong></td>
<td>- Main characters are introduced with some description or detail provided.</td>
<td>- Both main and active supporting characters are mentioned. - Main characters are not clearly distinguished from supporting characters.</td>
<td>- Inconsistent mention is made of involved or active characters. - Characters necessary for advancing the plot are not present.</td>
</tr>
<tr>
<td><strong>Character development</strong></td>
<td>- Main character(s) and all supporting character(s) are mentioned. - Throughout story it is clear that child can discriminate between main and supporting characters (e.g., more description of and emphasis on main character[s]). - Child narrates in first person using character voice (e.g., “You get out of my tree,” said the owl).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mental states</strong></td>
<td>- Mental states of main and supporting characters are expressed when necessary for plot development and advancement. - A variety of mental state words are used.</td>
<td>- Some mental state words are used to develop character(s). - A limited number of mental state words are used inconsistently throughout the story.</td>
<td>No use is made of mental state words to develop characters.</td>
</tr>
<tr>
<td><strong>Referencing</strong></td>
<td>- Child provides necessary antecedents to pronouns. - References are clear throughout story.</td>
<td>- Referents/antecedents are used inconsistently.</td>
<td>- Pronouns are used excessively. - No verbal clarifiers are used. - Child is unaware listener is confused.</td>
</tr>
<tr>
<td><strong>Conflict resolution</strong></td>
<td>- Child clearly states all conflicts and resolutions critical to advancing the plot of the story.</td>
<td>- Description of conflicts and resolutions critical to advancing the plot of the story is underdeveloped. OR - Not all conflicts and resolutions critical to advancing the plot are present.</td>
<td>- Random resolution is stated with no mention of cause or conflict. OR - Conflict is mentioned without resolution. OR - Many conflicts and resolutions critical to advancing the plot are not present.</td>
</tr>
<tr>
<td><strong>Cohesion</strong></td>
<td>- Events follow a logical order. - Critical events are included, while less emphasis is placed on minor events. - Smooth transitions are provided between events.</td>
<td>- Events follow a logical order. - Excessive detail or emphasis provided on minor events leads the listener astray. OR - Transitions to next event are unclear. OR - Minimal detail is given for critical events. OR - Equal emphasis is placed on all events.</td>
<td>- No use is made of smooth transitions.</td>
</tr>
</tbody>
</table>
### The Narrative Scoring Scheme

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Proficient</th>
<th>Emerging</th>
<th>Minimal/Immature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusion</td>
<td>-Story is clearly wrapped up using general concluding statements such as &quot;and they were together again happy as could be.&quot;</td>
<td>-Specific event is concluded, but no general statement is made as to the conclusion of the whole story.</td>
<td>-Child stops narrating, and listener may need to ask if that is the end.</td>
</tr>
</tbody>
</table>

Scoring: Each characteristic receives a scaled score of 0–5. Proficient characteristics = 5; Emerging = 3; Minimal/Immature = 1. Scores between (i.e., 2 and 4) are undefined; use judgment. Scores of zero and NA are defined below. A composite is scored by adding the total of the characteristic scores. Highest score = 35.

A score of zero is given for child errors (such as telling the wrong story, conversing with examiner, not completing/refusing task, using wrong language and creating inability of scorer to comprehend story in target language, abandoned utterances, unintelligibility, poor performance, or components of rubric are in imitation-only).

A score of NA (nonapplicable) is given for mechanical/examiner/operator errors (such as interference from background noise, issues with recording such as cut-offs or interruptions, examiner quitting before child does, examiner not following protocol, or examiner asking overly specific or leading questions rather than open-ended questions or prompts).