



The Index of Narrative Microstructure: A Clinical Tool for Analyzing School-Age Children's Narrative Performances

Laura M. Justice
Ryan P. Bowles
University of Virginia, Charlottesville

Joan N. Kaderavek
University of Toledo, Toledo, OH

Teresa A. Ukrainetz
University of Wyoming, Laramie

Sarita L. Eisenberg
Montclair State University, Montclair, NJ

Ronald B. Gillam
University of Texas at Austin

Purpose: This research was conducted to develop a clinical tool—the Index of Narrative Microstructure (INMIS)—that would parsimoniously account for important microstructural aspects of narrative production for school-age children. The study provides field test age- and grade-based INMIS values to aid clinicians in making normative judgments about microstructural aspects of pupils' narrative performance.

Method: Narrative samples using a single-picture elicitation context were collected from 250 children age 5–12 years and then transcribed and segmented into T-units. A T-unit consists of a single main clause and any dependent constituents. The narrative transcripts were then coded and analyzed to document a comprehensive set of microstructural indices.

Results: Factor analysis indicated that narrative microstructure consisted of 2 moderately related factors. The Productivity factor primarily comprised measures of word output, lexical diversity,

and T-unit output. The Complexity factor comprised measures of syntactic organization, with mean length of T-units in words and proportion of complex T-units loading most strongly. Principal components analysis was used to provide a linear combination of 8 variables to approximate the 2 factors. Formulas for calculating a student's performance on the 2 factors using 8 narrative measures are provided.

Conclusions: This study provided a method for professionals to calculate INMIS scores for narrative Productivity and Complexity for comparison against field test data for age (5- to 12-year-old) or grade (kindergarten to Grade 6) groupings. INMIS scores complement other tools in evaluating a child's narrative performance specifically and language abilities more generally.

Key Words: narrative development, narrative assessment, language assessment, school-age language

Disclosure Statement

Ronald B. Gillam is an author of and receives royalties for sales of the Test of Narrative Language (Gillam & Pearson, 2004), which was used for data collection in this research.

Narratives provide an exceptionally rich source of linguistic data for researchers and practitioners who are interested in studying expressive language competencies at the discourse level for children and adolescents. Numerous studies have shown that the narrative skills of

children in certain risk groups (e.g., those with language impairment and those with cognitive disabilities) differ significantly from those of typical children (e.g., Boudreau & Hedberg, 1999; Hayes, Norris, & Flaitz, 1998; Hemphill, Uccelli, Winner, Chang, & Bellinger, 2002; Kaderavek & Sulzby, 2000; Liles, Duffy, Merritt, & Purcell, 1995; Ripich & Griffith, 1988; Scott & Windsor, 2000; van der Lely, 1997). Of particular relevance to speech-language pathologists are additional findings showing that some narrative measures can reliably differentiate between children with and without language impairment (e.g., Liles et al., 1995; Scott & Windsor, 2000). Experts have emphasized the importance of including narrative analysis, a type of language sample analysis, as a routine part of language assessment for persons with suspected or identified language impairment (see Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003).

Presently, there are no “gold standard” guidelines that identify the most salient outcome variables to be studied during narrative assessment. Nonetheless, current best practice suggestions emphasize the importance of studying an individual’s narrative performance at two levels, namely *macrostructure* and *microstructure* (e.g., Hughes, McGillivray, & Schmidek, 1997; Owens, 1999; Paul, 2001). Macrostructural analysis examines children’s narrative abilities in terms of higher order hierarchical organization. It typically focuses on children’s inclusion of story grammar components and the complexity of episode structure. During macrostructural analysis, the speech-language pathologist might, for example, determine the number of episodes (i.e., segments that include an initiating event, attempt, and consequence) contained in a child’s narrative (e.g., Liles et al., 1995; Merritt & Liles, 1989), along with use of other narrative elements such as appendages, orientations, and evaluations (Labov, 1972; Ukrainetz et al., 2005). Microstructural analysis, by contrast, considers the internal linguistic structures used in the narrative construction, such as conjunctions, noun phrases, and dependent clauses. Discriminant analysis has shown macro- and microstructural variables to represent two distinct underlying areas of narrative competence (Liles et al., 1995; also see Hayes et al., 1998). The present research focuses on the latter aspect of narrative performance, namely the study of children’s narrative performance at the microstructural level.

Theoretical perspectives on narrative development suggest that narrative analysis should examine both macro- and microstructural aspects of performance. Specifically, Bock (1982) and Hargrove, Frerichs, and Heino (1999) have suggested that narrators have a *finite linguistic capacity*. This idea is exemplified in Crystal’s (1987) “bucket” theory, which suggests that there are trade-offs in language parameters across various language tasks. For example, there could be trade-offs in the extent of syntactic precision as children produce language that is conceptually advanced. Application of this theory to narrative production suggests that a child who is attempting to produce increasingly complex narrative macrostructure may demonstrate reduced complexity at a microstructural level, and vice versa. Or, put another way, to produce a narrative that is both well organized and aesthetically pleasing that also contains complex structural features

may pose a particular challenge (e.g., Baltaxe & D’Angiola, 1992; Copmann & Griffith, 1994; Gillam & Johnston, 1992; Kaderavek & Sulzby, 2000; Liles, 1985, 1987; Liles et al., 1995; McFadden & Gillam, 1996; Merritt & Liles, 1989; Paul & Smith, 1993; Scott & Windsor, 2000; Southwood & Russell, 2004; Strong & Shaver, 1991). Thus, it is important for speech-language pathologists to have tools by which to qualify and quantify a child’s narrative performance in both macro- and microstructure.

Our interest in the microstructural aspect of narration was derived from several studies showing microstructural indices to be particularly sensitive for characterizing a child’s linguistic competence and for identifying children with clinically depressed language skills, such as children with language impairment (see Feagans & Short, 1984; Liles et al., 1995; Ripich & Griffith, 1988; Scott & Windsor, 2000; van der Lely, 1997). Liles et al. (1995), for instance, showed that a single factor representing microstructural variables more readily differentiated children with language impairment from those without impairment relative to a factor representing macrostructure variables. Although there are a number of tools available for macrostructural analyses (e.g., Strong, 1998), relatively few are available for microstructural analyses. With this research, we developed a clinical tool that can be used by practitioners to derive an estimate of microstructural performance relative to age- and grade-based peers.

Narrative Analysis and Its Clinical Uses

There is a rapidly accumulating body of research suggesting that narrative assessment is a valid, sensitive, and potentially less biased language analysis tool relative to norm-referenced standardized language assessments (e.g., see Craig & Washington, 2000; Muñoz et al., 2003; Thompson, Craig, & Washington, 2004; Washington, Craig, & Kushmaul, 1998). Two more common approaches to interpretation of narrative outcome data are norm referencing and criterion referencing (Johnson, 1995). Norm referencing occurs when a student’s narrative data are compared against those of same-age or same-grade peers; it is used to determine whether the student’s performance is sufficiently different from age- or grade-based expectations to warrant the diagnosis of impairment. This approach is diagnostic, in that the narrative data are used as an identification tool (Stockman, 1996). Norm referencing can also be used as a progress-monitoring tool in intervention to document a child’s developing skills, specifically whether their skills are becoming more similar to those of typical peers during intervention. Criterion referencing occurs when a student’s narrative data are compared against specific criterion-level benchmarks. Here, interpretation involves identifying the specific level of performance on a given task. Criterion-referenced data are often used to support diagnoses derived from norm referencing, to identify a specific focus for intervention, and to monitor an individual’s progress during intervention (McCauley, 1996).

Because few standardized tests of narrative discourse have historically been available, clinical professionals have often

utilized research reports of the distributional characteristics of narrative indices to interpret children's performance for norm- and criterion-referenced purposes (cf. Greenhalgh & Strong, 2001; Hughes et al., 1997; Liles et al., 1995; Loban, 1976; Strong, 1998). This literature includes numerous studies detailing the age- and grade-related changes evident in both macro- and microstructural elements of children's narratives from preschool into adolescence (e.g., Dickinson, Wolf, & Stotsky, 1993; Liles, 1985; McCabe & Peterson, 1991; McCabe & Rollins, 1994; Stein, 1982), as well as the differences seen in narrative ability when comparing children from various subpopulations (e.g., Hemphill et al., 2002; Kaderavek & Sulzby, 2000; Liles et al., 1995). Johnson (1995) noted that published research was particularly useful for professionals to draw upon for analyzing narrative outcome data given the proliferation of both single studies and published reviews of narrative development in recent decades. When using published norms, professionals compare some measure of productivity or complexity derived from a child's narrative performance (e.g., mean length of T-units, with T-units representing a single main clause and any dependent constituents, including clauses and phrases) against available norms. Norms are used to identify the presence or severity of a disorder, often employing a specific cut-point (e.g., 10th or 25th percentile) for identifying children whose performance differs substantially from same-age peers. Published norms may also be useful for setting specific age- or grade-based criterion levels for criterion referencing.

There are some challenges in the clinical practice of translating microstructural narrative measures obtained from a student against published data for the purpose of norm referencing and criterion referencing. An initial challenge is that clinicians must find and select the published data by which to interpret the student's narrative measures and to set criterion standards. Clinicians who search through recent primary research looking for narrative data will often find summative group scores that either reflect a small subset of linguistic variables across a broad range of ages (e.g., Greenhalgh & Strong, 2001; Loban, 1976) or a broader set of linguistic variables across relatively few age levels (e.g., Liles et al., 1995; Strong, 1998). In turn, these may not reflect a comparable age range for the narrative measures being studied. Hughes et al. (1997) aggregated data from a variety of small and large studies conducted during the 1970s and 1980s that provide a larger age spectrum for interpreting narrative data. Likewise, Strong (1998) provided field test data from 78 children age 8–10 years (39 typically developing and 39 with language impairment). However, because each of these studies examined a different set of linguistic features, varied in specific elicitation procedures, and drew a relatively small number of participants primarily from local convenience samples, the clinical usefulness may be limited. Thus, as Strong recommends, speech-language pathologists may need to collect local field test data to develop local norms for use with narrative assessment, likely to be a challenge for many professionals due to constraints of time, funding, materials, and access to children.

As an alternative to collecting their own field test data, speech-language pathologists may look to reference databases

serving as companions to the Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 2000) software program. Narrative references are included in the SALT database for several large cohorts of children, including 200 kindergarten through third-grade students from ethnically and linguistically diverse schools in San Diego, CA, and 348 children age 3–13 years from Wisconsin (this sample did not include 8-, 10-, and 12-year-olds). Although these databases provide valuable references for interpreting children's narrative skills, their utility may be compromised if clinicians do not use the same elicitation protocols used by the developers of the databases. Additionally, a second challenge is that clinicians must decide which narrative measures to use when norm referencing and criterion referencing. The possibilities are numerous, including total number of T-units, mean length of T-units (in words or morphemes), mean length of dependent clauses, total number of words, total number of different words, and so forth. A professional who turns to the available published norms might, for instance, find that a student's narrative contains a below-average number of different words but a higher-than-average number of T-units. In such cases, it is not clear whether the student's narrative skills should be considered below average, above average, or somewhere in the middle.

Aims of the Present Work

This research was conducted to resolve the "clinical dilemmas" identified above by designing a clinical tool—the Index of Narrative Microstructure (INMIS)—that can be used by professionals for microstructural analysis of children's narratives. INMIS has several desirable properties, making it amenable to field-based use by clinical professionals. First, the stimulus used to elicit children's narratives is available in a commercially available standardized test of narrative performance; thus, speech-language pathologists can readily use the same elicitation stimulus as that used in INMIS development. Second, INMIS requires the analysis of a select number of microstructural aspects of narrative performance, all of which can be readily calculated by hand or using commercially available language-analysis computer software. Third, field test data provide normative references by which to compare a child's narrative performance for the age range of 5 to 12 years or, alternatively, for the grade range of kindergarten to sixth grade. As one component of a comprehensive language assessment, INMIS may serve as a screening or diagnostic tool to quantify and qualify a student's language ability and possibly to monitor narrative growth over a course of intervention. To achieve this broader purpose, two specific aims were addressed.

The first aim was to examine the dimensionality of microstructural elements of narrative performance. There are diverse ways to characterize the microstructural aspects of narratives, and it is not clear whether measures of these various characteristics represent a single underlying dimension of microstructural performance or rather represent several dimensions. Often, microstructural variables are differentiated into those representing productivity (i.e., volume of output in words or syntactic units, such as total number

of words) and complexity (i.e., grammatical complexity, such as mean length of T-units), but to our knowledge these theoretical distinctions have not been examined empirically. If one or two dimensions of microstructural performance could be identified empirically, this may present a more parsimonious approach to microstructural analysis, as professionals could compare a child's performance on these dimensions against published norms and use these dimensions as a means for monitoring narrative growth.

The second aim was to design a tool derived from analysis of the dimensionality of microstructural performance that could be used to analyze school-age children's narrative performance. By calculating children's scores on the factors identified in the first aim, clinical professionals could compare a child's microstructural performance against that of age- or grade-based peers derived from our field test data of children. Quantitative estimates of children's microstructural aspects of narrative production can serve as an important component of a comprehensive language assessment battery.

Method

Participants

The 250 children in this study were participants in the standardized norming procedures for the Test of Narrative Language (TNL; Gillam & Pearson, 2004). The norming procedures took place at urban, suburban, and rural public and independent elementary schools in nine states. Schools were recruited individually by a larger group of norming collaborators. From this group, five university researchers involved in the norming process formed a collaborative network to develop a comprehensive database of transcribed narratives from a random sample of students involved in norming at their sites. Thus, the development of this database was supported by, but independent of, the test norming process. In developing the database, each of the five collaborators was to randomly select about 50 transcripts from their sample for full transcription. Although it would have been desirable to fully transcribe all narratives collected at each site, the personnel and financial investments for doing so would have been prohibitive. The number of transcripts per site ranged from 35 transcripts to 66 transcripts as a function of local resources (e.g., access to transcribers). Nonetheless, the data set used in this study represents a broad cross-section of geographic regions across the United States to include the Southwest (Texas), the Midwest (Ohio), the West (Wyoming), the Northeast (New Jersey), and the Mid-Atlantic coast (Virginia).

The transcribed narratives were collected from 250 children age 5–12 years (128 male, 122 female; see Table 1), with a mean of 50 participants ($SD = 13.7$) per state. There were 54 kindergarteners, 29 first graders, 31 second graders, 47 third graders, 30 fourth graders, 33 fifth graders, and 26 sixth graders. The children attended seven schools, of which four were general public (Wyoming, New Jersey, Virginia), two were independent/private (Ohio, Texas), and one was a charter school (Texas). The percentage of students in the schools qualifying for free/reduced lunch (as a general indi-

TABLE 1. Participant characteristics ($n = 250$).

Age group	<i>n</i>	Age		Age range	Boys/girls
		<i>M</i>	<i>SD</i>		
5-year-old	29	5.7	0.21	5.0–5.9	13/16
6-year-old	38	6.4	0.26	6.0–6.9	14/24
7-year-old	31	7.4	0.3	7.0–7.9	15/16
8-year-old	41	8.5	0.29	8.0–8.9	23/18
9-year-old	35	9.5	0.28	9.0–9.9	18/17
10-year-old	28	10.4	0.31	10.0–10.9	16/12
11-year-old	25	11.3	0.2	11.0–11.7	15/10
12-year-old	23	12.2	0.2	12.0–12.8	14/9

cator of a school population's socioeconomic status) ranged from 1% to 76%, averaging 28% across the schools with a median of 18%. This was below the national average of 39.9% (U.S. Department of Education, 2003).

Data were collected using full population sampling; therefore, no exclusionary criteria were applied to the participation of individual students in the norming process with the exception of provision of informed consent by caregivers, presence on the assigned day of testing, and completion of the majority of the narrative tasks. Thus, although the sample was primarily children from general education who received no special services (95.2%, $n = 238$), about 5% of the sample was currently receiving special education for mild-to-moderate disabilities (e.g., attention deficit disorder, articulation impairment). The sample reflected the ethnic/racial distributions of the schools from which the students were drawn. In total, the sample was 80% non-Hispanic White ($n = 200$), 10% Hispanic White ($n = 26$), 8% Black ($n = 19$), 1% Asian/Pacific Islander ($n = 3$), and 1% "other" ($n = 2$). These demographics differed from that of the nation at large, in which 60.3% of elementary-grade pupils are non-Hispanic White, 17.1% are Hispanic White, and 17.2% are Black (U.S. Department of Education, 2006).

General Procedure

Narrative collection. Prior to collection of the data, five of the six authors of this study met at a central location to review the assessment protocol. Narratives were then collected over a 6-month period from children in their school environments by teams of test administrators working at each site. Test administrators included university researchers as well as trained, supervised research assistants who were either graduate or undergraduate speech-language pathology majors. A total of three spoken narratives were collected from each student following the standardized administration protocol designed by Gillam and Pearson (2004) for the TNL. All children were tested in individual 20-min sessions in a variety of niches within their schools (e.g., an empty classroom, private office, or conference room). The test administration instructions (Gillam & Pearson, 2004) provided explicit guidelines on allowable prompts and the sequence of the narrative tasks to ensure uniformity across administration contexts and personnel. Because a principal goal of the data collection procedures was the norming of a standardized test, adherence

to the manualized instructions was strictly followed for all participants.

Administration of the TNL involved the students' production of three spoken narratives that varied according to the stimuli used to elicit the narrative. Namely, children produced narratives in three formats in the following order: (a) with no picture cues (story retelling), (b) with five sequenced pictures (story generation), and (c) with a single picture depicting a fictional event (story generation). For each task, a sample story was produced by the examiner, the children were asked a series of comprehension questions about the story, and then the children were asked to produce their own story. No prompts were allowed, with the exception of several scripted examiner queries when children paused for long periods of time or when children appeared to be finished (e.g., "Is there anything else you would like to add?"). Children's narratives were tape-recorded in their entirety and were subsequently transcribed in university laboratories following a protocol established by the research collaborators.

The data presented in the current study are derived specifically from the narratives children produced in the third presentation format, which involved a single picture depicting a fictional event. These narratives represented children's performance on the most open-ended of the narrative tasks presented, and they provided the least constraints to variability in microstructural production. In this task, the children were first told a story by the examiner while looking at a single illustration of two children looking at a dragon guarding a treasure chest in a cave. The children were told to listen carefully to the story, as they would be asked to make up their own story after listening to the examiner's story. The scripted model story featured a fictional event (that of children finding a treasure guarded by a dragon) with a fully elaborated story grammar (i.e., containing details of setting, characters) with an initiating event, internal responses by the characters, a plan, an attempt, a consequence, and a resolution. It also contained models of complex syntax, including elaborated noun phrases, past tense verb phrases, and lexical, pronominal, and reference cohesion. Following the model story, examiners provided a second illustration depicting two children watching an alien family walk off a spaceship that had landed in a park. The children were instructed to produce their own story to go with the picture. Participants were reminded that stories have "a beginning, things that happen, and an ending" and were instructed to tell the best story possible. During their production, no reinforcement or praise was provided. When the participants appeared to be finished with their story, the examiner asked whether they wanted to add any additional information. When the children's story production appeared complete, they were thanked and returned to their classrooms. Three examples of narratives produced by the children are provided in Appendix A.

Narrative transcription. The children's narratives were transcribed by trained research personnel at university laboratories according to the conventions of SALT (Miller & Chapman, 2000). We used an iterative process to check the accuracy of transcripts and T-unit segmentation. Following the initial transcription, a second examiner independently listened to each audiotape while simultaneously examining the transcript in its entirety for errors in transcription or segmentation.

The original and second examiner then discussed discrepancies until they reached agreement on each discrepancy.

At the time of transcription, each narrative was segmented into T-units (Hunt, 1965) for analytical purposes. A T-unit consists of a single main clause and any dependent constituents, including clauses and phrases. By this definition, the following narrative is five T-units in length (each T-unit is bracketed): "[That boy is hiding with the girl.] [He grabs her and pulls her.] [They run now.] [They are running.] [And they tell their mom and dad.]"

Children's sentences comprising a series of successive main clauses, linked by coordinating or subordinating conjunctions, were segmented according to Hunt's (1965) directions, also described in Hughes et al. (1997). T-unit segmentation is a common tool for parsing narrative utterances (spoken or written) into reliable units. Utterances were included if they had a predicate structure, even if the subject was omitted due to ellipsis. Judgments of clausal subordination and coordination followed guidelines presented in Justice and Ezell (2002), except for infinitival verb constructions, which were coded as clausal units rather than phrasal units. During transcription, all concatenatives (*wanna*, *gonna*) were expanded into infinitival clauses (*want to*, *going to*) because of concerns about transcription reliability in differentiating concatenatives from infinitives (e.g., *wanna* vs. *want ta* vs. *want to*). The participants were all school-age, well past the age at which children typically develop productive control of infinitival clauses (Eisenberg, 1997; Menyuk, 1969). Thus, counting all concatenative forms as two words versus a single nongenerative form should not have affected data analyses.

Narrative Measures

We selected a set of narrative indices that reflected both productivity and complexity as identified in the extant literature (e.g., Baltaxe & D'Angiola, 1992; Copmann & Griffith, 1994; Feagans & Short, 1984; Gillam & Johnston, 1992; Liles, 1985, 1987; McFadden & Gillam, 1996; Merritt & Liles, 1989; Muñoz et al., 2003; Paul & Smith, 1993; Ripich & Griffith, 1988; Schneider & Winship, 2002; Scott & Windsor, 2000; Southwood & Russell, 2004; Strong & Shaver, 1991). We did not include measures that required a relatively fine-grained and sophisticated analysis of microstructure, such as identifying the number of grammatically acceptable complex T-units, examining the adequacy (i.e., completeness) of lexical and reference cohesive ties, and quantifying the quality of causal coherence (e.g., Gutierrez-Clellan & Iglesias, 1992; McFadden & Gillam, 1996). Although we appreciate that such measures provide significant information concerning children's narrative performance, particularly measures of the adequacy of cohesive ties (e.g., conjunctions), we selected measures that seemed most amenable to reliable field-based use by clinical professionals. For the most part, the measures selected can be readily computed using commercial language analysis software. Protocols used to compute each of the variables are described in Appendix B.

The three indices for productivity included: total number of words (TNW), total number of different words (NDW), and total number of T-units (LENGTH). TNW and NDW are

frequently used measures of lexical diversity reflecting differences in child vocabulary use (Klee, 1992; Miller, 1991; Muñoz et al., 2003; Owen & Leonard, 2002; Strong, 1998; Watkins, Kelly, Harbers, & Hollis, 1995). LENGTH was included as a measure of story output because language output improves with language maturity/ability and chronological age (Atkins & Cartwright, 1982; Evans & Craig, 1992; Southwood & Russell, 2004; Stalnaker & Craghead, 1982; Strong, 1998; Wren, 1985).

Six indices were included to represent structural complexity during narrative production: mean length of T-units in words (MLT-W), mean length of T-units in morphemes (MLT-M), total number of T-units that contained two or more clauses (COMPLEX), total number of coordinating conjunctions (COORD), total number of subordinating conjunctions (SUBORD), and proportion of complex T-units (PROPCOMPLEX). Inclusion of the mean length of utterance in morphemes and words reflected Brown's (1973) seminal research indicating that children's linguistic growth can be readily monitored by studying the length of utterances in morphemes and words. Examining the number of morphemes and/or words within specific linguistic units (e.g., utterances and clauses) is a predominant approach to monitoring grammatical complexity of narrative performance in pre-school and school-age children (e.g., Chapman, Seung, Schwartz, & Kay-Raining Bird, 1998; Curenton & Justice, 2004; Gillam & Johnston, 1992; Kaderavek & Sulzby, 2000; Liles et al., 1995; Muñoz et al., 2003; Peterson, Jesso, & McCabe, 1999; Scott & Windsor, 2000; Strong, 1998). The four measures of more complex syntax—namely, COMPLEX, COORD, SUBORD, and PROPCOMPLEX—were included because clausal embedding and conjunction use (for both coordination and subordination) are important cohesive devices in narrative production and show developmental effects as children mature linguistically (e.g., Bloom, Lahey, Hood, Lifter, & Fiess, 1980; Clancy, Jacobsen, & Silva, 1976; Curenton & Justice, 2004; Liles et al., 1995).

Analytical Approach

Two aims were addressed in this study: (a) to examine the dimensionality of microstructural elements of narrative performance and (b) to design a tool derived from analysis of the dimensionality of microstructural performance that can be used to analyze school-age children's narrative performance. To accomplish the first aim, we used exploratory factor analysis to identify the dimensions accounting for variance and covariance among the microstructural variables. Statistically, factor analysis is a technique used for data reduction that reduces a larger number of overlapping variables to a smaller set of factors; theoretically, the factors that emerge from this procedure identify the specific constructs (i.e., unobserved latent constructs) that represent dimensions of a given skill or behavior (Green, Salkind, & Akey, 2000). We thus used factor analysis to identify the specific constructs represented by a relatively large number of measures of microstructural narrative performance. To accomplish the second aim, we estimated factor scores for the dimensions of narrative microstructural performance that were identified when addressing the first

aim. We developed a formula for estimating factor scores for an individual's observed microstructural measures and confirmed the validity of the formulas based on procedures from Grice (2001) to ensure that our formulas provided valid estimates of an individual's true factor scores.

Results

Descriptive Statistics

Raw descriptive data for the nine microstructural variables by age group are presented in Table 2, and a table of correlation coefficients is presented in Table 3. Examination of the descriptive data presented in Table 2 shows a general linear increase in the means for all narrative measures through the age of 10 years, with a plateau in performance evident at this latter age. For the 11- and 12-year-olds, the descriptive data show a decrease in performance on all measures, with the 12-year-olds performing less well than the 11-year-olds on all measures except MLT-M and MLT-W. Also of note are the relatively large standard deviations for all measures, indicating large heterogeneity in performance within each of the age groups. Examination of the correlational data in Table 3 shows variability in the strength of associations among the dependent measures. Although the correlations among the three Productivity measures (TNW, NDW, LENGTH) were consistently strong, the correlations among the Complexity measures ranged from weak (e.g., MLT-W with COMPLEX) to strong (e.g., COMPLEX with COORD). Interestingly, some correlation coefficients representing the relationship between Productivity and Complexity measures were quite strong (e.g., TNW with COMPLEX).

Dimensionality of Microstructural Measures

The first aim of this research was to examine the dimensionality of measures of narrative microstructure, comprising nine measures: TNW, NDW, LENGTH, MLT-W, MLT-M, COMPLEX, COORD, SUBORD, and PROPCOMPLEX. We conducted an exploratory factor analysis in SAS with principal factor extraction, squared multiple correlation communalities estimates, and promax rotation (power = 4). The exploratory factor analysis yielded two clear factors based on the Kaiser-Guttman rule (eigenvalues of correlation matrix > 1), eigenvalues greater than average in the correlation matrix with estimated communalities on the diagonal, and a scree plot. The factor structure of the first model may have been influenced by the high correlation (.99) between MLT-W and MLT-M; thus, MLT-M was omitted as a variable in a rerun of the factor analysis. The results were similar to the initial factor analysis, but the reduced analysis had a somewhat clearer factor structure; therefore, we report only results with the eight variables (omitting MLT-M but retaining MLT-W). The two factors accounted for 76% of the variance in the narrative microstructure correlation matrix, and the correlation between the factors was .39. Eigenvalues are presented in Table 4, and standardized factor loadings are presented in Table 5. (Additional details are available by request.) The first factor consisted of the

TABLE 2. Means and standard deviations for narrative measures by age group.

Age	n	TNW		NDW		LENGTH		MLT-W		MLT-M		COMPLEX		COORD		SUBORD		PROPCOMPLEX	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
5	29	68	47	39	20	8.5	5.4	6.8	1.7	7.6	1.8	3.1	3.2	0.7	1	0.8	1.4	.33	0.2
6	38	77	54	43	22	9.6	6	7.5	1.6	8.3	1.6	3.5	2.8	1	1.3	0.5	0.8	.37	0.2
7	31	96	74	52	28	11.3	9.1	8.5	3.8	9.5	4.3	4.6	4.3	0.6	0.8	0.9	1.2	.38	0.2
8	41	137	77	69	27	15.8	8.9	8.1	1.4	9	1.5	7.6	5.2	2.2	1.6	1.3	1.4	.45	0.2
9	35	162	96	79	30	17.3	9.6	8.4	1.4	9.4	1.6	8.9	6.1	2.2	1.9	1.5	1.4	.51	0.2
10	28	237	196	101	49	21.5	14.5	8.9	2.1	10	2.4	12.2	9.8	3.2	2.5	2.6	2.6	.55	0.2
11	25	167	70	84	27	18.8	8.2	8.7	1.4	9.6	1.6	9.2	4.3	2.4	2.3	1.7	1.6	.5	0.2
12	23	148	95	67	36	14.6	8.7	8.8	1.9	9.7	2.1	7.3	4.4	2.2	1.8	1.5	2	.5	0.2

Note. TNW = total number of words; NDW = total number of different words; LENGTH = total number of T-units; MLT-W = mean length of T-units in words; MLT-M = mean length of T-units in morphemes; COMPLEX = total number of T-units that contained two or more clauses; COORD = total number of coordinating conjunctions; SUBORD = total number of subordinating conjunctions; PROPCOMPLEX = proportion of complex T-units.

variables NDW, TNW, LENGTH, COMPLEX, and, to a lesser degree, COORD and SUBORD. The second factor consisted of MLT-W and PROPCOMPLEX, with SUBORD having a small factor loading. To confirm this factor structure, we also split the sample at random and by gender, and reran the factor analysis. Results were consistent in all cases.

Clearly, this factor structure did not precisely match the hypothesized productivity and complexity groupings, which would have resulted in loadings of TNW, NDW, and LENGTH on one factor (Productivity), and MLT-W, COMPLEX, COORD, SUBORD, and PROPCOMPLEX loading on a second factor (Complexity). Nonetheless, the factor structure showed that the microstructural aspects of narrative production do not represent a single unidimensional construct and rather comprise two separate and only moderately related dimensions. For the present purposes, we retain the use of the terms Productivity and Complexity to describe these dimensions, with the refined perspective that the former references word output and diversity (NDW, TNW), T-unit output (LENGTH, COMPLEX), and, to a lesser degree, conjunction output (COORD, SUBORD), whereas the latter references syntactic organization (MLT-W, PROPCOMPLEX) and, to a lesser degree, use of subordinate conjunctions (SUBORD).

We then estimated factor scores using the correlation-preserving formula of ten Berge, Krijnen, Wansbeck, and

Shapiro (1999). The factor score formulas, scaled to a mean of 0 and standard deviation of 1, for the variables in their original metrics (i.e., unstandardized) are:

$$\begin{aligned} \text{Productivity} = & -1.60 + (-0.0010 \times \text{MLT-W}) \\ & + (-0.21 \times \text{PROPCOMPLEX}) + (0.017 \times \text{NDW}) \\ & + (-0.00054 \times \text{TNW}) + (0.014 \times \text{COORD}) \\ & + (0.0072 \times \text{SUBORD}) + (0.0094 \times \text{LENGTH}) \\ & + (0.068 \times \text{COMPLEX}). \end{aligned}$$

$$\begin{aligned} \text{Complexity} = & -2.84 + (0.27 \times \text{MLT-W}) \\ & + (0.85 \times \text{PROPCOMPLEX}) + (0.012 \times \text{NDW}) \\ & + (-0.0027 \times \text{TNW}) + (0.028 \times \text{COORD}) \\ & + (0.026 \times \text{SUBORD}) + (-0.085 \times \text{LENGTH}) \\ & + (0.14 \times \text{COMPLEX}). \end{aligned}$$

We confirmed the validity of the factor score estimates using Grice's (2001) procedure (see Grice, 2002, for the general SAS script). The determinacy indices for Productivity and Complexity were .985 and .903, respectively. Values substantially above .71 indicate good factor determinacy; thus, the present values are quite high. The validity coefficients were .985 and .902, also substantially above Gorsuch's (1983, p. 260) .80 cutoff. Finally, the univocality of .393 was close to the factor correlation of .391, also indicating good validity of the factor score estimates. Thus, we conclude that the factor score formulas provide valid estimates of the true factor

TABLE 3. Intercorrelations among narrative measures.

Variables	1	2	3	4	5	6	7	8	9	10
1. TNW	—	.95**	.79**	.17**	.16*	.78**	.55**	.54**	.32**	.39**
2. NDW		—	.82**	.25**	.24**	.81**	.62**	.59**	.38**	.50**
3. LENGTH			—	.07	.06	.92**	.58**	.47**	.25**	.36**
4. MLT-W				—	.99**	.21**	.24**	.33**	.58**	.27**
5. MLT-M					—	.20**	.23**	.33**	.57**	.27**
6. COMPLEX						—	.63**	.54**	.52**	.40**
7. COORD							—	.46**	.37**	.38**
8. SUBORD								—	.34**	.30**
9. PROPCOMPLEX									—	.32**
10. Age										—

**Significant at the .01 level (two-tailed).

TABLE 4. Two-factor solution of narrative microstructure.

Factor	Initial eigenvalue	% of variance	Rotation sums of squared loadings
1	4.73	59.1	4.39
2	1.37	17.2	2.00
3	0.61	7.6	
4	0.52	6.5	
5	0.40	5.0	
6	0.29	3.6	
7	0.05	0.6	
8	0.03	0.4	

scores. For details on the meaning and interpretation of these statistics, the reader is encouraged to consult Grice (2001).

Table 6 provides the distributional statistics for the Productivity and Complexity factors as identified using the factor score estimates. These data provide an index of microstructural performance on the Productivity and Complexity factors for 5–12-year-old children. Alternatively, Table 6 also provides values as a function of grade for kindergarten through sixth graders. We refer to these values as INMIS scores, the use of which provides clinical professionals a protocol for interpreting children’s microstructural narrative performance on two dimensions.

The INMIS data provided in Table 6 show that children’s narrative performance was not normally distributed. Median scores are generally below the mean scores, and the skewness statistics are substantially greater than 0 for most ages and grades as well as for the complete sample (most were more than 2 *SEs* from 0). This is particularly evident for Complexity. The distribution of INMIS scores for specific age/grade cohorts and the entire sample more generally indicates that the left tail of the distribution is thicker than the right tail, showing that children’s performance clusters toward the lower side of the distribution and that neither Complexity nor Productivity is normally distributed. We include in Table 6 the skewness and kurtosis statistics for the two narrative variables at each age and grade level. To aid in interpreting these statistics, it is helpful to know that a normal distribution has a skewness statistic of 0. A positively skewed distribution (positive skewness statistic) occurs when scores bunch up at the low end of the scale, and a negatively skewed distribution occurs when scores bunch up at the high end of the scale.

TABLE 5. Standardized regression coefficient factor loadings for the two-factor solution.

Variable	Factor	
	Productivity	Complexity
TNW	.921	-.057
NDW	.937	.023
LENGTH	.987	-.199
MLT-W	-.106	.831
COMPLEX	.892	.082
COORD	.596	.159
SUBORD	.494	.248
PROPCOMPLEX	.146	.694

A normal distribution has a kurtosis statistic near 0. A leptokurtic distribution (positive kurtosis statistic) is significantly too tall, and a platykurtic distribution (negative kurtosis statistic) is significantly too flat. In short, these results show that narrative performance—at least in terms of microstructural indicators—is not a normally distributed ability, and that children’s performance clusters in the lower end of the distribution.

In Table 6 we also provide age- and grade-based 10th and 25th percentiles. Because of the small sample sizes within grade and age groups and the lack of a normal distribution, we report a weighted moving average in order to smooth the distributions. For the extreme ages and grades, the weighted average is equal to 0.67 times the group norm plus 0.33 times the adjacent group norm. For nonextreme groups, the weighted average is equal to 0.5 times the group norm plus 0.25 times each of the adjacent norms.

Discussion

Overview of Major Findings

Our first aim was to examine the dimensionality of microstructural elements of school-age children’s narrative performance using data from a sample of 250 children in kindergarten through sixth grade sampled from five states. Factor analysis conducted on eight variables indicated a clear two-factor solution, thus suggesting that microstructural performance is best conceived as two dimensions that are only moderately related. However, the two factors did not match the a priori hypothesized factor structure groupings, which hypothesized Productivity to comprise TNW, NDW, and LENGTH and Complexity to include MLT-W, COMPLEX, COORD, SUBORD, and PROPCOMPLEX. Rather, the first factor was accounted for by NDW, TNW, LENGTH, COMPLEX, and, to a lesser degree, COORD and SUBORD, and the second factor was accounted for by MLT-W, PROPCOMPLEX, and, to a lesser degree, SUBORD. Although we maintained the terms Complexity and Productivity to reference the bidimensionality of narrative microstructure, the results provided evidence that the variable groupings often used in the literature (cf. Boudreau & Hedberg, 1999; Feagans & Short, 1984; Scott & Windsor, 2000) to theoretically organize microstructural measures do not empirically conform to the underlying factor structure. The present research (a) confirms that narrative microstructure is a multidimensional construct and (b) provides an empirical basis for organizing measures of microstructure into scientifically validated categories.

The second aim for this work was to design a clinical tool by which professionals could analyze children’s narrative microstructural performance. We noted earlier that clinicians who desire to utilize narrative measures in language assessment face the dual challenges of (a) deciding which microstructural measures to use to evaluate narrative performance from among the numerous possibilities detailed in the literature and (b) identifying data sources against which one might compare a child’s microstructural performance to estimate a child’s narrative abilities. Subsequent to identifying a clear two-factor model of microstructural performance, we

TABLE 6. Index of Narrative Microstructure: distributional statistics by age and grade.

Group	<i>n</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	Skewness	Kurtosis	10%	25%
Productivity								
5-year-old	29	-0.74	-0.84	0.55	1.83	4.26	-1.30	-1.04
6-year-old	38	-0.64	-0.79	0.58	1.28	1.73	-1.25	-1.02
7-year-old	31	-0.42	-0.57	0.80	1.91	5.81	-1.09	-0.85
8-year-old	41	0.09	-0.07	0.84	0.79	0.70	-0.82	-0.53
9-year-old	35	0.35	0.21	0.93	1.52	3.47	-0.54	-0.25
10-year-old	28	0.95	0.49	1.37	2.05	6.30	-0.43	-0.04
11-year-old	25	0.47	0.44	0.77	0.04	-0.11	-0.55	-0.05
12-year-old	23	0.21	-0.08	0.87	0.92	1.27	-0.69	-0.22
Kindergarten	54	-0.71	-0.82	0.56	1.67	3.29	-1.27	-1.04
First grade	29	-0.69	-0.86	0.58	1.12	0.93	-1.19	-0.93
Second grade	31	-0.05	-0.18	0.78	1.33	3.48	-0.97	-0.65
Third grade	47	0.22	0.05	0.93	1.54	3.22	-0.67	-0.36
Fourth grade	30	0.64	0.44	0.85	0.16	-1.23	-0.52	-0.13
Fifth grade	33	0.66	0.38	1.35	2.23	7.53	-0.60	-0.08
Sixth grade	26	0.34	0.41	0.82	0.66	0.94	-0.67	-0.14
Full sample	250	0.00	-0.18	1.00	1.57	5.23	-1.04	-0.76
Complexity								
5-year-old	29	-0.70	-0.81	0.79	-0.01	-0.73	-1.55	-1.24
6-year-old	38	-0.47	-0.51	0.71	-0.05	-0.56	-1.34	-1.07
7-year-old	31	-0.12	-0.38	1.30	3.52	15.05	-1.05	-0.80
8-year-old	41	0.00	0.02	0.76	-0.12	0.07	-0.89	-0.56
9-year-old	35	0.25	0.16	0.91	-0.13	-0.43	-0.80	-0.44
10-year-old	28	0.64	0.41	1.23	1.01	0.85	-0.67	-0.38
11-year-old	25	0.28	0.14	0.71	0.13	-0.36	-0.60	-0.25
12-year-old	23	0.37	0.27	0.92	0.68	2.44	-0.64	-0.16
Kindergarten	54	-0.63	-0.63	0.77	-0.05	-0.74	-1.42	-1.08
First grade	29	-0.14	-0.42	1.28	3.78	17.55	-1.21	-0.85
Second grade	31	-0.18	-0.24	0.76	0.73	2.06	-0.97	-0.67
Third grade	47	0.10	0.05	0.78	-0.00	-0.24	-0.88	-0.58
Fourth grade	30	0.37	0.30	0.99	-0.02	-0.90	-0.81	-0.44
Fifth grade	33	0.49	0.27	1.17	1.05	1.71	-0.58	-0.23
Sixth grade	26	0.45	0.41	0.77	1.40	3.77	-0.36	-0.07
Full sample	250	0.00	-0.03	1.00	1.25	4.94	-1.19	-0.63

used a factor score estimation technique (ten Berge et al., 1999) to develop a linear equation for each factor that could be used to compute complexity and productivity scores for a given child.

INMIS scores may be calculated and interpreted in the following way. First, clinicians calculate the eight microstructural measures on a given narrative sample, and then clinicians enter these numbers into the formulas provided. Second, a child's scores on the two factors are compared against field test reference data based on age or grade level. In this way, clinicians can estimate whether a child's complexity and productivity performance is similar to that of his or her peers, selecting age- or grade-based peers for reference. INMIS values for both age and grade are included, given that both may serve as reasonable marker variables for narrative competence (e.g., Gutierrez-Clellen & Iglesias, 1992; McCabe & Peterson, 1991). A number of current educational policies presume grade to serve as a reasonable marker variable for specific narrative accomplishments (e.g., Virginia Department of Education, 2003), yet it is also plausible that narrative performance may "vary as much horizontally throughout a population of one grade as it does vertically through the grades" (Moffett, 1968, p. 54). Thus, clinical professionals can select those values that are most amenable to their needs.

Example of INMIS Application

To guide clinical use of INMIS, we provide here an example of its use. We use the formula provided previously for calculation of complexity and productivity values as well as data from Table 6 showing distributional statistics on the INMIS Complexity and Productivity values for the full sample (*n* = 250) and for each age/grade cohort. In this illustration, we randomly selected the performance data from one 5-year-old in the larger data set; her performance yielded the following values for the eight microstructural variables: MLT-W = 5.58, PROPCOMPLEX = 0.22, NDW = 30, TNW = 67, COORD = 0, SUBORD = 0, LENGTH = 9, and COMPLEX = 2. These values were inserted into the INMIS formulas to yield the following values:

$$\begin{aligned}
 \text{Productivity} &= -0.96 = -1.60 + (-0.0010 \times 5.58) \\
 &\quad + (-0.21 \times 0.22) + (0.017 \times 30) + (-0.00054 \times 67) \\
 &\quad + (0.014 \times 0) + (0.0072 \times 0) \\
 &\quad + (0.0094 \times 9) + (0.068 \times 2). \\
 \text{Complexity} &= -1.45 = -2.84 + (0.27 \times 5.58) \\
 &\quad + (0.85 \times 0.22) + (0.012 \times 30) + (-0.0027 \times 67) \\
 &\quad + (0.028 \times 0) + (0.026 \times 0) \\
 &\quad + (-0.085 \times 9) + (0.14 \times 2).
 \end{aligned}$$

The data in Table 6, which provide INMIS scores based on age level, indicate that the 10th and 25th percentiles for Productivity are -1.30 and -1.04 , respectively, both below this child's value of -0.96 . The child's Productivity score, at -0.96 , is about 0.5 *SD* below the mean of Productivity scores for 5-year-olds, but it is relatively close to the median score. We therefore consider this child to show average performance in Productivity. By contrast, her score on the Complexity measure, at -1.45 , is below the 25th percentile compared with the -1.24 INMIS value, and it is only slightly above the 10th percentile score of -1.55 . Her Complexity score is about 1 *SD* below the mean of scores for 5-year-old children. We thus conclude that this child's microstructural narrative performance is typical in its productivity but lower than average for its complexity. Such between-child comparisons allow one to identify a child's relative standing within an age- or grade-based peer group. Although McFadden (1996) correctly notes that a "child's position within a distribution provides no inherent point at which he or she becomes language impaired" (p. 4), understanding how a child's narrative performance compares with that of his or her peers is a necessary component of language assessment. For this reason, we supplied INMIS values denoting the 25th and 10th percentiles at each age and grade (see Table 6), both of which provide indication of a child's relative standing among peers. Obviously, the 10th percentile provides a more stringent benchmark and is considered by some experts to serve as an indicator of impairment. The 25th percentile, on the other hand, is often used as an indicator of general risk, particularly as applied to identifying school-age children at risk for literacy difficulties (see Torgesen, 2000).

As shown in the above case example, INMIS outcomes may serve as a complement to other norm- and criterion-referenced assessment data to provide a useful guide for estimating a child's language abilities for diagnostic purposes. Additionally, the within-child comparison of Productivity and Complexity performance provides a tool for intralinguistic referencing, in which a child's performance on various language tasks creates a profile of her strengths and weakness. As with our case example, clinicians can use INMIS to identify relative strengths and weaknesses in narrative microstructure, which may help to set clinical goals. For this child, for instance, narrative intervention can be focused primarily on complexity rather than productivity.

Additional Research Findings

In addressing the two primary aims of this work, several additional and unexpected findings warrant comment. First, the performance data showed a developmental increase in the means on nearly all microstructural measures through age 10, followed by a drop in performance for the two older age groups. On the one hand, these data showed narrative microstructure to exhibit a developmental increase in productivity and complexity in the early elementary grades, from 5 to 10 years of age. On the other hand, these data also suggested that older children—those who are 11 and 12 years old—were producing narratives that structurally look like

those of younger children. Narrative microstructure peaked with the 10-year-old children, and the performance of the 11- and 12-year-olds looked similar to that of 8- and 9-year-olds.

Although this was an unexpected result, the finding that older children may have a reduced interest in producing an elaborated narrative is consistent with analyses of children's self-evaluations of narrative quality. Kaderavek, Gillam, Justice, Ukrainetz, and Eisenberg (2004) examined children's self-evaluations of their own narrative performance, finding that 11- and 12-year-old children provided self ratings that were significantly lower than those of younger children. While older students seemed aware of their relatively poor performance on the narrative tasks, the question nonetheless remains as to why performance dropped for the older pupils. It is plausible that the decrease in narrative microstructure is an outcome of testing, skill, or motivation, and does not represent a decline in narrative ability. Additionally, it is plausible that use of a fictional task is not the optimal approach for eliciting narratives in older pupils and that other narrative tasks (e.g., use of expository tasks) may access the most sophisticated language that older pupils can demonstrate. These findings suggest that researchers need to determine whether measures of narrative performance provide valid estimates of language skill for these older students, to include examination of performance across various narrative tasks. It also suggests that INMIS may not be a valid estimate of older children's narrative performance and thus should be used with caution.

Second, we observed considerable variability in performance at each age and grade group as shown by the large standard deviations for all microstructural indices; moreover, we found that microstructural measures were not normally distributed and rather clustered toward the lower end of the distribution. This variability is possibly an artifact of using language sampling as a context for linguistic analyses, in which children's productions can vary substantially relative to more controlled tasks. Nonetheless, numerous studies of narrative performance for preschoolers through adolescence show remarkable variability in complexity and productivity indices between students, even when defined narrowly by age, grade, or ability. As a few examples, Liles and colleagues (1995) reported an average length of 28 utterances for elementary-grade students with language impairment, with a standard deviation of 22 utterances. Hemphill and colleagues (2002) reported an average length of 72 words and a standard deviation of 63 words for 4-year-old children with a history of congenital heart defects. Our findings were similar to these earlier reports. As shown in Table 6, for example, the distributional properties of the narrative variables were consistently skewed for each grade with the exceptions of fourth and sixth grade. A similar finding occurred for the kurtosis statistics. Again, with the exception of fourth grade and sixth grade, most of the distributions were leptokurtic. In short, the present research shows that children's microstructural performance is not normally distributed among an age or grade level, and that performance is skewed toward lower values. While the lack of normal distribution for scores does not preclude their use for normative and criterion-referenced

purposes, future theoretical research on narrative performance will undoubtedly provide some rationale for these distributional irregularities.

Limitations

There are several important limitations to the present research. First, the narrative samples in this study represented only one type of narrative, an elicited self-generated fictional narrative. Scripts, personal narratives, and retellings are also important narrative genres that speech-language pathologists and researchers utilize for practice and research. The extent to which our findings may generalize to these other narrative forms is presently unknown.

Second, the children's narratives were elicited with a single picture, representing only one way to elicit a quality fictional narrative. We did not, in this study, establish a naive listener condition (Strong, 1998), nor did we provide children with opportunities to practice the narrative task, both of which may have affected performance. Additionally, the use of a single-picture stimulus to elicit a fictional narrative is considered one of the more difficult tasks along the continuum of narrative production (Hughes et al., 1997). Alternative stimuli for obtaining a fictional narrative are the use of wordless picture books, sequenced pictures, requests to retell a story or movie, or simply asking the child to "tell me a good story." As such, the data in this study reflect children's abilities at a fairly sophisticated narrative task, and thus they likely account for some of the distributional properties of the narratives we studied. Further inquiry into the distributional properties and developmental trajectories of narrative skills as measured by other tasks (e.g., retellings) is an important area of narrative-analysis inquiry.

Third, it remains unknown whether INMIS values differentiate children with typical language skills from those with language impairment. Rather, the present research showed how the relative standing of a child can be determined for productivity and complexity using INMIS scores, but it does not indicate the specificity and sensitivity of INMIS for identifying children with language impairment. An examination of INMIS use for children with language impairment is an important next step in our own planned programmatic research on narrative microstructure.

The fourth limitation warranting note is our relatively small sample size. Typically, the development of norms requires larger sample sizes than that used in the present research; thus, the values provided in this work are best considered field test data. On this note, professionals should use INMIS field test references with caution, as one part of a more comprehensive evaluation of language ability.

Conclusion

In sum, recent advances in research and practice have resulted in a heightened awareness of narrative as an area of importance when considering the language achievements of children in the elementary grades. Practitioners are increasingly responding to this awareness by including narrative performance as an area of focus within assessment and intervention. The present study provides a clinical tool—INMIS—

that practitioners can use to examine two aspects of child's microstructural performance: complexity and productivity. Using INMIS field test values available for 6–12-year-olds or, alternatively, kindergarten through sixth graders, professionals can identify a child's standing relative to peers and study within-child variability on the two factors.

Acknowledgments

The data collection activities in this study were supported by grants from Pro-Ed Corporation to the authors. Pro-Ed had no involvement in the preparation of the present work. The authors are grateful to the children and the numerous collaborators who participated in data collection. Gratitude is also expressed to Nikki Franklin, Jessica Kingsley, and Sarah Greene, all of whom assisted the first author with preparation of the data for this study. Portions of this work were presented at the 2003 Annual Convention of the American Speech-Language-Hearing Association in Chicago.

References

- Atkins, C. P., & Cartwright, L. R. (1982). An investigation of the effectiveness of three language elicitation procedures on Head Start children. *Language, Speech, and Hearing Services in Schools, 13*, 33–36.
- Baltaxe, C. A., & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders, 22*, 1–19.
- Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language, 7*, 235–261.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.
- Boudreau, D. M., & Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *American Journal of Speech-Language Pathology, 8*, 249–260.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Chapman, R., Seung, H., Schwartz, S. E., & Kay-Raining Bird, E. (1998). Language skills of children and adolescents with Down syndrome II: Production deficits. *Journal of Speech, Language, and Hearing Research, 41*, 861–873.
- Clancy, P., Jacobsen, T., & Silva, M. (1976). The acquisition of conjunction: A cross-linguistic study. *Papers & Reports on Child Language Development, 12*, 71–80.
- Copmann, K. S. P., & Griffith, P. L. (1994). Event and story structure recall by children with specific language disabilities, language impairments, and normally achieving children. *Journal of Psycholinguistic Research, 23*, 196–231.
- Craig, H., & Washington, J. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research, 43*, 366–379.
- Crystal, D. (1987). Towards a "bucket" theory of language disability: Taking account of interaction between linguistic levels. *Clinical Linguistics and Phonetics, 1*, 7–22.
- Curenton, S. M., & Justice, L. M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools, 35*, 240–253.

- Dickinson, D., Wolf, M., & Stotsky, S.** (1993). Words move: The interwoven development of oral and written language. In J. Gleason (Ed.), *The development of language* (3rd ed. pp. 369–420). New York: Macmillan.
- Eisenberg, S.** (1997). Investigating children's language: A comparison of conversational sampling and elicited production. *Journal of Psycholinguistic Research*, 26, 519–538.
- Evans, J. L., & Craig, H. K.** (1992). Language sampling collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech and Hearing Research*, 35, 343–353.
- Feagans, L., & Short, E. J.** (1984). Developmental differences in the comprehension and production of narratives by reading-disabled and normally achieving children. *Child Development*, 55, 1727–1736.
- Gillam, R. B., & Johnston, J. R.** (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research*, 35, 1303–1315.
- Gillam, R. B., & Pearson, N.** (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.
- Gorsuch, R. L.** (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Green, S. B., Salkind, N. J., & Akey, T. M.** (2000). *Using SPSS for Windows: Analyzing and understanding data*. Upper Saddle River, NJ: Prentice Hall.
- Greenhalgh, K. S., & Strong, C. J.** (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools*, 32, 114–125.
- Grice, J. W.** (2001). Computing and evaluating factor scores. *Psychological Methods*, 6, 430–450.
- Grice, J. W.** (2002). *Resources for computing and evaluating factor scores*. Retrieved January 4, 2006, from <http://psychology.okstate.edu/faculty/jgrice/factorscores/>
- Gutierrez-Clellen, V. F., & Iglesias, A.** (1992). Causal coherence in the oral narratives of Spanish-speaking children. *Journal of Speech and Hearing Research*, 35, 363–372.
- Hargrove, P. M., Frerichs, J., & Heino, K.** (1999). A format for identifying interactions among measures of communication skills: A case study. *Language, Speech, and Hearing Services in Schools*, 30, 11–25.
- Hayes, P. A., Norris, J., & Flaitz, J. R.** (1998). A comparison of oral narrative abilities of underachieving and high-achieving gifted adolescents: A preliminary investigation. *Language, Speech, and Hearing Services in Schools*, 29, 58–171.
- Hemphill, L., Uccelli, P., Winner, K., Chang, C., & Bellinger, D.** (2002). Narrative discourse in young children with histories of early corrective heart surgery. *Journal of Speech, Language, and Hearing Research*, 45, 318–331.
- Hughes, D., McGillivray, L., & Schמידek, M.** (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
- Hunt, K. W.** (1965). *Grammatical structure written at three grade levels*. Champaign, IL: National Council of Teachers of English.
- Johnson, C. J.** (1995). Expanding norms for narration. *Language, Speech, and Hearing Services in Schools*, 26, 326–341.
- Justice, L. M., & Ezell, H. K.** (2002). *The syntax handbook: Everything you learned about syntax ... but forgot*. Eau Claire, WI: Thinking Publications.
- Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., & Eisenberg, S.** (2004). School-age children's self-assessment of oral narrative production. *Communication Disorders Quarterly*, 26, 37–48.
- Kaderavek, J. N., & Sulzby, E.** (2000). Narrative production by children with and without specific language impairment: Oral narratives and emergent readings. *Journal of Speech, Language, and Hearing Research*, 43, 34–49.
- Klee, T.** (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, 12, 28–40.
- Labov, W.** (1972). *Language in the inner city*. Philadelphia: University of Pennsylvania Press.
- Liles, B. Z.** (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research*, 28, 123–133.
- Liles, B. Z.** (1987). Episode organization and cohesive conjunctions in narratives of children with and without language disorders. *Journal of Speech and Hearing Research*, 30, 185–196.
- Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L.** (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, 38, 415–425.
- Loban, W.** (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.
- McCabe, A., & Peterson, C.** (Eds.). (1991). *Developing narrative structure*. Hillsdale, NJ: Erlbaum.
- McCabe, A., & Rollins, P.** (1994). Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology*, 3(1), 145–56.
- McCauley, R. J.** (1996). Familiar strangers: Criterion-referenced measures in communication disorders. *Language, Speech, and Hearing Services in Schools*, 27, 122–131.
- McFadden, T. U.** (1996). Creating language impairments in typically-achieving children: The pitfalls of "normal" normative sampling. *Language, Speech, and Hearing Services in Schools*, 27, 3–9.
- McFadden, T. U., & Gillam, R.** (1996). An examination of the quality of narratives produced by children with language disorders. *Language, Speech, and Hearing Services in Schools*, 27, 48–57.
- Menyuk, P.** (1969). *Sentences children use*. Cambridge, MA: MIT Press.
- Merritt, D. D., & Liles, B. Z.** (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders*, 54, 429–438.
- Miller, J. F.** (1991). *Research on child language disorders: A decade of progress*. Austin, TX: Pro-Ed.
- Miller, J. F., & Chapman, R. S.** (2000). *Systematic Analysis of Language Transcripts (Version 6.1a)* [Computer software]. Madison: University of Wisconsin—Madison, Waisman Research Center, Language Analysis Laboratory.
- Moffett, J.** (1968). *Teaching the universe of discourse*. Boston: Houghton.
- Muñoz, M. L., Gillam, R. B., Peña, E. B., & Gulley-Faehnle, A.** (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools*, 34, 332–342.
- Owen, A. J., & Leonard, L. B.** (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, 45, 927–937.
- Owens, R. E.** (1999). *Language disorders: A functional approach to assessment and intervention*. Boston: Allyn and Bacon.
- Paul, R.** (2001). *Language disorders from infancy to adolescence: Assessment and intervention* (2nd ed.). St. Louis, MO: Mosby.

- Paul, R., & Smith, R. L.** (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36*, 592–598.
- Peterson, C., Jesso, B., & McCabe, A.** (1999). Encouraging narratives in preschoolers: An intervention study. *Journal of Child Language, 26*, 49–67.
- Ripich, D. N., & Griffith, P. L.** (1988). Narrative abilities of children with learning disabilities and nondisabled children: Story structure, cohesion, and propositions. *Journal of Learning Disabilities, 21*, 165–173.
- Schneider, P., & Winship, S.** (2002). Adults' judgments of fictional story quality. *Journal of Speech, Language, and Hearing Research, 45*, 372–383.
- Scott, C. M., & Windsor, J.** (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*, 324–339.
- Southwood, F., & Russell, A. F.** (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research, 47*, 366–376.
- Stalnaker, L. D., & Creaghead, N. A.** (1982). An examination of language samples obtained under three experimental conditions. *Language, Speech, and Hearing Services in Schools, 13*, 121–128.
- Stein, N.** (1982). What's in a story: Interpreting the interpretations of story grammars. *Discourse Processes, 5*, 319–335.
- Stockman, I. J.** (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27*, 355–366.
- Strong, C. J.** (1998). *The Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.
- Strong, C. J., & Shaver, J. P.** (1991). Stability of cohesion in the spoken narratives of language-impaired and normally developing school-aged children. *Journal of Speech and Hearing Research, 34*, 95–111.
- ten Berge, J., Krijnen, W., Wansbeck, T., & Shapiro, A.** (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and Its Applications, 289*, 311–318.
- Thompson, C. A., Craig, H. K., & Washington, J. A.** (2004). Variable production of African American English across oracy and literacy contexts. *Language, Speech, and Hearing Services in Schools, 35*, 269–282.
- Torgesen, J. K.** (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*, 55–64.
- Ukrainetz, T. A., Justice, L. M., Kaderavek, J. N., Eisenberg, S. L., Gillam, R. B., & Harm, H. M.** (2005). The development of expressive elaboration in fictional narratives. *Journal of Speech, Language, and Hearing Research, 48*, 1363–1377.
- U.S. Department of Education.** (2003). *Mini-digest of education statistics* (NCES 2006-017). Washington, DC: Author.
- U.S. Department of Education.** (2006). *2003 Reading Assessment*. [Electronic version]. Retrieved October 15, 2005, from <http://nces.ed.gov/programs/coe/2004/section1/table.asp?tableID=38>
- van der Lely, H. K. J.** (1997). Narrative discourse in grammatical specific language impaired children: A modular language deficit? *Journal of Child Language, 24*, 221–256.
- Virginia Department of Education.** (2003). *English standards of learning curriculum framework*. Richmond, VA: Author.
- Washington, J., Craig, H., & Kushmaul, A.** (1998). Variable use of African American English across two language sampling contexts. *Journal of Speech, Language, and Hearing Research, 41*, 1115–1124.
- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W.** (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research, 38*, 1349–1355.
- Wren, C. T.** (1985). Collecting language samples from children with syntax problems. *Language, Speech, and Hearing Services in Schools, 16*, 83–102.

Received July 17, 2005

Accepted January 21, 2006

DOI: 10.1044/1058-0360(2006/017)

Contact author: Laura Justice, Preschool Language & Literacy Lab, Curry School of Education, Box 400873, University of Virginia, Charlottesville, VA 22904-4273.
E-mail: ljustice@virginia.edu

Appendix A

Three Examples of Narratives (Parsed Into T-Units)

6-year-old female

One day two little kids went somewhere/
And then they (they they) went to somewhere to play/
And then they saw a planet coming down with aliens with a alien dog too/
And the kids got scared/
The one kid (the small) says don't go/
And they (were) had stuff in their hands/
The dad did/
The mom had the kid/
And the kid had a dog/
And (the alien) the mom alien looked/
And her (her) put her hand out/
And the other one that drive from there waved back/
And then the kids comed/
They started unpacking/
And they were doing it again/
They were there again/

8-year-old female

One Saturday morning when Tom and April woke up they decided to go to the park together/
And (when they were at the park > um) when they were at the park (they they were) they
were going to have a picnic at the park/
And when they got there they set down the stuff/
And then they walked around for a little bit/
And then out of the sky they saw this spaceship thing land/
And they saw (uh) a man alien a woman alien and a little baby alien with an alien dog/
April wanted to go meet them/
And Tom tried to hold her back because he didn't want her to get in trouble with the aliens/
And then (and then) they went back home/
And they told the mom and dad/
And then when they brought them back there the aliens were nowhere to be seen/
The end/

11-year-old female

I'll call this (the) the aliens/
One day when John and (Jill) Jill were walking to the picnic in the park (there) they saw a big
huge space shuttle with weird words on it/
And then when they looked again (that's th*) they saw it opening/
And when it opened alien people came out of it with an alien dog and a little girl a mom a dad
and another girl (coming) about to come out of it/
Jill wanted to go over there/
And she thought it was just some people dressed in a costume/
(And didn't yeah I say his name was John) and John he hurry up/
(And oh and um) John didn't want to go over there/
And Jill grabbed him and XX XX hand/
And Jill was like pulling him and almost (like) about to grab her/
And (when they) when Jill grabbed (um) John (they heard) the aliens heard (um) John and Jill/
And (they) they say hi/
And (like) the aliens say hi in like a crazy voice/
And then the aliens asked them (where a good spot to place um I mean) if there was a
house for rent anywhere/
And they said no/
And so the aliens they slept in the woods over night/
Then when they slept at night (J*) Jill and (J*) John went to where they were sleeping at/
And they looked in their XX/
And (all there were were jus* there were just cos*) they found out that they were real people
and that they had on costumes to trick people/

Appendix B

Narrative Microstructure Variables and Techniques for Calculation

Variable	Technique for calculation
Total number of words (TNW)	The total number of words in the child's spoken narrative (generated by Systematic Analysis of Language Transcripts [SALT])
Total number of different words (NDW)	The total number of different words in the child's spoken narrative (SALT-generated)
Total number of T-units (LENGTH)	The raw number of T-units in the child's spoken narrative (SALT-generated)
Mean length of T-units in words (MLT-W)	The average length of T-units in words in the child's spoken narrative (SALT-generated)
Mean length of T-units in morphemes (MLT-M)	The average length of T-units in morphemes in the child's spoken narrative (SALT-generated)
Total number of complex T-units (COMPLEX)	The total number of T-units containing an independent clause and at least one dependent clause in the child's spoken narrative. Any T-unit with 1 or more dependent clauses was hand coded as COMPLEX, subsequent to which SALT analysis summarized the frequency of this code within individual narratives. ^a
Total number of coordinating conjunctions (COORD)	The raw frequency for use of 7 coordinating (<i>for, and, nor, but, or, yet, so</i>) conjunctions when used to coordinate two clauses in a T-unit (hand-coded to document coordinating function for each term, then SALT-generated)
Total number of subordinating conjunctions (SUBORD)	The raw frequency for use of 26 subordinating conjunctions when used to subordinate 2 clauses in a T-unit (<i>since, though, unless, until, when, where, whereas, also, besides, then, however, still, that, therefore, wherever, whether, while, why, thus, after, although, as, as well as, because, if, rather</i> ; hand-coded to document subordination for each term, then SALT-generated)
Proportion of complex T-units (PROPCOMPLEX)	COMPLEX was divided by LENGTH using SPSS

^aSeven percent of transcripts were randomly selected and double-coded to check the hand-coding procedure used to compute the COMPLEX, COORD, and SUBORD variables. Item-by-item interrater agreement was 97% ($SD = 3.9\%$, range = 89%–100%); the hand coding was consequently deemed reliable.
