Early Childhood Research Quarterly 25 (2010) 218–234

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol)

Laura M. Justice[a,*], Ryan Bowles[b], Khara Pence[c], Carolyn Gosse[d]

[a] School of Teaching and Learning, The Ohio State University, 356 Arps Hall, Columbus, OH 43210, United States
[b] Michigan State University, United States
[c] Washington DC, United States
[d] University of Virginia, United States

ARTICLE INFO

Article history:
Received 21 January 2009
Received in revised form 29 October 2009
Accepted 6 November 2009

Keywords:
Narrative assessment
Language development
Measurement
Language evaluation

ABSTRACT

Analysis of children's spoken narratives represents a potentially informative approach to language assessment within early childhood settings. Yet, narrative assessment is not readily amenable to at-scale use given the time needed to collect, transcribe, and analyze a child's narrative sample and the lack of consensus regarding what aspects of narrative expression ought to be examined (e.g., language form, language content). The purpose of this study was to describe a direct assessment of children's language abilities within a narrative context, the Narrative Assessment Protocol (NAP), which examines five aspects of language: sentence structure, phrase structure, modifiers, nouns, and verbs. In this study, we present findings regarding internal consistency, test–retest reliability, construct validity, and the concurrent and predictive validity of the NAP. NAP scores from 262 3–5-year-old children participating in preschool programs were assessed for these purposes. Findings indicated that the NAP exhibits reasonable psychometric properties across the areas addressed, to include significant concurrent and predictive relations with a norm-referenced measure of general language ability. Although more research is needed, preliminary findings indicate that the NAP provides professionals with a valid and informative assessment approach for examining children's language skills within a narrative context; such information may be useful for establishing and monitoring children's language growth within preschool programs or language interventions.

Professionals who work with young children within preschool classroom settings, including early childhood educators and literacy coaches, are increasingly interested in identifying effective and efficient approaches for screening and evaluating children's language abilities. Indeed, some federal programs that provide funding to preschool programs, such as Early Reading First, require grant recipients to implement ongoing measures of children's language ability to inform the design of classroom instruction and to identify children who may require additional learning supports (see Bryan, Ergul, & Burstein, 2008). Such initiatives are largely based on research findings showing that young children's language abilities serve as an important foundation for and predictor of their later reading skills, social and behavioral competence, and general academic performance (e.g., Catts, Fey, Zhang, & Tomblin, 2001; Pankratz, Plante, Vance, & Insalaco, 2007; Storch & Whitehurst, 2002). Of particular interest to many professionals today is increasing the availability of measures that reliably and validly represent children's abilities to use language at the discourse level within developmentally appropriate and naturalistic contexts. This is particularly important in light of trends whereby children's language abilities are represented exclusively by contrived single-word vocabulary measures that do not represent children's abilities to use language in context.

* Corresponding author.
E-mail address: justice.57@osu.edu (L.M. Justice).

0885-2006/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2009.11.002

In the current manuscript, we present a language assessment tool, the Narrative Assessment Protocol (NAP), designed for early educators and other professionals to use to assess children's expressive language skills within the context of a narrative task (e.g., the telling of a story). Use of the NAP involves eliciting a fictional narrative from a child using a wordless picture book, and coding the spoken narrative on a standardized score form for inclusion of specific features representing sentence structure (e.g., use of complex sentences), phrase structure (e.g., use of elaborated noun phrases), modifiers (e.g., use of adverbs), nouns (e.g., use of possessive forms), and verbs (e.g., use of past tense verbs). Although children's narratives can be video-recorded so that coding can take place offline (after the fact), the NAP was designed for real-time/online use (coding occurs simultaneous to the child's production of the narrative) by persons who have completed a training program designed to promote reliable implementation of the NAP coding system (available at no charge at preschoollab.com). The NAP was designed to provide a reliable and valid assessment of children's expressive language skills within a narrative context, thereby offering professionals a potentially more authentic glimpse of children's strengths and needs in language expression.

A narrative, by definition, refers to one's production of a fictional or real account of an experience or event that is temporally sequenced (Engel, 1995). Narrative is generally referenced as a type of discourse, along with conversation and exposition, because it involves production or comprehension of multiple sentences or utterances that unfold over time. Narratives can be elicited from children as a way to assess their discourse-level language skills as well as use of specific language forms or functions within a more naturalistic and ecologically valid context than more constrained tasks (e.g., producing sentences in imitation). In this regard, narrative assessment is often viewed as an authentic approach to language assessment, as children's language skills are examined in a highly contextualized task that represents "realistic learning demands" and circumstances (Schraeder, Quinn, Stockman, & Miller, 1999). Assessment of a child's spoken narrative therefore typically involves elicitation of a fictional or real recount and then documenting specific features, typically differentiated into narrative macrostructure and microstructure.

Narrative macrostructure refers to the general, global characteristics of a narrative, such as the thematic organization of main ideas (Hughes, McGillivray, & Schmidek, 1997). When one thinks about children's narratives, it is typically features of macrostructure that first come to mind, such as whether a narrative includes traditional story grammar elements of setting, character, and plot. Macrostructural assessment is based on the perspective that children's narrative abilities are influenced by their "mental representations of events and the verbalizations of such scripts" (Berman, 1995, p. 287). Therefore, assessment of narrative macrostructure may involve analysis of a child's use or understanding of causal networks, event representations (i.e., scripts), and traditional story grammar. Macrostructure analysis may also consider the extent to which the child's narrative (a) leads sequentially up to a high point (called *high point analysis*), (b) is coherent (*dependency analysis*; e.g., McCabe & Peterson, 1984), and/or (c) contains evaluative devices and other elements of "artfulness" (Ukrainetz, Justice, Kaderavek, Eisenberg, & Gillam, 2005). Macrostructural features of children's narratives are multiply determined, influenced by general language ability, a variety of cognitive processes (e.g., event recall, causal thinking), and cultural experiences (e.g., Brice-Heath, 1983; Trabasso & van den Broek, 1985).

It is also possible to assess narratives for containment of more micro-level properties, often referred to as narrative microstructure. The assessment of a narrative's microstructure investigates its more internal properties, such as the number of words and sentences it contains, the way in which specific cohesive devices (e.g., conjunctions) are used to link words and sentences, and its overall grammatical complexity (e.g., number of complex sentences). This is sometimes referred to as form/function analysis (Berman, 1995), because it concerns how children's developing language forms (e.g., sentences, words) are used to meet narrative functions. Typically, analysis of a narrative's microstructure adheres to the conventions of language sampling, in that it involves first parsing the sample into smaller units, most often the T-unit (also called a C-unit); a T-unit is one independent clause and any dependent phrases and clauses. Then, the narrative is analyzed to determine such features as percentage of complex T-units (T-units with 2+ clauses), mean number of words per coordinate clause, use of specific phrase and clause structures, and percentage of grammatically correct T-units (Justice et al., 2006).

Of the two paradigms commonly used to examine children's narrative abilities – macrostructure and microstructure – the NAP is best characterized as a tool for assessing features of narrative microstructure, as it documents the presence of syntactic, morphological, and lexical structures that children use in the process of constructing a narrative. Measures of narrative microstructure are generally seen to offer general approximates of a child's expressive language ability within naturalistic and functional contexts (as compared to the more contrived contexts characteristic of many standardized language assessments), yet they are under-utilized as an assessment approach because they typically must rely on the methods of language sampling (i.e., capturing the sample, transcribing it, parsing it, and then analyzing it). On the contrary, macrostructural analysis typically involves documenting presence of specific global structures, and therefore does not necessarily require use of language sampling methods. Importantly, indices derived from assessment of children's narrative expression that focus on microstructural features (e.g., grammatical complexity) serve as particularly robust predictors of children's later language and reading comprehension (Pankratz et al., 2007), making it important for educators to have access to tools that allow them to assess these features of children's narratives.

As an assessment of children's language abilities within a narrative context, the NAP was designed specifically for use by professionals who work with preschool-aged children as a companion to tools that examine other domains of language competence. Language is a multidimensional construct comprising, minimally, two modalities (comprehension, expression) and three domains (vocabulary, grammar, narrative) of development (see Tomblin, Records, & Zhang, 1996, for discussion of this matrix of language skill). Although children's abilities to express themselves within the context of narration is considered

an important domain of children's language development (see Boudreau & Hedberg, 1999; Curenton & Justice, 2004; Liles, 1987), and narrative expression is represented in multidimensional models of language ability (e.g., Tomblin et al., 1996), there are too few standardized approaches available for assessing children's language skills within a narrative context for preschool-aged children. This presents an important limitation to current language assessment practices, given research findings showing that measures of narration collected during the preschool years exhibit excellent predictive validity with later measures of language and reading comprehension (Pankratz et al., 2007) and can be validly and reliably used with children of diverse cultural and linguistic backgrounds (Peña et al., 2006).

Historically, the practice of narrative assessment has relied upon traditional language sampling procedures, whereby a narrative sample is elicited, transcribed, and then analyzed for diagnostic and instructional/therapeutic purposes (Damico, Oller, & Storey, 1983; Hux, Morris-Friehe, & Sanger, 1993; Washington & Craig, 2004). Narrative assessment has not gained a foothold in the battery of assessments frequently employed in early childhood settings, attributable in part to the length of time required for narrative transcription and the lack of an efficient and valid coding system for quantifying narrative content at both the macro- and microstructure levels (Justice et al., 2006). Nonetheless, given the potential importance of measures of language ability within narrative contexts to serve as an indicator of a child's current and future language performance, there appears to be a clear need for a tool that mainstreams this type of assessment. This need is at least implicitly supported by references to narrative outcomes as an important element of preschool instructional practices within various state early learning standards. For instance, Ohio's Early Learning Content Standards for English Language Arts include standards that target children's narrative expression (e.g., "retell information from informational text," and "retell or re-enact events from a story;" Ohio Department of Education, 2008, pp. 26–27), as does Connecticut's Preschool Curriculum Framework ("retell information from a story;" State of Connecticut State Board of Education, 2006, p. 24).

Some efforts to standardize narrative assessment practices have made assessment of children's narratives more accessible to educators and allied professionals. These efforts have included, for instance, the story retelling task of Culatta, Page, and Ellis (1983), which was used at scale in epidemiological research on the incidence of specific language impairment (Tomblin et al., 1996), and the more recent commercial publication of the norm-referenced *Test of Narrative Language* (Gillam & Pearson, 2004). Neither, however, was designed for use with preschool-aged children, precluding their use within early childhood settings. An exception is *The Renfrew Bus Story* (Cowley & Glasgow, 1994; Renfrew, 1969), a standardized narrative task from Britain that has been reliably used with children of preschool age (e.g., Bishop & Edmundson, 1987; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). This task involves analyzing two aspects of a child's retelling of a story, information (i.e., semantic content of the narrative) and length (i.e., syntactic complexity of the narrative). Although scores from this narrative task are valid indicators of children's current language skill and are strongly predictive of children's future language and reading performance (Pankratz et al., 2007), the need for professionals to transcribe children's narrative productions to derive the length score undermines the scalability of this tool into preschool settings in which educators may have neither the resources nor the time to conduct transcription.

The NAP was designed to contribute to efforts to make narrative assessment a potentially routine part of language screening and assessment within preschool programs. Our particular interest was designing a tool with potential for at-scale use, or *scalability*, a concept that refers to extending the use of a practice to large numbers of individuals (or organizations) (see McDonald, Keesler, Kauffman, & Schneider, 2006). With this goal in mind, we sought to develop a valid and reliable tool that could potentially be used by large numbers of professionals at low cost, in terms of both financial resources and the time needed for implementation. Three characteristics of the NAP design make it amenable to at-scale use. First, the NAP is scored while listening to a child produce a spoken narrative (typically from a videotape), with no transcription required. This is an important characteristic, given that transcription requires specialized skills and software, and the products of transcription can be unreliable (Gavin & Giles, 1996).

Second, the NAP requires relatively little time to administer, with coding of the narrative lasting only slightly longer than the narrative itself. Total administration and scoring time needed, if standardized NAP procedures are followed, is less than 20 min (8 min to collect the narrative and about 10 min to score it later from video). Given that many early childhood professionals express concerns over the use of lengthy assessments with young children (Gardner, Froud, McClelland, & van der Lely, 2006; Sturner et al., 1994), this is a valuable feature.

Third, the NAP was designed to be cost effective, requiring few specialized materials or test forms. Developed with the support of federal funds, administration of the NAP requires a commercially available storybook (approximate cost, $7), the elicitation protocol and administration directions (Appendices A and B, respectively), score forms [Appendix C (Long Form) and Appendix D (Short Form)], and, if desired, a video camera or tape recorder and recording media.

Additionally, self-study materials are available via the internet to help potential NAP users become reliable users of this tool (www.preschoollab.com). The self-study materials include a set of audio-taped narratives on which to practice NAP coding as well as a separate set on which one can determine his/her reliable use of the NAP coding system. Although the efficacy of these self-study materials for use outside of the laboratory setting has not been assessed (and is not the goal of the present study), extant research does suggest the viability of (a) online self-study materials for increasing teachers' knowledge of assessment principles (e.g., Huai, Braden, White, & Elliot, 2006) and (b) didactic lessons coupled with practice opportunities for increasing teachers' reliability in administering pupil assessments (e.g., Kourtessis, Tsigilis, Maheridou, Ellinoudis, & Kiparissis, 2008).

The goal of the present article is to provide evidence for the utility of the NAP as an assessment tool for examining young children's language abilities within a narrative context. Specifically, this study was conducted to determine the

reliability and validity of the NAP. First, concerning reliability, we determined whether NAP scores can be reliably determined across different coders (i.e., inter-rater reliability) and whether NAP scores are invariant to the particular method of coding (i.e., online from the child's narrative video recording, or offline from a transcript). Second, concerning validity, we examined the factorial structure of children's NAP scores collected at two time points in an academic year (fall and spring), and factorial invariance (Meredith & Horn, 2001) to determine whether the nature of narrative assessment remains the same across the two occasions. Additionally, criterion-related validity of the NAP was assessed to determine whether NAP scores provided information about children's more general language ability both concurrently and predictively.

1. Method

1.1. Participants

Participants were drawn from a study of 339 children ranging in age from 40 to 66 months ($M = 52$, $SD = 5.5$) who were enrolled in a year-long study of language and literacy development among preschool-aged children. Full study procedures appear in Cabell et al. (submitted for publication). From among those enrolled in this study, we included in the present study only those children from whom spoken narratives were collected on video in the fall of their preschool year. A spoken narrative was not collected from all children enrolled in the parent study because their parents did not provide video consent as a specific addendum to consent agreements. In addition to procuring written caregiver consent for their children's participation in this study, children were asked to provide verbal assent at the start of each assessment session; 12 children (from among the larger study's 339 participants) did not provide verbal assent and were not assessed. In a few instances, narratives were collected but video recording equipment or recording media malfunctioned. For this reason, narratives were only available for 77% ($n = 262$) of those children enrolled in the parent study.

Of these 262 children, 49% were girls and 51% were boys. Parents reported children's race/ethnicity to be White/Caucasian for 111 children (42% of sample), Black/African American for 92 children (35%), Hispanic/Latino for 16 children (6%), Multi-Racial for 16 children (6%), Asian for 1 child (<1%), and another or unspecified race or ethnicity for 26 children (10%). According to parental report, the majority of children ($n = 198$; 93%) spoke English in the home; this information was unavailable for 49 children.

The children attended preschool programs prioritizing eligibility for at-risk pupils, including state-funded preK programs and Head Start programs. Individual programs were free to establish risk using local methods, although participation tended to prioritize children who exhibited developmental disabilities or vulnerabilities due to poverty, neglect, or abuse. For the most part, children were eligible because they resided in low-income homes based on community standards, with 80% of children residing in homes with annual household incomes less than $50,000 per year (all sources of income). Of 209 children for whom information about maternal education was available, 75% of children's mothers had a high school diploma as their highest degree earned; only four mothers had a bachelor's or graduate degree. About 14% of children in the sample had identified disabilities and qualified for special education services in their schools.

1.2. General procedures

As a part of the parent study, children were administered a comprehensive portfolio of standardized language and literacy measures within approximately 4–6 weeks of the beginning of the academic year. This same assessment battery was repeated in a 4-week window just prior to the end of the school year in the following spring. Assessments were administered to children over two 20–30-min sessions. In an effort to control for order effects, assessments were not given in any specific order; rather, when a child was given a particular assessment was based on when a child could be pulled from a classroom for a test session and which assessments were being implemented at that time.

All assessments were administered by trained research assistants in quiet locations within children's schools. Prior to conducting any assessments, each research assistant completed a comprehensive standardized training protocol that involved review of test materials, completion of quizzes on the test materials, and observation of assessments in the field by senior research staff. For the present purposes, assessments of interest comprised a standardized assessment of general language ability, the *Clinical evaluation of language fundamentals preschool, second edition* (*CELF Preschool-2*; Wiig, Secord, & Semel, 2004), and collection of a spoken fictional narrative. These measures represented the only two assessments used to examine children's language ability in the larger study; other measures examined literacy skills (e.g., alphabet knowledge, print-concept knowledge).

1.2.1. CELF Preschool-2

The CELF Preschool-2 is a standardized, norm-referenced language assessment used to identify, diagnose, and evaluate language disorders in children ages 3–6 years. Three Core Language subtests were administered in this study: Sentence Structure, Word Structure, and Expressive Vocabulary. Scaled scores for each of the three subtests are based on a mean of 10 and a standard deviation of three. The Core Language composite standard score, derived from the three subtest scores, is based on a mean of 100 and a standard deviation of 15. According to the test authors, the Core Language composite standard score can be used to discriminate the performance of average or above-average language users from the performance of

children with language disorders (Wiig et al., 2004). This measure demonstrates adequate internal consistency, test–retest reliability, inter-rater reliability, and validity (Wiig et al., 2004).

1.2.2. Spoken fictional narrative

A spoken fictional narrative using the wordless storybook *Frog Where Are You?* (Mayer, 1969) was elicited from each child. This storybook has been used in a number of studies involving collection of a narrative sample (Botting, 2002; Boudreau & Hedberg, 1999; Eriksson, 2001; Greenhalgh & Strong, 2001; Miles & Chapman, 2002; Norbury & Bishop, 2003; Pearce, McCormack, & James, 2003; Restrepo, 1998). This wordless picture book presents an illustrated story sequence involving an explicit episodic structure (e.g., initiating event, plan, goal attempts, and consequences) as well as many of the traditional story grammar components (e.g., characters, setting). The plot surrounds a young boy and his dog as they engage in a search for the boy's missing frog. In this story, the dog and the boy encounter several animals during their search and eventually find a family of frogs, from which the boy selects a single frog to take home with him. To elicit the narrative, each child was advised that he/she would be asked to tell a make-believe story using the wordless storybook after first previewing each page of the book under the guidance of the examiner. A script was used to ensure that examiners collected the narrative sample in a reliable manner (see Appendix A). All narratives were videotaped in their entirety for later analysis.

1.3. Narrative Assessment Protocol (NAP)

Using the video collected from each child during elicitation of the spoken fictional narrative, each child's narrative was coded following NAP coding procedures on a NAP score sheet, namely the Long Form available in Appendix C. (The Short Form, which appears in Appendix D, was developed in response to results of factor analysis, presented subsequently.) Trained research assistants coded each narrative from videotape using standard desktop computers, media software (e.g., Windows Media Player), and headphones. When coding a videotaped narrative using the NAP, coders could pause the video as many times as needed to take notes or to re-listen to particular segments. In general, coding of a narrative sample was completed in about 10 min.

Research assistants coded narratives using the NAP procedures following completion of a standardized training protocol designed to achieve reliability. The entire training protocol is available on the web (www.preschoollab.com) so that assistants could pursue training on a self-study basis. The primary difference between the procedures available online to the public versus those followed by assistants in the lab is that the latter features videotapes of children producing spoken narratives, whereas the former features the audio-taped derivatives of these narratives, to preserve confidentiality of children from whom narratives were elicited. The protocol that trainees followed to become reliable NAP coders was as follows.

First, trainees reviewed background information and completed a series of written lessons on phrases, clauses, and sentence types presented in a basic primer on syntax (Justice & Ezell, 2002). These lessons were designed to introduce coders to linguistic terminology that may be somewhat unfamiliar to them (e.g., clauses, phrases, verbs, nouns). Second, trainees listened to two children's narrative samples while reviewing the accompanying transcripts which contained NAP scores. Third, trainees completed five practice sessions for which they scored children's narratives (from video) using the NAP coding system; subsequently, the trainee studied the corresponding transcripts to verify and correct their own codes. These practice sessions included a range of opportunities for trainees to hear various NAP items (e.g., samples of elaborated noun phrases, tier-two nouns, prepositional phrases). Fourth, trainees completed reliability sessions for which they scored children's narratives (again from video) until they achieved a specified criterion accuracy level; specifically, trainees were considered reliable coders when they achieved overall 83% agreement (within one value) to master-coded transcripts.

The NAP coding protocol requires one to document whether and how often a child uses 18 distinct language forms in his/her spoken narratives. The NAP score sheet (Appendix C) was used to code for the presence of these 18 items as organized into five sets (i.e., indicators): (a) Sentence Structure (four items), (b) Phrase Structure (three items), (c) Modifiers (two items), (d) Nouns (three items), and (f) Verbs (six items). To arrive at these five sets and 18 items, we conducted a comprehensive survey of the research literature for description of syntactic, morphological, and lexical structures that were commonly included in analysis of narrative microstructure. The NAP Coding Catalog, which provides a detailed description with examples for each of the 18 items, is presented in Appendix E; coders referred to the Coding Catalog as needed while coding videotapes. For each item, the coder scored the frequency of occurrence on a rating scale comprising 0 (no occurrence), 1 (one occurrence), 2 (two occurrences), and 3+ (three or more occurrences). For Modifiers, Nouns, and Verbs, an occurrence was scored only for unique usages (i.e., if a child used the pluralized noun *frogs* twice in a narrative, the occurrence was only scored once). For Sentence Structure and Phrase Structure, occurrences were scored even if they were not a unique usage (i.e., if a child used the elaborated noun phrase *an empty plate* twice in a narrative, each occurrence would be scored).

Individual indicator scores were created by summing the scores for all items in the indicator category, and maximum scores varied according to the number of items corresponding to each indicator (e.g., the maximum score for Sentence Structure is 18, whereas the maximum score for Modifiers is 6). A sum score was created by summing all items on the NAP (range 0–54). The coefficient alpha for the Sum Score is .83, indicating a reasonably high level of internal consistency among the 18 NAP items.

Table 1
Comparison of NAP scores from online and offline coding ($n = 31$).

Measure	Online scoring (Video)		Offline scoring (Transcript)	
	M	*SD*	*M*	*SD*
NAP indicator				
Sentence structure	2.7	2.3	2.7	2.1
Phrase structure	5.1	1.9	5.5	1.4
Modifier	0.2	0.5	0.8	1.0
Nouns	2.5	1.7	2.6	1.6
Verbs	8.0	3.7	8.0	3.6
NAP sum score (Long Form)	18.5	7.3	19.5	7.2

2. Results

2.1. Reliability of the NAP

With respect to reliability, goals of this research included determining whether NAP scores can be reliably determined across different coders (i.e., inter-rater reliability) and whether NAP scores are invariant to the particular method of coding (i.e., online from the child's narrative video recording, or offline from a transcript).

2.1.1. Inter-rater reliability

The inter-rater reliability of the NAP coding system was assessed using 33 narratives (12%) randomly selected from the fall corpus. These narratives were double scored by two NAP coders working independently of one another. Drawing from reliability procedures used in large-scale observational work (La Paro, Pianta, & Stuhlman, 2004), agreement on each of the 18 individual items occurred when the two coders were within one scale point of one another. For instance, if one coder identified a narrative as containing two compound nouns and the other coder identified the same narrative as containing three compound nouns, the coders were considered to be in agreement based on "within-one" procedures (La Paro et al., 2004). An agreement percentage was computed for each of the 18 items for each of the 33 narratives; agreement scores ranged from 82% (adverbs) to 100% accuracy (e.g., pluralized nouns), with an average agreement score of 93% for the individual NAP items. These values are similar to those reported recently when independent coders assessed teacher–child interactions during classroom-based instruction (Pianta, Mashburn, Downer, Hamre, & Justice, 2008) and are similar to that of conventional narrative coding practices (e.g., Glasgow & Cowley, 1994; Peña et al., 2006). The level of agreement between the two coders suggests that the NAP can be reliably applied by coders following training.

2.1.2. Reliability of online and offline coding

An important feature of the NAP is that it was designed to be reliably scored online from video or audio rather than from transcribed narratives. We studied whether NAP scores are invariant to the particular method of coding-scored from video as the child produces a narrative versus scored on a verbatim written transcript of the child's narrative—for a randomly selected subset of narratives ($n = 31$, or 12%) from the fall of the year. Each of these scripts was transcribed in its entirety using the *Systematic Analysis of Language Transcripts, Research Version 9* software (Miller & Iglesias, 2006). Narratives were segmented into T-units during transcription following procedures outlined in Justice et al. (2006). A T-unit consists of a single main clause and any dependent constituents, including clauses and phrases (Hunt, 1965). Transcription was completed by research assistants who had completed a comprehensive SALT training tutorial, similar to the procedures used to train reliable NAP coders. Following the initial transcription, a second examiner independently watched each videotaped narrative while simultaneously examining the transcript in its entirety for errors in transcription or segmentation; thus, 100% of narratives were checked for accuracy. Each of these transcripts was coded for the 18 NAP indicators by two reliable NAP coders in a collaborative processes designed to achieve 100% accurately coded transcripts, so that these codes could serve as a gold-standard comparison against the online coding procedures.

A comparison of scores from online versus offline NAP coding is presented in Table 1. The difference in total NAP sum scores when scored online from video ($M = 18.5$, $SD = 7.4$) versus offline from transcripts ($M = 19.5$, $SD = 7.2$) was negligible. With the exception of modifiers, $t(30) = 2.8$, $p < .01$, in which a slightly higher number was identified on video, there were no statistically significant differences between NAP scores for offline and online scores (using a series of paired samples *t* tests; all $ps > .05$).

We also assessed the correlational relations between online and offline scores. A Bonferroni corrected alpha of .008 for six comparisons (.05/6) was used to control for Type I error. Five of the six correlations were statistically significant: Sentence Structure, $r = .8$, $p < .001$, Phrase Structure, $r = .64$, $p < .001$, Nouns $r = .75$, $p < .001$, Verbs, $r = .89$, $p < .001$, and for the sum score, $r = .92$, $p < .001$. The correlation between online and offline scoring of Modifiers was low and was not statistically significant, $r = .04$, $p = .834$. These data show that NAP scores derived from online scoring are similar to and strongly associated with those derived from offline scoring of transcribed narrative samples, with the exception of Modifiers.

2.2. Validity of the NAP

2.2.1. Construct validity

In the present study, we assessed two forms of validity of the NAP: construct validity and criterion-related validity. To assess construct validity, we considered the factor structure of the NAP and its invariance over time. We expected the NAP to reflect a single factor of narrative skills that is stable in structure. We first performed an exploratory factor analysis for all 18 items using the EFA procedure in Mplus (Muthén & Muthén, 2006). For both the fall and spring scores, a single factor clearly emerged based on a scree plot and examination of the eigenvalues. The first eigenvalue was high (4.84 for fall, 4.59 for spring), and much higher than the second eigenvalue (1.49 and 1.36). We repeated the analysis considering the items to be categorical rather than continuous. Results confirmed the single factor solution, so we report only the continuous results.

Examination of the factor loadings indicated that several items did not load strongly on the single narrative skill factor. Six items did not meet a .4 factor loading cutoff in both fall and spring: two sentence structure items, one phrase structure, one modifier, one noun, and one verb. Removal of these items did not impact the factor structure, and their removal would lessen the burden on coders. Therefore, we compiled the remaining 12 items in a Short Form of the NAP (NAP Short Form), which appears in Appendix D. The correlation between Long and Short Form scores on the NAP, based on the fall assessment time point, was strong $r = .99$, $p < .001$; except where noted, we report results for the Short Form only subsequently in this manuscript.

We examined factorial invariance across the two time points on the NAP Short Form using a confirmatory approach in Mplus (Meredith & Horn, 2001). We considered models with a single factor within each occasion for each of the 12 indicators, allowing uniquenesses to correlate for the same indicator across fall and spring. Weak factorial invariance (i.e., equal factor loadings) was supported, as there was no statistically significant increase in misfit ($\Delta\chi^2 = 13$, $\Delta df = 17$, $p = .76$, RMSEA = .045 and .044). Imposing strong factorial invariance by constraining indicator intercepts to be equal introduced significant misfit ($\Delta\chi^2 = 74$, $\Delta df = 18$, $p < .01$) but the RMSEA (.055) was still within adequate range (<.08; Browne & Cudeck, 1993). Imposing strict factorial invariance introduced significant further misfit ($\Delta\chi^2 = 90$, $\Delta df = 18$, $p < .01$), but the RMSEA was still within adequate range (.066). Overall, we conclude that the NAP has strong construct validity as a unidimensional measure with 12 items, and shows structural stability from fall to spring, although the strength of the invariance is not entirely clear. Wackwitz and Horn (1971) found that unit-weighting salient indicators is an effective estimate of factor scores, particularly with a single factor (Wackwitz & Horn, 1971). Therefore, we employ a NAP sum score for the remainder of this article, which also maximizes practical usefulness of the instrument. The coefficient alpha for the sum score is .84 in the fall and .83 in the spring for the Short Form (.83 and .82 for the Long Form), indicating a high level of consistency among the NAP items.

2.3. Criterion-related validity

Descriptive statistics regarding children's fall NAP scores appear in Table 2. Table 3 presents descriptive statistics regarding children's performance on the standardized test of general language ability in the fall and spring of the academic year (CELF Preschool-2). The normative language data indicate that this group of children received standard scores relatively lower than age-based peers, with a mean of 84.22 ($SD = 18.4$) on the Core Language Composite at the fall assessment. These findings are not surprising given that the children were participants in preschool programs designed to serve children experiencing risk.

We studied the criterion-related validity of the NAP by examining concurrent relations between children's performance in fall on the NAP and the CELF Preschool-2, and predictive relations between children's performance in fall on the NAP and spring on the CELF Preschool-2. Table 3 also provides correlations between fall NAP scores (both the Long Form and Short Form) and the CELF subtest raw scores and Core Language Composite at fall and spring. In the fall, NAP scores based on both the Long and Short Forms were consistently and moderately correlated with the CELF scores; correlations for the NAP Sum Score and the Fall Core Language Composite were medium in size, $r = .35$, $p < .001$ and $r = .34$, $p < .001$ for the Long and Short Forms, respectively. Correlations for the NAP Sum Score and the Spring Core Language Composite slightly higher, $r = .43$, $p < .001$ and $r = .44$, $p < .001$ for the Long and Short Forms, respectively. As benchmarks, previous research involving preschool-aged children reported concurrent correlations of .12 to .36 between measures derived from conversational language samples (including narrative tasks) and performance on standardized language measures (Ukrainetz & Blomquist, 2002), and predictive correlations of .30 to .39 between measures derived from narrative samples and performance on standardized language measures (Pankratz et al., 2007). Consequently, the present findings show children's performance on the two measures (NAP and CELF Preschool-2) to exhibit concurrent and predictive relations with standardized language measures that compare favorably to previous reports.

2.4. Exploratory analyses of the diagnostic potential of the NAP

Although the intent of this research was not to assess the diagnostic accuracy of the NAP, for exploratory purposes we examined whether children's scores on the NAP would differ as a function of their performance on norm-referenced language measures and the presence of a disability. If so, this might suggest the potential for future exploration of the NAP's diagnostic potential. We conducted two types of exploratory analyses. First, NAP fall sum scores were compared for children with low standard scores (≤ -1 *SD* of the mean on the Core Language Composite of the CELF Preschool-2) on the fall norm-referenced

Table 2
Mean NAP scores in fall of year.

NAP indicator (Max Score-Long Form)	*n*	*M*	*SD*
Sentence structure (12)	262	1.74	2.10
Compound sentence[a]	262	.13	.44
Complex sentence	262	.92	1.16
Negative sentence	262	.55	.94
Interrogative[a]	262	.14	.49
Phrase structure (9)	262	4.54	2.27
Elaborated noun phrase	262	1.84	1.23
Compound noun[a]	262	.36	.74
Prepositional phrase	262	2.34	1.12
Modifiers (6)	262	.29	.61
Adverb[a]	262	.05	.28
Advanced modifiers	262	.24	.53
Nouns (9)	262	2.02	1.68
Pluralized noun	262	1.24	1.06
Possessive form[a]	262	.10	.34
Tier-two noun	262	.68	.87
Verbs (18)	262	5.47	3.76
Auxiliary verb + Main verb	262	1.39	1.25
Copula	262	1.32	1.24
Irregular past tense verb	262	1.68	1.26
Regular past tense verb	262	.91	1.13
Tier-two verb	262	.13	.38
Compound verb[a]	262	.03	.20
Total NAP score: Long Form (54 points max)	262	14.06	8.19
Total NAP score: Short Form (36 points max)	262	13.24	7.62

[a] Items not included in Short Form.

Table 3
Standardized language scores in fall and spring of year and correlations with NAP scores.

	Measures			
	Sentence structure	Word structure	Expressive vocabulary	Core language composite
CELF subtests				
Fall CELF *M* (*SD*)	7.7 (3.18)	7.0 (3.0)	7.6 (3.14)	84.2 (18.4)
Spring CELF *M* (*SD*)	8.5 (3.12)	7.7 (3.24)	9.0 (8.01)	89.2 (16.19)
Fall assessment				
n	244	241	240	237
Fall Sum Score: NAP Long Form	.36**	.48**	.41**	.35**
Fall Sum Score: NAP Short Form	.35**	.48**	.42**	.34**
Spring assessment				
n	218	216	216	214
Fall Sum Score: NAP Long Form	.40**	.49**	.45**	.43**
Fall Sum Score: NAP Short Form	.39**	.50**	.46**	.44**

CELF subtest scores from the *Clinical evaluation of language fundamentals preschool, second edition* (Wiig et al., 2004). Correlational analyses conducted with CELF raw scores for individual subtests and standard score for composite.

** $p \leq 0.01$.

assessment of general language ability as compared to children with scores in the average range (>−1 *SD* of the mean). Fifty-four percent of the sample had low standard scores ($M = 69.4$, $SD = 22$), whereas 46% were in the average range ($M = 97.3$, $SD = 9.1$). NAP fall sum scores were significantly lower for children with low standard scores ($M = 11$, $SD = 7.3$) compared to those with average scores, ($M = 15.8$, $SD = 7.2$), $t(260) = 5.23$, $p < .001$ ($d = 0.66$), and the difference was consistent with a medium to large effect size.

Second, NAP fall sum scores were compared for children who received special education services ($n = 37$) to those who did not ($n = 224$; data unavailable for one child); the former group represented 14% of children in this study. NAP fall sum scores were significantly lower for children who received special education services ($M = 10.5$, $SD = 6.9$) compared to those who did not ($M = 13.6$, $SD = 7.7$), $t(259) = 2.34$, $p < .05$ ($d = 0.41$); the difference was consistent with a medium-sized effect.

3. Discussion

The Narrative Assessment Protocol examines a number of microstructural features of children's spoken narratives and was designed for use by professionals who work within preschool settings. Presently, there is great interest in the systematic implementation of language assessments within preschool programs to promote differentiated instruction within the general curriculum and to identify children who may require more intensive language development supports. This inter-

est is driven, in part, by response-to-intervention frameworks in which children who do not make adequate progress in language and literacy achievement within the general curriculum are provided supplemental, specialized learning opportunities (Justice, 2006). Federal and state initiatives often require preschool programs receiving public funds to develop and implement assessment approaches that examine children's language growth over time and that identify children who may need additional instructional supports in this area of development.

Presently, there are few options available to professionals who work with preschool children who desire to include narrative analysis within their preschool assessment portfolios within response-to-intervention or other assessment frameworks. The rationale for use of narrative assessment as a complement or alternative to traditional language measures has been well articulated in the literature. For instance, some experts contend that virtually no other single assessment approach can reveal so much about an individual's language skills (McCabe & Rollins, 1994). The examination of a child's fictional story can provide telling information about skills in syntax, morphology, vocabulary, and phonology. In this regard, the collection and analysis of a narrative is also viewed as a highly efficient approach to language assessment. Moreover, experts contend that assessment of children's language skills must broaden beyond the use of decontextualized and non-authentic test situations to include examination of children's language expression within naturalistic discourse-level contexts (Laing & Kamhi, 2003; Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003).

To date, the research community has made limited progress toward developing narrative assessment tools that have potential for at-scale use. This is because the majority of approaches to narrative assessment described within the literature require time-consuming transcription and coding of narrative samples (Botting, Faragher, Smikin, Knox, & Conti-Ramsden, 2001; Pearce et al., 2003; Young, Diehl, Morris, Hyman, & Bennetto, 2005). The NAP was designed specifically to bypass the need for transcription and thus make use of the narrative context a more viable option for use within preschool programs, with two potential uses: (1) to provide an in-depth examination of children's strengths and weaknesses in language skills so as to guide instructional planning, and (2) to determine whether children's language skills have improved over time for outcome assessment. Importantly, children's language skills are assessed within the context of an ecologically valid and contextualized narrative task.

Both the tool and the data presented in this study provide a useful opportunity for educators to broaden their use of language assessments with young children. For instance, the data provided in Table 2 show how 262 preschoolers performed on the 18 NAP items; these data may be useful for comparing a given child's use of specific linguistic forms within a narrative context to a broader sample. Importantly, findings of the present research demonstrated that the NAP has reasonable psychometric characteristics for use as a narrative assessment tool to achieve the aforementioned purposes. Concerning reliability, findings showed that NAP scores were highly consistent when applied by two independent coders observing the same narratives. Findings also showed a high level of congruence between coders' NAP scores when applied to online (real-time) narratives observed on video and those applied independently to verbatim transcribed narratives. In other words, NAP scores derived from online analysis of videotaped narratives (involving no transcription) did not differ in any practical sense from NAP scores derived from transcripts of those same narratives. Indeed, the correlation between NAP sum scores scored from video and those scored on transcripts was .92 ($p < .001$). The most salient difference in online and offline coding involved identification of modifiers (adverbs, adjectives), which were coded at significantly higher rates on transcripts compared to video. We speculate that modifiers are particularly troublesome because they occur so infrequently across these transcripts; for instance, adverbs averaged only .05 instances per narrative across the sample. With so few instances occurring across the narratives, it may be that coders are not primed to recognize modifiers when they do occur, although working from written transcripts (rather than more fleeting videos) seems to provide an advantage to the ease of identifying these structures. It is also possible that the training materials do not provide sufficient opportunities to observe modifiers across practice sessions, given their scarcity, and thus coders are not as prepared to identify these structures as compared to others when working from video.

Nonetheless, with the exception of modifiers, findings do suggest the viability of coding children's narratives online rather than transcription. Free from the constraints of transcription, this finding is an important one, as it increases the possibility that narrative assessment can be used in at-scale applications by individuals who do not have the resources needed for transcription (e.g., specialized software, trained transcribers). This finding is an important one given that the NAP was designed to bypass the need for transcription of narrative samples prior to analyses. To our knowledge, this is the first study showing that online coding of microstructural features of children's narratives can be done reliably. Moreover, that NAP scores based on video were so similar to those based on verbatim transcripts shows that there is no need for narrative samples to be transcribed for the NAP protocol to be used.

Additional goals of this study concerned assessing elements of the validity of the NAP, to include examining the factorial structure of children's NAP scores across two time points, to determine whether the nature of NAP scores remained the same over time, and examining criterion-related validity of the NAP in relation to a commonly used standardized and norm-referenced test of children's general language abilities. Factor analysis based on a confirmatory approach using children's NAP scores from fall and spring indicated that NAP scores exhibit stability over time and that the measure has strong construct validity as a univariate measure. Assessment of factor loadings for individual items indicated that six items had relatively low loadings on a single narrative skills factor, and these items were eliminated in a Short Form of the assessment tool. The availability of a Short Form improves the utility of this tool, as it lessens the number of items to be scored by coders from 18 (Long Form) to 12 (Short Form). Importantly, reliability data presented in this research were based on the Long Form; it is possible that the Short Form may make this tool more viable for a wider variety of end-

users (e.g., educators with very little prior training in language assessment), a possibility that should be explored in future research.

The investigation of concurrent and predictive correlations between NAP sum scores (Long and Short Forms) and subtest scores and Language Composite from the CELF Preschool-2 provide strong evidence that both represent similar constructs and, by extension, that the NAP serves as an index of children's language abilities, albeit in a more naturalistic context. The correlation between the NAP Short Form sum score and the concurrent Core Language Composite of the CELF Preschool-2 was .34; the correlation between the NAP Short Form sum score and children's later composite scores was .44. The relationships seen among scores provide reasonable evidence that NAP scores provide a representation of children's language abilities more generally. Given that the correlations were moderate in size, it seems that NAP scores and CELF Preschool-2 scores assess related but different aspects of children's language abilities. The exploratory analyses that examined NAP scores for children with low initial language skills (as assessed on the norm-referenced CELF Preschool-2) and children receiving special education services also indicated that NAP scores were significantly different as a function of these child characteristics. Although exploratory in nature, these results suggest that narrative assessment should be explored as a component of diagnostic assessments for young children.

Because few if any standardized measures are available to assess preschool-aged children's language expression within a narrative context, the evidence presented here suggests that the NAP may fill this gap to provide descriptive-developmental information about children's narrative skills specific to the microstructural domain. The descriptive data it provides may be useful to classroom teachers and allied professionals as they seek to provide high-quality language instruction within the preschool setting and to address standards of learning that specifically focus on skills within the domain of expressive narration (e.g., "Retell or re-enact events from a story;" Ohio Department of Education, 2008, pp. 26–27). Understanding children's strengths in narration as well as their unique limitations may be important for educators as they work to differentiate instruction so as to align instructional approaches to children's distinct needs within the domain of language development. For instance, for the 4-year-old child whose narrative is assessed using the NAP and who is found to use no elaborated noun phrases in his spoken narrative, a preschool teacher may subsequently increase the frequency with which this child hears models of elaborated noun phrases during various classroom activities (Justice, Mashburn, Pence, & Wiggins, 2008).

There are several limitations to this work that warrant discussion. First, this study involved only children residing in a single state, all of whom attended need-based preschool programs. In some analyses, the sample size was smaller, involving only a subset of children. It is unclear whether the findings of this work may be appropriately generalized to other children, including those reared in more advantaged circumstances or those from linguistic or cultural backgrounds that differ from those of our sample. For instance, its appropriateness for children who are learning English as a second language has not been ascertained.

Second, we have argued in this work that the NAP may be scored online (via video) and that this features makes the NAP a potentially scalable assessment tool. However, the NAP coders trained for this study were all undergraduate or graduate students in a university research lab, pursuing degrees in education or communication disorders. An important future goal of research on the NAP, given our interest in its scalability, is determining whether field-based professionals can use the tool reliably following use of the online self-study materials. Research findings have raised concerns about the level of knowledge concerning language and literacy that educators in the field exhibit (e.g., Cunningham, Perry, Stanovich, & Stanovich, 2004), and it is also the case that many early childhood educators may have had little formal training on assessment, language structures, and language development. While these circumstances do pose challenges to the potential scalability of this tool, future research can explore various approaches to promoting educators' reliable use of this and other assessment practices, such as direct workshop experiences and expanded practice opportunities.

Third, this report did not assess the diagnostic or screening accuracy for the NAP. It is unclear whether children who perform very poorly on the NAP at a specific point in time (e.g., scores ≤ -2 *SD* of the mean) are at an elevated risk for future language and/or reading comprehension problems. In other words, we did not assess the sensitivity and specificity of this tool with respect to identifying children exhibiting risk in language development. This is an important avenue for future research, as tools with high levels of diagnostic accuracy can be used to identify children for supplemental interventions within response-to-intervention frameworks.

Finally, the tool described in this study examines only several features of children's narrative expression, and thus offers only a partial glimpse into this important aspect of children's language development. The NAP does not examine aspects of macrostructure that may be particularly informative to early childhood professionals, such as a child's inclusion of story grammar elements. Even within the framework of microstructure, the NAP does not examine narrative features that may also be quite important to examine, including coherence of the narrative as represented through lexical and grammatical ties. Of additional note is that the NAP examines narrative expression, and does not attempt to study narrative comprehension. We are hopeful that by identifying these limitations of the NAP as an index of narrative ability, we might stimulate research that pushes the field toward identifying not only what features of narrative ability ought to be assessed, but also how these features can be reliably assessed in scalable methods.

To sum, the NAP provides early educators, including classroom teachers, a viable tool for scaling up the use of narrative assessment within preschool settings. The NAP provides descriptive-developmental information of five aspects of children's language abilities within a narrative context, including sentence structure, phrase structure, modifiers, nouns, and verbs. The tool can be used reliably by trained coders who score children's fictional narratives in real time from videotape. The research findings reported here show that NAP scores serve as valid indicators of children's narrative abilities specifically

and language abilities generally; yet, the tool can be administered at very low costs and with few specialized tools. In this regard, the NAP presents a viable option for implementing narrative analysis within preschool settings and taking narrative assessment to scale. Further development of this tool using sophisticated measurement approaches may focus on assessing its diagnostic accuracy so as to broaden the purposes it may serve. Coupling this tool with complementary measures that assess other aspects of children's narrative skill, such as those that capture global properties of narrative macrostructure, may provide the field of early childhood education a particularly rich way to represent children's emerging language skills within naturalistic and authentic contexts.

Acknowledgements

The findings reported here are based on research supported by the U.S. Department of Education Institute of Education Sciences through grant R305G050005 to The Ohio State University. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Department of Education.

Appendix A. Narrative Assessment Protocol (NAP) Narrative Elicitation Protocol (Words in bold are spoken to the child)

1. Introduce the task

   Show the cover of the book to the child. Say: **We are going to look at a storybook together. It is a special storybook that doesn't have any words in it. It only has pictures. The first time we look at the book, look carefully at the pictures. The second time, we will look at each page again, and you will tell me a story about the pictures**.
2. Look at the pictures silently

   Hold the book open so that your hands do not cover the pictures. Begin the task by saying: **Let's look at the pictures together. Be sure to look carefully at each page.** Open the book to the first picture. Say to the child: **Look at this picture.** Draw the child's attention to the picture by *circling* the page with your finger. Allow about 5–10 s for the child to look at each page before turning to the next page. If it seems that the child is off task or not looking at both pages of the book, you may prompt the child by saying:
   - **Look at this page** or
   - **Now look at this one**.

   Do not comment about the pictures in any way. If the child offers comments or asks questions, gently remind him/her that you are just looking at the pictures this time, and he/she can tell the story the next time you look at the book.
3. Elicit a narrative from the child

   Close the book and hand it to the child. Say: **Now I want you to tell me a make-believe story using the pictures in this book. Make up a story for me that tells about the pictures in this book. Try to make the story as long as you can. Use the pictures in this book to tell me a story**. If the child is hesitant to tell a story, or tells you "I don't know how to read," you may prompt the child by saying:
   - **This book doesn't have any words in it, so you can tell any story you want about the pictures. Tell me a story about the pictures.**

   During the story production, you may provide prompts such as:
   - **Tell me about this page** or
   - **What about this page?**

   As the child produces a story, you may also repeat *exactly* what the child says about a picture. This provides an acknowledgement to the child that you are listening, and also will help future coders to understand an utterance that may be unintelligible on video/audio. *Do not change* what the child says in your repetition, even if there are errors. For example, if the child says "Frog goed in the water," you repeat "Frog goed in the water." *Do not correct the utterance* (e.g., "Frog went in the water."). At the end of the book, say: **Is there anything else you want to add to your story?**

Appendix B. Narrative Assessment Protocol (NAP): Administration Directions

Directions:

1. Begin coding the child's narrative after the examiner provides directions for the child to begin telling a story. Refer to the Coding Catalog for specific descriptions of each NAP item along with examples.
2. Each time a child uses one of the NAP items, make notes in the space provided and circle the corresponding number (0, 1, 2, 3+) to track the frequency of occurrence. For Modifiers, Nouns, and Verbs, record unique usages in the spaces provided (1, 2, 3) and score only these unique usages. For instance, if a child uses the adverb "slowly" twice, only code the first usage as an Adverb item. Unique usages are not required for Sentence Structure and Phrase Structure.
3. Do not code off-topic discourse, such as conversation with the examiner about the camera or other conversation not pertaining to the book. Do not code questions about the book addressed to the examiner (e.g., Is that a frog?)

4. If a child self-corrects, score the corrected form.
5. If a child uses items that are not grammatically correct (per Standard American English) but contain specific items required by an item, give credit for these items. For example, if the child says "They was looking out the window" or "They be looking out the window," give the child credit for **auxiliary verb + main verb**. As another example, if the child says "He don't know it's a tree" or "He not know it's a tree," give the child credit for **negative sentence** because both constructions convey negation.
6. If the examiner recasts or expand a child's utterance to a more sophisticated form, and then the child imitates the more sophisticated utterance, do not give the child credit for the more sophisticated utterance.
7. If the child repeats an utterance because the examiner asks for clarification, do not code the repeated utterance. For example, if the child says "He said, where are you frog?" and the examiner says "What?" and the child repeats "He said, where are you frog?," do not code the repetition. If the child repeats without prompting from the examiner, code the repeated utterance.

Appendix C. Narrative Assessment Protocol (NAP): Long Form

Child Name: Child Age:	Coder: Narrative Collection Date: NAP Scoring Date:

NAP Items	Frequency				Examples
Sentence Structure					
Compound sentence*	0	1	2	3+	She likes it so she'll buy it.
Complex sentence	0	1	2	3+	That boy who hit me is mean.
Negative sentence	0	1	2	3+	The frog can't go there.
Interrogative sentence*	0	1	2	3+	Frog, are you hiding in my boot?
Phrase Structure					
Elaborated noun phrase	0	1	2	3+	*The little dog* saw the frog.
Compound noun*	0	1	2	3+	*The dog and the boy* got it.
Prepositional phrase	0	1	2	3+	The boy looked *in his boot.*
Modifiers					
Adverb* 1. 2. 3.	0	1	2	3+	He was *really* angry.
Advanced modifier 1. 2. 3.	0	1	2	3+	The frog was in the *filthy* water.
Nouns					
Pluralized noun 1. 2. 3.	0	1	2	3+	The five frog*s* got in.
Possessive noun* 1. 2. 3.	0	1	2	3	The boy*'s* hat was lost.

Appendix C (*Continued*)

Tier-two noun 1. 2. 3.	0	1	2	3+	She was the *doe* in the woods.
Verbs					
Auxiliary verb + main verb 1. 2. 3.	0	1	2	3+	The boy *is yelling* at the dog.
Copula 'be' verb + … 1. 2. 3.	0	1	2	3+	The frog *was* here.
Irregular past tense verb 1. 2. 3.	0	1	2	3+	The dog *fell*.
Regular past tense verb 1. 2. 3.	0	1	2	3+	The dog walk*ed*.
Tier-two verb 1. 2. 3.	0	1	2	3+	The frog *whirled* around.
Compound verb* 1. 2. 3.					The frog *danced and ran*.

*These items are omitted on the Short Form.

Appendix D. Narrative Assessment Protocol (NAP): Short Form

Child Name:	Coder:
Child Age:	Narrative Collection Date:
	NAP Scoring Date:

NAP Items	Frequency				Examples
Sentence Structure					
Complex sentence	0	1	2	3+	That boy who hit me is mean.
Negative sentence	0	1	2	3+	The frog can't go there.
Phrase Structure					
Elaborated noun phrase	0	1	2	3+	*The little dog* saw the frog.
Prepositional phrase	0	1	2	3+	The boy looked *in his boot*.
Modifiers					
Advanced modifier 1. 2. 3.	0	1	2	3+	The frog was in the *filthy* water.
Nouns					
Pluralized noun 1. 2. 3.	0	1	2	3+	The five frog*s* got in.
Tier-two noun 1. 2. 3.	0	1	2	3+	She was the *doe* in the woods.
Verbs					
Auxiliary verb + main verb 1. 2. 3.	0	1	2	3+	The boy *is yelling* at the dog.
Copula 'be' verb + … 1. 2. 3.	0	1	2	3+	The frog *was* here.
Irregular past tense verb 1. 2. 3.	0	1	2	3+	The dog *fell*.
Regular past tense verb 1. 2. 3.	0	1	2	3+	The dog walk*ed*.
Tier-two verb 1. 2. 3.	0	1	2	3+	The frog *whirled* around.

Appendix E. Narrative Assessment Protocol (NAP): Coding Catalog

INDICATOR/Items	Description	Example(s)
I. Sentence structure		
1. Compound sentence[a]	Sentences containing a *coordinating conjunction* used to join two or more independent clauses (e.g., ***for***, ***nor***, ***but***, ***or***, ***yet***, ***so***). Do not score ***and***.	She likes it so she'll buy it. The owl came out but the boy ran. He came but the frog didn't.
2. Complex sentence	Sentences containing two verbs (minimally) and a *subordinating conjunction* (e.g., ***when***, ***that***, ***who***, ***which***). They may also use infinitives.	That boy who hit me is mean. She ate the one that I didn't like. I'll call when we get there.
3. Negative sentence	Sentences using ***no*** or ***not*** to negate the meaning of the **Sentence**s including yes-no questions that use "do insertion" or questions that use inversion.	The frog can't go there. He is not alright. That isn't it. Frog, are you hiding in my boot? Is he coming? Why did you do it?
4. Interrogative sentence[a]		
II. Phrase structure		
1. Elaborated noun phrase	Phrase consisting of one or more **modifiers** (*determiner*, *initiator*, *adjectival*) preceding a singular or plural noun. Exclude ***a***, ***an***, and ***the***.	The little dog looked for the frog. The boy made an angry face. It's by the black door.
2. Compound noun[a]	Coordinating conjunction used to join nouns as subjects or objects (e.g., ***and***, ***nor***, ***but***, ***or***).	The dog and the boy are coming. The hat or mittens in there.
3. Prepositional phrase	Phrase consisting of the object of the preposition, the preposition, and any associated adverbs or adjectives.	He and I are on it. The frog hopped out the window. Look on the table.
III. Modifiers[b]		
1. Adverb[a]	A word ending in -ly that modifies a verb. (Other adverb forms not counted.)	He is really happy. He got me really. It's nearly over.
2. Advanced modifier	Single-word adjectives or adverbs that add significant precision to a noun or verb.	The frog was in the filthy water. The girl seems frustrated. The dangerous cat got it.
IV. Nouns[b]		
1. Pluralized noun	Noun with plural marker attached (e.g., /***s***/, /***z***/, /***Iz***/) or an irregular plural.	The five frogs got it. The families are all coming. He jumped in puddles.
2. Possessive form[a]	Noun with possessive marker attached (e.g., /s/, /z/, /Iz/).	The boy's hats are in that. The frog's hands are pointy. Her boy's frog is missing.
3. Tier-two noun	Noun that is a more complex, precise, or interesting form of a common noun. Proper nouns are not included.	She was the doe in the woods. The frog jumped into the barrel. She put on the bonnet.
V. Verbs[b]		
1. Auxiliary verb + main verb	"Helping verb" used in contracted or uncontracted form in present progressive or past progressive verb construction.	The boy is yelling at the dog. I am going home. That dog's coming.
2. Irregular past tense verb	Verb that marks past tense irregularly.	She swam there. The dog fell. The boy ran there.
3. Reg past tense verb	Verb with regular past tense marker attached: /Id/ or /t/.	The dog walked. The girl jumped. The boy tickled it.
4. Tier-two verb	Verb that is more precise or complex form of a verb (Beck, McKeown, & Kucan, 2002) or is a synonym for a basic-level, all general purpose verb (e.g., make, do, go.).	The frog whirled around. The frog sneaked out. Please don't annoy her.
5. Compound verb[a]	Coordinating conjunction used to join verbs (e.g., ***and***, ***nor***, ***but***, ***or***). Do not count verbs repeated for emphasis.	The frog danced and ran. The boy walked and talked. Eat and drink it.

[a] These items do not need to be scored for the Short Form.

[b] These indicators require unique usages for each scoring.

References

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life*. New York: Guilford Press.

Berman, R. A. (1995). Narrative competence and storytelling performance: How children tell stories in different contexts. *Journal of Narrative and Life History*, *5*, 285–313.

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*(2), 156–173.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching & Therapy*, *18*, 1–22.

Botting, N., Faragher, B., Smikin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: What differentiates good and poor outcome? *Journal of Child Psychology and Psychiatry*, *42*(8), 1013–1020.

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*(3), 249–260.
Brice-Heath, S. (1983). *Ways with words*. New York: McGraw-Hill.
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–159). Newbury Park, CA: Sage.
Bryan, T., Ergul, C., & Burstein, K. (2008). Curriculum-based measurement. In L. M. Justice, & C. Vukelich (Eds.), *Achieving excellence in preschool literacy instruction* (pp. 317–338). NY: Guilford Press.
Cabell, S., Justice, L. M., Piasta, S., Curenton, S., Wiggins, A. K., & Pence Turnbull, K. (submitted for publication). The impact of a classroom-based language intervention on preschoolers' oral language and emergent literacy skills.
Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32*(1), 38–50.
Cowley, J., & Glasgow, C. (1994). *The Renfrew Bus Story: Language screening by narrative recall*. Circle Pines, MN: American Guidance.
Culatta, B., Page, J. L., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools*, *14*(2), 66–74.
Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Stanovich, P. (2004). Disciplinary knowledge of K-3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia*, *51*, 139–168.
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.
Damico, J., Oller, J. W., & Storey, M. E. (1983). The diagnosis of language disorders in bilingual children: Surface-oriented and pragmatic criteria. *Journal of Speech and Hearing Disorders*, *48*, 385–394.
Engel, S. (1995). *The stories children tell*. New York: W.H. Freeman and Company.
Eriksson, M. (2001). Narratives validate communicative development inventories. *Applied Psycholinguistics*, *22*, 45–60.
Gardner, H., Froud, K., McClelland, A., & van der Lely, H. K. J. (2006). Development of the grammar and phonology screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language & Communication Disorders*, *41*(5), 513–540.
Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech & Hearing Research*, *39*(6), 1258–1262.
Gillam, R. B., & Pearson, N. A. (2004). *Test of narrative language*. Austin, TX: Pro-Ed.
Glasgow, C., Cowley, J. (1994). *The Renfrew Bus Story-American Edition*. Unpublished manuscript. Centreville, DE: Centreville School.
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*(2), 114–125.
Huai, N., Braden, J., White, J., & Elliot, S. N. (2006). Effect of an internet-based professional development program on teachers' assessment literacy for all students. *Teacher Education and Special Education*, *29*, 244–260.
Hughes, D. L., McGillivray, L. R., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal*, *54*(4), 305–309.
Hux, K., Morris-Friehe, M., & Sanger, D. D. (1993). Language sampling practices: A survey of nine states. *Language, Speech, and Hearing Services in Schools*, *24*(2), 84–91.
Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools*, *37*(4), 284–297.
Justice, L., & Ezell, H. K. (2002). *The syntax handbook*. Eau Claire, WI: Thinking Publication.
Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology*, *15*(2), 177–191.
Justice, L. M., Mashburn, A., Pence, K., & Wiggins, A. (2008). Experimental evaluation of a comprehensive language-rich curriculum in at-risk preschools. *Journal of Speech, Language, and Hearing Research*, *51*, 1–19.
Kourtessis, T., Tsigilis, N., Maheridou, M., Ellinoudis, T., & Kiparissis, M. (2008). The influence of a short intervention program on early childhood and physical education teachers' ability to identify children with developmental coordination disorders. *Journal of Early Childhood Teacher Education*, *29*, 276–286.
La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal*, *104*(5), 409–426.
Laing, S., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools*, *34*, 44–55.
Liles, B. Z. (1987). Episode organization and cohesive conjunctives in narratives of children with and without language disorder. *Journal of Speech & Hearing Research*, *30*(2), 185–196.
Mayer, M. (1969). *Frog, where are you?* New York: Dial Books.
McCabe, A., & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research*, *13*, 457–480.
McCabe, A., & Rollins, P. (1994). Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology*, *3*, 45–56.
McDonald, S., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up effective interventions. *Educational Researcher*, *35*, 15–24.
Meredith, W., & Horn, J. L. (2001). The role of factorial invariance in measuring growth and change. In L. M. Collins, & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 201–240). Washington, DC: APA.
Miles, S., & Chapman, R. S. (2002). Narrative content as described by individuals with down syndrome and typically developing children. *Journal of Speech, Language, and Hearing Research*, *45*(1), 175–189.
Miller, J. F., & Iglesias, A. (2006). *Systematic analysis of language transcripts (SALT)*. Madison, WI: University of Wisconsin.
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.
Muthén, B. O., & Muthén, L. K. (2006). *Mplus (Version 4.2)*. Los Angeles, CA: Muthén & Muthén.
Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313.
Ohio Department of Education. (2008). *Early learning primary content standards for English language arts*. Columbus, OH: Ohio Department of Education.
Pankratz, M. E., Plante, E., Vance, R., & Insalaco, D. M. (2007). The diagnostic and predictive validity of the Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools*, *38*(4), 390–399.
Pearce, W. P., McCormack, P. F., & James, D. G. H. (2003). Exploring the boundaries of SLI: Findings from morphosyntactic and story grammar analysis. *Clinical Linguistics and Phonetics*, *17*(4–5), 325–334.
Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., et al. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1037–1057.
Pianta, R. C., Mashburn, A., Downer, J., Hamre, B., & Justice, L. M. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*, 431–451.
Renfrew, C. (1969). *The bus story: A test of continuous speech*. Oxford, UK.
Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1398–1411.

Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J. (1999). Authentic assessment as an approach to preschool speech-language screening. *American Journal of Speech-Language Pathology*, *6*, 195–200.
State of Connecticut State Board of Education. (2006). *The Connecticut Framework: Preschool Curriculum Framework*. Hartford, CT: State of Connecticut State Board of Education.
Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.
Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, *41*, 407–418.
Sturner, R. A., Layton, T. L., Evans, A. W., Heller, J. H., Funk, S. G., & Machon, M. W. (1994). Preschool speech and language screening: A review of currently available tests. *American Journal of Speech-Language Pathology*, *3*, 25–36.
Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, *39*(6), 1284–1294.
Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.
Ukrainetz, T., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy*, *18*, 59–78.
Ukrainetz, T., Justice, L. M., Kaderavek, J., Eisenberg, S., & Gillam, R. (2005). Artful storytelling: The development of expressive elaboration in fictional narratives. *Journal of Speech, Language, and Hearing Research*, *48*, 1363–1377.
Wackwitz, J. H., & Horn, J. L. (1971). On obtaining the best estimates of factor scores within an ideal simple structure. *Multivariate Behavioral Research*, *6*, 389–408.
Washington, J. A., & Craig, H. K. (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology*, *13*, 329–340.
Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals—Preschool* (second ed. (CELF preschool-2)). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.
Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, Speech, and Hearing Services in Schools*, *36*(1), 62–72.