Storytelling from pictures using the Edmonton Narrative Norms Instrument

Évaluer grâce au « Edmonton Narrative Norms Instrument » une histoire contée à partir d'images

Phyllis Schneider
Denyse Hayward
Rita Vis Dubé

Abstract

This paper describes the development of an instrument for collecting story samples and local norms, the Edmonton Narrative Norms Instrument (ENNI), and presents results of one measure of storytelling ability used with the ENNI data, story grammar units (SGU). The purpose of the present study was to examine the measure's ability to detect developmental changes in story production and discriminate between children with and without language impairments. Participants were 377 children aged 4-9 (300 with typical development, 77 with language impairments). Each child told stories while looking at picture stimuli (with no oral model). The SGU measure revealed a significant age trend. The measure correctly classified children aged 4-8 into groups 80.8% of the time. Thus the ENNI shows promise as part of the speech-language pathologist's battery of instruments.

Abrégé

Cet article décrit l'élaboration d'un instrument de collecte d'échantillons d'histoires et de normes locales, baptisé l'Edmonton Narrative Norms Instrument (ENNI). Il présente les résultats d'une mesure de capacité à conter une histoire faite à partir de l'ENNI, soit les unités de grammaire de l'histoire. Cette étude visait à examiner si cette mesure peut déceler des changements de la capacité à raconter une histoire et si elle peut faire la différence entre les enfants ayant un trouble du langage et ceux qui n'en ont pas. Elle a porté sur 377 enfants de 4 à 9 ans (300 ayant un développement type et 77 ayant un trouble). Chaque enfant a raconté des histoires à partir d'images (sans modèle oral). Les mesures d'unités de grammaire de l'histoire ont montré une tendance importante par âge. La mesure a correctement classé les enfants de 4 à 8 ans en groupes dans 80,8 % des cas. Ainsi, l'ENNI semble un outil prometteur pour le travail des orthophonistes.

Key words: narratives; stories; children; language impairments; language assessment

*Marilyn Kertoy served as Acting Editor for the review of this article.

Phyllis Schneider, *Ph.D.,*
Department of Speech Pathology and Audiology, University of Alberta
Edmonton, Alberta Canada

Denyse Hayward, *Ph.D.,*
J. P. Das Developmental Disabilities Centre, University of Alberta
Edmonton, Alberta Canada

Rita Vis Dubé, *Ph.D.,*
Toronto District School Board
Toronto, Ontario Canada

Introduction

Story tasks have become a common feature of clinical assessment and intervention in the field of speech-language pathology. Their use is recommended in many commonly-used textbooks (e.g., Lahey, 1988; Lund & Duchan, 1993; Paul, 2001). There are many good reasons for the use of stories in the clinic. For example, as compared to most tests of language which assess the use of words and sentences in isolation, stories require children to combine words and sentences for a particular purpose. Thus they provide information about how well children can use their discrete language skills to communicate. In addition, stories are a part of everyday life in interaction with others, in educational contexts, and in recreation (e.g., books, television, and film). Oral stories are considered to be a form of literate language and to serve as a bridge between oral and written language styles (Westby, 1999). Evidence supporting stories as bridge between oral and written styles comes from several studies that found that, in contrast to conversation, children's stories had features characteristic

of written language such as longer mean length of utterances (MacLachlan & Chapman, 1988; Wagner, Nettelbladt, Sahlén, & Nilholm, 2000), more syntactically complex language (Westerveld, Gillon, & Miller, 2004), and more phrasal expansions (Wagner, Nettelbladt, Sahlén, & Nilholm, 2000).

The ability to produce and understand stories has been found to be impaired in children with learning disabilities, even when the children have not previously been found to have problems in basic language skills (Ripich & Griffith, 1988; Roth & Spekman, 1986). Similar problems have been identified in children with specific language impairments and those with language/learning disabilities (Bishop & Adams, 1992; Liles, 1985, 1987; Merritt & Liles, 1987, 1989; Paul & Smith, 1993). Story production scores have been found to predict academic achievement of children with typical development (Griffin, Hemphill, Camp, & Wolf, 2004; O'Neill, Pearce, & Pick, 2004), children with learning disabilities (Feagans & Appelbaum, 1986) and children at risk for language impairments (Fazio, Naremore, & Connell, 1996). For children with early language impairments, preschool story production predicts later oral language development (Bishop & Edmondson, 1987), while story comprehension has been found to predict reading comprehension (Bishop & Adams, 1990). Thus story tasks appear to discriminate between children with impairments and those without in a real-life language context (that is, one that is more like everyday language use than an assessment context focusing on discrete language skills), making story tasks potentially valuable ones for assessment and intervention.

To be clinically useful, tasks and materials used for story assessment need to have normative information associated with them. Without normative information, it is impossible to determine with certainty whether a particular child is telling stories as we should expect for the child's age. Although normative information is available for some narrative tasks and some limited age ranges (e.g., Cowley & Glasgow, 1994; Hughes, McGillivray, & Schmidek, 1997; Strong, 1998), we felt that it would be useful to have normative information on children from preschool to school age using a set of carefully designed story stimuli.[1] We designed the Edmonton Narrative Norms Instrument (ENNI) to collect local norms using story stimuli designed for the purpose. Local norms have been proposed as a valuable resource for the assessment of children's storytelling ability (Hughes, McGillivray, & Schmidek, 1997). We began with local norms in order to develop the ENNI and its measures on a comprehensive set of data representative of the Edmonton demographic profile. If the ENNI proved useful for this local setting, its usefulness for other settings (geographic and linguistic) could eventually be explored.

We decided to use a task that would require children to formulate stories from pictures rather than to retell orally presented stories, as formulation appears to provide more information about children's independent storytelling abilities. Using oral and pictorial stories that were structurally comparable, Schneider (1996) and Schneider and Dubé (1997, 2005) found that children tell stories that vary in quality depending on how they were originally presented. Schneider (1996) found that children with language impairments aged 5 to 9 provided more story information when they retold stories they had heard than when they had to formulate stories from pictures without hearing an oral version. A similar pattern of results was obtained in another study for children with typical development in Kindergarten and Grade 2. Children in both grades provided more story information (Schneider & Dubé, 2005) and better referential cohesion (Schneider & Dubé, 1997) when retelling oral stories than when formulating stories from pictures alone. Both types of tasks provide valuable information about children's storytelling abilities, but each provides different insights into these abilities. Picture stories reveal the child's ability to formulate the story as opposed to the ability to recall a story formulated by someone else as in retell tasks. Thus picture story tasks appear to tax children's independent abilities to a greater extent than story retell tasks and may therefore detect problems that would not be evident in retell tasks. To gain a complete picture of a child's storytelling abilities, it is important to assess storytelling along a continuum of difficulty. Since oral retell tasks are already available (Cowley & Glasgow, 1994; Strong, 1998), we wanted to develop a picture-based instrument that would require formulation to increase the available story instruments with normative information.

To enable valid and reliable scoring of the stories told from pictures, it was important that the stimuli clearly depicted stories that fit some model of a good story. Stimuli for the ENNI were carefully designed according to a model of story knowledge, story grammar.

The Story Grammar Model

The story grammar model describes the information that adults identify as essential to "good" stories, and that adults and older children typically include in their stories (Stein & Policastro, 1984). Although different researchers have posited somewhat different schematic organizations (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977), there is agreement on the basic components of the model. Stories consist of sets of sequentially related categories of units and each category refers to different types of information that serve specific functions in the story. Table 1 describes these categories and their definitions.

Within the model, there is at least one central character who is motivated to carry out some type of goal-directed action. A story revolves around an attempt or attempts by the central character(s) to attain a goal and the story ends with an outcome in which the goal may or may not be successfully achieved.

There are two major components to the story grammar model: structural patterns and story grammar units or elements. Structural patterns describe the overall content and organization of stories; the basic pattern that

Table 1

Story Grammar Units

Setting:
Characters in the story
Location, activity, and/or habitual state or characteristic ('he was always hungry'; 'she liked to read')
Initiating Event: event that sets off the story's events; will cause the protagonist to respond in some way
Internal Response: reaction of protagonist to the initiating event
Internal Plan: of protagonist to deal with the Initiating Event
Attempt to obtain the goal
Outcome or Consequence of Attempt
Reaction: how the character(s) feel or think about the outcome, or how they react physically (e.g. run away)

would be considered adequate is called a *complete episode*, with some patterns being less well developed and others being elaborations of the complete episode. Story grammar units are the categories of information that are typically provided in a certain order within episodes. They can be considered core story content that would typically be included in good stories. However, some units, such as initiating events, attempts, and outcomes, are typically included more often than others, such as units describing inner thoughts and feelings (Hughes, McGillivray, & Schmidek, 1997; Stein & Glenn, 1979).

After hearing stories corresponding to a complete episode according to the story grammar model, children as young as age 4 recall stories according to the idealized schema (Stein & Albro, 1997; Stein & Glenn, 1979). Information that fits one of the story grammar unit categories is more likely to be included in children's story retellings than information that does not fit into one of the categories (Goldman & Varnhagen, 1986). When children and adults are asked to retell a story that does not conform to the story grammar model, their retellings contain changes that result in closer conformity to the model (Stein & Glenn, 1979). Furthermore, when asked to retell stories in which specific categories were omitted, adults and children add information that corresponds to the omitted components in their narrations (Mandler & Johnson, 1977; Stein & Glenn, 1979).

In general, the experimental evidence supports the distinctions among the units of the model. Thus the story grammar model appears to be a valid representation of how individuals organize story information in order to encode, understand and retrieve stories. The acquisition of story schema knowledge appears to develop as a function of age with older children's stories approximating the competency observed in adults.

The story grammar model has been recommended for clinical use by several authors (e.g., Hedberg & Stoel-Gammon, 1986; Hughes, McGillivray, & Schmidek, 1997). Researchers have applied the model to compare stories told by children with and without language impairments, using a variety of elicitation techniques such as story-stem completion (e.g., Merritt & Liles, 1987), retell of orally presented stories (Ripich & Griffith, 1988; Schneider, Williams, & Hickmann, 1997), and formulations from film viewings (e.g., Liles, 1987; Merritt & Liles, 1987) and pictures (e.g., Ripich & Griffith, 1988).

Recently, the clinical utility of story grammar analyses has been questioned. Liles, Duffy, Merritt, and Purcell (1995) reanalyzed data from previous studies by Liles and her colleagues and found that their measures derived from the story grammar model did not contribute to predictions of group membership (typical development or language disorder) in a discriminant function analysis. However, this lack of discrimination may have been related to the choice of story grammar measures. Their measures were percent of possible episodes (complete and incomplete) included in story retell and total number of episodes (complete and incomplete). A complete episode was defined as one containing three Story Grammar units: an initiating event or internal response, an attempt, and a consequence; an incomplete episode was defined as an episode containing one or two of these units. The inclusion of both complete and incomplete episodes in the scores may have obscured differences in the amount of story content included in children's stories. That is, two children with the same number of complete and incomplete episodes could have very different numbers of story grammar units. Scoring based on story grammar units included in a story may be more sensitive to language status in that it is a more direct measure of amount of basic information included in the stories.

Research Questions

The current study investigated whether story grammar units would provide a measure of the development of storytelling from ages 4 to 9, specifically in terms of an increase in the amount of information included in stories. We also wanted to determine whether and to what degree story grammar unit scores would discriminate between groups of children with and without previously identified language impairments. Because the goal of this paper was to investigate the usefulness of the story grammar analysis for normative purposes, the research questions focused on examining developmental

trends, differences between typical development and language impairment groups, and discrimination between these groups.

1) Are developmental trends evident in the data for amount of story grammar information? That is, does the amount of story grammar information increase with age?

2) Are there differences between the groups in the amount of story grammar information included in children's stories within each age group?

3) Do story grammar scores discriminate between children with and without language impairments?

4) Do story grammar scores correlate with a standardized test of language?

Methods

Participants

A total of 377 children, ages 4;0 through 9;11, served as participants. The sample consisted of two subgroups within every age group: a wide range of children with typical development and children previously identified as having a language impairment. Because we wanted the instrument to be useful for language assessment, we considered it essential to include children with language impairments in the normative sample. If groups with special needs are excluded from normative samples, then the interpretation of data from children from excluded groups is difficult. This is because if a child from an excluded group receives a score that was received by any children in the normative sample, even if significantly below the mean, then that child has scored similarly to children with no identified problems (Ukrainetz McFadden, 1996). Because we intended the norms to be particularly useful for professionals interested in language impairment, special care was taken to include a sample of children previously identified as having a specific language impairment. The term *specific language impairment* (SLI) refers to a problem in language that is not due to another condition such as general developmental delay or sensory impairment. We included children with language impairments who had a language/learning disability, as well as those who had Attention Deficit Disorder with or without hyperactivity if controlled through medication. Our definition does rule out children who may be receiving services for language impairments who have other conditions; thus, the participants are not representative of the full range of children receiving language services in Edmonton. However, as a first step, we decided to focus on the SLI population to make the best use of our resources. A commonly-cited estimate of the prevalence of specific language impairment is 7.4% of the population, from a study by Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien (1997) of Kindergarten children. Although other estimates exist and will vary according to the measures used to define SLI, we adopted Tomblin et al.'s estimate due to its relative recency and its focus on specific language impairments rather than language impairments more broadly defined. To assure as representative a subsample as possible without overrepresenting children with specific language impairments, the subsample was oversampled (i.e., sampled at more than 7.4% of the total), with subsequent weighting of subsample data when calculating norms.

Sample size for the children with typically developing language was 50 per age group (one-year intervals), with equal numbers of boys and girls. The goal for children with language impairments was 15 per age group; the obtained sample varies from 10 to 17 children per age group. Gender was left to vary in this group. As expected, there were more boys than girls (48 of 77 – 62%) in the group with language impairments. Sample information is summarized in Table 2.

Children in the school-age range were chosen from children attending Kindergarten through Grade 4 in Edmonton public and separate schools. The younger children were chosen from those attending preschools, daycare centres, and Kindergarten programs in Edmonton. Schools were randomly selected from areas across Edmonton to assure a cross-section of socioeconomic groups. In all, 34 schools and 13 daycares, preschools and independent Kindergarten programs were visited to collect the data. All participants spoke English as a first language at home.

To identify potential children with typical development for the study, each teacher was asked to refer two children in the upper level of achievement, two children from the middle level, and two children in the lower level (one boy and one girl at each level). In all cases, the children who were referred for the typical development sample were not to have speech or language difficulties or any other diagnostic label such as attention deficit disorder, learning disability, or autism. All children who were referred to the typical development group, whose parents gave permission, and who verbally assented to participate were included in the study.

The subsample of children with language impairments was obtained with the cooperation of three sites: a public school serving children with language/learning disabilities; a rehabilitation hospital that has several programs for children with language impairments; and Capital Health Authority, which serves preschool and school-aged children throughout the city. Sites were asked to refer children with a rating of 2 to 5 on Capital Health's Severity Rating Scale, which ranges from 1 (mild) to 5 (severe). Children could be referred if they had fine or gross motor delays, attention deficit disorder with or without hyperactivity (ADD/ADHD) with medication, a diagnosed learning disability, or mild or moderate speech disorder. Sites were asked not to refer children who had received diagnoses of mental retardation, ADD/ADHD without medication, autism, hearing impairment, severe visual impairment that would result in inability to see pictures even with correction, or severe speech impairments that would preclude accurate orthographic transcription of their stories. IQ test information was not collected; the speech-language pathologists referring

Table 2

Number, Age and Socioeconomic Status Information on the Normative Sample

Age Group	Language Group	N	N Boys	Mean Age	Age SD	Age Range	Mean SES	SES SD	SES Range
4	TD	50	25	4.60	.24	4.04 - 4.97	47.38	13.58	23.70 - 82.91
	SLI	12	9	4.66	.23	4.18 - 4.97	47.17	10.80	34.45 - 70.27
5	TD	50	25	5.51	.27	5.01 - 5.98	46.64	12.12	24.11 - 73.38
	SLI	14	8	5.41	.26	5.07 - 5.85	46.52	12.00	25.53 - 63.64
6	TD	50	25	6.56	.29	6.04 - 6.95	48.31	14.75	25.53 - 101.53
	SLI	11	6	6.64	.26	6.13 - 6.95	40.26	13.97	26.36 - 60.73
7	TD	50	25	7.54	.28	7.01 - 7.98	45.13	13.65	24.11 - 101.32
	SLI	13	10	7.56	.23	7.15 - 7.92	42.42	13.30	23.70 - 65.43
8	TD	50	25	8.58	.28	8.01 - 8.99	45.04	11.55	23.70 - 75.87
	SLI	17	10	8.70	.26	8.11 - 8.96	42.42	7.40	32.78 - 60.73
9	TD	50	25	9.49	.28	9.02 - 9.99	48.79	12.04	25.56 - 80.32
	SLI	10	5	9.50	.21	9.10 - 9.82	48.71	9.66	27.60 - 60.73
Total	**TD**	**300**	**150**	**7.05**	**1.72**	**4.04 - 9.99**	**48.66**	**12.95**	**23.70 - 101.53**
sample	**SLI**	**77**	**48**	**7.09**	**1.71**	**4.18 - 9.82**	**44.43**	**11.27**	**23.70 - 70.27**

children for the study were asked to refer children for whom they had no concerns regarding cognitive abilities.

Demographic information was collected on the families of participating children to permit description of socioeconomic status and ethnic composition of the sample. The purpose of collecting demographic information was to ensure a sample representative of the Edmonton population.

Socioeconomic status (SES) was estimated from parents' occupations using the Blishen Scales (Blishen, Carroll, & Moore, 1987). Based on Canadian census information, this index reflects equally weighted components of education and income level by occupation. Scores of the Blishen scale range from 17.81 (newspaper carriers and vendors) to 101.74 (dentists) with a mean of 42.74 (SD=13.28). SES is reported for each age group in Table 2.

Ethnic composition corresponded closely to the range of ethnic diversity in the city of Edmonton according to Statistics Canada data (Statistics Canada, no date). Data collection was conducted throughout the school year, with care taken to collect data from the full age range throughout the year so that no one age group was sampled at a different point in the school year than another age group.

All children in the SLI group and 88 children in the TD group (29%) were given the Clinical Test of Language Fundamentals (CELF), using either the CELF-Preschool (Wiig, Secord, & Semel, 1992) for children under 6 years of age, or the CELF-III (Semel, Wiig, & Secord, 1995) for children aged 6 and over. The CELF tests were chosen because they are very commonly used in Edmonton to assess children's language abilities. The children in the TD group who were not tested on the full CELF were tested on two subtests of the CELF test appropriate to their age groups. Subtests from the CELF-P were Linguistic Concepts and Recalling Sentences in Context. Subtests from the CELF-III were Concepts and Directions and Recalling Sentences, which are analogous to the CELF-P subtests used. These two subtests were chosen in order to have information on one receptive subtest and one expressive subtest for all the children. The CELF-P manual recommends these two subtests for use as language screening ("Quick Tests"), with the rest of the test administered if either of the subtest scores falls below 1 standard deviation (Wiig, Secord, & Semel, 1992). Means for the typical development and specific language impairment groups are reported in Table 3. There were an additional 19 children in the language impairment group who attained a score of 85 on both the receptive and expressive language total score of the CELF-P or CELF-3; these children were excluded from the sample.

Materials

Six original picture sets with animal characters were used to elicit stories, two each at three levels of complexity. The stories were controlled in pairs and systematically varied across levels for length, amount of story information, and number and gender of characters. Table 4 summarizes the characteristics of each story set. These picture stories were designed to provide a range of narrative complexity. At the low end, the simple stories depict a single episode with two characters; at the complex end, the complex stories depict a complex episode with multiple attempts at a goal and four characters. Scripts controlling for these factors for stories to be portrayed by the pictures were written by Dubé (2000) for her doctoral research investigating the language skills of Deaf children. A panel of narrative experts was asked to comment on the scripts with regard to their narrative structure and their appropriateness for children; the stories were revised based on comments from the panel. The black and white line drawing pictures were then drawn by a professional cartoonist following the scripts. The pictures were given once again to the panel of narrative experts, as well as to a panel of teachers of Deaf children. Both panels approved the pictures as appropriate for research with children. The pictures may be viewed on and downloaded from the ENNI website (Schneider, Dubé, & Hayward, 2004b).

The pictures for each story were placed in page protectors in a binder. Each story was in its own binder.

Procedure

Three research assistants were employed to collect the data. None were speech-language pathologists. The second author (who is a registered speech-language pathologist) administered the standardized testing. She was blind to the performance of the children in the storytelling production. The research assistants and second author were not blind to the status of the children (typical development or language impairment), as many children were tested in schools for children with language and learning disabilities.

Each child was seen individually in the child's school, preschool, or daycare in two sessions. The storytelling task was administered in the first session. When administering each story, the examiner first went through all the pages so that the child could preview the story, after which the examiner turned the pages again as the child told the story. The examiner turned the page when the child appeared to be finished telling the story for a particular picture. The examiner held the binder in such a way that she could not see the pictures as the child told the story, which meant that the child needed to be explicit if the examiner was to understand the story; the child could not legitimately use pointing in lieu of language when telling the story. Instructions informed the child that the examiner would allow the child to see all the pages first, and then the child would be asked to tell the story to the examiner. The instructions emphasized that the examiner would not be able to see the pictures, so the child would have to tell a very good story in order for the examiner to understand it.

The child was first given a training story consisting of a single episode story in five pictures with a main character (a boy) and a minor character (a man). The purpose of the training story was to familiarize the child with the procedure and to allow the examiner to give more explicit prompts if the child was having difficulty with the task, such as providing the story beginning (e.g., "Once upon a time … there was a …") or encouraging the child to go beyond labelling ("You've told me who is in the pictures – now can you tell me a STORY about the pictures?"). After the training story was administered, the two story sets were given. Administration of the story sets was counterbalanced, with half of the children telling stories from Set A first and the other half telling stories from Set B first. For the sets A and B stories, the examiner was restricted to less explicit assistance than in the training story such as general encouragement, repetition of the child's previous utterance, or if the child did not say anything, a request to tell what was happening in the story. Stories were audiorecorded using JVC minidisk recorders.

In the second session, children participated in a comprehension task involving the pictures in the set A stories (this task will not be discussed in this paper). After that, the CELF-P or CELF-3 test or subtests were administered.

Transcription

Children's story retellings were transcribed in full using the CHAT transcription system from the CHILDES database (MacWhinney, 2000; MacWhinney & Snow, 1990). The CHILDES database is a collection of transcripts from many researchers of primarily children's language samples in a number of languages. CHILDES also provides a system for analysing transcripts using the CLAN program, called CHAT, which was used for the analyses of storytelling described below. The transcripts were divided into communication units (C-units), each of which consisted either of one independent clause plus any dependent clauses associated with it or of a sentence fragment. Initial transcriptions were made by the research assistants of tapes they had collected themselves. Then these transcripts were checked and put into CHAT format by students in the speech-language pathology master's program at the University of Alberta. The students were trained by the first author by practicing with training tapes. Transcripts were checked against the recordings by the primary investigator before being analysed. Another student in the speech-language pathology master's program transcribed 12.6% of the checked stories for reliability purposes; word-by-word reliability for the stories analysed in this paper was calculated to be 96.4%. Except for the research assistants who collected the data, all transcribers and checkers were blind to the status of the children (typically development or language impairment).

Table 3

Test scores by Test, Age Group and Language Group

Age Group	Language Group	CELF Subtest 1 M(SD)*	CELF Subtest 2 M(SD)**	No. receiving full CELF***	Receptive Language	Expressive Language	Total Language
4	TD	10.82 (3.32)	9.96 (2.38)	15	108.07 (14.00)	106.00 (9.86)	107.67 (12.13)
	SLI	4.33 (2.64)	5.42 (1.17)	12	78.33 (15.87)	77.83 (5.94)	76.83 (8.62)
5	TD	10.74 (2.63)	9.96 (2.79)	19	103.37 (9.51)	104.32 (11.78)	103.79 (8.51)
	SLI	5.00 (2.88)	4.43 (1.28)	14	79.86 (15.45)	74.00 (12.20)	76.21 (11.35)
6	TD	11.58 (3.03)	11.76 (3.32)	14	111.57 (12.40)	110.86 (11.05)	107.36 (19.93)
	SLI	5.72 (1.79)	5.27 (2.20)	11	80.79 (11.53)	79.01 (11.57)	78.70 (8.34)
7	TD	12.24 (3.26)	11.66 (2.79)	15	108.53 (24.99)	112.13 (14.91)	110.33 (19.43)
	SLI	6.38 (2.36)	4.31 (1.50)	13	81.62 (12.86)	69.69 (11.87)	74.00 (11.05)
8	TD	12.16 (2.92)	10.84 (2.74)	10	109.70 (10.48)	105.70 (16.44)	107.50 (12.94)
	SLI	7.47 (2.38)	5.00 (1.80)	17	83.24 (16.55)	70.18 (9.42)	76.29 (11.94)
9	TD	11.84 (2.80)	11.14 (2.60)	15	107.87 (14.26)	97.73 (11.56)	102.80 (12.22)
	SLI	8.10 (2.56)	5.40 (1.96)	10	80.00 (13.16)	70.70 (11.98)	73.50 (11.27)
Total	**TD**	**10.78 (2.98)**	**9.96 (2.58)**	**34**	**105.44 (11.75)**	**105.06 (10.84)**	**105.5 (10.28)**
CELF-P	**SLI**	**4.69 (2.74)**	**4.88 (1.31)**	**26**	**79.15 (15.35)**	**75.77 (9.84)**	**76.50 (9.99)**
Total	**TD**	**11.96 (3.00)**	**11.35 (2.88)**	**54**	**109.35 (16.65)**	**106.61 (14.34)**	**106.94 (16.53)**
CELF-III	**SLI**	**6.94 (2.40)**	**4.96 (1.84)**	**51**	**81.66 (13.66)**	**72.06 (11.34)**	**75.68 (10.75)**

* Subtest 1: Ages 4-5, CELF-P Linguistic Concepts; Ages 6-9, CELF-III Concepts and Directions

**Subtest 2: Ages 4-5, CELF-P Recalling Sentences in Context; Ages 6-9, CELF-III Recalling Sentences

*** A subsample of children in the TD group and all children in the SLI group were given the full CELF appropriate to their age groups

Coding for Story Grammar Units

The Story Grammar (SG) model was used to analyze the stories as well as to design the materials in order to capture the elements that need to be included in the story for it to be considered an adequate story. Because the stories were presented in pictorial form only, the child did not hear an oral version of the story to which the child's version could be compared. The SG model provides a principled way to determine whether the child included important story information. The main concern was whether or not a child was telling a story that would be understandable to the listener. Some children may tell stories that include much more detail; while these might be preferred by some listeners on esthetic grounds over stories with less detail, scoring for assessment purposes

Table 4

Characteristics of the story sets

Story	Number of Episodes	Setting	Number of Characters	Character Description	No. of Pages
A1	1	Swimming pool	2	Young female elephant young male giraffe	5
A2	2	same	3	same as A1 plus adult male elephant lifeguard	8
A3	3	same	4	same as A2 plus adult female elephant	13
B1	1	Park	2	young male rabbit young female dog	5
B2	2	same	3	same as B1 plus adult female rabbit doctor	8
B3	3	same	4	same as B2 plus adult male rabbit balloon-seller	13

A1 was the simple story used in this study; A3 was the complex story.

should focus on basic story information that all stories should contain.

Two of the six ENNI stories from Set A were used in the Story Grammar analysis: a simple, one-episode story with two characters, and a complex, 3-episode story with four characters. We chose these stories for initial analysis to explore the usefulness of Story Grammar for obtaining developmental trends and differentiation between groups.[2] Stories were coded for the information that corresponds to a story grammar (SG) unit. Scoring protocols listing what would count as each SG unit were created and used to score the stories. The scoring protocol for the simple story is included in the Appendix. Both protocols are available from the website (Schneider, Dubé, & Hayward, 2004b).

As can be seen in the protocol, three SG units that are considered core units – Initiating Event, Attempt, and Outcome – were given two points rather than one. These units were awarded more points to reflect their more central nature. We felt that this would be potentially useful when using the protocol for clinical purposes with an individual child, as the scoring would give more weight to the core information expected to be included in each story, and would differentiate scores that consisted of core information from those that included only characters and setting information, for example. To check the effect that this scoring had on the results, all statistics were performed on both the adjusted and unadjusted data and the same pattern of results was obtained. We report here only the results using the adjusted data. Maximum possible scores (adjusted data) were 12 for the simple story and 37 for the complex story.

The first two authors scored the stories using the protocols. The first author was completely blind to the group membership of the participants. While the second author had seen the participants for a question task and standardized testing, she scored the stories without identifying information on the transcripts. Thus both scorers were blind to group membership when scoring. To determine interscorer agreement, each scorer scored the same 20% of the stories. Cohen's kappa was computed for agreement on each story using the procedure described by Bakeman & Gottman (1986). This measure takes into account differences between scorers on individual scoring categories; it adjusts for frequencies of different categories and thus it corrects for agree-ments expected by chance (Bakeman & Gottman, 1986). The kappa for A1 was .92, and for A3 kappa was also .92. These kappas are significant at $p < .001$ and indicate excellent interscorer agreement (Landis & Koch, 1977).

The Story Grammar measure was the total points awarded for SG units included in the story. There were two SG scores, one for the simple story and one for the complex story.

All statistics were computed using the SPSS program (SPSS Inc., 2004).

Results

Developmental trends

The Story Grammar model that was used to develop the picture story sets and the scoring protocol predicts that story grammar knowledge should increase with age. When a theory or model predicts increases in measures of a construct due to age, empirical evidence of such increases can be taken as evidence of construct validity (Ventry & Schiavetti, 1986). Thus an increase in SG scores across our age range would provide evidence of construct validity.

Trend analysis was used to investigate the first research question (Are developmental trends evident in the data for amount of Story Grammar information?). This technique can be used to test for developmental trends in scores (e.g., increase or decrease) across a number of age groups, as well as periods of stabilization of scores. Trend analysis was selected as the statistical technique rather than a simple ANOVA model because trend analysis can

identify gradual increase over an ordered set of groups (in this case, age groups) rather than simply testing for main effects and differences among groups. Trends can be in different patterns such as linear or quadratic. In other words, it is possible to examine to what extent scores increase in regular increments across the age groups, or show a different pattern across the age groups.

The data for the children with specific language impairments were first weighted within each age group so that they represented 7.4% of the age group, using the formula (N_{Total} * .074) / N_{SLI}. The weighted data for the simple story are displayed in Figure 1. Trend analysis for the simple story revealed significant linear and quadratic trends: Linear, $F(326) = 161.52$, $p < .0001$; Quadratic, $F(325) = 96.15$, $p < .0001$. The linear trend suggests an increase in scores with increasing age. As is apparent by looking at Figure 1, the quadratic trend is significant because of the similarity of scores in the three higher age groups; although not quite reaching a ceiling, the scores appear to have levelled off. It appears that growth in scores for the simple story occurs mainly in the younger children.

Weighted data for the complex story are displayed in Figure 2. Trend analysis again revealed significant linear and quadratic trends: Linear, $F(326) = 196.47$, $p < .0001$; Quadratic, $F(325) = 114.41$, $p < .0001$. Once again the linear trend indicates an increase in scores with increasing age. In this case, the quadratic trend appears to be related to a levelling off of scores in the last two age groups (8- and 9-year-olds).

Thus, for both the simple and complex stories, there is a developmental trend in the amount of story information children include in their stories. The trend appears to level off earlier for the simple story than for the complex one.

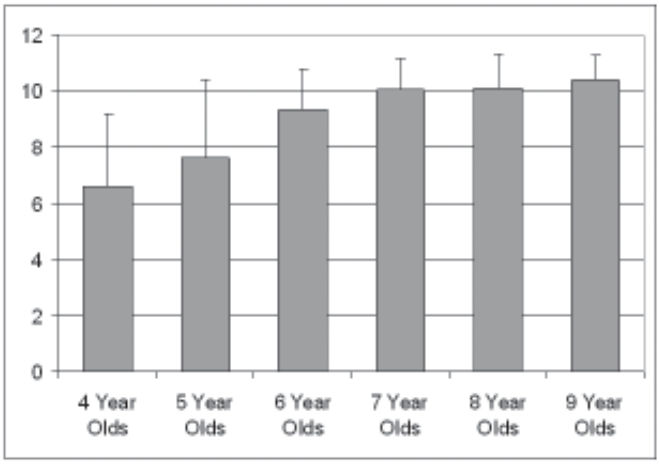

Figure 1. Means for the simple story. Scores for the children with language impairments are weighted so that they represent 7.4% of each age group. The lines on the top of each bar indicate the standard deviation for that bar.

Group differences

Construct validity can also be evidenced by demonstrating that differences between distinct groups that are predicted by a model can be demonstrated empirically (Ventry & Schiavetti, 1986). Previous research using the Story Grammar model has indicated that storytelling abilities differ between children who are typically developing and children with language impairments, presumably related to Story Grammar knowledge (e.g., Liles, 1987; Merritt & Liles, 1987). Thus we posed our second research question: Are there differences between the groups in the amount of story grammar information included in children's stories within each age group? An affirmative answer would provide further evidence of construct validity.

Table 5 displays the results by both age group and language status. Because the main focus of interest was differences between children with typical development and children with language impairments, a priori planned comparisons were made of the two language status groups' scores within each age group using Student's *t* tests with correction for unequal variances and for multiple comparisons.[3] These tests revealed that for both the simple and complex stories, children with typical development (TD) had significantly higher scores than children with specific language impairment (SLI) except for both stories at age 9, for which the two groups were not different on either story, and for the simple story at age 7, for which the difference between the two groups failed to meet the adjusted alpha of .017. Effect sizes for each comparison are also provided in Table 5. Effect sizes were above .80 and thus may be considered large (Murphy & Myors, 1998), with the exception of the simple story for 5-year-olds (which showed a moderate effect size) and both comparisons for 9-year-olds.

Discrimination between groups

Group comparisons can tell us about differences between groups, but not how representative the differences are of individuals. Extreme scores from a small subset of individuals may cause a statistic to be significant, even when differences in variance have been accounted for in the test. We were interested in examining to what extent narrative scores would characterize the group with language impairments and whether the groups could be distinguished by their narrative scores. To examine whether SG scores discriminated between the TD and SLI groups, a discriminant analysis was performed using scores for the two stories. Discriminant analysis can be used to predict to which group particular cases belong (Klecka, 1980). Thus it can be used to investigate whether and to what extent a measure classifies participants into the correct group (in this case, TD or SLI). The 9-year-old group was omitted from this analysis due to the lack of statistically significant difference between the two language groups at

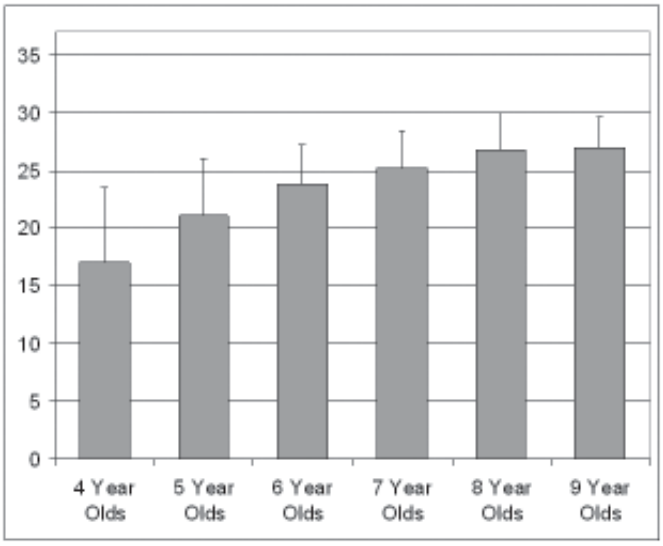

Figure 2. Means for the complex story. Scores for the children with language impairments are weighted so that they represent 7.4% of each age group. The lines on the top of each bar indicate the standard deviation for that bar.

this age reported above. The lack of a difference in this age group indicates that the ENNI is unlikely to discriminate between groups at this age. We did include the data for the 7-year-olds despite the nonsignificant result for the simple story because scores for the complex story yielded a significant difference between the groups.

To permit calculation of discriminant analysis on the entire data set, *z*-scores were first computed for simple and complex SG scores within each age group. Then the *z*-scores for children aged 4-8 were entered into the discriminant analysis. The use of *z*-scores (a type of standardized score) essentially controls for the effect of age because each child's score represents the child's score relative to the distribution of scores in his or her own age group.

As part of the analysis, a discriminant function is calculated, which is a mathematical formula that combines the predictor variables to discriminate between the groups (Brace, Kemp, & Snelgar, 2003). If the discriminant function is statistically significant, then the predictor variables are successfully discriminating between groups. The value of the discriminant function was significantly different for the TD and SLI groups, Wilks $\lambda = .73$, $\chi^2 = 99.96$, $df = 2$, $p < .0001$. Correlations between predictor variables and the discriminant function indicate that both SG scores contribute to the prediction of group membership, although the correlation for the complex story is higher than that for the simple story (.96 versus .69).

Discriminant analysis yields measures of specificity (in this case, the percentage of children with typical development who are identified as such by the measures), sensitivity (in this case, the percentage of children with specific language impairment who are identified as such), and overall accuracy (the percentage of all children correctly identified). Specificity was 83.6% and sensitivity was 70.1%. Overall, 80.8% of children were correctly classified.

By computing an odds ratio, the difference in the two groups can be highlighted. Using the scores from the simple and complex story, children in the SLI group are almost 12 times more likely to be categorized as being in the SLI group than are children in the TD group (odds ratio [OR] = 11.98, confidence interval [CI] = 6.44-22.30). Odds ratios can also highlight the effects of the story scores separately and together in classifying children. If we consider each story separately, the simple story scores yield an OR of 4.49 (CI – 2.55-7.91), meaning that

Table 5

Test scores by Test, Age Group and Language Group

Age Group	Simple Story						Complex Story					
	TD		SLI		p-value*	Effect size *d*	TD		SLI		p-value*	Effect size *d*
4	6.78	(2.53)	4.58	(2.58)	.002	.86	17.68	(6.11)	10.33	(6.47)	< .001	1.17
5	7.82	(2.62)	5.79	(3.56)	.011	.65	21.74	(4.21)	14.71	(7.04)	.002	1.21
6	9.5	(1.18)	7.36	(2.58)	.011	1.07	24.1	(3.16)	20.18	(4.38)	< .001	1.03
7	10.2	(.76)	8.62	(2.63)	.026	.92	25.52	(2.76)	21.69	(4.99)	.01	.94
8	10.24	(1.06)	8.94	(1.92)	.008	.84	27.4	(2.01)	20.76	(5.72)	< .001	1.55
9	10.44	(.68)	9.8	(2.15)	.188	.40	27.10	(2.50)	25.90	(4.07)	.19	.35

* Student's t tests, one-tailed, corrected for unequal variances where appropriate, with adjusted alpha of .017

children in the SLI group are more than 4 times as likely to be classified as such by their simple story scores. The complex story scores result in an OR of 7.85 (CI-4.34-14.20), indicating that children in the SLI group are nearly 8 times as likely to be classified as being in the SLI group than are the children in the TD group. Thus the two scores together produce a greater likelihood that children will be categorized into the correct group more than either individual story score. However, all odds ratios are significant at $p < .05$.

Correlations with CELF scores

Concurrent validity is often estimated by correlating scores with those of a standardized test that purports to test similar skills. Since there was no other test of narrative skills at the time the ENNI was normed, scores from the Clinical Evaluation of Language Fundamentals were used to estimate concurrent validity. Both the CELF-P and the CELF-3 provide Receptive Language, Expressive Language and Total Language composite scores. As described earlier, 4- and 5-year-olds in the TD group were given either the full CELF-P or two of its subtests; children aged 6 to 9 in the TD group received either the full CELF-3 or two subtests. All of the children in the SLI group received the full CELF appropriate to their age group. Table 6 reports the correlations for the simple and complex story standardized scores with the CELF standard scores. Hammill, Brown, and Bryant (1992) considered a test to have evidence of validity if at least half of the correlation coefficients reported for validity are significant at .05 and reach or exceed .35 in magnitude. Of the correlations in Table 6, 24 of the 30 correlations were at or above .35, and all were significant at .01. These results suggest that a degree of concurrent validity exists for the ENNI.

Discussion

This study investigated whether story grammar units would prove to be a useful measure of the development of story-telling from ages 4 to 9 in terms of amount of information included in stories. The trend analysis indicated that there is a develop-mental trend for number of story grammar units, to age 7 for the simple story and to age 8 for the complex story. Thus it appears that story grammar units can provide useful information about the development of story-telling in younger children. For older children, story grammar unit scores do not increase. However, the results cannot be interpreted to mean that storytelling skills are completely acquired by age 8. Other aspects of storytelling that are not captured by the story grammar model are likely to continue to develop after this time. Some story features that are not captured by story grammar are evaluation (e.g., comments about the story or the characters), perspective-taking, and amount and type of description. The story grammar units score does not capture a feature of the structural pattern aspect of the model, namely, the elaboration of simple episodes into interactive ones in which more than one character's motivations are highlighted. In addition, stories that are more complex than the stories used in the current study could cause older children to have difficulty providing all story units. Other aspects of storytelling such as cohesion are likely to continue to develop through these ages.

We were also interested in whether and to what degree story grammar units would discriminate between groups of children with and without previously identified language impairments. Within each age group, children

Table 6

Correlations between ENNI scores and CELF scores.
All correlations (Pearson's r) are significant at < .01 (two-tailed).

CELF-P	Receptive Language*	Expressive Language*	Total Score*	Linguistic Concepts**	Recalling Sentences in Context**
Simple story	.48	.46	.49	.38	.26
Complex story	.63	.71	.70	.54	.49
CELF-3					
	Receptive Language*	Expressive Language*	Total Score*	Concepts and Directions**	Recalling Sentences**
Simple story	.34	.43	.38	.31	.33
Complex story	.46	.45	.45	.39	.37
All children					
	Receptive Language*	Expressive Language*	Total Score*	Subtest 1***	Subtest 2***
Simple story	.37	.44	.41	.32	.30
Complex story	.51	.53	.52	.43	.40

*Composite scores - available for a portion of the Typical Development group and the entire Specific Language Impairment group.
**These subtests were given to all children.
***Subtest 1 was Linguistic Concepts (4 and 5 year olds) or Concepts and Directions; Subtest 2 was Recalling Sentences in Context (4 and 5 year olds) or Recalling Sentences

in the Typical Development (TD) group attained higher scores than the children in the Specific Language Impairment (SLI) group, with the exception of age 9. Discriminant analysis indicated that z-scores for story grammar units correctly classified children into TD or SLI 80.8% of the time for children aged 4 to 8 (83.6% for TD and 70.1% for SLI).

We did not include the SG scores from the 9-year-olds in our sample because of the lack of a significant difference in the group comparisons. Because of this lack of difference, there was no reason to expect the scores to discriminate between groups at this age group. We would recommend use of the SG measure with caution at this age. While it may reveal storytelling problems for a particular 9-year-old child who does score lower than most in the distribution, the results of this study suggest that many children at this age who have language impairments will still score in the normal range on this measure.

The results of the discriminant analysis indicate that the story grammar measure would not be sufficient in itself to identify a language impairment at any age. According to Plante and Vance (1994), test accuracy should be at least 80% for use in diagnosis of language impairments. This suggestion implies that individual tests should be able to categorize the majority of children accurately, which in turn implies that most children with language impairments share impairments in the skills measured by the tests. However, children with language impairments are a heterogeneous group whose impairments can be in a number of different areas (Paul, 2001). Thus, unless a test purports to be representative of a wide range of skills, it is unlikely to capture all impairments of any sample of children with "language impairments". In practice, clinicians do not rely on a single test to categorize a child as having a language impairment, using instead some combination of tests, subtests, and informal measures.

In our sample, it is important to remember that the two pre-existing groups were identified not on the basis of storytelling ability, but rather on the presence or absence of language impairment. In this heterogeneous group, not all of these children would necessarily have difficulties with storytelling skills. Although the presence of storytelling problems in groups of children with language impairments has been well documented, the prevalence of such problems within the population of children with language impairments has not yet been described. Insofar as the children in the current study are representative of children in the population, the results of the discriminant analysis suggest that prevalence is around 70%, at least when narrative tasks are administered with materials and procedures similar to those used in the current study and a story content measure is used. Not all children identified as having a language impairment scored low on this narrative task, suggesting that some children with language impairments may have deficits that do not impair their ability to include story content when formulating stories from pictures. Given that the ENNI is a new instrument requiring additional evidence of validity and reliability, the suggestion of a prevalence rate of around 70% can only be considered an initial estimate that will require additional data. It should be noted as well that 16% of the children in the TD group were misclassified using SG scores, indicating that some children with no known language difficulties provided relatively low amounts of story information.

Evidence for validity and reliability of the ENNI

In order to be a useful assessment instrument, the ENNI must show evidence that it is valid, that is, that it measures what it purports to measure, and that it is reliable, that is, that it measures accurately and consistently. Some of the findings from the current study provide evidence of the ENNI's validity and reliability.

Construct validity. The stories represented in the pictures drawn for this study were constructed according to the Story Grammar model, as discussed earlier. This model predicts that story knowledge increases with age. Previous research has found differences between children with and without language impairments in storytelling, with Story Grammar knowledge proposed as the underlying skill difference (e.g., Liles, 1985, 1987; Merritt & Liles, 1987, 1989). Thus we expected to find an increase in our Story Grammar measure with age, as well as differences between the children with typical development and those with language impairments. The trend analyses showed growth to age 7 or 8 in Story Grammar scores, and the group comparisons showed differences between groups except for age 9. Across the age groups for which there were group differences, ENNI scores discriminated children in the TD and SLI groups in 80.8% of the sample. Thus we believe that there is evidence of construct validity of the ENNI.

Concurrent validity. Correlations with composite and subtest scores of two versions of the Clinical Evaluation of Language Fundamentals indicate that the ENNI correlates at a statistically significant level to scores from the CELF-P and CELF-3.

Reliability. At present, the only evidence of reliability is the interscorer reliabilities of .92 for each story measure that was reported earlier. A limitation of this finding is that the scorers had developed the scoring protocol and were able to discuss the scoring system at length before scoring transcripts for reliability, which maximized the likelihood of agreement between them. It is important to establish whether scorers can use the ENNI reliably when simply following written scoring instructions. Future studies will focus on investigating reliability amongst potential users of the instrument, such as speech-language pathologists. It will also be important to investigate test-retest reliability, in order to establish whether performance is stable at a given age. Findings of adequate

test-retest reliability would provide greater confidence in the age trends reported in this paper.

The complex story from the Set B stories has not yet been analysed. If scoring of the Set B complex story correlates well with the Set A complex story, it will provide evidence of alternate-form reliability.

The ENNI and the Test of Narrative Language

Since the process of developing the ENNI began, another instrument for narrative assessment has been developed: The Test of Narrative Language (TNL; Gillam & Pearson, 2004). The TNL is normed on children aged 5-12 from a number of states in the United States. It consists of four storytelling tasks, one involving oral retell of a simple script-based story and the others involving pictures. One of the picture sets depicts a story with one main protagonist and repeated attempts at a goal; it appears to be intermediate in complexity between the ENNI simple story with two main characters but a simple episode structure and the ENNI complex story with four main characters and repeated attempts at a goal. The TNL's other two picture storytelling stimuli consist of a single picture each. One picture is used for a comprehension task, while the other is used to elicit the production of a story. Storytelling from a single picture requires children to formulate more of the story than is depicted; it requires formulation with less structuring than picture sequence tasks and thus should prove quite useful for older children. It appears that together the ENNI and TNL offer a wide range of story formulation tasks with picture support. Unfortunately, at the current time these two instruments are normed on children from different geographic locations. It would be very interesting to see correlations among storytelling scores when the instruments are used with the same children.

Clinical implications

Results of this study indicate that the ENNI could be useful for determining whether a child aged 4-8 includes story content in ways similar to peers when formulating stories from picture stimuli. Over 70% of children previously assessed as having a language impairment were discriminated by their Story Grammar unit scores. The ENNI would be useful in helping to determine which aspects of language use are impaired after determination of impairment has been made. Specifically, the SG unit score would help to determine whether a child can provide basic story information from pictures without an oral story model. For initial assessment of a language impairment, it would be necessary to use other measures of language as well, given that almost 30% of children with a diagnosis of language impairment in our study did not appear to have difficulty with our measure of story content.

Future analyses

In the current set of analyses, the simple story appeared to develop at an earlier age and contributed somewhat less to the discriminant analysis than did the complex story. It may be the case that complex stories are more useful for assessment purposes in the entire age range. Development of the Set B complex story analyses will help to determine whether using two complex stories discriminates better between children with and without language impairments than using a simple and complex story.

More information is needed on the reliability and validity of the ENNI as an assessment tool. Future studies should focus on establishing its reliability and validity in a variety of ways.

Content analyses such as the Story Grammar unit analysis used in the current study capture one aspect of storytelling skills. Other skills also show promise for discriminating between children with and without language impairments. For example, the way that children link parts of their stories together for the listener (cohesion) is an important skill that has been shown to vary by age (Schneider & Dubé, 1997) and language status (Liles, 1987; Schneider, 2001). Cohesion may thus prove useful in discriminating among groups in a normative sample. Stories can also be analyzed as language samples, with measures such as number of different words and mean length of communication unit. More information on these measures in a standardized context such as the ENNI would be very helpful in assessment if the measures discriminate.

References

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis.* Cambridge, UK: Cambridge University Press.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1927-1050.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research, 35,* 119-129.

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

Blishen, B. R., Carroll, W. K., & Moore, C. (1987). The 1981 socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology, 24,* 465-488.

Brace, N., Kemp, R., & Snelgar, R. (2003). *SPSS for psychologists: A guide to data analysis using SPSS for Windows* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cowley, J., & Glasgow, C. (1994). *The Renfrew Bus Story* (American Ed.). Centreville, DE: The Centreville School.

Dubé, R. V. (2000). An instrument for language assessment of Deaf children: American Sign Language and English. Unpublished dissertation, Faculty of Rehabilitation Medicine, University of Alberta.

Fazio, B. B., Naremore, R. C., & Connell, P. J. (1996). Tracking children from poverty at risk for specific language impairments: A 3-year longitudinal study. *Journal of Speech and Hearing Research, 39,* 611-624.

Feagans, L., & Appelbaum, M. I. (1986). Validation of language subtypes in learning disabled children. *Journal of Educational Psychology, 78,* 358-364.

Gillam, R. B., & Pearson, N. A. (2004). *The Test of Narrative Language.* Austin, TX: Pro-Ed.

Goldman, S. R., & Varnhagen, C. K. (1986). Memory for embedded and sequential story structures. *Journal of Memory & Language, 25,* 401-418.

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2005). Oral discourse in the preschool years and later literacy skills. *First Language, 24,* 123-147.

Hammill, D. D., Brown, L., & Bryant, B. R. (1992). *A consumer's guide to tests in print* (2nd Ed.). Austin, TX: Pro-Ed.

Hedberg, N. L., & Stoel-Gammon, C. (1986). Narrative analysis: Clinical procedures. *Topics in Language Disorders, 7,* 58-69.

Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment.* Eau Claire, WI: Thinking Publications.

Klecka, W. R. (1980). *Discriminant analysis.* London: Sage Publications.

Lahey, M. (1988). *Language disorders and language development.* New York: Macmillan.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Liles, B. Z. (1985). Production and comprehension of narrative discourse in normal and language disordered children. *Journal of Communication Disorders, 18,* 409-427.

Liles, B. Z. (1987). Episode organization and cohesive conjunctives in narratives of children with and without language disorder. *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.

Lund, N. J., & Duchan, J. F. (1993). *Assessing children's language in naturalistic contexts* (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall.

MacLachlan, B. G., & Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders, 53,* 2-7.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (*Third Ed.) Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language, 17,* 457-472.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111-151.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research, 30,* 539-552.

Merritt, D. D., & Liles, B. Z. (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders, 54,* 438-447.

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests.* Mahwah, NJ: L. Erlbaum Associates.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test - Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language, 24,* 149-184.

Paul, R. (2001). *Language disorders: From infancy through adolescence* (2nd Ed.). St. Louis: 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.

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

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.

Roth, F.P., & Spekman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally-achieving students. *Journal of Speech and Hearing Disorders, 51,* 8-23.

Rumelhart, D.E. (1975). Notes on a schema for stories. In D.G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 185-210). New York: Academic Press.

Schneider, P. (1996). Effects of pictures versus orally presented stories on story retellings by children with language impairment. *American Journal of Speech-Language Pathology, 5,* 86-96.

Schneider, P. (2001). Microstructure analyses: Referential cohesion. Presented as part of seminar: "Development and implementation of a narrative norms project," American Speech-Language-Hearing Association Convention, New Orleans, LA, November.

Schneider, P., & Dubé, R. V. (1997). Effect of pictorial versus oral story presentation on children's use of referring expressions in retell. *First Language, 5(3),* 283-302.

Schneider, P., & Dubé, R.V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology, 14,* 52-60.

Schneider, P., Dubé, R. V., & Hayward, D. (2004a). Introducing the Edmonton Narrative Norms Instrument. Canadian Association of Speech-Language Pathologists and Audiologists Conference, Ottawa, ON, May 2004.

Schneider, P., Dubé, R. V., & Hayward, D. (2004b). The Edmonton Narrative Norms Instrument. Available from http://www.rehabmed.ualberta.ca/spa/enni/.

Schneider, P., Williams, B., & Hickmann, M. (1997). The awareness of referential cohesion in children with and without language/learning disabilities. *Journal of Speech-Language Pathology and Audiology, 21* (1), 8-16.

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals III.* San Antonio, TX: The Psychological Corporation.

SPSS Inc. (2004). *SPSS for WIndows 13.0.* Chicago: SPSS Inc.

Statistics Canada (no date). *Canada dimensions: The people* [on-line]. Retrieved September, 2003 from http://www.statcan.ca.

Stein, N. L., & Albro, E. R. (1997). Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In M. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 5-44). Mahwah, NJ: L. Erlbaum Associates.

Stein, N.L., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing, Vol. 2: Advances in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.

Stein, N. L., & Policastro, M. (1984). The concept of a story: A comparison between children's and teachers' viewpoints. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and the comprehension of text* (pp. 113-155). Hillsdale, NJ: Lawrence Erlbaum Associates.

Strong, C. J. (1998). *The Strong Narrative Assessment Procedure.* Eau Claire, WI: Thinking Publications.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology, 9,* 77-110.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Ukrainetz McFadden, T. (1996). Creating language impairments in typically achieving children: The pitfalls of "normal" normative sampling. *Language, Speech, and Hearing Services in Schools, 27,* 3-9.

Ventry, I. M., & Schiavetti, N. (1986). *Evaluating research in speech pathology and audiology* (2nd Ed.). New York: Macmillan.

Wagner, C. R., Nettelbladt, U., B. Sahlén, B., & Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders, 35,* 83-93.

Westby, C.E. (1999). Assessing and facilitating text comprehension problems. In Catts, H.W., & Kamhi, A.G. (Eds.), *Reading disabilities: A developmental language perspective* (pp. 154-223). Boston: Allyn and Bacon.

Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech Language Pathology, 6,* 195-208.

Wiig, E.H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool.* San Antonio, TX: The Psychological Corporation.

Author Note

Funding for this study was provided by the Children's Health Foundation of Northern Alberta. The authors would like to thank Marilynn McAra, Livia Tamblin, and Linda Kaert for their assistance in data collection, and Jess Folk-Farber, Rhonda Kajner, Roxanne Lemire, Marlene May, Michelle Millson, Ignatius Nip, Michelle Trapp, and Kathy Wagner for their assistance with other aspects of the study.

A version of portions of this paper describing the rationale for and design of the ENNI appears on the ENNI website (Schneider, Dubé, & Hayward, 2004b). The Story Grammar analysis described in this paper was present at the CASLPA conference in 2004 (Schneider, Dubé, & Hayward, 2004a).

Correspondence concerning this article should be addressed to: Phyllis Schneider, Department of Speech Pathology and Audiology, University of Alberta, 2-70 Corbett Hall, Edmonton, AB T6G 2G4.
E-mail: phyllis.schneider@ualberta.ca.
Web page: www.rehabmed.ualberta.ca/spa/enni

Footnotes

[1] Since the ENNI was developed, another instrument, the Test of Narrative Language, has been published. This test will be discussed in relation to the ENNI in the Discussion section.

[2] The full data set is being used in other analyses (including cohesion, syntax, and word use; results will be reported in subsequent papers.

[3] To balance the risk of Type 1 and 2 errors, an adjusted alpha was calculated for each story as follows: Overall alpha was set at .10, which was then divided by the number of comparisons (6), for an adjusted alpha per comparison of .017.

Received: November 11, 2004
Accepted: June 20, 2006

Appendix

Scoring Protocol for the Simple Story

Edmonton Narrative Norms Instrument
Story Grammar Scoring Sheet for Story A1

Child's Name: Age: Date:

Please read the section of the Manual on scoring SG units before using this sheet.

SG Unit	***Acceptable [child need only have one alternative per unit to get credit for that unit]***	**Score**
Character 1	giraffe / male/ boy (or any type of animal such as horse) [not acceptable: pronoun]	0 1
Character 2	elephant / female / girl (or any type of animal such as cow) [not pronoun]	0 1
Setting	swimming pool had a ball / playing with ball / want to play ball	0 1
Initiating Event	ball goes in water/pool/sand/mud ball is in water they see a ball	0 2
Internal Response	one / both want to get ball elephant says, e.g., "look what happened," "what am I going to do?" Elephant upset / sad [not: he/she/they want to go swimming]	0 1
Internal Plan	giraffe decides to / thinks he will get the ball	0 1
Attempt	giraffe jumps in pool / swims toward ball / tries to get ball [not: giraffe swimming (without goal); giraffe falls in water]	0 2
Outcome	giraffe gets ball / gives ball to elephant [not: elephant gives ball to giraffe, unless it is noted as unexpected, e.g.; 'but instead, Elephant gets it and gives it to him']	0 2
Reaction of Giraffe	giraffe is happy / proud / smiles giraffe says "You're welcome" giraffe's teeth are chattering / giraffe is cold/wet	0 1
Reaction of Elephant	elephant is happy / is grateful / says thank you elephant hugs the ball [not: holds/has the ball]	0 1
Reaction both or unknown	'they' are happy/in love [code only as replacement for Reaction of Character 1 or 2; there should not be more than 2 reactions total]	0 1
	Total raw score:	
	Standard score:	