LSHSS

Who Does What to Whom: Introduction of Referents in Children's Storytelling From Pictures

Phyllis Schneider
Denyse Hayward
University of Alberta, Canada

ABSTRACT: **Purpose:** This article describes the development of a measure, called First Mentions (FM), that can be used to evaluate the referring expressions that children use to introduce characters and objects when telling a story.
Method: Participants were 377 children ages 4 to 9 years (300 with typical development, 77 with language impairment) who told stories while viewing 6 picture sets. Their first mentions of 8 characters and 6 objects were scored as fully adequate, partially adequate, inadequate, or not mentioned. Total FM scores were compared across age and language groups.
Results: There were significant differences for age and language status, as well as a significant Age × Language interaction. Within each age group except age 9, children in the typical development group attained higher scores than children in the group with language impairment.
Conclusion: These results suggest that the FM measure is a useful tool for identifying whether a child has a problem with introducing referents in stories.

KEY WORDS: narratives, storytelling, child language, language impairment, language assessment

Story tasks have become a common feature of clinical assessment and intervention in the field of speech-language pathology. Stories are a part of everyday communication, both at home and in school. They can provide a more holistic language context than most tests of language, which assess the use of words and sentences in isolation, because stories require children to combine words and sentences for a particular purpose (Schneider, Hayward, & Dubé, 2006). Thus, they provide information about how well children can use their discrete language skills to communicate. 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). Support for this claim comes from several studies finding that, unlike conversation, children's stories have linguistic complexity that is characteristic of written language (MacLachlan & Chapman, 1988; Wagner, Nettelbladt, Sahlén, & Nilholm, 2000; Westerveld, Gillon, & Miller, 2004).

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 that are representative of his or her age. Children's ability to provide story content and structure has been well documented in previous research (e.g., Liles, 1985b; Shapiro & Hudson, 1991; Stein, 1988), and normative information about story content and structure knowledge is becoming available (e.g., Schneider et al., 2006). Less well understood are age expectations for other important aspects of storytelling, such as referential cohesion, or the use of referring expressions to talk about characters, objects, and other entities in discourse.

Our purpose in this article is to present a measure that can be used to evaluate how children introduce referents in stories. The stories were elicited from children using the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2009). First, we review the notion of referential cohesion and how it has been measured in previous studies. We then describe our measure for evaluating referent introduction, First Mentions (FM), and present results using the measure on data from the ENNI normative sample.

Cohesion

The term *cohesion* refers to the use of various linguistic means to link utterances together into a unitary text (Halliday & Hasan,

1976; Hickmann & Schneider, 2000). Two aspects of cohesion that have been studied extensively in recent years are *referential cohesion,* whereby text is connected by linked references within the text (e.g., *the girl…she*), and *conjunctive* or *connective cohesion,* whereby connectives are used to tie clauses and sentences together within a text (e.g., The girl went into the cave. *But then* she ran out.). Cohesion is pertinent to any type of discourse or text, including conversation, stories, and expository text. This article focuses on referential cohesion in the context of stories.

Referential cohesion. *Referring expressions* are linguistic forms that are used to refer to *referents* such as animate beings (*an elephant, Ella, she*), objects (*a toy, it*), and other entities such as places (*the park, there*) and concepts (*an idea*). Appropriate use of referring expressions in discourse contributes to referential cohesion. When a story is told, referring expressions must be used to mention referents for the first time in such a way that the listener understands that they are new to the story. Thereafter, referring expressions are used to continue to refer to the referents in a way that allows the listener to recognize them as the same referents that were introduced earlier. Referring expressions can be considered adequate if they are appropriate for the listener's knowledge, shared physical context, and preceding linguistic context. For example, an indefinite noun phrase such as *an elephant* is adequate for introducing a new character in a story in the absence of a shared physical context because indefinite articles signal that the referent is not known to the listener (Givón, 1992). In contrast, *the elephant* or *she* would only be adequate for mentioning the character later on in the story, or if the referent (or a representation of it) were in the context shared by the speaker and listener, because definite forms signal that referents are expected to be known to the listener. The following is an example of adequate first and second mention of a character and an object in the absence of a shared context:

> *An elephant* was bouncing *a ball. She* accidentally dropped *it.*

The establishment of referents in discourse when there is no shared context is termed *endophora,* which is reference that is accomplished through language, as shown in the above example. The alternative way to refer to a character or item is by using expressions that "point" directly to referents in the extralinguistic context, which is termed *exophora*. The following example illustrates the difference in adequate reference when the extralinguistic context is shared:

> (Context: adult says to child as they look at a picture in a book of an elephant bouncing a ball) Look, *she*'s bouncing *it*!

In this example of exophoric reference, the adult is able to point directly to the shared extralinguistic context with the referring expressions and thus can use the pronouns *she* and *it*. The references are understood because the child can refer to the context and relate the referring expressions to the pronouns. In contrast, endophoric reference essentially creates the referents linguistically for the listener (e.g., *There is an elephant bouncing a ball*).

Referent introduction (i.e., first mentions of referents) is an important aspect of all types of connected discourse, from casual conversation to more planned and tightly structured forms such as narratives and expository text. Regardless of the discourse type, it is necessary to introduce referents in a way that is comprehensible to the audience. The current article focuses on referent introduction in stories told by children.

Development of Referent Introduction

Children's ability to introduce and maintain referents in narratives develops gradually throughout the early school years (e.g., Hickmann, 1991, 1997, 2003; Kail & Hickmann, 1992; Karmiloff-Smith, 1987; Peterson, 1993; Schneider & Dubé, 1997; Villaume, 1988; Warden, 1981; Wigglesworth, 1990). Young children frequently introduce referents in a confusing way, tending to use referring expressions that are exophoric, even when the listener does not have access to the extralinguistic context and thus cannot understand exophoric referents (Kail & Hickmann, 1992). In referent introductions, young children often use pronouns and definite noun phrases—forms that are generally inadequate for first mentions of referents—which suggests that they are using them to refer directly to the extralinguistic context (i.e., exophorically) rather than endophorically (Hickmann, 1995; Karmiloff-Smith, 1987). Thus, a young child might introduce referents in an utterance such as "She's bouncing it" in the absence of a shared context. Note that the issue here is not mastery of particular forms such as pronouns and noun phrases; young children can use these forms adequately in less complex contexts (Hickmann, 1995; Wittek & Tomasello, 2005). Rather, the young child struggles with the appropriate choice of these forms in the context of extended discourse, in which the current knowledge state of the listener must be constantly monitored (Kail & Hickmann, 1992).

Kindergartners correctly introduce referents more often when they are retelling a fictional story they just heard than when they are formulating the story themselves from pictures for a listener who cannot see the pictures (Schneider & Dubé, 1997), suggesting that when kindergartners must choose referents based on listener knowledge, they tend to use less adequate forms. Pratt and MacKenzie-Keating (1985) obtained similar results with videotaped versus orally presented fictional stories. In their study, children in first and third grade made higher proportions of referential errors in their retellings after viewing videotapes of puppets presenting the story through dialogue than they did after listening to a story. Even when telling stories about their personal experiences, which is believed to be an easier task for children than telling fictional stories (McCabe, 1996), young preschool-age children did not adequately introduce one in five referents (Peterson & Dodsworth, 1991). Younger school-age children (age 7) continue to introduce referents inadequately, using definite forms to a greater extent than older children. After age 9 or so, children introduce characters in simple stories in ways similar to adults (Vion & Colas, 1998). Older children (age 11 and up) and adults tend to use endophoric reference even when the context is shared, for example, when pictures are visible to the listener (Kail & Hickmann, 1992). Studies have found similar results in a number of languages (e.g., English, French, German, Mandarin), with minor variations in development due to language-specific means of referring (Hickmann & Hendriks, 1999; Hickmann, Hendriks, Roland, & Liang, 1996; Wong & Johnston, 2002).

Schneider (2008) compared the referent introductions of 9-year-old typically developing (TD) children to those of a group of 10 adults ages 25–33. A significant difference between the 9-year-olds and adults was found, with a large effect size ($\eta_p^2 = .85$).

These results suggest that the ability to introduce referents in stories continues to develop for some time after age 9.

First Versus Subsequent Mentions

Young children may also use inadequate referring expressions when referring to characters after the first mention. However, preschool and young school-age children have more difficulty with first mentions of referents, or referent introductions, than with subsequent mentions (Pratt & MacKenzie-Keating, 1985; Schneider & Dubé, 1997), at least with relatively short stories with few characters. Difficulty in subsequent mentions is likely to vary with story complexity. When telling short, simple stories, for example, it may not be necessary to refer to characters more than once or twice after the referent introduction. The complexity of the character set can also impact the child's ability to refer to characters correctly later in the story. For example, multiple male characters can be introduced as "another boy" but then may need to be differentiated as "the boy who came second" or "the guy with the hat" in subsequent mentions, whereas in a simple story with one character of each gender, characters could be referred to subsequently as *he* and *she*. An additional factor that can impact a child's ability to use referring expressions correctly is the way the speaker chooses to tell the story; one speaker may limit subsequent references to characters and objects and make them adequate; another speaker may try to tell a more complete story and thus have more difficulty keeping track of referents when referring to them. With first mentions, we can expect speakers to refer to each character and object once. With subsequent mentions, speakers will vary in whether and how often they refer to each established referent.

Studies of Reference and Cohesion by Children With Language Impairment

Studies of the use of cohesion ties have used a number of approaches to quantify differences between children with and without language impairment (LI). Some studies have looked at frequencies and distributions of cohesive ties used by children in their stories (Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001; Liles, 1985a, 1985b; Strong & Shaver, 1991; Vallance, Im, & Cohen, 1999). However, as pointed out by Scott (1988), frequencies and relative distributions of cohesive ties have sometimes failed to differentiate between children with and without LI. Additionally, given that there are many ways to use cohesive ties in a story, it is not clear whether or how frequencies of cohesive ties are related to the quality of the story.

Other studies have considered the adequacy of referential and connective cohesive ties, that is, whether or not the ties contributed to cohesion (Girolametto et al., 2001; Liles, 1985a, 1985b; Liles, Duffy, Merritt, & Purcell, 1995; Paul, Hernandez, Taylor, & Johnson, 1996; Paul & Smith, 1993; Schneider, 1996). Adequacy is determined by identifying "incomplete/erroneous" ties versus "complete" ones. An incomplete reference tie is one that refers back to a referent that was never introduced; an erroneous reference tie is one that points to the wrong referent (e.g., *he* instead of *she*) or that has several possible antecedents (e.g., *he* when there were several males mentioned previously). An erroneous conjunctive tie is one that expresses a relation between sentences that is not appropriate for the context (e.g., "He liked it. *But* he was happy"). By focusing on adequacy of ties for cohesion, these studies contribute to our understanding of story quality and provide potentially useful information for planning intervention. However, combining different types of cohesive ties into a single count obscures important differences among them. Use of inadequate (i.e., incomplete or erroneous) reference ties can make a story very confusing and is typically a major reason for difficulty in understanding a young child's story unless the listener already knows the story. Children do not appear to use misleading conjunctive ties as frequently as misleading reference ties; in fact, it is common for young children to simply leave out conjunctive ties or to use simple additive ones (e.g., *and, then*), which continue to be used in similar frequencies at least through age 9 (Peterson & McCabe, 1987). Speakers have a range of choices regarding when and how to connect clauses with conjunctions. It is impossible to tell a comprehensible story without using referential ties to refer to people and/or objects, but it is perfectly possible to tell an acceptable and understandable story with few or no conjunctive ties. Thus, the use of referential and conjunctive ties will have different impacts on the story's perceived quality. In addition, for purposes of treatment planning, because intervention for referential cohesion would be conducted in a different manner than intervention for conjunctive cohesion, it would be helpful to know more precisely in which area a child was having difficulty. Thus, a measure that focuses solely on referential ties rather than on both referential and conjunctive ties would be potentially useful for clinical purposes.

On the other hand, it is possible to restrict the range of cohesive devices too narrowly by focusing on particular linguistic forms, thereby missing the overall picture of referential cohesion. Finestack, Fey, and Catts (2006) analyzed pronoun use, including the percentage of complete (i.e., adequate) pronoun reference, by children in Grades 2 and 4 and found no differences between children with and without LI. In a study of children ages 8–11 with fetal alcohol spectrum disorder (FASD), Thorne, Coggins, Olson, and Astley (2007) analyzed stories for pronominal and nominal reference separately and found that the rate of ambiguous nominal reference discriminated between children with FASD and age-matched children without FASD. These studies focused on the use of one or two particular linguistic forms that can be used to refer in discourse. However, referential cohesion is best considered in terms of function rather than particular linguistic forms. For example, a speaker might choose to refer to a previously introduced character with a pronoun, a definite noun phrase, or a proper name; all three expressions could be equally adequate in some contexts, whereas not all would be adequate in others. To capture the referential ability of an individual, the focus needs to be on whether a form that is adequate for a particular point in a narrative is selected from among the forms that are available to the individual. Mastery of referential cohesion is the ability to introduce and maintain referents in a comprehensible way in a discourse context, rather than mastery of individual linguistic forms.

Other research studies have looked at referential cohesion separately but comprehensively, focusing on the percentage of referential cohesive ties that are adequate (i.e., not incomplete or erroneous; Boudreau & Chapman, 2000; Klecan-Aker, 1985; Norbury & Bishop, 2003; Strong, 1998). These studies found differences on referential adequacy (RA) scores between children with and without LI (and Down syndrome in the case of Boudreau & Chapman, 2000). This method of analyzing referential cohesion is commonly used in research on TD children (Hickmann, 1991;

Hickmann, Hendriks, Roland, & Jiang, 1996; Schneider & Dubé, 1997; Tsai & Chang, 2008). Using this method, any referring expression that is not fully adequate for its occurrence in a story is considered inadequate, and the percentage of RA is calculated as the number of adequate referring expressions divided by the total number of referring expressions. Frequently, adequacy of referring expressions used to introduce characters is assessed separately from adequacy of those used for subsequent mentions (as in Schneider & Dubé, 1997). Use of the percentage of adequate references has been recommended as a part of narrative assessment (Hughes, McGillivray, & Schmidek, 1997; Strong, 1998).

RA measures focus on the function of referring expressions in context and therefore have advantages over other measures that either combine referential and other cohesion types or focus on a single type. However, in grouping all expressions that are not fully adequate as inadequate, RA measures fail to make some important distinctions in degree of less-than-adequate references. In measures of adequacy that involve percentages, introductions with definite noun phrases (e.g., *the elephant*) are typically included with inadequate expressions such as pronouns because both wrongly signal to the listener that the referent should already be known. However, the nouns in definite noun phrases at least allow the listener to understand what is being referred to, even if the determiner may confuse the listener as to whether the referent is new or not. Therefore, to investigate children's growing competence with referent introduction, it would be useful to distinguish between partially adequate and fully inadequate expressions. In the current study, we developed a measure that would incorporate such distinctions.

Another limitation of the RA approach is that it is typically calculated based on the number of referents a child chooses to mention. The result is that some children will receive an RA score based on a smaller number of referents than other children. The child with a larger number of attempted referents could have more problems with adequacy simply because the child is attempting to tell a more complete story, whereas the child who attempts only a small number of referents may achieve a high RA score and will thus appear more competent than the first child. A measure that controls for the number of referents would better distinguish between these children.

To date, there is no normed narrative instrument that includes a measure of referential cohesion. The likely reason is that it is difficult to specify the rules for determining adequacy of subsequent mentions. Adequacy of subsequent mentions depends on the length of a story and the number and order of referents mentioned. However, the rules for adequate first mentions are more straightforward than those for subsequent mentions. In addition, if analysis is restricted to first mentions, the analysis can include the same set of referents for all children. For these reasons, and because first mentions appeared to discriminate well among age and language groups in earlier studies (Schneider & Dubé, 1997; Schneider & Hayward, 2004), we decided to develop a scoring system for first mentions as our measure of referential cohesion.

RA can vary greatly depending on the complexity of the narrative stimuli used to elicit stories. For example, using three of Mercer Mayer's frog stories as stimuli with TD 8- to 10-year-olds, Strong (1998) obtained mean percentages of referential cohesion errors varying from 6% to 27%. The variation appears to be due to story differences, with the story having the most characters (*Frog Goes to Dinner;* Mayer, 1974) having the highest mean percentage of problem references. Given this variation, we felt that it was very important to use stories that controlled for the number and type of referents.

In the current study, we investigated whether an FM score would be a useful measure of the development of cohesion in storytelling from ages 4 to 9 using stories that controlled for the number and type of referents. We wanted to determine whether and to what degree FM scores would reveal differences between groups of children with and without previously identified LI. Because the goal of this article was to investigate the usefulness of the FM measure for assessment purposes, the research questions focused on examining developmental trends and group differences, as well as the measure's validity. The research questions were:

- Are there significant effects for age and language status (TD versus LI) on FM scores?
- Are there significant effects for age and language status on RA scores?
- Do FM scores correlate significantly with a standardized test of language at a level sufficient to indicate concurrent validity?

METHOD

Participants

Participants consisted of 377 children ages 4 through 9 whose stories form the ENNI database. Within each 1-year interval, there were 50 TD children (25 boys and 25 girls) and a smaller sample of children with LI. The target sample for children with LI was 15 per age group; the obtained sample varied 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, or 62%) in the LI group. Sample information is summarized in Table 1.

The younger TD study participants were chosen from children attending 13 preschools, day care centers, and kindergarten programs in Edmonton. The older TD participants were chosen from children attending kindergarten through Grade 4 in 34 Edmonton public and separate schools. Schools were randomly selected from areas across Edmonton to ensure a cross-section of socioeconomic groups. All participants spoke English as a first language at home; information about other languages spoken in the home was not collected.

To identify potential TD children for the study, we asked teachers to refer two children in the upper level of achievement, two children in the middle level, and two children in the lower level (one boy and one girl at each level). We asked teachers to refer children who did not have suspected or identified speech or language difficulties or any other diagnostic label such as attention deficit disorder with or without hyperactivity (ADD/ADHD), learning disability, or autism. Information and parental consent forms were sent to the homes of children referred by the teachers. Children whose parents returned the forms were included in the study.

We obtained the sample of children with LI 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 LI; and Capital Health Authority, which serves preschool- and school-age children throughout the city. We

Table 1. Number, age, and socioeconomic status (SES) information for the study participants in the typically developing (TD) group and the group with language impairment (LI).

		N		*Age*			*SES*		
Age	*Language status*	*Total*	*Boys*	M	SD	*Range*	M	SD	*Range*
4	TD	50	25	4.60	.24	4.04–4.97	47.38	13.58	23.70–82.91
	LI	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
	LI	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
	LI	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
	LI	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
	LI	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
	LI	10	5	9.50	.21	9.10–9.82	48.71	9.66	27.60–60.73

requested the speech-language pathologists (SLPs) at these sites to refer children who had a diagnosis of LI. Children could be referred if they had concomitant fine or gross motor delays, ADD/ADHD with medication, a diagnosed learning disability, or mild or moderate speech disorder, in addition to LI. (Information regarding concomitant conditions of children referred to the study was not collected.) 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 an inability to see pictures even with correction, or severe speech impairment that would preclude accurate orthographic transcription of their stories. IQ test information was not collected; the SLPs referring children for the study were asked to refer children for whom they had no concerns regarding cognitive abilities. Because we do not have IQ scores to confirm that our participants were in the normal range of cognitive ability, we will refer to our participants in this group as having LI rather than specific language impairment.

To ensure that our sample was representative of the Edmonton population, we collected demographic information on the families of participating children. Socioeconomic status (SES) was estimated from parents' occupations using the *Socioeconomic Index for Occupations in Canada* (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 index range from 17.81 (newspaper carriers and vendors) to 101.74 (dentists), with a mean for Canada of 42.74 (SD = 13.28). SES information is reported for each age and language group in Table 1. Blishen SES scores were compared across age and language groups using an analysis of variance (ANOVA). Results revealed no significant differences in SES for age, $F(5, 358) = .99$, $p = .43$, or language, $F(1, 358) = 1.84$, $p = .18$.

Ethnic information was collected on a parent checklist based on Statistics Canada categories of visible minorities. According to Statistics Canada data (Statistics Canada, n.d.), the ethnic composition of the sample corresponded closely to the range of ethnic diversity in the city of Edmonton.

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 were tested on two subtests of the Clinical Test of Language Fundamentals (CELF), using either the CELF–Preschool (CELF–P; Wiig, Secord, & Semel, 1992) for children < 6 years of age or the CELF—Third Edition (CELF–3; Semel, Wiig, & Secord, 1995) for children ≥ 6. We chose to use the Linguistic Concepts and Recalling Sentences in Context subtests from the CELF–P because the test authors recommend them for use in screening (Wiig et al., 1992). We used the Concepts and Directions and Recalling Sentences subtests from the CELF–3 because they are analogous to the CELF–P subtests used. Subtest means for the TD and LI groups are reported in Table 2. The purpose of administering the CELF subtests was to obtain language information for the TD children. In addition, 88 participants in the TD group (29%) and all of the participants in the LI group were given the core subtests of the CELF–P or CELF–3 that make up the composite scores for receptive language, expressive language, and total language. Collection of core CELF data permitted calculation of correlations as a measure of concurrent validity using composite scores and gave us comparable information across the group of children with LI, who presumably would have been tested on a variety of language measures for their initial diagnosis. CELF composite score data are reported in Table 3.

In addition to the 77 children in the LI group, there were 19 children who were referred to the study as having LI who attained a score > 85 on both the receptive and expressive language total score of the CELF–P or CELF–3; these children were excluded from the sample. The decision was made to exclude these children because we did not have other information to back up the referral information indicating LI. The LI sample thus included only children who scored ≤ 85 on the receptive language score, the expressive language score, or both. Children in the TD group were not excluded on the basis of their CELF score. First, although these children's scores may reflect an unidentified LI, it is possible for a child with functional language skills to score below the normal range on a standardized test. Additionally, only the two subtest scores were available for most of the children in the TD group, which would not be adequate for identifying LI without supporting information. Finally, eliminating the children from the TD group who had the lowest CELF scores would potentially bias the sample in the direction of greater differences between the groups on the ENNI. Consequently, some of the children in the TD group had

Table 2. Clinical Test of Language Fundamentals (CELF) subtest scores by age, language, and subtest.

Age	Language status	*CELF Subtest 1*[a]			*CELF Subtest 2*[b]		
		M	SD	*Range*	M	SD	*Range*
4	TD	10.82	3.32	3–16	9.96	2.38	5–18
	LI	4.33	2.64	3–11	5.42	1.17	4–7
5	TD	10.74	2.63	3–15	9.96	2.79	3–16
	LI	5.00	2.88	3–11	4.43	1.28	3–7
6	TD	11.58	3.03	6–17	11.76	3.32	5–17
	LI	5.72	1.79	4–9	5.27	2.20	3–10
7	TD	12.24	3.27	4–17	11.66	2.79	5–17
	LI	6.38	2.36	3–11	4.31	1.50	3–7
8	TD	12.16	2.92	4–17	10.84	2.74	4–16
	LI	7.47	2.38	4–13	5.00	1.80	3–9
9	TD	11.84	2.80	6–17	11.14	2.60	5–16
	LI	8.10	2.56	4–13	5.40	1.96	3–8
Total CELF–P	TD	10.78	2.98	3–16	9.96	2.58	3–18
	LI	4.69	2.74	3–11	4.88	1.31	3–7
Total CELF–3	TD	11.96	3.00	4–17	11.35	2.88	4–17
	LI	6.94	2.40	3–13	4.96	1.84	3–10

Note. The CELF–Preschool (CELF–P; Wiig, Secord, & Semel, 1992) was administered to the children who were < 6 years of age. The CELF–Third Edition (CELF–3; Semel, Wiig, & Secord, 1995) was administered to the children who were ≥ 6 years of age.

[a]Children < 6 years of age were given the Linguistic Concepts subtest of the CELF–P. Children ≥ 6 years were given the Concepts and Directions subtest of the CELF–3. [b]Children < 6 years of age were given the Recalling Sentences in Context subtest of the CELF–P. Children ≥ 6 years were given the Recalling Sentences subtest of the CELF–3.

Table 3. Composite CELF–P and CELF–3 scores by group.

Age	Language status	N *receiving full CELF*[a]	*Receptive language*[b]			*Expressive language*[c]			*Total language*		
			M	SD	*Range*[d]	M	SD	*Range*[d]	M	SD	*Range*[d]
4	TD	15	108.07	14.00	75–131	106.00	9.86	94–130	107.67	12.13	90–137
	LI	12	78.33	15.87	50–114	77.83	5.94	65–85	76.83	8.62	68–99
5	TD	19	103.37	9.51	81–116	104.32	11.78	85–133	103.79	8.51	88–118
	LI	14	79.86	15.45	61–108	74.00	12.20	50–92	76.21	11.35	58–96
6	TD	14	111.57	12.40	88–128	110.86	11.05	94–128	111.29	12.03	91–129
	LI	11	80.79	11.53	50–98	79.01	11.57	61–100	78.70	8.34	63–92
7	TD	15	108.53	24.99	65–143	112.13	14.91	78–139	110.33	19.43	70–138
	LI	13	81.62	12.86	50–96	69.69	11.87	50–86	74.00	11.05	51–90
8	TD	10	109.70	10.48	94–125	105.70	16.44	86–131	107.50	12.94	90–129
	LI	17	83.24	16.55	54–106	70.18	9.42	50–82	76.29	11.94	55–95
9	TD	15	107.87	14.26	88–139	97.73	11.56	80–118	102.80	12.22	86–122
	LI	10	80.00	13.16	53–100	70.70	11.98	50–90	73.50	11.27	55–85
Total CELF–P	TD	34	105.44	11.75	75–131	105.06	10.84	85–133	105.50	10.28	88–137
	LI	26	79.15	15.35	50–114	75.77	9.84	50–92	76.50	9.99	58–99
Total CELF–3	TD	54	109.35	16.65	65–143	106.61	14.34	78–139	107.96	14.67	70–138
	LI	51	81.66	13.66	50–106	72.06	11.34	50–100	75.68	10.75	51–95

[a]29% of the children in the TD group and all of the children in the LI group were given the full CELF appropriate to their age group (<6 years of age, CELF–P; ≥6 years, CELF–3). [b]Receptive language for the CELF–P is a composite of the Linguistic Concepts, Basic Concepts, and Sentence Structure subtests. Receptive language for the CELF–3 is a composite of the Sentence Structure, Concepts and Directions, and Word Classes subtests. [c]Expressive language for the CELF–P is a composite of the Recalling Sentences in Context, Formulating Labels, and Word Structure subtests. Expressive language for the CELF–3 is a composite of the Word Structure, Formulated Sentences, and Recalling Sentences subtests. [d]Ranges for the LI group may end above 85, the cutoff for this group, because individuals only had to score ≤85 on *either* the expressive language or receptive language score to be included. Thus, a child's score could have been >85 on one of the two tests, and this will be reflected in the range.

subtest scores < 1 *SD* on the CELF–P or CELF–3. Specifically, 15 children (from 1 to 6 per age group) had standard scores < 7 on the Linguistic Concepts (CELF–P) or Concepts and Directions (CELF–3) subtest, 11 children (1–3 per age group) had scores < 7 on the Recalling Sentences in Context (CELF–P) or Recalling Sentences (CELF–3) subtest, and six children (0–2 per age group) had scores < 7 on both subtests (either Linguistic Concepts and Recalling Sentences in Context or Concepts and Directions and Recalling Sentences). Of the 160 children in the TD group who were given the full CELF, five (from 0–3 per age group) had a receptive language composite score < 85, three (0–2 per group) had an expressive language composite score < 85, and two (both 7-year-olds) had a total language score < 85.

Materials

Stimuli for the current study were the two story picture sets of the ENNI, which was originally developed by Dubé (2000). The stories were created according to story grammar principles (Stein & Glenn, 1979) and depict information that is considered to be essential to good stories. Each story set contains three stories in pictures (no text), with two main animal characters of different species, a young male and a young female, introduced in the first story in the set (five pages long, single basic episode). These characters appear in all of the stories in their set. The second story (eight pages, two episodes) introduces a third character who is an adult animal (the same type of animal as one of the main characters), and the third story (13 pages, three episodes) introduces a fourth character in addition to the previous three (another adult of the same type of animal as the third character, opposite gender). Thus, the stories increase in referential difficulty; the first two animals can be distinguished in a number of simple ways (e.g., gender, type of animal), whereas the later characters are more difficult to differentiate when referring to them. If all characters were introduced in a single story, that story would be very long and complex. Having three stories in a set with increasing complexity permitted us to administer each story individually and to analyze them individually and as a set. For the current analysis, we analyzed referent introductions across the entire story set. For other analyses, we analyzed individual stories (Schneider et al., 2006).

Each story page was put into its own plastic page protector, and each story was put into its own binder, permitting each page to be presented separately. The pictures in which the targeted characters and objects first appeared are provided in Appendix A. The full picture sets may be viewed and downloaded from the ENNI Web site (http://www.rehabmed/ualberta.ca/spa/enni).

As the ENNI was designed as a storytelling task rather than a retell task, no verbal story is presented to children; rather, children are asked to tell the story to the examiner from the pictures.

Procedure

Data were collected by three female research assistants. Children were seen in their school or preschool settings. The child was first shown 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 provide 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 labeling ("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 A and B story sets, 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.

Each child was presented with each story, one page at a time, before being asked to tell the story. Then the child was again presented with the story page by page and was asked to tell the story to the examiner, who held the story binder in such a way that she could not see the pictures. The child was reminded before each story that the examiner would not be able to see the pictures. Inability of the examiner to see the pictures was established so that pointing or other exophoric reference would not be possible for indicating referents, and children would need to refer endophorically to introduce referents adequately. The procedure was repeated for each of the six stories, with a brief break between the two story sets. Stories were audio-recorded using JVC digital minidisk recorders.

Children's story retellings were transcribed in full using the Codes for the Human Analysis of Transcripts transcription system from the Child Language Date Exchange System (CHILDES) database (MacWhinney, 2000). 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 analyzing transcripts using the Computerized Language Analysis program, which was used for the analyses of storytelling described below. Before being analyzed, all transcripts were checked against the recordings by the first author. A research assistant transcribed 5% of the checked stories for transcription reliability purposes. Word-by-word reliability was calculated to be 96.5%.

Development of the FM Scoring

To develop the FM scoring, we examined the introduction to all eight characters and six of the objects depicted in the stories. We chose the objects for scoring on the basis of the likelihood that they would be mentioned by speakers. Preliminary analyses of stories indicated that some objects such as *picnic basket* (Set B, story B2) were mentioned less consistently than other objects, and in fact could be left out without making the story deficient. For example, a child could mention that the characters in the second story of set B were "having a picnic," without mentioning a basket. In contrast, in the complex story of set B (B3), it would not be possible to tell the story as depicted without mentioning the balloon that one of the characters lets loose. To ensure that the target referents were those that were likely to be mentioned, we chose referents that were mentioned by the majority of the oldest participant groups. The six objects selected for the analysis (three from each story set) were each mentioned by 98%–100% of the 8- and 9-year-olds in the TD group, as were all eight story characters. Only the first mentions of target referents were scored. Re-introductions of characters were not scored (e.g., the elephant and giraffe were scored only in the first story of Set A, even though they might be introduced by children as though they were new characters in the second and third stories.)

We developed a scoring system for the measure in which a score of 3 indicated a fully adequate referring expression for its context

(e.g., indefinite determiner plus noun, as in "There was *an elephant* bouncing *a ball*"; name, as in "*Ella the elephant* was bouncing a ball"; possessive pronoun plus noun when the possessor has been introduced, as in "She bounced *her ball*"), a score of 2 indicated a less than adequate expression that was still partially informative (e.g., definite determiner plus noun, as in "*The elephant* was bouncing *the ball*"), and a score of 1 indicated an inadequate referring expression (e.g., personal or demonstrative pronoun, as in "*She* was bouncing *that*"; use of definite determiner with a noun that had been used for a previously introduced character, as in "*the elephant*" for the third character). Referents that were omitted altogether received a score of 0. We developed the point system to provide a more graduated scoring than systems using an adequate/inadequate dichotomy, such as the RA measure described earlier. Scoring was not dependent on the use of a particular term; for example, the giraffe character could be referred to as a *horse, zebra,* or *cow,* and the ball could be referred to as a *balloon* or *egg*. Scoring was dependent on the appropriateness of the linguistic form for first mention (indefinite or definite determiner, pronoun, etc.). The complete FM measure is available on the ENNI Web site (http://www.rehabmed/ualberta.ca/spa/enni), including examples of scoring for 1, 2, and 3 points for each of the 14 target referents. Scoring criteria for the first three referents are provided in Appendix B.

Each child's FM score was calculated by counting the total number of points awarded for the 14 targeted referents. For example, if a child received scores of 3 for six referents, scores of 2 for three referents, and scores of 1 for three referents, omitting the last two referents, the child would receive a total FM score of 27. The maximum possible score was 42.

For purposes of comparison with scoring used in previous studies, we also calculated an RA score for the 14 referents used in the FM scoring. To do this, we divided the number of adequate referring expressions (i.e., those scored as 3 in the FM scoring) by the total number of the 14 referents mentioned by each child. We used the number of referents mentioned by each child rather than the total number used for the FM score (14) because RA scores are typically calculated on referents that are actually mentioned by each individual rather than on a predetermined number of potential referents. Note that, to keep the FM and RA scoring comparable, we included only the 14 referents targeted by the FM system in the RA scoring rather than scoring all referents mentioned by the child, as is common in previous research. For example, the child in the previous example referred to 12 of the 14 referents and used an adequate referring expression for six of them, which would be an RA score of 6/12 or 50%. Note that this would result in the same RA score for children who mentioned 7 of 14 referents adequately as those who mentioned 2 of 4 adequately. This is the procedure that is typically used in studies of RA (and in fact is one of the limitations of the measure, as discussed earlier).

Data Scoring and Reliability

The first author scored the transcripts using the FM measure. To check scoring reliability, the second author scored 20% of the transcripts (entire story sets from 20% of the children, randomly chosen). Cohen's kappa was computed; this statistic takes into account differences between scorers on each item as well as the probability of agreement by chance on individual items, and thus is considered a more rigorous way to calculate reliability than point-to-point reliability for multi-item scoring systems (Bakeman & Gottmann, 1986). A Cohen's kappa of .85 was obtained, indicating excellent reliability (Landis & Koch, 1977).

Table 4. Means, standard deviations, and effect sizes for First Mention scores by age and language.

	TD group		*LI group*			
Age	M	SD	M	SD	*Effect size* (η_p^2)	*Effect size (Cohen's* d*)*
4	30.04	6.56	20.92	7.66	.23	1.28
5	34.92	3.97	26.43	5.57	.40	1.76
6	37.02	4.43	31.00	5.69	.20	1.18
7	39.58	2.86	34.92	3.64	.29	1.42
8	39.96	2.42	34.47	3.78	.43	1.73
9	39.94	2.05	37.40	4.43	.12	.74

Note. The effect size describes the difference between the diagnostic groups.

RESULTS

First Mentions

An ANOVA was used to investigate the first research question: Are there significant effects for age and language status on FM scores? Means and standard deviations for the FM data are displayed in Table 4 by age and language. There was a significant main effect for age, $F(5, 376) = 54.48$, $p < .001$, $\eta_p^2 = .43$. Because Levene's test of equality of variances yielded a significant difference, Games-Howell post hoc tests were used to look at differences between age groups. The post hoc tests revealed that the 4-year-old age group differed from all the other age groups, as did the 5- and 6-year-old age groups; 7-, 8-. and 9-year-olds differed from the younger age groups but did not differ from one another. It appears that the FM measure used with the ENNI stories reveals development in adequacy of referent introduction between the ages of 4 and 7. The main effect for language (TD vs. LI) was also significant, $F(1, 376) = 110.91$, $p < .001$, $\eta_p^2 = .25$. In addition, there was a significant Age × Language interaction, $F(5, 376) = 3.09$, $p = .01$, $\eta_p^2 = .04$.[1]

To investigate differences between TD children and LI children within each age group, post hoc comparisons were made of the two groups' FM scores within each age group using least squares difference (LSD) tests with Bonferroni correction for multiple comparisons and corrections for unequal variances where appropriate. The two groups were significantly different within each age group at $p < .0001$, with effect sizes (η_p^2) ranging from .20 to .43, with the exception of the 9-year-old group, in which the two

[1]Inspection of the data distribution, confirmed with a single-sample Kolmogorov-Smirnov test, revealed negative skewness in the three oldest age groups, indicating that more children had scored in the higher ranges than in the lower ranges in these age groups. To investigate the effect of deviation from normality on the results, nonparametric tests were conducted and yielded the same pattern of results: Kruskal-Wallis revealed a significant difference for age, $\chi^2 = 123.5$, $df = 375$, $p < .00001$; 2-sample Kolmogorov-Smirnov test yielded a significant difference between the TD and LI groups within each age group (p values ranging from .00001 to .024) except for the 9-year-olds. Because the overall pattern was essentially the same, we report the parametric test results in this article.

language groups did not differ, $p = .11$, $\eta_p^2 = .12$. These results indicate that the FM measure yields significant differences between children with and without LI in the age range of 4 to 8 years.[2]

We examined the CELF–3 results for the 9-year-old children in the TD group who had obtained scores > 1 *SD* below the mean to see whether these scores had affected the result for that age group. Two children who had obtained low expressive language scores (80 and 82) scored 39 and 40 on the FM measure, which was very close to the group mean. A third child who had been tested on two subtests of the CELF–3 obtained a score of 5 on Recalling Sentences and an FM score of 34. A comparison of the groups with these children's FM scores removed yielded nearly identical results as the original test ($p = .09$).

RA

To answer the second question, we conducted the same analyses using the RA measure (i.e., percentage of referent introductions that were fully adequate). Means and standard deviations for the RA data are provided in Table 5. Results for the ANOVA were similar to those for FM in that there was a main effect for age, $F(5, 376) = 28.34$, $p < .001$, $\eta_p^2 = .28$, and for language, $F(1, 376) = 71.13$, $p < .001$, $\eta_p^2 = .16$. However, there was no significant Age × Language interaction, $F(5, 376) = .96$, $p = .44$, $\eta_p^2 = .01$. Post hoc tests with Games-Howell were similar to those for FM except that there was no difference between the 5- and 6-year-old age groups. As with the FM measure, children in the TD group had higher scores within each age group than children with LI, except in the 9-year-old group. Effect sizes were smaller in these analyses than in the FM analyses, indicating that less variance was accounted for in the RA analyses. Thus, although the overall pattern of results was similar for the two measures, the FM measure provided a bit more information than the RA measure, namely, higher effect sizes, an interaction between age and language, and differences among the younger age groups.[3]

Correlations Between FM Scores and CELF Scores

The final research question concerned correlation of the FM scores with scores from the two CELF tests, the CELF–P and CELF–3. Correlations of FM scores with these standardized tests could be considered evidence of concurrent validity. We adopted the criteria of Hammill, Brown, and Bryant (1992), which stated that correlations between tests can be considered to show concurrent validity if at least half of the correlations are > .35 in magnitude and significant at .05 or above. To test this, we calculated correlations between FM scores and the CELF subtest scores that had been collected from all of the children. We also calculated correlations with composite scores for the children who had received the entire CELF test (165 children, 88 in the TD group and all 77 children in the LI group). As can be seen in Table 6, all of the correlations between the FM scores and the CELF scores exceeded .35 and were statistically significant above .05. These results indicate that the FM measure shows concurrent validity with the CELF tests.

Table 5. Means, standard deviations, and effect sizes for referential adequacy scores (percentage of referent introductions that were fully adequate) by age and language.

	TD group		*LI group*			
Age	M	SD	M	SD	*Effect size* (η_p^2)	*Effect size (Cohen's* d*)*
4	.50	.21	.30	.20	.13	.98
5	.64	.19	.43	.18	.17	1.13
6	.72	.22	.48	.25	.15	1.02
7	.85	.15	.69	.20	.14	.91
8	.87	.14	.59	.15	.40	1.93
9	.86	.12	.75	.23	.08	.60

Note. The effect size describes the difference between the diagnostic groups.

DISCUSSION

The purpose of this study was to develop a measure of referential cohesion using first mentions of characters and objects in stories told from the ENNI. The measure was applied to data collected for the local normative sample of the ENNI. Results of the scoring were examined to see whether there were differences among age groups. ANOVAs revealed that FM scores did increase between the ages of 4 and 7 but did not appear to change from ages 7 to 9. Thus, it appears that in these fairly simple stories, children's ability to use adequate referring expressions to introduce characters and objects gradually improves until age 7, when it appears to be mastered by the majority of children. However, the FM measure captures differences in children of different language abilities beyond age 7. Within-age comparisons of scores from children with and without LI revealed that the two groups' FM scores were significantly different in each age group except age 9. In our data, the means for ages 7–9 were very close, with the standard deviation decreasing over this range. This suggests that the children with TD had reached a plateau on this measure and in these simple stories,

Table 6. Correlations of First Mentions scores to CELF scores.

Test	N	*Receptive language*	*Expressive language*	*Total language*
CELF–P composite score	60	.71	.66	.74
CELF–3 composite score	105	.51	.50	.53
All CELF scores (Preschool + 3)	165	.58	.55	.58
Subtests	N	*Subtest 1*	*Subtest 2*	
CELF–P subtests	126	.59	.47	
CELF–3 subtests	251	.48	.52	
All CELF subtests (Preschool + 3)	377	.51	.50	

Note. All correlations were significant at $p < .000001$.

[2]The ENNI Web site provides a table that can be used to obtain standard scores for each age group based on the raw total FM score. This table is available at http://www.rehabmed.ualberta.ca/spa/enni/pdf/FM%20norms.pdf.

[3]To check for potential effects of the distribution of the proportion data, we conducted further analyses using transformed RA scores (arcsine transformation). ANOVA and post hoc tests showed the same pattern of results using the transformed data.

allowing the children with LI to catch up by age 9. It is possible that children with LI catch up in their referring abilities by age 9, at least when telling stories that are comparable in complexity to the ENNI stories. However, it must be noted that the 9-year-old sample contained only 10 children with LI, and therefore had limited power to show a difference in this age group. Further studies would be needed to address the question of whether and when differences are no longer found using the FM measure with children with LI.

A related study (Schneider, 2008) compared the FM scores of the ENNI sample of 9-year-old TD children to a group of 10 adults ages 25–33. The goal was to explore whether the FM measure would show differences between adults and children in referent introduction. Despite the lack of change between ages 7 and 9 found in the current study, the Schneider (2008) study found a significant difference between the FM scores of the 9-year-olds and those of the adults, with a large effect size. These results suggest that development of referent introduction is not completed by age 7 but continues to develop into adulthood.

Correlations of FM scores with scores from the CELF–P and CELF–3 indicated good concurrent validity. Despite the fact that these tests do not target the ability to provide context-appropriate referents in discourse, scores from these tests correlated with the FM scores. Stories provide a context for children to use the language skills that are directly tested in tests of discrete language skills, along with their ability to monitor listener knowledge and keep track of the story thus far, to provide referents to meet listener needs.

Similar to previous research on cohesion in stories reviewed earlier (e.g., Hickmann, 1991, 1997, 2003; Kail & Hickmann, 1992; Karmiloff-Smith, 1987; Liles, 1985a, 1985b; Peterson, 1993; Schneider & Dubé, 1997), we found that older children achieved higher scores on referent introduction than younger children, and children with TD had higher scores than children of the same age with LI. The advantages that the FM score has over previous measures of cohesion adequacy are that it is limited to one type of cohesion, namely, referential; it focuses on adequacy of expressions for cohesion within stories rather than on mastery of particular linguistic forms; and it facilitates the qualitative analysis of referential error types for intervention planning.

Examples from our normative data illustrate the difference between FM and RA measures and make it clearer why FM might be a better measure of the skill. One 6-year-old TD child introduced seven of the 14 referents with definite noun phrases (e.g., *the elephant, the airplane*) and introduced one with a pronoun. He scored just within 1 *SD* of his age group mean with the FM measure (based on 33 of 42, or 79% possible raw score points) but 1.5 *SD*s below the mean with the RA measure (with a raw RA score of 43%). His story began, "The elephant and the giraffe play with the ball." The three referents were scored as inadequate (0) in the RA scoring but were awarded 2 points each in the FM scoring. In contrast, consider a 6-year-old child with LI who earned a higher RA score than FM score. He obtained a score in the normal range for RA (57%), but his FM score (31/42, or 74%) was 1.24 *SD*s below the mean. This child introduced five referents with expressions that were scored 1 (mainly pronouns) using FM scoring and one with a definite determiner that was scored 2. In the RA scoring, these expressions were scored 0. Note that this child's RA score would be no different than that of a child who used six definite determiners and no pronouns. A story in which referents are introduced with definite noun phrases is easier for listeners to follow than a story in which pronouns are used to introduce referents, and thus, the FM scoring appears to reflect the quality of referent introduction better than the RA scoring.

We believe that the FM measure is potentially useful for identifying problems in establishing characters and objects in stories. By evaluating reference to a pre-established set of target characters and objects, the measure avoids a major limitation of measures such as the RA measure that are calculated on all referents introduced by children. With a consistent set of referents, it is not possible for two children to achieve the same score based on different numbers of attempts—children attempting fewer referents will obtain a lower score. In addition, the FM measure takes into account different degrees of referential inadequacy by awarding more points for referring expressions that provide some information (e.g., definite determiner + noun) than for those providing minimal information (e.g., pronouns).

Clinical Implications

The results of this study suggest that the FM measure is a useful tool for assessing an important aspect of narrative ability, namely, referential cohesion. The FM measure could be used to determine whether a child is having difficulty with this aspect of storytelling. In conjunction with other measures such as Story Grammar, the clinician could determine the type and overall severity of a child's discourse problems. Working on referring expressions in intervention could be very helpful in making a child's storytelling more comprehensible. The FM measure provides information about types of errors that can be used in planning intervention. If a child uses pronouns to introduce referents, for example, the SLP could encourage the child to provide nouns. If the child uses definite noun phrases, the SLP could teach the child to use the indefinite article and, to emphasize that a new referent is being mentioned, the SLP could teach the child to use it within introductory phrases such as "Once there was a…" and "then along came a…".

The FM measure evaluates only the way that referents are introduced. As mentioned earlier, previous work has suggested that children have more difficulty with referent introduction than with subsequent mentions, at least for simple stories. In addition, first mentions are easier to evaluate because referents can be expected to be introduced once in a story while they may be referred to any number of times subsequently, and the rules for evaluating subsequent mentions are more difficult to specify. Nevertheless, it is possible that some children would have greater difficulty with subsequent mentions than introductions. The FM measure considers only a portion of referential cohesion. When planning intervention, we recommend attention to a child's overall competence with referential cohesion.

The FM scoring as described here is specific to the ENNI stories: Scoring was based on the particular referents that were pictured in the ENNI stimuli, and evaluation of the results was made with reference to the ENNI local normative database. It would not be advisable to use the FM measure in its current form with other story sets unless normative data were available for those stories as well. In future research, FM scoring could be adapted to other stories by describing criteria for selected referents in the stories and collecting normative data for the appropriate age range. It would also be possible to make the task more appropriate for older children by choice of target referents. For example, characters who were similar in many ways could be incorporated into stories, making

it more difficult to distinguish them in referring expressions, as in the film used in studies by Liles (1985a) and in the pear stories study (Chafe, 1980), both of which depicted a large number of young male characters. Future research could focus on FM scores that can be expected using more referentially complex stories at different ages.

Although the FM scoring as presented here is tied to the ENNI stories, we believe that the principle of attending to type of referential error can be applied more broadly when working with children on storytelling. It is important to note not only the frequency of errors made by children when introducing characters and objects, but also the types of errors. As noted earlier, knowing the type of errors children make can guide the way clinicians help children and may also help in selecting appropriate stories for children (e.g., number of potential confusable characters in a story).

The current study did not investigate whether subtypes of LI would be related to adequacy of referent introduction. It is possible that children with receptive language problems might score differently than children with expressive language problems only. Additionally, it would be interesting to investigate how children with other developmental disorders, such as autism spectrum disorder, would score on the FM measure. Future studies could include samples of children with different patterns of communication impairment.

Future research should be conducted to explore referential ability in the context of other narrative measures. It is expected that measures of a range of narrative ability, including story information, cohesion, and semantic and syntactic measures, would provide a complete picture of a child's competence in narrative contexts. As studies of adults' judgments of stories have shown (McCabe & Peterson, 1984; Schneider & Winship, 2002), different narrative measures contribute to adults' overall assessment of the quality of a story, and no one measure appears to capture everything that contributes to perceptions of story quality. In order to be able to provide effective narrative intervention, it is necessary to have measures that will pinpoint the exact nature of a particular child's difficulty with storytelling. It is important to look at stories from a number of different perspectives in order to characterize a child's narrative ability and to design an intervention that targets his or her areas of difficulty.

ACKNOWLEDGMENTS

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.

REFERENCES

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

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.

Boudreau, D. M., & Chapman, R. S. (2000). The relationship between event representation and linguistic skill in narratives of children and adolescents with Down syndrome. *Journal of Speech, Language, and Hearing Research, 43,* 1146–1159.

Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

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

Finestack, L. H., Fey, M. E., & Catts, H. W. (2006). Pronominal reference skills of second and fourth grade children with language impairment. *Journal of Communication Disorders, 39,* 232–248.

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. S. (2001). Children with a history of expressive vocabulary delay: Outcomes at 5 years of age. *American Journal of Speech-Language Pathology, 10,* 358–369.

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics, 30,* 5–56.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Longmans Group.

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

Hickmann, M. (1991). The development of discourse cohesion: Some functional and cross-linguistic issues. In G. Piéraut-le Bonniec & M. Dolitsky (Eds.), *Language bases…discourse bases: Some aspects of contemporary French-language psycholinguistics research* (pp. 157–185). Amsterdam, Netherlands: John Benjamins.

Hickmann, M. (1995). Discourse organization and the development of reference to person, space and time. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 194–218). Oxford, UK: Blackwell.

Hickmann, M. (1997). Information status and grounding in children's narratives: A crosslinguistic perspective. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 221–243). Mahwah, NJ: Erlbaum.

Hickmann, M. (2003). *Children's discourse: Person, space and time across languages*. Cambridge, UK: Cambridge University Press.

Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language, 26,* 419–452.

Hickmann, M., Hendriks, H., Roland, F., & Liang, J. (1996). The marking of new information in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language, 23,* 591–619.

Hickmann, M., & Schneider, P. (2000). Cohesion and coherence anomalies and their effects on children's referent introduction in narrative retell. In M. Perkins & S. Howard (Eds.), *New directions in language development and disorders* (pp. 251–260). New York, NY: Plenum.

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

Kail, M., & Hickmann, M. (1992). French children's ability to introduce referents in narratives as a function of mutual knowledge. *First Language, 12,* 73–94.

Karmiloff-Smith, A. (1987). Function and process in comparing language and cognition. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 185–202). Orlando, FL: Academic Press.

Klecan-Aker, J. S. (1985). Syntactic abilities in normal and language deficient middle school children. *Topics in Language Disorders, 5,* 46–54.

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

Liles, B. Z. (1985a). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28,* 123–133.

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

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.

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* (3rd ed.). Mahwah, NJ: Erlbaum.

Mayer, M. (1974). *Frog goes to dinner*. New York, NY: Puffin Pied Piper.

McCabe, A. (1996). Evaluating narrative discourse skills. In S. F. Warren & J. Reichle (Series Eds.), K. N. Cole, P. S. Dale & D. J. Thal (Vol. Eds.), *Communication and language intervention series: Vol. 6. Assessment of communication and language* (pp. 121–141). Baltimore, MD: Brookes.

McCabe, A., & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research, 13,* 457–480.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38,* 287–313.

Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late talkers: Early school age. *Journal of Speech and Hearing Research, 39,* 1295–1303.

Paul, R., & Smith, R. L. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36,* 592–598.

Peterson, C. (1993). Identifying referents and linking sentences cohesively in narration. *Discourse Processes, 16,* 507–524.

Peterson, C., & Dodsworth, P. (1991). A longitudinal analysis of young children's cohesion and noun specification in narratives. *Journal of Child Language, 18,* 397–415.

Peterson, C., & McCabe, A. (1987). The connective 'and': Do older children use it less as they learn other connectives? *Journal of Child Language, 14,* 375–381.

Pratt, M. W., & MacKenzie-Keating, S. (1985). Organizing stories: Effects of development and task difficulty on referential cohesion in narrative. *Developmental Psychology, 21,* 350–356.

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

Schneider, P. (2008, June). *Referent introduction in stories by children and adults*. Poster presented at the triennial meeting of the International Association for the Study of Child Language, Edinburgh, Scotland.

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., & Hayward, D. (2009). *The Edmonton Narrative Norms Instrument*. Available from http://www.rehabmed/ualberta.ca/spa/enni.

Schneider, P., & Hayward, D. (2004, June). *Measuring referring expressions in a story context*. Poster presented at the Symposium for Research in Child Language Disorders, Madison, WI.

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology, 30,* 224–238.

Schneider, P., & Winship, S. (2002). Adults' judgments of fictional story quality. *Journal of Speech, Language, and Hearing Research, 45,* 372–383.

Scott, C. M. (1988). A perspective on the evaluation of school children's narratives. *Language, Speech, and Hearing Services in Schools, 19,* 67–82.

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

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology, 27,* 960–974.

Statistics Canada. (n.d.). *Canada dimensions: The people*. Retrieved from http://www.statcan.ca.

Stein, N. L. (1988). The development of children's storytelling skill. In M. B. Franklin & S. S. Barten (Eds.), *Child language: A reader* (pp. 282–297). New York, NY: Oxford University Press.

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.

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

Strong, C. J., & Shaver, J. P. (1991). Stability of cohesion in the spoken narratives of language-impaired and normally developing school-aged children. *Journal of Speech and Hearing Research, 34,* 95–111.

Thorne, J. C., Coggins, T. E., Olson, H. C., & Astley, S. J. (2007). Exploring the utility of narrative analysis in diagnostic decision making: Picture-bound reference, elaboration, and fetal alcohol spectrum disorders. *Journal of Speech, Language, and Hearing Research, 50,* 459–474.

Tsai, W., & Chang, C. (2008). "But I first…and then he kept picking": Narrative skill in Mandarin-speaking children with language impairment. *Narrative Inquiry, 18,* 349–377.

Vallance, D. D., Im, N., & Cohen, N. J. (1999). Discourse deficits associated with psychiatric disorders and with language impairments in children. *Journal of Child Psychology and Psychiatry, 40,* 693–704.

Villaume, S. K. (1988). Creating context within text: An investigation of primary-grade children's character introductions in original stories. *Research in the Teaching of English, 22,* 161–182.

Vion, M., & Colas, A. (1998). L'introduction des référents dans le discours en français: Contraintes cognitives et développement des compétences narratives [Introducing referents in French: Cognitive constraints and development of narrative skills]. *L'année psychologique, 98,* 37–59.

Wagner, C. R., Nettelbladt, U., 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.

Warden, D. A. (1981). Learning to identify referents. *British Journal of Psychology, 72,* 93–99.

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

Wigglesworth, G. (1990). Children's narrative acquisition: A study of some aspects of reference and anaphora. *First Language, 10,* 105–125.

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

Wittek, A., & Tomasello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics, 26,* 541–558.

Wong, A. M.-Y., & Johnston, J. R. (2002). The development of discourse referencing in Cantonese-speaking children. *Journal of Child Language, 31,* 633–660.

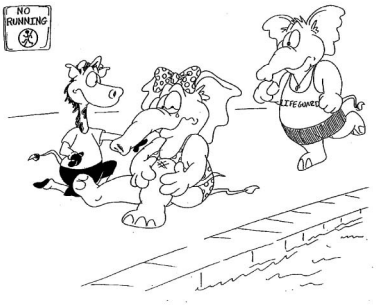
Received June 8, 2009
Revision received October 5, 2009
Accepted February 3, 2010
DOI: 10.1044/0161-1461(2010/09-0040)

Contact author: Phyllis Schneider, Department of Speech Pathology and Audiology, University of Alberta, 2-70 Corbett Hall, Edmonton, Alberta, T6G 0G2, Canada. E-mail: phyllis.schneider@ualberta.ca.

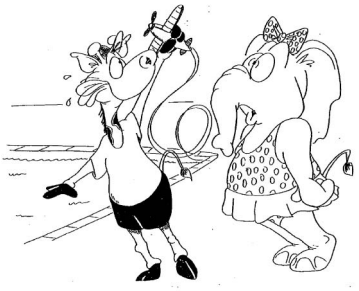
APPENDIX A. PICTURES IN WHICH THE TARGET REFERENTS ARE INTRODUCED IN THE EDMONTON NARRATIVE NORMS INSTRUMENT (ENNI; SCHNEIDER, DUBÉ, & HAYWARD, 2009).

Set A

Giraffe, elephant, ball



Second elephant



Airplane

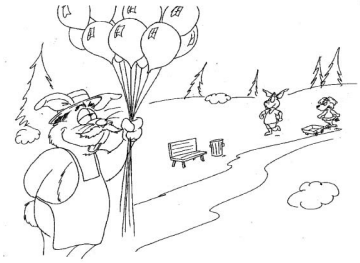
Third elephant, net

Set B

Rabbit, dog, sandcastle

Second rabbit

Balloon



Third rabbit

Balloons (end)

Note. Copyright 2009 by Wooket. Reprinted with permission.

APPENDIX B. EXAMPLE OF SCORING CRITERIA FOR THE FIRST MENTIONS MEASURE

Character	*Score as 3*	*Score as 2*	*Score as 1*
Giraffe – Story A1	**a/this** _____ (e.g., *a giraffe, this cow*) **name** (e.g., *Gerry, Geegee*) **possessive + noun** (e.g., *her friend* if "she" already introduced) **another animal** **the other animal** (if C mentioned 2 animals and one animal mentioned previously)	**the/that** _____ (e.g., *the giraffe*) **a [invented word],** e.g., *a geegee* **someone/somebody** **possessive + noun** (if other character not yet introduced) **another/the other** _____ (e.g., *the other animal* if no animal mentioned previously)	**pronoun** (*he, she, it*) **the [invented word]**, e.g., *the geegee* (an invented **name** would be scored as 3)
Elephant – Story A1	**a/this** _____ (e.g., *a elephant*) **name** (e.g., *Ellie*) **possessive + noun** (e.g., *her friend* if "she" already introduced) **another** _____ (*e.g., another animal* if other character introduced as animal) **the other** _____ (e.g., *the other animal* if C mentioned 2 animals and one animal mentioned previously)	**the/that** _____ (e.g., *the elephant*) **a [invented word]** **someone/somebody** **possessive + noun** (if other character not yet introduced) **another/the other** _____ (e.g., *the other animal* if no animal mentioned previously)	**pronoun** (*he, she, it*) **the [invented word]** (an invented **name** would be scored as 3)
Ball – Story A1	**a/this** _____ (e.g., *a ball, a balloon, an orange*) **possessive + noun** (e.g., *her ball, the elephant's ball*) **the ball** if character is "playing ball"	**the/that**_____ **vague or empty term**, e.g., *a thingy/something/whatchacallit* **a [invented word]**	**pronoun** (*it*) **the [invented word]**

Note. A score of 3 indicates that the expression is fully adequate for first mention, a score of 2 indicates that the expression is less than adequate but still partially informative, and a score of 1 indicates that the expression is inadequate for first mention. Full scoring information is available at http://www.rehabmed.ualberta.ca/spa/enni.