Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Communication Disorders 39 (2006) 232–248
Journal of Communication Disorders

Pronominal reference skills of second and fourth grade children with language impairment

Lizbeth H. Finestack [a,*], Marc E. Fey [a], Hugh W. Catts [b]

[a] *Department of Hearing and Speech, University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, Kansas 66160-7605, United States*
[b] *Department of Speech-Language-Hearing: Sciences and Disorders, The University of Kansas, Lawrence, United States*

Received 11 May 2005; received in revised form 11 November 2005; accepted 21 December 2005

Abstract

Pronominal referencing was evaluated in a sample of 569 children comprising four diagnostic subgroups: typical language (TL), specific language impairment (SLI), nonspecific language impairment (NLI), and typical language with low nonverbal IQ (LNIQ). Participants generated oral narratives in second grade and again in fourth grade. The narratives of the females in the TL group included a significantly higher rate of pronominal references than the narratives of both the males in the TL group and the females in the NLI group. A higher percentage of complete pronominal references was found in the TL group compared to the SLI group. These findings suggest that pronominal referencing measures are not sensitive enough to differentiate school-aged children with typical language development from those with language impairment.

Learning outcomes: The reader will: (1) become familiar with narrative language measures and (2) learn how groups of children with varying language and cognitive profiles perform on measures of referential cohesion.

1. Introduction

Language weaknesses of school-aged children who have difficulties acquiring language are evident when these children are asked to generate and/or retell oral

* Corresponding author.
E-mail address: lfinestack@kumc.edu (L.H. Finestack).

0021-9924/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcomdis.2005.12.003

narratives. Compared to children with typical language development (TL), children with language impairment (LI) have difficulties producing narratives at the global or macrostructural level as well as the sentence or microstructural level. At the macrostructural level, children with LI tend to include fewer story components (e.g., setting, initiating events, and direct consequences) (Merritt & Liles, 1987), fewer complete narrative episodes (Liles, 1987; Merritt & Liles, 1987), and fewer pieces of relevant information (Paul & Smith, 1993). Moreover, children with LI demonstrate poorer overall narrative quality based on development of characters, setting, plot complexity, ending, and language sophistication (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004). At the microstructural level, children with LI exhibit a lower average number of different word roots (Fey et al., 2004; Paul & Smith, 1993), a shorter mean length of utterance per T-unit (Paul & Smith, 1993), fewer sentences per narrative (Liles, 1985c), reduced grammatical accuracy (Fey et al., 2004; Gillam & Johnston, 1992; Norbury & Bishop, 2003), poorer use of cohesive conjunctives (Liles, 1987), a lower percent of cohesive adequacy (Paul & Smith, 1993), fewer correct uses of cohesive ties (Baltaxe & D'Angiola, 1992), and a greater number of incomplete and error ties (Liles, 1985a).

In a study of children's narrative generation skills, Fey et al. (2004) included the largest sample of children with LI to date. The sample of 538 children comprised four subgroups (i.e., children with typical language and cognitive development [$N = 262$], children with LI, but typical nonverbal IQ [$N = 111$], children with LI and below normal nonverbal IQ [$N = 75$], and children with typical language development, but below normal nonverbal IQ [$N = 90$]). Fey et al. confirmed previous findings of macro- and microstructural performance differences between groups of children with and without LI. Specifically, second and fourth grade children with LI demonstrated language weaknesses compared to children with typical development on the macrostructural narrative quality measure and the microstructural measures of number of different words, clausal density (i.e., average number of clauses per C-unit), and percent of grammatical C-units. However, Fey et al. did not include a measure of cohesion, one of the microstructural measures that has consistently differentiated groups of children with and without LI and was found by Liles et al. to be among the most discriminating measures (Liles, 1985b; Merritt & Liles, 1987; Purcell & Liles, 1992). The current investigation was designed to examine pronoun use among the children in the Fey et al. sample.

According to Halliday and Hasan (1976), cohesion is one of the defining elements of text that establishes continuity in a spoken or written narrative. There are five distinct types of textual cohesion, including reference, substitution, ellipsis, conjunction, and lexical cohesion. For each type, interpretation relies on the establishment of a relationship between the cohesive device and another entity either within or outside the text. A number of studies have examined these cohesion types in the narratives of children with LI compared to children with typical language development and have found that children with LI use a lower percentage of complete cohesive ties (Baltaxe & D'Angiola, 1992; Paul & Smith, 1993; Strong & Shaver, 1991), a higher percentage of incomplete cohesive ties (Liles, 1985a; Strong & Shaver, 1991), and a smaller total number of ties (Strong & Shaver, 1991).

When examined individually, not all types of textual cohesion identified by Halliday and Hasan (1976) are equally efficient in differentiating groups of children with impairments from those with typical development. For example, in an investigation of all five types of cohesion among children with typical language development and autism, Baltaxe and D'Angiola (1992) only found group differences based on reference and ellipsis. In another study, Liles (1985a) initially aimed to identify and analyze all of Halliday and Hasan's cohesion types among children with typical language development and children with LI; however, due to low rates of occurrence, comparative reference, ellipsis, and substitution were excluded from analyses. Group differences were found for only pronominal referencing and lexical cohesion. Based on these studies, the one cohesive element that most consistently appears to differentiate children with LI from their peers is referential cohesion.

Despite the seemingly reliable differences found between children with LI and children developing typically based on referential cohesion skills, the generalizability of these findings is limited. Previous investigations have included relatively small samples of a particular subgroup of mostly clinically referred children with LI. Of the studies including measures of cohesiveness referenced thus far, the LI group sample sizes ranged from 8 to 23 children. Almost all of the studies included children with LI that could be subcategorized as children with specific language impairment (SLI). That is, these children demonstrated LI, without any evidence of cognitive impairment based on measures of performance IQ. Additionally, the children with SLI included in these studies were receiving speech-language services and may only represent the most severely affected children with SLI.

1.1. Current investigation

The current investigation sought to examine the use of pronominal cohesion by children with different language and cognitive profiles using the same large groups of children that participated in the Fey et al. study (2004). Both this and the Fey et al. studies included two diagnostic groups commonly found on the caseloads of speech-language pathologists: (a) children with below average language, but typical nonverbal IQ (i.e., those with SLI), and (b) children with below average language and nonverbal IQ (i.e., those with nonspecific LI, or NLI). Also included in these studies were two control groups that did not demonstrate language impairment: (a) children with typical language and nonverbal IQ (i.e., those with TL) and (b) children with typical language, but below average nonverbal IQ (i.e., those with low nonverbal IQ, or LNIQ). Two cohesion measures, frequency and completeness of pronominal referencing, were selected as the basis for comparison due to their use in differentiating children with different language abilities in studies with smaller sample sizes.

This study aimed to answer two questions based on an oral story generation task for second and fourth grade children across the four subgroups studied by Fey et al. (2004). The two questions asked for these four subgroups were:

1. Do children with TL, LNIQ, SLI, and NLI use a significantly different rate of pronominal references, measured in terms of pronominal references per C-unit?

2. Do children with TL, LNIQ, SLI, and NLI use a significantly different number of complete pronominal references, measured in terms of the percent of all pronominal references used that are judged to be complete?

These questions were considered in terms of both the grade level and sex of the children.

Based on findings from earlier studies, it was expected that the two groups with LI (i.e., the SLI and NLI groups) would use significantly fewer pronouns per C-unit (Strong & Shaver, 1991) and a smaller percentage of complete ties than the children without LI (i.e., the TL and LNIQ groups). Previous investigations have not included different subtypes of children based on cognitive and language skills; therefore, there were no empirically driven expectations for differential performance between the two groups of children with LI (i.e., SLI versus NLI) or the two groups of children with typical language development (i.e., TL versus LNIQ). Moreover, findings from previous investigations indicate that older children produce more references and demonstrate greater cohesion adequacy (Strong & Shaver, 1991); therefore, it was expected that, regardless of diagnosis, all of the children would produce more references and a higher percentage of complete referential ties in fourth grade than in second grade. It is often claimed that girls have a language advantage over boys due to girls' tendency to develop language skills earlier than boys (Moyano & McGillivray, 1988 [cited in Hughes, McGillivray, & Schmidek, 1997]). Although the evidence supporting this claim has been somewhat unclear (MaCaulay, 1978), Fey et al. (2004) found that for almost all of their narrative measures, the girls outperformed the boys. Therefore, it was expected that the girls would also outperform the boys on our narrative cohesion measures.

2. Method

2.1. Participants

The participants for this study came from the same sample included in the Fey et al.' (2004) investigation of oral and written narrative skills of children with LI. These participants had been involved in an epidemiological study of language impairments in kindergarten children (Tomblin et al., 1997). The epidemiological sample consisted of 7218 children who were all monolingual English speakers, had no evidence of sensory deficits or frank neurological disorders, and had no known diagnosis of autism or mental retardation when initially evaluated in kindergarten.

The children in the epidemiological study (Tomblin et al., 1997) were originally identified as having LI in kindergarten based on five composite scores from the *Test of Language Development-2:Primary* (TOLD-2:P) (Newcomer & Hammill, 1988) and a narrative retell task (Culatta, Page, & Ellis, 1983). The five composites represented overall language production and comprehension abilities, as well as the specific domains of vocabulary, grammar, and narration. Children who scored lower than −1.25 S.D.s, based on chronological age, on two or more of the composites were identified as having LI in the epidemiological study. From this sample, all of the children identified as having LI and a random sample of children without LI were recruited to participate in a series of follow-up

longitudinal investigations (e.g., Catts, Fey, Zhang, & Tomblin, 1999; Ellis Weismer et al., 2000). A total of 328 children with LI and 276 children with no evidence of LI were available for the follow-up studies. Of these 604 children, 569 completed a battery of language, IQ, and oral narrative tests in the second and fourth grades and, therefore, were included in the present investigation. Our sample was slightly larger than the Fey et al. sample ($n = 538$) because these investigators excluded children who did not produce written narratives in the second or fourth grades, whereas we did not.

2.1.1. Second grade group classifications

Fey et al. (2004) were interested in comparing the narrative skills of children with different language and cognitive profiles. Therefore, they opted not to use the multivariate standard for LI developed by Tomblin et al. (1996), which is itself based in part on the children's narrative skills. Tomblin et al. (1996) found that a single-test composite score of −1.14 S.D.s that contained no measure of narrative performance provided a good univariate representation of the Tomblin et al. multivariate criterion for LI. Thus, Fey et al. formed a composite from the *Clinical Evaluation of Language Fundamentals-3* (CELF-3) (Semel, Wiig, & Secord, 1995), the *Peabody Picture Vocabulary Test—Revised* (PPVT-R) (Dunn, 1981), and the Expressive Vocabulary subtest of the *Comprehensive Receptive and Expressive Vocabulary Test* (CREVT) (Wallace & Hammill, 1994) and adopted −1.14 S.D.s as their criterion for LI. We adopted the same method for identification of LI in the present investigation based on the children's second grade performance.

Additionally, performance IQ was indexed in second grade based on the entire Performance Scale of the *Wechsler Intelligence Scale for Children—Revised* (WISC-R) (Wechsler, 1974). As was the case for Tomblin et al. (1997) and Fey et al. (2004), children with IQ scores lower than −1 S.D. were identified as having nonverbal cognitive impairment.

Based on these two identification measures (i.e., LI and nonverbal cognitive impairment), each child was identified as falling into one of the following groups: typical language with typical nonverbal cognition (i.e., typical language), typical language with impaired nonverbal cognition (i.e., low nonverbal IQ), LI with typical nonverbal cognition (i.e., specific language impairment), and LI with impaired nonverbal cognition (i.e., nonspecific language impairment). Of the 569 children who completed the entire oral evaluation battery in second grade, 339 were identified as TL, 69 as LNIQ, 61 as SLI, and 100 as NLI. Table 1 presents each of the groups' means and standard deviations for language, nonverbal IQ, chronological age, and biological mother's education, as well as the groups' male:female ratios, and race/ethnicity composition.

2.1.2. Group comparisons

The nature of the selection criteria used for each group ensured significant group differences based on nonverbal IQ for the typical nonverbal cognition groups (i.e., TL and SLI) and the impaired nonverbal cognition groups (i.e., NLI and LNIQ) and on language performance for the language impaired groups (i.e., SLI and NLI) and the typical language groups (i.e., TL and LNIQ). However, *t*-tests revealed differences between groups with similar nonverbal IQ or language profiles. Specifically, the TL group's nonverbal IQ was significantly higher than that of the SLI group, $t(398) = 4.671$, $p < .001$, $d = .74$; and the

Table 1
Participant characteristics by group

Characteristic	TL, $N = 339$, mean (S.D.)	SLI, $N = 61$, mean (S.D.)	NLI, $N = 100$, mean (S.D.)	LNIQ, $N = 69$, mean (S.D.)
Language composite[a,b]	100.39 (11.99)	77.67 (6.04)	72.44 (9.27)	93.27 (8.43)
Nonverbal IQ[b,c]	103.50 (10.97)	96.70 (6.93)	75.34 (6.89)	79.81 (4.65)
Age (years; months)	8; 0 (0; 5)	7; 11 (0; 4)	8; 0 (0; 4)	8; 1 (0; 4)
Maternal education (years)[d]	13.97 (2.14)	13.25 (1.95)	12.37 (1.82)	12.82 (1.76)
Male:female ratio	1.44	1.44	1.00	1.04
White (%)	93	85	68	68
Black (%)	4	13	29	28
Native American (%)	1	0	0	1
Hispanic (%)	1	2	2	1
Asian (%)	1	0	1	1

[a] The language composite is based on standard scores from the CELF-3 (Semel et al., 1995), the PPVT-R (Dunn, 1981), and the Expressive Vocabulary subsets of the CRVT (Wallace & Hammill, 1994).

[b] The average standard scores are summarized here; *z*-scores were used for group comparison analyses.

[c] The nonverbal IQ score is a composite based on the standard scores of the Block Design and Picture Completion subtests of the WISC-R (Wechsler, 1974).

[d] The biological mother's level education was not available for all children; therefore, these averages are based on smaller samples. The sample sizes for this measure for each group are: TL = 335, SLI = 60, NLI = 97, and LNIQ = 66.

LNIQ group's nonverbal IQ was significantly greater than that of the NLI group, $t(167) = 4.700$, $p < .001$, $d = .76$. Moreover, the TL group's language performance was significantly higher than that of the LNIQ group, $t(406) = 4.701$, $p < .001$, $d = .69$, and the SLI group's language performance was significantly greater than that of the NLI group, $t(159) = 3.919$, $p < .001$, $d = .67$.

Three separate ANOVAs were completed to compare the groups based on measures not specifically included in the group selection criteria (i.e., age, biological mother's level of education attainment, and the male:female ratio) with alpha set at .05. Significant ANOVAs were followed by post hoc comparisons utilizing Tukey's Honest Significant Difference Test for unequal samples, with alpha set at .05. Main effects were found for both age, $F(3, 565) = 3.860$, $p = .009$, and maternal education, $F(3, 554) = 18.886$, $p < .001$. Post hoc group comparisons based on age failed to yield significant differences (all $p > .076$, d ranged from .174 to .489). It should be noted, however, that the children in the SLI group were on average 2 months younger than the children in the LNIQ group and 1 month younger than the children in the other groups. Post hoc comparisons revealed significantly greater levels of education attainment for the TL group compared to both the LNIQ ($p = .006$, $d = .524$) and NLI ($p < .001$, $d = .655$) groups. A main effect was not found for the male:female ratio ($p = .301$).

A Chi-square test for group differences in race revealed significant group differences: $\chi^2 = 64.09$, d.f. = 12, $p < .001$. Inspection of the ethnic compositions indicates that the TL group included considerably more children who were White and fewer children who were Black than either the LNIQ or NLI groups.

Group differences found on most of the pre-experimental measures were anticipated and appear to depict the natural characteristics of groups of children with similar language

and cognitive profiles. For example, although the performance IQs of samples of children with SLI fall in the average range, group IQs have been found to be lower than those of their typically developing peers (Stark & Tallal, 1981). Moreover, when compared to other children with LI, children with both LI and cognitive impairment have been found to demonstrate more severely impaired language abilities compared to children with LI, but without cognitive impairment (Bishop & Edmundson, 1987). Additionally, the higher level of maternal education attainment for the TL group was expected due to the influence of genetics and environmental factors on language development (Gilger, 1995). Group differences based on race were not expected and may reflect racial and/or cultural bias in the tests used for experimental grouping. There is no apparent reason to expect differences between White and Black children in their referencing of pronouns, however. Therefore, this difference between groups would not appear to be a significant factor in our analyses for this investigation.

2.2. Procedure

2.2.1. Task administration

The oral narrative samples were collected as part of a large test battery administered during two 2-h sessions in both the second and fourth grades. The oral samples were embedded in the middle of the testing session. Whether the narratives were collected during the first or second testing sessions was counterbalanced across participants. Testing was completed in specially designed vans parked at the participants' schools or homes by 1 of 17 trained examiners. All language tasks, including the narrative generation tasks, were conducted by a certified speech-language pathologist.

On the first day of testing, each participant was shown four sets of laminated colored pictures and asked to choose two sets, one to use for an oral narrative and one for a written narrative. Each picture set contained three pictures depicting a conflict and vague resolution. Two of the stories included only animal characters; the other two stories included both human and animal characters. These pictures can be viewed and downloaded at the following website: http://www.ku.edu/%7Esplh/research/catts1.html.

The examiner placed the pictures from one of the sets not selected by the child on the table and labeled pre-identified key elements (e.g., characters and setting elements). The examiner then read a pre-written model story based on the picture set to the child. Across story sets, the models were written to be approximately equal in number of C-units, words, grammatical complexity, and basic episodic structure. Next, the examiner presented the participant with one of the child's selected picture sets and asked the child to identify key elements in the story. If the child failed to identify any of the key elements, the examiner pointed the element out, and labeled it for the child. The child was then prompted to tell a story using the three pictures within the set. The examiner was instructed not to comment or intervene during the story telling. Examiner prompts were only allowed under two circumstances: (a) if the child merely labeled story elements, he or she was prompted to tell a story rather than just describe what was in the pictures, and (b) if the child stopped without obviously ending the story, he or she was asked, ''Is that all?''

All of the children's oral narratives were transcribed onto computer in SALT format (Miller & Chapman, 2000) by a graduate student in speech-language pathology.

One C-unit (i.e., an independent clause plus all of its associated dependent clauses) was transcribed per line. Stereotypic closings (e.g., ''The end.'') and comments not part of the story (e.g., ''What do you call that?'' ''I like horses.'') were coded on special lines and were not included in any of the analyses. Unintelligible utterances were also excluded from all analyses.

2.2.2. Dependent measures

For the two primary study analyses, two measures related to narrative cohesiveness were obtained: (a) rate of pronominal reference per C-unit and (b) the percentage of complete pronominal references. To obtain these measures, each child's narrative was coded using Computerized Profiling's Narrative Analysis Profile (NAP) (Long, Fey, & Channell, 2002), a computer module based on Halliday and Hasan's (1976) descriptions and recommended analysis of textual cohesion. Specifically, NAP was used to identify the personal pronominal references that appeared in the narratives, including both existential and possessive first, second and third person pronouns (see Table 2). The pronominal references coded included all of those identified by Halliday and Hasan except *one* and *one's*. These pronominal references were excluded due to their low frequency in American English children's narratives and their potential confusion with the more frequent use of this form as a quantifier. Other pronominal references excluded from analysis were expletive *it* (e.g., *It* is raining), *you* in social phrases (e.g., thank *you*), and pronouns appearing in idioms (e.g., He didn't make *it* (i.e., he died)). Additionally, only anaphoric references were included (i.e., references tied to explicit noun references, or R-expressions, in the preceding text); neither exophoric ties (i.e., references referring to referents outside of the text) nor cataphoric ties (i.e., references referring to referents in the following text) were included.

Once a pronominal reference was identified by NAP, a judgment was made regarding its completeness. A pronoun provides information about the referent's relationship to the speaker, gender, and number, but it does not specify a referent. For example, the pronoun *he* in the sentence, *He played on the swings*, expresses that the entity is one male who is not the speaker or listener, but without additional information, the referent is unknown. To interpret a pronoun, there must be a noun, or R-expression, to which a specific pronoun

Table 2
Pronominal references identified and coded

	Pronominal references	
	Existential	Possessive
First person	I Me	Mine My
	We Us	Ours Our
Second person	You	Yours Your
	We Us	Ours Our
Third person	He Him	His
	She Her	Hers Her
	It	Its
	They Them	Theirs Their

refers. If the sentence, *John went to the park*, is inserted before the sentence, *He played on the swings*, it is clear that *he* refers to the R-expression *John*. Therefore, to be coded as complete, a pronoun's corresponding R-expression had to be clearly and unambiguously identified. If there were more than one possible corresponding R-expression, the pronoun was coded as ambiguous. If there were no possible R-expressions to link to the pronoun, it was coded as failed.

2.3. Coding procedure and reliability

NAP automatically identified all of the relevant pronominal references in the narratives based on the SALT transcription. Thus, the identification of pronouns within the narratives was essentially errorless. A graduate student in speech-language pathology was trained to code the identified pronouns for completeness and served as the primary coder. All of the narratives were coded by the primary coder and a secondary coder (the first author) until the coders reached a level of at least 90% agreement based on 40 narrative samples coded independently. This level was met after scoring 110 narratives. Thereafter, the secondary coder randomly selected approximately 10 of every 40 stories coded by the primary coder to code independently. Both coders were blind to the children's group assignments.

Narratives coded by both the primary and secondary judges were compared for reliability. Of the 1028 narratives scored by the primary coder after reaching 90% reliability, 245 (i.e., 24%) were coded by the secondary coder for reliability. The coders agreed on 87% of the completeness codes. The coders met to discuss discrepancies and modified the pronominal codes, as they deemed appropriate.

2.3.1. Rescaling of scores on dependent variables

Each participant's oral narrative was based on one of four different story picture stimulus sets. To control for possible effects based on stimuli selection, both of the dependent variables were rescaled to a common standard (see Fey et al., 2004). Arbitrarily, the story ''John's Birds'' was used as the common standard; therefore, measures based on the other three stories were adjusted up or down to match the scale of ''John's Birds.''

3. Results

To determine if the rate of pronominal references per C-unit and the percentage of complete references differed across group assignments, two analyses of variance (ANOVAs) with repeated measures were completed. In each analysis, the between subject variables were Group (TL versus LNIQ versus SLI versus NLI) and Sex. Grade served as the repeated measure. Effects were judged to be reliable if a significant ANOVA (with alpha set at $p = .05$) was followed by a significant Tukey's Honestly Significant Difference Test for unequal samples (with alpha set at $p = .05$). To help meet the assumption of homogeneity of variance, the rates of references were square-root transformed and the percentages of complete references were arc sine transformed. The means and standard

Table 3
Means and standard deviations for pronominal reference rate

Group	Pronominal reference/C-unit,[a] mean (S.D.)		
	2nd	4th	Total
TL	.636 (.333)	.686 (.346)	.661 (.340)
Male ($n = 200$)	.572 (.297)	.645 (.332)	.608 (.317)
Female ($n = 139$)	.728 (.360)	.745 (.358)	.736 (.358)
LNIQ	.557 (.290)	.658 (.331)	.608 (.314)
Male ($n = 35$)	.593 (.321)	.687 (.379)	.640 (.352)
Female ($n = 34$)	.521 (.255)	.628 (.276)	.575 (.269)
SLI	.550 (.263)	.609 (.347)	.579 (.308)
Male ($n = 36$)	.577 (.266)	.606 (.354)	.592 (.312)
Female ($n = 25$)	.511 (.259)	.611 (.342)	.561 (.304)
NLI	.530 (.299)	.535 (.295)	.532 (.296)
Male ($n = 50$)	.493 (.255)	.566 (.313)	.530 (.286)
Female ($n = 50$)	.566 (.336)	.503 (.276)	.535 (.307)

[a] Based on untransformed rescaled values.

deviations for the untransformed rescaled values of each dependent measure appear in Tables 3 and 4.

The ANOVA for the rate of pronominal references per C-unit was significant for Group, $F(3, 561) = 10.593$, $p < .001$. Post hoc analysis revealed a significant difference between the TL and NLI groups, $p < .001$, $d = .41$. As illustrated in Fig. 1, this main effect was found to be driven by a significant Sex × Group interaction, $F(3, 561) = 4.354$, $p = .005$. Follow-up comparisons examining differences based on sex within each group indicated a

Table 4
Means and standard deviations for percentage of complete references

Group	Complete references (%),[a,b] mean (S.D.)		
	2nd	4th	Total
TL	84.18 (20.48)	85.83 (17.08)	85.00 (18.84)
Male ($n = 66$)	87.37 (17.06)	82.30 (19.13)	84.83 (18.23)
Female ($n = 69$)	81.13 (23.01)	89.20 (14.19)	85.17 (19.47)
LNIQ	84.30 (17.96)	83.01 (15.80)	83.65 (16.71)
Male ($n = 9$)	84.99 (20.00)	77.58 (17.81)	81.28 (18.76)
Female ($n = 11$)	83.74 (17.09)	87.44 (13.14)	85.59 (15.00)
SLI	65.30 (23.12)	77.94 (23.98)	71.62 (24.12)
Male ($n = 12$)	58.18 (26.00)	79.14 (27.89)	68.66 (28.46)
Female ($n = 8$)	75.98 (13.09)	76.14 (18.24)	76.06 (15.34)
NLI	78.22 (18.55)	81.39 (20.16)	79.80 (19.21)
Male ($n = 10$)	80.88 (14.85)	81.14 (22.23)	81.01 (18.40)
Female ($n = 12$)	76.00 (21.56)	81.60 (19.28)	78.80 (20.20)

[a] Based on untransformed rescaled values.
[b] Only children who produced narratives with five or more pronominal references are included in group comparison analyses.

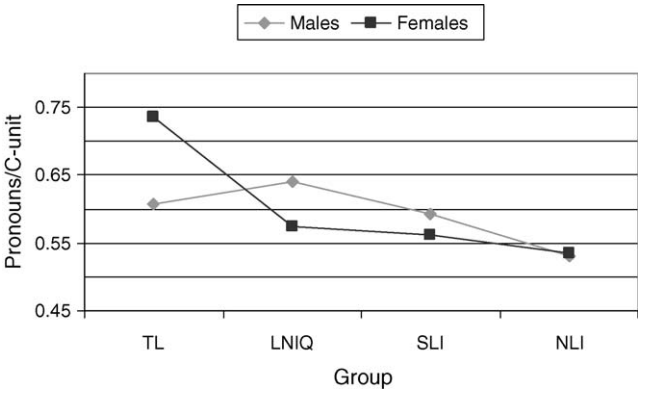

Fig. 1. Group × Sex interaction.

significant difference between the males and females only in the TL group, $p < .001$, $d = .38$ (all other $p > .987$, $d < .17$). Comparisons within sex across diagnostic groups indicated a significant difference between only the TL and NLI groups for the females, $p < .001$, $d = .621$. Although the differences between the females in the TL group and the females in both the SLI and LNIQ groups were not statistically reliable (both $p > .190$), medium effect sizes were found for both comparisons (SLI: $d = .52$; LNIQ: $d = .454$). There were no other significant main effects or interactions found for referencing rate (all other $p > .437$).

The finding that the TL females outperformed the TL males based on pronominal referencing rate was not anticipated. Therefore, a post hoc analysis was completed to determine whether these two groups differed in ways other than sex. Individual *t*-tests examining potential differences between these two groups based on their language composite scores, performance IQ, age, and mothers' education failed to indicate any significant differences (all $p > .082$, $d < .19$). Thus, there is no evidence that these groups differed in any other ways that would have contributed to the advantage of the females in the TL group on this measure.

To help maintain the assumption of normality for analyses of the percentage of complete references, only narrative samples with five or more pronouns were included. This criterion resulted in dramatically reduced sample sizes: TL = 135, LNIQ = 20, SLI = 20, and NLI = 22. The ANOVA for the percentage of complete references was significant for Group, $F(3, 189) = 4.827$, $p = .003$. There were no other significant main effects or interactions (all other $p > .074$). Post hoc analysis revealed a significant difference between the TL and SLI groups, $p = .031$, $d = .67$. Although there were no other statistically reliable differences between the other groups (all other $p > .152$), the comparison between the LNIQ and SLI groups yielded a medium effect size ($d = .53$; all other $d < .34$).

The finding of a difference in complete pronoun references between the TL and SLI groups but no difference between the TL and NLI groups was not anticipated. In fact, because the complete NLI group was lower than the SLI group in both IQ and language (see Section 2) and because the complete NLI group tended to use pronouns less than the other

groups, this finding seems extraordinary. The answer to this irregularity seems to lie in the reduction of our samples to exclude all children who failed to generate stories with at least five pronouns. For example, approximately 80% of the children with NLI were excluded for this analysis, whereas only about 65% of the SLI children were excluded. The result of eliminating so many low functioning children in the NLI group was that there were no longer differences between the SLI and NLI groups based on language performance ($p = .982$, $d = .197$). Similarly, there were no longer differences based on this measure between the TL and LNIQ groups ($p = .179$, $d = .593$). Because children who produced shorter stories and fewer pronouns were excluded from our analyses, it is likely that we overestimated the pronoun referencing skills of each of our groups. It appears, however, that this overestimation may have had the greatest impact on the NLI group, making it not significantly different from the SLI and TL groups.

4. Discussion

In this investigation, significant differences based on rate of pronominal use were found between the children with NLI and the female children with TL. The children with SLI could not be differentiated from the children with TL or NLI based on rate of pronominal referencing. In a previous study using children with SLI who seemed to closely resemble the children with SLI in our investigation, Strong and Shaver (1991) found that children with LI used significantly fewer cohesive ties and fewer T-units compared to the children with TL, reflecting an overall decreased referencing rate. Based on this finding, we predicted that both the children with SLI and the more severely impaired children identified as having NLI would use fewer pronouns per C-unit.

Closer examination of the Strong and Shaver study (1991) reveals differences in participant characteristics that may help to explain why our children with SLI did not demonstrate referencing rate differences. All of the children with LI included in the Strong and Shaver study had previously been identified as LI by a school speech-language pathologist and performed at least 1 S.D. below the mean on two or more tests of expressive and/or receptive language. Strong and Shaver's language criterion for LI may have been slightly more liberal than our criterion for SLI (−1.14 S.D. on our language composite); however, it is possible that the children with LI in Strong and Shaver's study were more severely affected than the children in our SLI group. Our study was based on a sample of children who were tested and then diagnosed as LI whether or not they were exhibiting signs of LI in the classroom. Thus, our epidemiologically based sample of children with SLI may have included children with less obvious symptoms of LI than the clinic-referred sample studied by Strong and Shaver. In contrast, our more severely affected LI group, the NLI group, exhibited a lower rate of pronominal referencing than children with TL, as was predicted based on findings of Strong and Shaver.

Contrary to our expectations, the rate of pronominal referencing did not significantly increase with grade for the children with TL or for any of the clinical groups. Thus, it appears that the rate of pronominal referencing is relatively stable from second to fourth grade. However, Strong and Shaver (1991) reported significant increases in the use of cohesion devices for children the same age as ours. This difference may be attributed to

differences in study tasks. In our investigation, we read a pre-written model story to the children and then asked them to generate their own narrative based on an entirely different set of pictures. In the Strong and Shaver study, each participant viewed wordless story pictures while listening to a pre-recorded audio-taped story. The examiner then prompted the participants to retell the story to a listener unfamiliar with the story without picture cues. This retell task demanded a greater amount of the participants' memory than our task did. However, Strong and Shaver chose to use a story retell task to reduce task difficulty by decreasing cognitive processing required in narrative generative tasks and to ensure that they were measuring the child's best performance. Thus, it is possible that regardless of grade level, our generative task was too difficult for our participants.

Two other points of difference between Strong and Shaver's investigation and our own involve the cohesion measures used for study analyses. First, Strong and Shaver (1991) included all types of cohesion (i.e., reference, substitution, ellipsis, lexical, and conjunction), whereas we only investigated pronominal references. Although pronominal referencing seemed to be the most frequent and best cohesion measure for study analyses based on previous reports (e.g., Baltaxe & D'Angiola, 1992; Liles, 1985a; Roth, Spekman, & Fye, 1995), it is possible that one or more of the cohesion measures other than pronominal referencing are more likely to increase with age beyond second grade. Second, we used the rate of referencing per C-unit as the dependent measure. Although Strong and Shaver found differences between their groups of 8- and 10-year-old children for both the total number of references and the total number of C-units, they did not directly analyze the rate of references per C-unit. Based on the means provided by Strong and Shaver, we calculate the overall rates of references per C-unit for the 8- and 10-year-olds to be 3.99 and 4.13, respectively. The differences across grades reported in the Strong and Shaver study may reflect differences across ages in the overall length of narrative retells more than increases in the general tendency to make use of pronouns. If there is a real difference in cohesion use across grades in the Strong and Shaver study, it appears to be rather small.

Perhaps because of their small sample sizes, other studies of pronominal referencing among children with language impairments have not examined differences between sexes. Sex was included in this analysis due to indications of a storytelling advantage for girls over boys (Fey et al., 2004; Moyano & McGillivray, 1988 [cited in Hughes et al., 1997]). In this study, based on rate of pronominal referencing, an advantage was found for the girls; however, this advantage only held true for the TL group. There was not a clear advantage for the girls in the clinical groups. The significant difference of referencing rate found between our NLI and TL groups was driven by the performance of the girls in the TL group, as indicated by our significant Group × Sex interaction. The higher performance by the girls on this measure may be a subtle indication of the girls higher level of language functioning observed more generally in the Fey et al. study on measures such as number of different words, number of C-units, and narrative quality.

Our analysis of the completeness of referential ties indicated that the children in the SLI group included significantly fewer complete references than did the children with TL in their oral narratives. This finding was anticipated based on outcomes from previous investigations (e.g., Baltaxe & D'Angiola, 1992; Liles, Duffy, Merritt, & Purcell, 1995; Paul & Smith, 1993; Strong & Shaver, 1991). We expected that the children with NLI would also demonstrate reduced pronominal referencing, but this was not the case. Previous investigations identified

group differences based on referencing accuracy for children with profiles resembling our SLI group (e.g., Liles et al., 1995; Strong & Shaver, 1991) as well as for children with varying language and cognitive profiles. For example, Baltaxe and D'Angiola (1992) found that their group of children with autism produced significantly fewer complete pronominal references than their language matched group of children with typical development. Paul and Smith's (1993) examination of cohesion adequacy of 4-year-old children revealed significant group differences for children identified as having expressive language delays and children with typical language development.

Contrary to our predictions, the children in the NLI group did not differ significantly from the children with TL. The explanation for this finding became clear when we inspected the constituency of our groups after they were reduced to include only those children with at least five pronouns in their narratives. This led to the elimination of approximately 80% of our lower functioning NLI participants from this analysis. The result was the elimination of the significant language performance differences between the SLI and NLI groups that existed in the full sample. Consequently, although our procedure probably led to a general overestimation of second and fourth grade children's pronoun referencing skills, this overestimation appears to have been greatest for the NLI group. Future studies of pronoun referencing would do well to ensure that children tell longer or more stories that include at least five and, ideally, many more pronouns that then can be examined for referential completeness.

As was found with rate of pronominal referencing, there were no differences in the completeness of pronominal references between the children's narratives produced in the second and fourth grades. Strong and Shaver (1991) found a trend for an increase in complete cohesive ties especially between the 8-year-old children and the older 9- and 10-year-old children in their study. It may be that our participants, particularly those in the TL group, were approaching the ceiling level, even at the second grade. This especially high performance probably reflects an overestimation of these children's abilities, because many children who produced fewer than five pronouns were eliminated from our analysis. Still, based on all the evidence currently available, it appears that any gains made in pronominal referencing from approximately second to fourth grade, even for children with typical language development, are relatively small.

Our analysis of pronominal referencing completeness did not demonstrate performance differences based on Sex. There was no indication of an advantage for girls over boys for this measure, as was the case for referencing rate for the TL group. The inconsistency of the girls' tendency to outperform the boys across subgroups and across referencing measures is consistent with MaCaulay's (1978) position that the advantage in language skills for girls over boys is not as consistent as is often presumed.

4.1. Conclusions

Contrary to previous study findings (especially Liles et al., 1995), our investigation demonstrates that pronominal referencing skills do not clearly differentiate groups of children with typical language development from all groups of children with LI. However, we found that pronominal referencing, whether measured according to rate or

completeness, is a difficult language area for some school-aged children. Thus, although it may not be productive to examine pronominal referencing skills to identify children with LI, this is a potential area of weakness that clinicians should consider when they evaluate and treat school-aged children.

Acknowledgements

This study was supported by a grant from the National Institute on Deafness and Other Communication Disorders (1-P50-DC02726) and center grant #HDO258 from the National Institute on Child Health and Human Development. We also thank Bruce Tomblin, Paula Buckwalter, Marlea O'Brien, Connie Ferguson, Jodi Schwartz, Amy Kundel, and Kelli Stanfield. Portions of the results reported in this manuscript were presented at the 2005 ASHA Convention, Philadelphia.

Appendix A. Continuing education questions

1. Which of the following narrative measures was not studied by Fey et al. (2004)?
   a. number of different words
   b. use of cohesive devices
   c. narrative quality
   d. clausal density
   e. percent of grammatical C-units
2. Which of the following is not a type of cohesion?
   a. reference
   b. substitution
   c. ellipsis
   d. semantic
   e. lexical
3. An example of a complete pronominal reference is:
   a. Ann gave *her* a picture.
   b. Seth and Brad went to a movie.
      He liked it.
   c. Kathy likes puzzles.
      He does a crossword everyday.
   d. Dad wants to make soup.
      He needs to go to the grocery store first.
   e. It was raining outside so *his* brother brought an umbrella.
4. This study found that:
   a. children used a higher rate of referencing in fourth grade than in second grade
   b. children with SLI used pronominal referencing less than children with TL
   c. children with SLI used fewer complete references than children with TL
   d. children with NLI used fewer complete references than children with TL
   e. girls with TL used more complete references than boys with TL

5. One conclusion drawn from this study is:
   a. pronominal referencing is a reliable measure for differentiating children with TL and SLI
   b. children with LI do not have difficulty using pronominal referencing in oral narratives
   c. speech-language pathologists working with children with LI do not need to focus on pronominal referencing difficulties
   d. by fourth grade, children with typical and impaired language abilities master the use of pronominal referencing in their oral narratives
   e. pronominal referencing is an area of weakness for some children with LI

References

Baltaxe, C. A., & D'Angiola, N. (1992). Cohesion in the disclosure interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders, 22*(1), 1–21.

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

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331–361.

Culatta, B., Page, J., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools, 14*, 66–74.

Dunn, L. M. (1981). *Peabody picture vocabulary test—revised*. Circle Pines, MN: American Guidance Service.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 865–878.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, B. J., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language and Hearing Research, 47*(6), 1301–1318.

Gilger, J. W. (1995). Behavioral genetics: Concepts for research and practice in language development and disorders. *Journal of Speech and Hearing Research, 38*, 1126–1142.

Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35*(6), 1303–1315.

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

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

Liles, B. Z. (1985a). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*(1), 123–133.

Liles, B. Z. (1985b). Narrative ability in normal and language disordered children. *Journal of Speech and Hearing Research, 28*, 123–133.

Liles, B. Z. (1985c). Production and comprehension of narrative discourse in normal and language disordered children. *Journal of Communication Disorders, 18*(6), 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*(2), 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*(2), 415–425.

Long, S. H., Fey, M. E., & Channell, R. W. (2002). *Computerized profiling (version 9.4.1)*. Milwaukee, WI: Computerized Profiling.

MaCaulay, R. K. S. (1978). The myth of female superiority in language. *Journal of Child Language, 5*(2), 353–363.

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*(4), 539–552.

Miller, J. F., & Chapman, R. (2000). *Salt: Systematic analysis of language transcripts*. Madison: Language Analysis Laboratory, Waisman Center, University of Wisconsin. [computer software].

Newcomer, P., & Hammill, D. (1988). *Test of language development-2:Primary.* Austin, TX: Pro-Ed.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*(3), 287–313.

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*(3), 592–598.

Purcell, S. L., & Liles, B. Z. (1992). Cohesion repairs in the narratives of normal-language and language-disordered school-age children. *Journal of Speech and Hearing Research, 35*(2), 354–362.

Roth, F. P., Spekman, N. J., & Fye, E. C. (1995). Reference cohesion in the oral narratives of students with learning disabilities and normally achieving students. *Learning Disability Quarterly, 18*(1), 25–40.

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical evaluation of language fundamentals* (3rd ed.). Psychological Corporation.

Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46*(2), 114–122.

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*(1), 95–111.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39*(6), 1284–1294.

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 and Hearing Research, 40*(6), 1245–1260.

Wallace, G., & Hammill, D. (1994). *Comprehensive receptive and expressive vocabulary test*. Austin, TX: Pro-Ed.

Wechsler, D. (1974). *Wechsler intelligence scale for children—Revised*. New York: The Psychological Corporation.