University of Alberta

A Comparison of Three Narrative Language Assessments

by

Allison Lorraine Laing

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Speech Language Pathology

Department of Speech Pathology and Audiology

Edmonton, Alberta
Fall, 2008

Your file Votre référence
ISBN: 978-0-494-47286-6
Our file Notre référence
ISBN: 978-0-494-47286-6

Abstract

Three tests of narrative language, *the Renfrew Bus Story, the Edmonton Narrative Norms Instrument (ENNI)* and *the Test of Narrative Language (TNL)*, were administered to 36 typically developing five year olds in order to establish concurrent validity and determine if the scores on the three measures provide a similar picture of a child's ability or if they differ for the same students (macrostructure analysis). Language complexity measures including Mean Length of Communication Unit (MLCU) and Complexity Index (CI) were also calculated and compared across the three tests (microstructure analysis). Macrostructure analysis revealed no significant differences between test scores. Significant correlations were found only between the TNL and the other two tests, Microstructure analysis indicated that children scored significantly higher on MLCU and CI for tests that included story retell tasks. Clinicians should be cognizant of these differences when assessing narrative language abilities.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Schneider for her continuous support. Throughout my academic career, Dr. Schneider provided me with endless opportunities to challenge myself and has contributed greatly to my professional growth. Her guidance, wisdom and patience contributed greatly to the success of my research. Thank you Dr. Schneider for putting up with my millions of questions, my many thesis drafts, and my many last minute requests. Thank you most of all for believing in me.

I would also like to thank my thesis committee for the time they spent reviewing my thesis, providing insightful comments and supporting me in this endeavor.

This thesis would not have been possible without the help of the following people. Thank you to my fellow graduate students Karen Bylsma, Carrie Timgren and Melissa Harasem for your countless hours of transcription, scoring and number crunching. Also, thank you to the participating day cares, parents and children- without them I wouldn't have had any numbers to crunch!

Last, but not least I would like to thank my family. I would like to thank my parents for supporting me in my academic pursuits and helping me out financially. Also, thank you to my new husband, Dylan, for still wanting to marry me after all this!

Table Of Contents

Introduction 1

Methods 10
Research Design 10
Participants 10
Materials 11
Procedures 15
Outcome Measures 16
Reliability 18
Data Analysis 19

Results 20
Macrostructure Analysis 20
Microstructure Analysis 24

Discussion 28
Macrostructure Analysis 28
Microstructure Analysis 31
Limitations 33
Future Research 34
Clinical Implications 35

Conclusion 36

Bibliography 37

Appendix A 42

List of Tables

Table 1. Summary of Test Structure for the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language 9

Table 2 Demographic Characteristics of Research Participants..... 11

Table 3. Descriptive statistics for standard test scores on the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36) 20

Table 4. Pearson Correlations (and r^2) between scores on the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36) 21

Table 5. Descriptive Statistics for MLCU of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36).. 24

Table 6. Pearson Correlations (and r^2) between MLCU of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36) 24

Table 7. Descriptive Statistics for Complexity Index of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36) 26

Table 8. Pearson Correlations (and r^2) between the Complexity Index of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36) 26

List of Figures

Figure 1: Summary of Children's Performance Classified as Average (A), Below Average (-) or Above Average by the Bus Story, the Edmonton Narrative Norms Instrument, the Test of Narrative Language, the Kaufman Brief Intelligence Test, and the Clinical Evaluation of Language Fundamentals Preschool-Second Edition. 23

Figure 2: Error bars showing Mean Length of Communication Unit with 95% Confidence Intervals for the Bus Story, the Test of Narrative Language and the Edmonton Narrative Norms Instrument 25

Figure 3: Error bars showing mean Complexity Index with 95% Confidence Intervals for the Bus Story, the Test of Narrative Language and the Edmonton Narrative Norms Instrument 27

Introduction

In recent years, the value of assessing narrative language abilities in children has gained recognition (Hughes et. al, 1997) and as a result several tests of narrative language skills have appeared. These tests use different methods to assess narrative language abilities such as story retell tasks, story generation tasks and use of pictures. However, the use of different methods has been shown to affect the quality of the stories children tell (e.g., Merritt & Liles, 1989; Ripich & Griffith, 1988; Schneider, 1996; Schneider & Dubé, 2005). Therefore, it is not yet known whether these tests provide a similar picture of a child's narrative language abilities. The purpose of the current study is to determine whether or not three different tests of narrative language provide similar assessments of a child's narrative abilities.

Narrative language assessments have become useful tools for measuring higher level language abilities in children. A narrative is defined as a time-ordered sequence of events that typically involves a goal directed behavior (Westby, 2005). Producing a narrative requires the ability to integrate and organize separate linguistic units for a specific purpose. Conceptual knowledge of temporal relations, causal relations, and human intentionality (motivation) is therefore required in order to comprehend and produce narratives. A good narrative is cohesive and structured according to a story framework or schema. Both comprehension and production of narratives requires retrieving a mental representation of story structure. This representation must either be held in memory while new information is being added or be modified to accommodate information that does not match (Stein & Glenn, 1979).

The concept of story structure is described by Westby (1995) as a measure of story macrostructure (overall content and organization). Westby (1995) noted that making use of story grammar knowledge (schema that represents a frequent organizational pattern of elements independent of specific content) facilitates recognition of overall themes and provides scaffolding for assimilating information and making inferences. Thus, story grammars facilitate comprehension of narrative text. Stein and Glenn (1979) proposed that story grammar knowledge is knowledge about story structure that adults recognize as essential for producing a good story.

They found that the children in their study appeared to expect a certain pattern of information when listening to a story. Stories that violated this structure were re-organized to fit the expected pattern by changing the order of presented categories, adding new categories, adding logical relations or deleting inconsistent or poorly organized information. The story grammar model developed by Stein and Glenn (1979) and used in this study has been used throughout the literature as a measure of overall story content and organization (e.g., Ripich & Griffith, 1988; Roth & Spekman, 1986; Schneider & Dubé, 2005). The components of the Story Grammar Model are outlined in Appendix A.

In addition to an analysis of overall content and organization (macrostructure), narrative assessments must also include an analysis of the relationships among the parts of a story (microstructure) (Hughes, McGillivray, & Schmidek, 1997). Microstructure analysis can include measures of cohesion, grammatical units and lexical diversity (Hughes, et al., 1997). Cohesion analysis examines the linguistic features that link the parts of a narrative together. Stein and Glenn (1979) describe three independent types of relations used to connect episodes within a story: *THEN* relations connect episodes occurring in succession, *AND* relations connect simultaneously occurring episodes, and *CAUSE* relations connect episodes with a direct causal link between them. These relations correspond to Hughes et al. (1997) category of conjunctive markers. Other cohesive markers may include referential expressions, lexical associations such as synonyms, and substitutions (Hughes, et al., 1997). Additional microstructural elements may include an analysis of subordinate clauses, T-units (main clause and all subordinate clauses attached to it), and lexical diversity. The concept of lexical diversity in narratives can be related to a literate language style. Literate language is generally associated with written language and compared to oral language is decontextualized, more formal and more complex (contains more expansions). Oral narratives, however, contain many of these literate language features and are therefore thought to bridge the gap between oral and literate language (Westby, 2005).

Narrative language abilities have in fact been found to predict later linguistic skills and academic achievement. Bishop and Edmundson (1987) used a number of

language measures on 4-year-olds with language impairments in an attempt to distinguish between transient and persistent impairment. Although their results supported the idea that many preschool language impaired children make progress and eventually catch up with their peers, it was found that certain factors seemed to predict later patterns of disorder. They concluded that good outcomes in these children were more likely in the absence of nonverbal impairments and when a smaller range of functions were impaired. The best predictor of language outcome in this study, however, was the ability to retell a simple story as measured using *The Bus Story: A test of continuous speech* (Renfrew, 1969). Poor performance on this measure was only partially explained by productions containing few words or unintelligible utterances. Analysis of children's responses indicated that they had difficulty with story sequencing and relating the pictures to one another, despite the presence of the stimulus items in the re-tell condition. Fazio and Naremore (1996) found similar results in their longitudinal study of children from poverty at risk for specific language impairment. The results of their study indicated that story-retelling was the single best predictor of academic status in kindergarten children receiving academic remediation, as compared to standardized-test discrepancy scores, rote-memory ability and invented-morpheme learning.

The previously mentioned study by Bishop & Edmundson (1987) demonstrated that in addition to predicting later achievement, narrative language abilities can help distinguish between typically developing children and those with language impairment. Additional research by Paul and Smith (1993) used *The Bus Story* (Renfrew, 1977) to analyze narrative language abilities in normally developing children, children identified with slow expressive language development, and children with late developing language, all at 4 years of age. They found that the 4-year-olds with slow expressive language development had significantly lower scores compared to normally developing peers on all language measures taken on the narrative sample, including information score (amount of story content recalled), mean length of utterance per T-unit, cohesive adequacy, and lexical diversity. Further analysis of the data led the authors to conclude that these deficits were not a

result of syntactic and grammar difficulties, but actually involved difficulties with higher level aspects of language such as formulation, organization and retrieval.

Feagans and Short (1984) found differences in narrative abilities between normally developing children and children with reading disabilities. This study involved a story retell task based on a script-like narrative and comprehension of the story was assured prior to recall. The difficulty in paraphrasing story narratives by children with reading disabilities was evident by the fact that they produced fewer action units, fewer complex sentences and more non-referential pronouns (pronouns with no clear noun reference) in their story productions.

Differences in narrative language abilities have also been found between children with learning disabilities and normally achieving children. Roth and Spekman (1986) used a modified version of the story grammar model (Stein & Glenn, 1979) to analyze spontaneously generated oral stories in children 8-14 years old. The results indicated that the stories produced by children with learning disabilities were shorter and contained fewer units of meaning (propositions). Differences were also found in the number of complete episodes-those containing an initiating event, attempt and outcome. Children with learning disabilities produced significantly fewer complete episodes and were less likely than normally achieving peers to include Attempts and Responses and Minor Setting (story context) information within their stories. Children with learning disabilities, however, produced more propositions containing Initiating Event information. In terms of inter-episodic relations, children with learning disabilities were less likely to use complex temporal relations such as causality and simultaneity to connect their episodes, as evidenced by significantly lower use of CAUSE and AND relations in their stories.

The results of the previous studies support the idea that narrative language ability is a useful set of skills that should be included within any child language assessment. In addition to measuring higher level language abilities, predicting academic achievement and distinguishing between normally developing children and those with language impairment, narrative assessment provides a simple, appealing method to obtain a language sample in a natural yet constrained context. However, as evident from the previous studies, a variety of stimuli and elicitation procedures, both

formal and informal, have been used to assess narrative abilities and it has been found that differences in elicitation techniques have an effect on the quality of narratives produced. Merritt and Liles (1989) compared story retelling and story generation tasks in children with language disorders and typically developing children 9-11 years of age. In the story generation task, children were provided with the beginning of a story and were asked to complete it. They found that both groups produced significantly more clauses, story components (story grammar units) and complete episodes in the story retelling tasks. However, the results also indicated that regardless of the task, both groups produced stories consistent with a structure of story grammar and the authors concluded that children approached each task with a similar organizational plan. The authors went on to discuss the benefit of using story retelling tasks over story generation tasks in a clinical setting in order to obtain a more representative sample of a child's language. They also found that samples obtained from story retelling tasks can be scored more reliably because of the availability of a story model for comparison. Story retelling tasks are also thought to be better for evaluating story comprehension abilities for the same reason (Merritt & Liles, 1989).

Ripich and Griffith (1988) also compared story retelling of an oral story (without pictures) and story generation abilities in children with and without learning disabilities. The story generation task required children to tell a story from a picture sequence. Their results are similar to those of Merritt and Liles (1989) in that they found children with learning disabilities ages 7 through 12 were able to organize narratives according to a story structure. They did find, however, that children with learning disabilities omitted more events across the different stories used and that the omissions were greatest in the story generation conditions. The authors speculated that it is more difficult to generate a story according to a set of pictures than it is to retell a story and the two tasks provide different information. Analysis of story retellings provides information regarding the salience of events, use of propositions and cohesive markers, and memory for story structure. Formulation of stories without a model, on the other hand, provides information on the ability to organize and produce cohesive text, which may provide a more realistic picture of a child's

narrative language abilities (Ripich & Griffith, 1988). The research by Ripich and Griffith (1988), however, raised the question of stimulus presentation. They found that the presence of pictures reduced the amount of information provided in both retelling and generation tasks for younger children and those with learning disabilities. The authors speculated that the presence of pictures creates the assumption of shared knowledge between the story teller and the listener, which has been found to decrease the amount of story content provided by children (Liles, 1985). Due to the fact that the oral and picture story tasks were not structured the same, however, these results need to be interpreted carefully.

Schneider (1996) investigated story presentation effects further by comparing story grammar units in a variety of retell and generation conditions for children with language impairments, aged 5-10. Each child was presented with four conditions: picture-only (child told the story from pictures with no oral model), oral followed by pictures (child had access to pictures for retell after hearing the story), oral with pictures (child had access to pictures during presentation of story and during recall) and oral-only condition (pictures were not present during the presentation or recall of the story). The results show a significant effect of presentation methods on narratives produced. Retellings from the pictures-only condition had fewer story grammar units and fewer complete episodes than the oral-only condition. However, the children in this study produced more original units (both relevant and irrelevant to the stimuli) and exhibited fewer indications of formulation problems (i.e. fewer mazes, false starts, repetitions, pauses and fillers) in the picture-only condition. While comparisons between the other conditions did not necessarily reach significance, the data showed a trend indicating that greater quantity and quality is produced in oral-only conditions. One possible explanation for these findings is that pictures actually divert attention from the oral presentation of stories and therefore affect comprehension of the story (Schneider, 1996).

Similar results were found in a follow-up study conducted with normally developing children in kindergarten and Grade 2 (Schneider & Dubé, 2005). Developmental differences were found in that kindergarten children recalled more information when oral presentation was combined with pictures as compared to the

picture-only condition, whereas recall for children in Grade 2 was better in both oral-only and combined conditions than in the picture-only condition. In other words, children at both grade levels produced more information in retellings based on oral presentation than on pictures alone, although younger children also seem to benefit from the presence of pictures. Differences in processing demands between story retelling (gist recall) and formulation (independent performance) might provide one explanation for these findings, but more research is needed in this area. What can be concluded from the above research is the importance of including a variety of narrative tasks and stimuli presentation methods when assessing a child's narrative language abilities.

In order for a narrative language assessment to provide an adequate picture of narrative language ability, it must include an analysis of both macro-and-microstructure elements across a variety of different tasks (retell and formulation) using a variety of presentation stimuli (oral, visual, or combination). Whether or not the tests are designed according to a model of story-telling should also be considered. Current tests of narrative language, however, vary with respect to how these factors are addressed. For example, *The Renfrew Bus Story* (Cowley & Glasgow, 1994) is an oral retell screening tool, normed on children aged 3;6-6;11. The test is based on The Bus Story (Renfrew, 1977), which was originally normed on children in the United Kingdom. The Renfrew Bus Story uses the same story but with pictures drawn for a North American audience; it was normed on children in the United States. The Renfrew Bus Story consists of a short story read aloud to the child while he/she follows along with a series of twelve pictures. The child is then asked to retell the story from the pictures and his/her response is recorded and transcribed. Overall content memory, a type of macrostructure analysis, is determined by matching the child's story against information selected from the originally presented story. The child receives a score based on the amount of information that was included and credit is given for responses that occur in proper sequence and responses that contain appropriate semantic meaning and appropriate pronoun referents. Therefore, the information score for The Renfrew Bus Story contains both macro- and microstructure elements. The test also yields a score for sentence length as a measure

of language complexity. A model of storytelling was not used in the development of the test stimuli or as a guide to scoring. As mentioned previously, the original Bus Story is considered a useful tool in differentiating between normal language peers and those with expressive language delay (Paul & Smith, 1993) and predicting persistent language impairment (Bishop & Edmundson, 1987).

The *Edmonton Narrative Norms Instrument* (ENNI) (Schneider, Dubé, & Hayward, 2004), on the other hand, was designed according to the Story Grammar Model (Stein & Glenn, 1979). Normed on children 4;0-9;11, the ENNI tests narrative ability through a story formulation task using sets (sequences) of picture stimuli that vary in range of complexity depending on the number of episodes and characters present. The Story Grammar Model guides the analysis of the stories (macrostructure) and children receive credit for the number of story grammar units provided. Microstructure analysis is completed by evaluating First Mentions as a measure of cohesion. First Mentions are considered a measure of a child's ability to use linguistic forms (referring expressions) to refer to objects, events, concepts, and so forth for the first time. Language complexity measures such as mean length of communication unit, complexity index, number of different words are also included within the ENNI. A recent investigation into the use of the ENNI to evaluate storytelling showed that developmental trends exist for the number of units included in children's stories (Schneider, Hayward, & Dubé, 2006).

Another standardized tool for narrative language assessment, *The Test of Narrative Language* (TNL) (Gillam & Pearson, 2004), considers both comprehension and production of narrative language. It is normed on children aged 5;0-11;11 and assesses both story retell and formulation using three different presentation conditions: no picture cues, sequenced picture cues and single picture cues. Analysis of the story retell task is similar to that used to obtain an information score in The Bus Story in that the child's response is matched to information in the original story that is considered important. In addition to story content, analysis of the story formulation tasks provide a general score of narrative production ability that includes measures of grammar, relationships (temporal/causal), story elements, vocabulary and creativity. Although the TNL was designed with reference to a model of narrative language, the

scoring/analysis is not guided by this model. Instead, items were chosen based on their ability to discriminate between children with and without language impairment.

The procedures used to assess narrative language abilities, including the standardized tests mentioned above (The Bus Story, The ENNI and The TNL), vary greatly in the theory behind their construction, in the development, complexity and presentation of the stimuli used to elicit narrative language, and in the models used to guide and interpret their data. Previous research has shown that these differences can affect a child's performance. In addition, these tests are all normed on populations of children from different geographic regions, making it difficult to generalize results across the three tests. Therefore, the purpose of the current study is to compare these three standardized tests of narrative language in order to answer the following questions: 1) Is there evidence of concurrent validity between the three tests? 2) Do children perform differently across the three tests? Table 1 summarizes the differences in structure between the three tests used in this study.

Table 1. Summary of Test Structure for the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language

Test	Type of task	Stimulus
1. The Bus Story	Retell	Picture sequence
2. The ENNI		
a. Story A1	Generation	5-Picture sequence
b. Story A3	Generation	13-Picture sequence
3. The TNL		
a. Task One	Retell	No pictures (oral only)
b. Task Two	Generation	Picture sequence
c. Task Three	Generation	Single picture

Examining concurrent validity between the three tests will help determine if they measure similar types of skills, which may have important implications for test interpretation by clinicians. Comparing the different tests will provide insight on how a child performs across a variety of contexts, which is necessary in order to set appropriate goals and guide intervention for children with language difficulties.

Methods

Research Design

To determine concurrent validity, relationships between the three tests were examined using a within-subject, correlation design. Analyses of variance were then used to compare each child's performance between the three tests.

Participants

Data were collected from 36 children (22 girls and 14 boys) from various kindergarten and preschool programs throughout the city of Edmonton, Alberta. The average age of the participants was 62.7 months (SD 3.89) and ranged from 57 to 73 months. Twenty-seven of the children were Caucasian, 5 were Chinese and 4 were of mixed ethnicity. Although 5 children came from homes that spoke a language in addition to English, all children spoke English as their first language. Information on the number of years of maternal education was collected for 31 of the 36 children. This was used as a measure of socio-economic status and ranged from 12 to 22 years with an average of 15.8 years. Parents of the 5 remaining children did not provide information regarding maternal education.

All of the children in the study were enrolled in programs for typically developing children and none of the children had been identified as having any developmental or language difficulties. Children's scores on the *Kaufman Brief Intelligence Test* (KBIT) (Kaufman & Kaufman, 2004), a measure of verbal and non-verbal intelligence, ranged from 80-132 with a mean of 109.14 and a standard deviation of 9.13. Seven children received scores greater than one standard deviation above the mean, indicating above average performance. Only one child received a score that was more than one standard deviation below the mean (80), indicating below average performance. Objective measures of language ability were obtained using the *Clinical Evaluation of Language Fundamentals-Preschool, Second Edition* (CELF Preschool-2) (Wiig, Secord & Semel, 2004). Scores on this measure ranged from 86-127 with a mean of 110.42 and a standard deviation of 10.17. Fourteen children scored more than one standard deviation above the mean, indicating above average performance. All remaining children performed within the average range (85-

115). Based on these scores, the children in the sample were considered to have average or above average language abilities. Table 2 summarizes the participant characteristics.

Table 2 Demographic Characteristics of Research Participants

Characteristic	Mean	Standard Deviation	Minimum	Maximum
1. Age (months) n=36	62.69	3.89	57	73
2. Maternal Education (years) n=31	15.84	2.83	12	22
3. KBIT scores n=36	109.14	9.13	80	132
Above Average n=7				
Average n=28				
Below Average n=1				
4. CELF P-2 scores n=36	110.42	10.17	86	127
Above Average n=14				
Average n=22				

Materials

Narrative Language Assessments

The following narrative assessments were used to obtain a picture of a child's storytelling abilities.

The Renfrew Bus Story (the Bus Story)

The Renfrew Bus Story (Cowley & Glasgow, 1994), the American version of The Bus Story (Cowley & Glasgow, 1994) was used in this study. The normative sample includes 418 normally developing children aged 3;6-6;11 from the Mid-Atlantic States, Florida and Illinois and includes a range of socio-economic levels. Pre- and post-test reliability coefficients, which vary from 0.5825 to 0.7918, are provided for the Information, Sentence Length and Complexity scores. Inter-tester reliability was also determined for the information score (0.92, 0.72, 0.70), sentence length (0.79, 0.83, 0.81) and complexity (0.22, 0.60, 0.33), indicating that caution is warranted when scoring for complexity. The British version of The Bus Story has been shown to be an effective tool in the identification of children with language delays (Bishop & Edmundson, 1987). Correlations between the British and American version were calculated for the information score (0.98) and sentence length (1.00) as

evidence of concurrent validity. Children's performance on these measures is expressed as a percentile and as a standard score with a mean of 100 and a standard deviation of 15. Only the Information score was used in this study.

The Edmonton Narrative Norms Instrument (the ENNI)

The ENNI (Schneider, et al., 2004) includes local normative data on 377 children 4-9 years of age from schools throughout Edmonton, Alberta. The normative sample consists of both typically developing children and children with specific language impairment and includes a range of socioeconomic levels. Word by word transcription reliability was calculated to be 0.97 between transcribers for the normative data. Schneider, Hayward, & Dubé, (2006a) also found good inter-rater reliability for the Story Grammar measure as indicated by a Cohen's kappa of 0.92 (significant at $p < 0.001$). A Cohen's kappa of 0.85 was computed as a measure of First Mentions reliability (Schneider & Hayward, 2006b). Additional research on inter-rater reliability for Story Grammar using untrained clinicians as raters found intraclass correlations of 0.92 and 0.96 for stories A1 and A3 respectively (Beswick, 2008). The Story Grammar scores of the ENNI were also found to correlate at a statistically significant level to scores from the CELF-P and CELF-3, indicating evidence of concurrent validity (Schneider, et al., 2006a). In addition, trend analyses showed an increase in story grammar scores with age until 7 years for the simple story and 8 years for the complex story, as well as differences between age groups (except for age 9). This shows evidence of construct validity when combined with the fact that ENNI scores discriminated between typically developing children and children with specific language impairment across the age groups. Using a set of measures derived from the ENNI, discriminant analysis revealed sensitivity of 80-94% and specificity from 94-100% within each age group, indicating good levels of accuracy in discrimination (Schneider, et al., 2006b). Normative data is included for both micro- and macrostructure language analyses including story information (Story Grammar units), referring expressions (First Mentions), Mean Length of Communication Unit, syntactic complexity (Complexity Index), Total Number of Words and Number of Different Words. Children's performance on these measures

is expressed as a z-score and a standard score with a mean of 10 and a standard deviation of 3.

Test of Narrative Language (the TNL)

The TNL was normed on 1,059 children ages 5-12 in 20 states across the United States. The normative sample is comprised of a range of socioeconomic levels, ethnicity and geographic areas and includes children with a range of abilities, such as speech and language disorders, learning disabilities, and other exceptionalities. The TNL provides three composite scores as measures of narrative language ability: narrative comprehension, oral narration and narrative language ability. Children's performance on these measures is expressed as percentile ranks, age equivalents and standard scores. Test-retest reliability was investigated using a separate sample of children with language disorders and was found to range between 0.81-0.85 for the different measures. Point to point agreement between trained and untrained scorers was expressed as a percentage of agreement between the scorers and was found to be 94% for narrative comprehension and 90% for oral narration scores. Inter-rater reliability using Cohen's kappa, however, indicated a wide range of coefficients (0.03-1.00). Additional information is also provided regarding the validity of the TNL, although the authors state that more research is needed. The TNL was found to correlate at a statistically significant level ($p<0.001$) with the Spoken Language Quotient of the *Test of Language Development-Primary: Third Edition* (TOLD-P:3). Coefficients ranged from 0.78 to 0.82 for the different TNL measures. Comparison to language sample measures also indicated statistically significant ($p<0.001$) correlations ranging in magnitude from 0.39-0.79. Taken together, these measures provide support for concurrent validity. Evidence of construct-identification validity was also provided. Statistically significant ($p<0.0001$) correlations were found when comparing the narrative comprehension and oral narration measures with age, which is consistent with the developmental nature of narrative language ability. In addition, the calculated values for sensitivity (0.92), specificity (0.87) and positive predictive value (0.88) provide evidence of value in using this test to identify children with language disorders.

General Measures of Language Ability and Intelligence

The following tests were used as objective measures to describe and categorize the participants in terms of their language abilities and verbal and non-verbal intelligence:

The Clinical Evaluation of Language Fundamentals Preschool-Second Edition (CELF Preschool-2)

The CELF Preschool-2 is an assessment tool used to identify, diagnose and evaluate language deficits in children 3 to 6 years of age. It outlines an assessment process with four levels that allows the examiner to identify if a language disorder exists, describe the nature of the disorder, evaluate classroom and literacy fundamentals and/or evaluate language and communication in context. As the focus of this study was on typically developing children, the CELF Preschool-2 was administered to describe the language abilities of the sample. Therefore, only Level One (determining if there is a disorder) of the assessment process was completed. This resulted in a Core Language score, which was found to be the best discriminator between performance of typically developing children and children with language disorders. The Core Language Score is a standard score with a mean of 100 and a standard deviation of 15. It is derived from a combination of scores received on three subtests: Sentence Structure (the child points to pictures in response to verbal directions), Word Structure (the child completes a sentence with the targeted structure(s)), and Expressive Vocabulary (the child identifies objects, people or activities portrayed in pictures). Children scoring within one standard deviation of the mean (85 to 115) were considered to have average language abilities.

Kaufman Brief Intelligence Test-Second Edition (KBIT)

The KBIT is an assessment tool used as a measure of verbal and nonverbal intelligence for individuals ranging from 4 to 90 years of age. It assesses a range of skills including word knowledge, range of general information, verbal concept formation, reasoning ability and solving problems. The KBIT yields a Verbal score, a Nonverbal score and an overall IQ Composite score, all with a mean of 100 and a

standard deviation of 15. It consists of three subtests: Verbal Knowledge (the child points to the picture that shows the meaning of a word or answers a question that was presented verbally), Matrices (the child is required to choose the picture that complements the stimulus picture, completes a visual analogy or solves a matrix), and Riddles (the child points to a picture or supplies a single word that answers the riddle). For the purposes of this study, the KBIT was used as an objective measure to further describe the sample. Children scoring within one standard deviation of the mean (85-115) were considered to have typical verbal and non-verbal intelligence for their age range.

Procedures

The three tests of narrative language (the Bus Story, the ENNI and the TNL) as well as the KBIT and the CELF Preschool-2 were administered according to the standardized procedures outlined in their respective manuals. Testing for each child was conducted by the same examiner over a period of one week, with 2-3 sessions ranging from 30 to 45 minutes in length to minimize fatigue. The children were tested at the same time of day on each visit in a quiet area to reduce distractions. Testing was conducted in a variety of settings including kindergarten classrooms, day cares and at children's homes, but the environment remained constant for each individual across the testing sessions.

The three tests of narrative language were always administered before the KBIT and the CELF Preschool-2, but all narrative tests were administered in 6 different counter-balanced orders to control for sequence and practice effects. To determine if the order of testing had an effect on performance, a three-by-six (Test by Order) mixed analysis of variance was conducted where Test was a within-subjects factor and Order was a between-subjects factor. Results indicated that the main effect of Order for test scores across the three tests of narrative language was not significant ($F_{(5,30)}$=0.851, p=0.53). There was also no significant main effect of Order on MLCU and CI for the three tests ($F_{(5,30)}$=0.584, p=0.71; $F_{(5,30)}$=0.809, p=0.55, respectively). Therefore, the groups were combined for further analysis.

Each session was audio-recorded using a mini-disc recorder. The responses for the three tests of narrative language were then transcribed and scored according to the standardized procedures outlined in the testing manuals. Raw scores were converted to standard scores according to the individual test specifications. Systematic Analysis of Language Transcripts (SALT) software was used to facilitate transcription and scoring and also allowed for additional analyses.

Outcome Measures

Outcome measures were obtained for both macrostructure analysis (narrative language ability) and microstructure analysis (language complexity).

Macrostructure Analysis (Narrative Language Ability)

Each test used slightly different measures to assess narrative language ability, but the underlying constructs were considered to be related. The following scores were used as outcome measures of narrative language ability:

The Renfrew Bus Story

Story Information-An analysis of story information provides a measure of the amount of information recalled (content memory) in a story retell task. Credit is given for responses that contained appropriate semantic meaning, occurred in proper sequence and contained appropriate pronoun referents. The child's responses are matched against content within the originally presented story using the test's scoring criteria.

The Edmonton Narrative Norms Instrument

Story Grammar Units-An analysis of story grammar units provides a measure of story content based on the information/elements that are thought to be characteristic of a 'good' story. Credit is given for the inclusion of specific story grammar units, which included Setting, Initiating Event, Internal Response, Internal Plan, Attempt, Outcome, and Reaction. The test provides criteria for identifying each unit within the two stories (A1 and A3). A score is derived for each story separately.

Test of Narrative Language

Oral Narration-The TNL provides a composite score that includes a variety of story-telling and linguistic elements. Credit is given for story content/information, temporal and causal relations, grammar, vocabulary, story coherence, and story elements.

Microstructure Analysis (Language Complexity Measures)

A narrative sample is essentially a language sample, which means it can be analyzed for language complexity measures. Each test contains some kind of language complexity measure (i.e., sentence length, first mentions, etc), but they are different for each test and embedded within the total score for the TNL. This makes it difficult to compare across the three tests. For the purposes of this study, two common analysis measures, the syntactic Complexity Index and the Mean Length of Communication Unit, were used to determine if one would obtain similar information when using the stories as language samples. Since a larger number of utterances is desirable for language analysis (Shipley & McAfee, 2004), all six stories of the ENNI and the three tasks of the TNL that involve narration by the child were combined to calculate the language complexity measures.

Complexity Index (CI) - Each set of stories was examined for the number of independent and dependent clauses. Each dependent clause was coded and counted in SALT. The CI was then calculated by dividing the number of independent and dependent clauses by the number of independent clauses.

Mean Length of Communication Unit (MLCU) - MLCU was calculated using SALT. It was computed by dividing the number of words in each transcript by the total number of utterances in the analysis set. MLCU does not include interrupted or abandoned utterances or utterances containing unintelligible portions.

Reliability

Inter-rater reliability was calculated for each of the measures used in this study. To determine reliability for the test scores, a graduate speech-language pathology student scored the three narrative tests from transcripts provided by the primary investigator and point-by-point reliability (percent agreement) was calculated. For the Bus Story and the TNL, percent agreement was calculated on all items by comparing the number of points given for each item. Agreement was established if the same number of points was awarded for each item. Disagreement occurred when a different number of points were rewarded for a particular item. For the ENNI, the Story Grammar units assigned by each scorer were compared. Scoring agreement was said to occur when the same Story Grammar units were identified for a particular child. Disagreement occurred when scorers identified different Story Grammar units or when one scorer assigned a Story Grammar unit that was not identified by the other scorer. Reliability for test scores was calculated on 20% of the data (7 children) that were randomly selected from the sample.

For the Bus Story, percent agreement for each child ranged from 84 to 97%. Overall percent agreement was 89%. Percent agreement for the TNL was calculated for each individual story and for all three stories across each child. Percent agreement for the Macdonald's story was 94%, for the Late story was 86% and for the Alien story was 72%. Percent agreement for each child across the three stories ranged from 79 to 89%. Overall percent agreement for the TNL was 86%. For the ENNI, percent agreement was calculated for each story. For Story A1, percent agreement for each child ranged from 67-100%. Overall percent agreement was 91%. For Story A3, percent agreement for each child ranged from 46 to 94% with an overall agreement of 88%.

Pearson's Correlation Coefficients were calculated as a measure of inter-rater reliability for the language complexity measures (MLCU and CI). A different graduate speech-language pathology student assisted with reliability for CI by identifying independent and dependent clauses in transcripts provided by the primary investigator. There was a significant positive correlation between scorers for the Bus Story, the ENNI and the TNL ($r=0.929$, $p<0.05$; $r=0.920$, $p<0.05$, 0.984, $p<0.01$,

respectively). This same graduate student also transcribed 15% of the data (5 children) from the audio recordings to determine MLCU reliability. Significant positive correlations ($p<0.01$) were found between scorers for the Bus Story, the ENNI and the TNL ($r=0.917$, $r=0.968$, $r=0.981$, respectively).

Data Analysis

Macrostructure Analysis

Pearson's correlation coefficients were calculated to determine the relationship between the test scores for the three narrative assessments. Each child's performance was also compared using analyses of variance to determine whether they performed significantly differently on any of the three tests. When significance was obtained, post hoc analysis was conducted using paired sample *t* tests. To adjust for multiple comparisons, a Bonferroni procedure was used in which alpha was divided by the number of tests in each set of comparisons (3) to yield an adjusted alpha of 0.017. Before the test score means could be compared, however, the Bus Story scores had to be transformed to a scale with a mean of 10 and a standard deviation of 3.

Microstructure Analysis

Pearson's correlation coefficients were also calculated to examine the relationship between the language complexity measures for the three tests. Differences in performance were examined using analyses of variance and significant results were subjected to post hoc analysis using paired sample *t* tests with a Bonferroni adjusted alpha level of 0.017.

Results

Macrostructure Analysis

Correlations

Descriptive information on the three tests is provided in Table 3. The ENNI consists of a total of six stories, which differ in the number of episodes and number of characters. Story Grammar scores are provided for two out of the six stories (A1 and A3). Since a composite score of the two stories is not available, both stories were included in the analysis. Standard scores for the Bus Story have a mean of 100 and a standard deviation of 15, whereas standard scores for the ENNI and the TNL have a mean of 10 and a standard deviation of 3. The Information standard score for the Bus Story was converted to the same scale as the ENNI and the TNL, based on the mean and standard deviation of the Bus Story normative sample. This resulted in a new mean and standard deviation for the study sample (10.63 and 1.66, respectively).

Table 3. Descriptive statistics for standard test scores on the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Test	*M*	*SD*	Minimum	Maximum	Skewness Ratio
1. Bus Story	106.33	13.00	76	127	-1.22
(transformed)	(10.63)	(1.66)			
2. ENNI_A1	9.92	2.78	1	15	-2.52
3. ENNI_A3	10.61	3.25	0	15	-3.69
4. TNL	10.67	2.82	6	18	1.91

The Skewness Ratio (skewness/standard error) indicated that the scores for the ENNI were not normally distributed due to the existence of an outlier (i.e. one child received a markedly lower score on the ENNI compared to the other children). Using non-parametric tests or removing the outlier, however, did not significantly alter the results. Therefore, parametric tests that are considered robust to the assumption of normality were used to analyze the data.

Table 4 shows that three of the six pairs of variables were significantly correlated. The Oral Narration score of the TNL was significantly correlated with the Story Grammar score of story A1 of the ENNI. There was also a significant positive

correlation between the two ENNI subtests and between the Information score of the Bus Story and the Oral Narration score of the TNL. Coefficients of determination (r^2) ranged from approximately 11-23 %, indicating a medium effect size (Corty, 2007).

Table 4. Pearson Correlations (and r^2) between scores on the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Variable	1	2	3	4
1. Information Score	--	.075	.039	.335* (0.112)
2. A1 Story Grammar	--	--	.483** (0.233)	.368* (0.135)
3. A3 Story Grammar	--	--	--	.141
4. Oral Narration	--	--	--	--

*p<0.05, **p<0.01

Differences

An analysis of variance was conducted to determine if the children's performance was different on any of the tests. The results indicated that test scores did not significantly vary across the three tests ($F_{(3,105)}$=0.955, p= 0.417).

The results of the macrostructure (narrative language ability) analysis show that the TNL correlates significantly with the two other narrative language assessment tools, which do not correlate with one another. These results show evidence of concurrent validity for narrative language skills, at least between the Bus Story and the TNL, each of which contain a story retell component. Evidence of similarity was also found between story A1 of the ENNI and the TNL, each of which contain a story generation component. The results also show that test scores do not significantly differ from one another, indicating that children as a group are performing similarly on the three tests. Taken together, these results suggest that the TNL measures similar types of skills as the ENNI and the Bus Story and children's storytelling abilities yield similar results across the three tests.

To further explore the patterns of results, children's performance on each test score was classified as average, below average or above average according to the

manual specifications. Children scoring within one standard deviation of the mean were considered average in comparison to their same age peers. As illustrated in Figure 1, only 11 of the 36 children (31%) were categorized the same across all conditions (test scores). The biggest concern for interpretation of test scores, however, is when a child is classified as below average on one test and above average on another. This occurs for only 3 of the 36 children (8%). Caution for interpretation of results should also be taken when a child is classified as average on one test and below average on another, which occurs for 6 out of 36 (17%) of the children in this study.

Figure 1: Summary of Children's Performance Classified as Average (A), Below Average (-) or Above Average by the Bus Story, the Edmonton Narrative Norms Instrument, the Test of Narrative Language, the Kaufman Brief Intelligence Test, and the Clinical Evaluation of Language Fundamentals Preschool-Second Edition.

ID	BUS	A1	A3	TNL
1	A	A	A	+
2	A	+	+	+
3	A	+	+	+
4	A	A	+	+
5	+	A	A	A
6	+	A	A	+
7	A	A	A	A
8	A	-	A	-
9	-	A	A	A
10	A	A	A	A
12	A	A	A	A
13	A	A	A	A
14	**+**	**-**	**-**	**A**
15	+	A	A	A
16	+	A	A	+
17	+	A	+	A
18	-	A	-	A
19	+	A	A	A
20	**+**	**A**	**-**	**+**
21	A	A	A	A
22	A	A	A	A
23	+	A	A	A
25	A	A	A	A
26	A	A	A	-
27	**+**	**-**	**A**	**A**
28	A	A	A	A
29	A	A	A	+
30	A	-	A	A
31	+	+	A	+
32	A	A	+	A
34	A	A	A	A
35	A	-	A	-
36	A	A	A	A
37	A	A	+	A
38	A	A	A	A
39	+	A	A	A

Microstructure Analysis

Correlations

Mean Length of Communication Unit

Descriptive information on the three tests is included in Table 5. Following the procedures outlined in the ENNI manual for analysis of language complexity measures, all six stories were combined as a language sample for analysis. Similarly, for the purposes of this study, language complexity measures were analyzed across all three TNL stories.

Table 5. Descriptive Statistics for MLCU of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Test	*M*	*SD*	Minimum	Maximum
1. Bus Story	7.70	1.27	3.60	10.21
2. ENNI	6.98	1.18	3.86	9.47
3. TNL	7.62	1.29	5.63	10.83

Table 6 shows that only one of the three variables was significantly correlated. There was a significant positive correlation between the Mean Length of Communication Unit of the ENNI and the TNL. MLCU of the ENNI accounts for approximately 12% of the variance in MLCU of the TNL, indicating a medium effect size (Corty, 2007).

Table 6. Pearson Correlations (and r^2) between MLCU of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Variable	1	2	3
1. Bus Story	--	.185	-.006
2. ENNI	--	--	.399* (0.159)
3. TNL	--	--	--

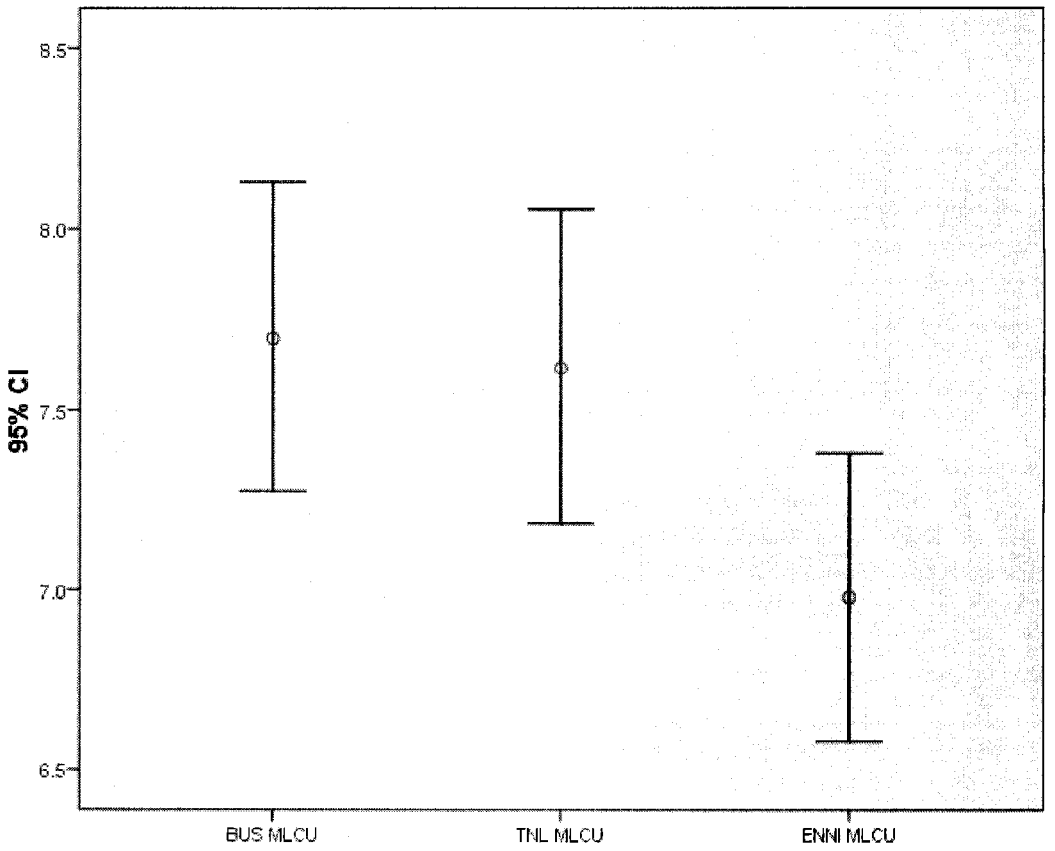
*p<0.05, **p<0.01

Differences

An analysis of variance found a significant effect for MLCU ($F_{(2,70)}$=4.446, p=0.015, *partial eta squared*=0.113). There was also a significant linear trend

($F_{(1,35)}$=7.651, p=0.009, *partial eta squared*=0.179) over the mean values for MLCU, illustrated in Figure 2. MLCU was longest for the Bus Story, shorter for the TNL and shortest for the ENNI. Paired sample t tests with Bonferroni correction for multiple comparisons were used for post hoc analysis. Results indicated that the mean MLCU of the ENNI was significantly lower than the mean MLCU of the Bus Story and the TNL (t=2.766, p=0.009, d=0.589 and t=2.825, p=.008, d=0.517, respectively). Cohen's d values indicate moderate effect sizes. No differences were found between the mean MLCU of the Bus Story and the TNL (t=0.276, p=0.784). The results show that while the ENNI correlated with the TNL on this measure, it differed from the other two tests in that the mean MLCU was significantly lower.

Figure 2: Error bars showing Mean Length of Communication Unit with 95% Confidence Intervals for the Bus Story, the Test of Narrative Language and the Edmonton Narrative Norms Instrument



Complexity Index

Descriptive information on the Complexity Index measure for the three tests is included in Table 7.

Table 7. Descriptive Statistics for Complexity Index of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Test	*M*	*SD*	Minimum	Maximum
1. Bus Story	1.49	0.21	1.00	2.07
2. ENNI	1.28	0.14	1.05	1.66
3. TNL	1.35	0.19	1.00	1.89

Correlations

Table 8 shows that only one of the three variables was significantly correlated. There was a significant positive correlation between the Complexity Index of the ENNI and the TNL. This variable accounts for approximately 21% of the variance, indicating a medium effect size (Corty, 2007).

Table 8. Pearson Correlations (and r^2) between the Complexity Index of the Bus Story, the Edmonton Narrative Norms Instrument and the Test of Narrative Language (n=36)

Variable	1	2	3
1. Bus Story	--	.074	.290
2. ENNI	--	--	.463** (0.214)
3. TNL	--	--	--

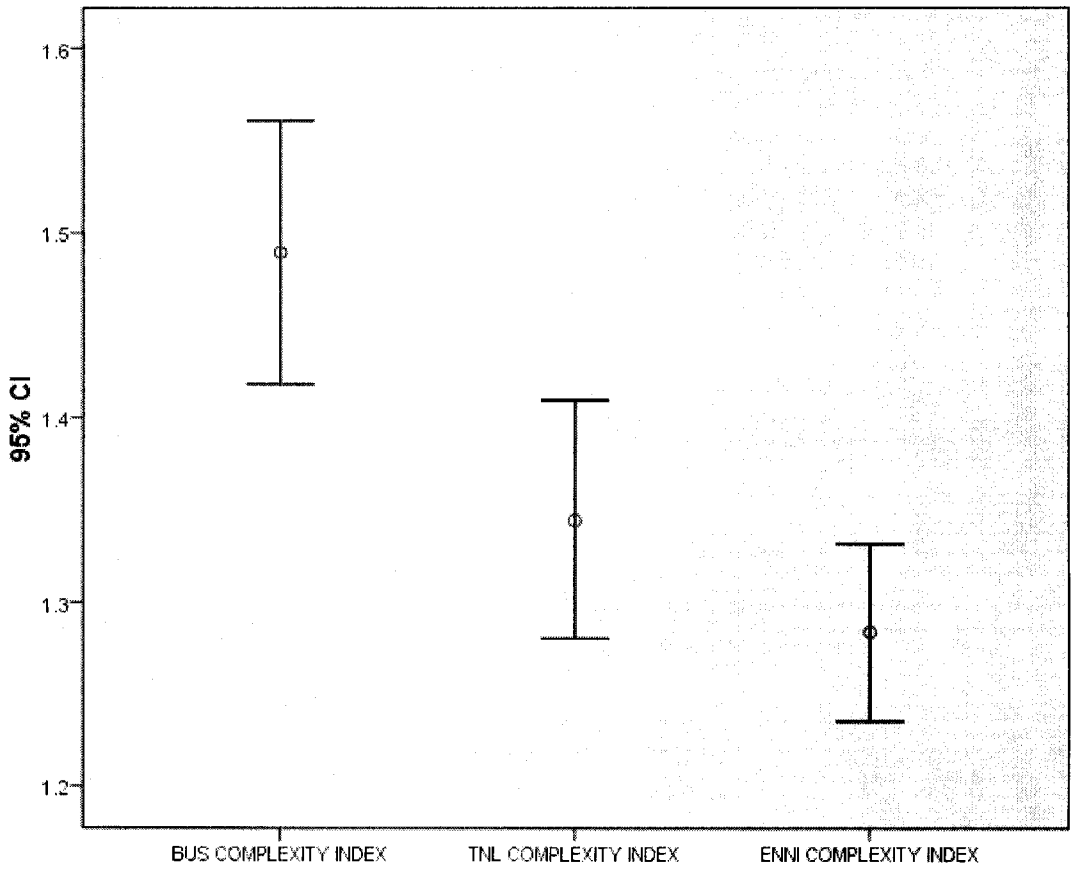
*p<0.05, **p<0.01

Differences

A significant effect for Complexity Index was found using an analysis of variance ($F_{(2,70)}$=16.132, p<0.0001, *partial eta squared*=0.315). There was also a significant linear trend ($F_{(1,35)}$=25.171, p<0.0001, *partial eta squared*=0.418) over the mean values for CI, as illustrated in Figure 3. Post hoc paired sample t tests with Bonferroni correction indicated that the mean CI of the Bus Story was significantly higher than the mean CI of the ENNI (1.28) and the TNL (1.34) (t=5.017, p<0.001, d=1.141; t=3.625, p=.001, d=0.720, respectively). Cohen's d values indicate medium

to large effect sizes for these differences. Differences between the mean CI of the ENNI and the TNL were not significant (t=2.063, p=0.047). These results show that the CI of the Bus Story was different from the other two tests and did not correlate with them.

Figure 3: Error bars showing mean Complexity Index with 95% Confidence Intervals for the Bus Story, the Test of Narrative Language and the Edmonton Narrative Norms Instrument



Discussion

Earlier studies have found that the quality of narratives produced by children is different depending on the type of task used to elicit the story. Differences have been found in narratives produced through story retell and story generation tasks (e.g., Merritt & Liles, 1989; Ripich & Griffith, 1988). Different stimulus types (i.e., oral vs. picture stimuli) have also been found to affect the quality of narratives produced (e.g., Schneider, 1996; Schneider & Dubé, 2005). Since narrative language assessment tools use a variety of elicitation techniques and stimuli, it is important to know how the tests are related to each other. Knowing how the tests compare has implications for clinicians who are using the tests to examine narrative language ability. Thus the current study compared three tests of narrative language abilities, which differ in their elicitation techniques and stimuli, to determine if they provide a similar picture of a child's narrative language ability.

Macrostructure Analysis (Narrative Language Ability)

The results of this study need to be interpreted carefully. The results show that test scores of narrative language ability for the three tests did not differ significantly from one another, suggesting that children as a group perform similarly on the three tests. However, correlations indicate evidence of concurrent validity only between the TNL and the Bus Story and the TNL and the ENNI. There was no significant relationship found between the ENNI and the Bus Story, which are two completely different tasks (e.g. generation vs. retell).

The results are consistent with previous research that found both typically developing children and children with language disorders perform differently on story retell and story generation tasks (Merritt & Liles, 1989). Performance differences were also found between oral only tasks and tasks that include pictures (Schneider, 1996; Schneider & Dubé, 2005). Therefore, it seems logical that the ENNI (story generation task) would not correlate with the Bus Story (a story retell task) and that the TNL, which contains both retell and generation tasks, would be related to both. Differences in performance may be explained by the greater processing demands

required for story generation tasks. Generating a story requires adding or modifying information within a story framework as well as generating unique ideas and stringing thoughts together into a cohesive story. It requires an ability to integrate and organize thoughts into a logical sequence, which also requires knowledge of temporal relations. These tasks are considered to be more difficult, yet may be more representative of a child's narrative ability compared to retell tasks that require the child to simply repeat a story told to the child.

Deciding which test provides a more representative picture of a child's narrative language abilities, however, depends on the purpose of the assessment. The ENNI Story Grammar score provides us with a picture of narrative ability in terms of identifying specific story grammar units, which gives us information about overall story content and organization. The Information score of the Bus Story provides information on the child's ability to recall salient information, memory of story structure, use of propositions and cohesive markers. The Oral Narration score of the TNL is a combination of narrative skills including story components, microstructural elements (temporal and causal relations, grammar, verb tense) and aesthetic qualities. Within the TNL, these skills are difficult to isolate without looking at the different tasks separately. For example, it is possible for a child scoring average on the TNL to do quite poorly on the retell task, but do very well on the generation tasks (or vice versa). Even though the overall score indicates average performance, unless we look at the different tasks, we may be missing important information about a child's specific skills. Because standard scores are not available for the different TNL tasks, comparisons across the three tasks and with the three narrative assessment tools could not be conducted. If standard scores were available for the individual TNL tasks, clinicians might be able to gain a more complete picture of a child's narrative abilities by using only one standardized assessment tool. Therefore, future research that examines the different TNL tasks separately would provide valuable information on how children are performing across these tasks and across the three tests

Although there was no difference between test scores of the TNL and both stories of the ENNI, a significant correlation was only found between the TNL and story A1, which is somewhat surprising considering that A1 and A3 are subtests of

the same test and are correlated with one another (indicating internal consistency for the ENNI). This may be explained by examining the stories in more detail. Story A1 of the ENNI is a simple story with two main characters and repeated attempts at a goal. Story A3 is more complex with four characters and repeated attempts at a goal. The first picture elicitation task of the TNL involves one main character with repeated attempts at the goal. However, the other picture elicitation task requires children to formulate a story from a single picture. This provides less structure for the child compared to sequenced pictured tasks and requires that the child generate a story beyond what is depicted in the picture. Given the young ages of children in the current study, who were closer to the lower end of the age range for the TNL, it may be that children performed better on the easier TNL task, resulting in a correlation with story A1, but less well on the more difficult TNL task, resulting in a lack of correlation with story A3.

The classification of a child's performance as average, above average or below average on the different tasks also requires some consideration. The most concerning results are for those children that are classified as above average on one test and below average on another. This was the case for three of the children in this study. Although child 14 could be considered an outlier, it is interesting to note that all three of these children were categorized as above average on the Bus Story, below average on one or both of the ENNI tasks, and average or above average on the TNL. This pattern of results might also be explained by the greater processing demands required for story generation tasks such as the ENNI and has serious clinical implications when identifying a child with language difficulties. Six children also scored within the average range on one test and below average on another test. While there does not appear to be a specific pattern to these results, they are clinically significant in determining whether or not the child has language difficulties. It is clear that some children have more difficulty with certain narrative tasks, which may be overlooked if limiting narrative language assessment to one type of task.

Microstructure Analysis (Language Complexity Measures)

The narratives produced by children through the three different tests were also used as language samples and analyzed for semantic and syntactic complexity to gain additional information about a child's language abilities. Comparisons between language complexity measures allow us to determine if we would obtain the same information when using the test stories as language samples. Previous research indicates that there is a difference in the quality of children's stories depending on the type of task used to elicit them (Merritt & Liles, 1989).

Mean Length of Communication Unit

The results indicated that MLCU was longest for the Bus Story and shortest for the ENNI, with the TNL falling in between. This supports previous research that children tend to produce more clauses and longer stories in story retell tasks when compared to story generation tasks (Merritt & Liles, 1989). It makes sense that the TNL, which includes both retell and generation tasks, would be somewhere in between for this measure. Only the difference between the ENNI and the other two tests was significant. It is possible that the greater processing demands required to formulate a story make it more difficult to generate longer, more complex sentences. These demands would be similar across story generation tasks and may explain the correlation found between the ENNI and the TNL, which both contain story generation tasks. For story retell tasks, however, the child simply has to remember the information provided. As well, the child has an adult model from which he/she can mimic certain language aspects. Therefore, the difference in MLCU between the ENNI and the Bus Story is not surprising. It may be that the story retell component of the TNL also increases overall MLCU, explaining the significantly higher MLCU on the TNL when compared to the ENNI. Story generation tasks in this test also provide an oral model before the child is asked to generate his/her own story, which may also contribute to a slightly higher MLCU. To better understand the effect of story presentation techniques, it would be beneficial to see how MLCU compares among the different stories of the TNL as well as with the other tests.

Knowledge of shared information may also help to explain the differences in MLCU between the ENNI and the other two tests. The administration of the ENNI is designed so that only the child can see the pictures. The child is instructed to tell the story in the pictures so that the examiner can understand what it happening. This puts even more demand on the child to accurately describe the picture and tell the story. This differs from conditions where the child assumes shared knowledge (i.e. both the examiner and the child can see the pictures or the story is retold after the examiner) and can focus more on their use of language. Again, more research would be needed to determine the effect that knowledge of shared information has on MLCU.

Complexity Index

The Complexity Index is a measure of the syntactic complexity of children's sentences. There was a significant positive relationship between the ENNI and the TNL on the Complexity Index, whereas the Bus Story did not correlate with either. Comparisons between the three tests show that the mean CI for the Bus Story is significantly higher than the other two tests. The significant mean differences between a strictly retell task (the Bus Story) and those containing generation tasks (ENNI and TNL) is not surprising. Children are believed to be capable of producing more complex sentences when repeating an adult model than when generating sentences on their own. Based on previous research, we anticipate that tests involving generation tasks would elicit similar levels of complexity and that they would be simpler than when children retell a story that they heard. Once again, it appears that language complexity may be affected by the greater processing demands required to generate stories.

To clarify the results of the microstructure analysis, one might look at the pattern differently. For both the MLCU and CI measures, the Bus Story is significantly higher than the ENNI. This fits with the previous explanation that language complexity may be higher for story retell tasks due to fewer processing demands. The TNL, on the other hand, which contains both story retell tasks and story generation tasks, shows a less consistent pattern. For the MLCU measure, the

TNL looks more like the Bus Story and is significantly higher than the ENNI. However, for the CI measure, the TNL looks more like the ENNI and is significantly lower than the Bus Story. These inconsistencies are difficult to interpret making it hard to determine a child's actual language complexity abilities. Without being able to analyze the different TNL tasks separately, it might be better to use assessment tools that look at retell and generation tasks independently to get a more comprehensive picture of a child's abilities.

Limitations

The representativeness of the study sample relative to all five year old children cannot be determined. Although attempts were made for random sampling, most of the children in the study attended programs in a small geographical area within the cities of Edmonton and Sherwood Park (a suburb of Edmonton) that were willing to participate. None of the centers were located in inner city areas or rural regions. Participation in the study was based solely on the willingness of daycares to allow the examiner to enter the site and on parental consent for their child to take part in the study, which may have biased the sample in some way. The results are also applicable only to five year olds, however, the limited age range of the Bus Story precludes comparisons across all three tests past age 6;11. The small sample size used in this study also limits the generalizability of the results.

In terms of concurrent validity, correlations between test scores of the TNL and the ENNI support the assumption that similar narrative language skills are assessed by these tools. Similarly the relationship found between the TNL and the ENNI on the language complexity measures indicates that these assessment tools are related in terms of language measures commonly applied to narrative samples. In other words, the results seem to provide evidence of concurrent validity between the ENNI and the TNL. Hammill, Brown and Bryant (1992) propose that tests should be considered to show evidence of concurrent validity only if at least half of the correlations are significant at 0.05 and reach or exceed 0.35 in magnitude. Although the Information score of the Bus Story and the Oral Narration score of the TNL were

significantly correlated, the value of *r* (.335) does not quite meet these criteria for concurrent validity.

Future Research

As mentioned previously, generalizations about the patterns observed in this study must be limited to five year old children with average or above average language abilities. It would be interesting to determine if similar patterns are established for older children and for children with language impairment. Previous research including children with language impairment indicate that these children score lower on the amount of story content recalled, mean length of utterance, cohesive adequacy and lexical diversity as a result of formulation, organizational and retrieval difficulties (Paul & Smith, 1993). Similarly, children with learning disabilities produce shorter stories with fewer complete episodes, propositions and complex temporal relations (Stein & Glenn, 1979). Further research is needed to determine if these children would perform differently on the three different tests. While all three tests provide evidence of their ability to discriminate between typically developing children and those with language impairment, their ability to predict later skills has not yet been demonstrated. Only the Bus Story has been investigated and shown to be a reliable predictor of later linguistic skills and academic achievement. Even though the results of this study indicate that children perform similarly on the three tests, different patterns of performance are observed. Therefore, further research is needed to determine if results of the ENNI and/or the TNL are also useful in predicting future academic achievement and linguistic skills.

The effect of comprehension questions on subsequent production should also be considered. The TNL is the only one of the three tests that has a comprehension component. In the retell task, the child is first asked a set of questions to determine comprehension before the child is asked to retell the story. Each story generation task is also preceded by oral presentation of a different story (accompanied with pictures) followed by a set of comprehension questions. This may assist the child by pointing out salient story structure components that should be included in the retell. Or it may separate the adult model (oral presentation) from the child's production with enough

of a delay that it has little/no effect on the child's language structures. Further research is needed to determine the effect that comprehension questions have on a child's performance.

Clinical Implications

The results of this study indicate that while overall children perform similarly on the Bus Story, the ENNI and the TNL, different kinds of skills are being measured by the different tests. The TNL has an advantage over the other tests in that it includes a range of narrative tasks. However, the structure of the TNL does not allow one to easily evaluate the different tasks separately. Using the Bus Story and the ENNI, on the other hand, would provide the clinician with separate scores for the different kinds of tasks (i.e. story retell vs. generation). The disadvantage of using two separate tests, however, is that they are normed on different populations. The current study points out some of similarities and differences among the tests and suggests that clinicians take them into consideration when choosing an assessment tool. The results also support previous research indicating differences in the quality of narratives produced through story retell versus story generation tasks. Both of these types of tasks provide valuable information about a child's narrative abilities. Both types of tasks reflect real world narrative demands, so we want to avoid choosing only the task that provides the "best" narrative performance. Therefore, it is important to use a variety of methods to obtain a complete picture of a child's specific narrative skills.

Although the ENNI is the only one of the three tests that includes language sample analysis procedures, the TNL was found to significantly ($p<0.001$) correlate with common language sample measures taken from conversational samples in their normative sample such as total number of words ($r=0.72$), number of different words ($r=0.79$), mean length of utterance ($r=0.66$), total number of clauses ($r=0.69$) and number of story grammar propositions ($r=0.78$) (Gillam & Pearson, 2004). Language sample analysis has also been conducted on the Bus Story in studies looking at narrative language abilities in typically developing children and those with language impairment (Paul & Smith, 1993). In other words, it is acceptable to use the stories

elicited by these tests as language samples and analyze them accordingly. The results of this study support previous research indicating that the elicitation technique has to be considered when using the stories as language samples due to the differences in performance found between the techniques.

Conclusion

Relationships were found between the TNL and the other two tests on measures of narrative language ability as well as language complexity measures. However, relationships were not found between the ENNI and the Bus Story on any of the measures examined in this study. These results are not surprising when we consider that the Bus Story and the ENNI are completely different tasks; previous research has demonstrated that these differences (retell versus story generation) yield differences in performance. The TNL, on the other hand, includes a combination of both tasks, which explains the similarities found between it and the other two tests. Comparisons indicate that children perform similarly on the three tests; however, closer examination indicates that different patterns of performance exist across the three tests. This indicates that different skills are being measured by the different tests and this needs to be considered when interpreting the results of an assessment and choosing intervention goals. Clinicians should thus be encouraged to use a variety of narrative assessments including both retell and generation tasks as well as different elicitation techniques and stimulus sets to obtain a comprehensive picture of a child's narrative language abilities.

Bibliography

Beswick, N. (2008). *Determination of the inter-rater reliability of the Edmonton Narrative Norms Instrument.* Unpublished manuscript. University of Alberta, Edmonton, AB.

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

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

Corty, E.W. (2007). *Using and interpreting statistics: A practical text for the health, behavioral, and social sciences.* St. Louis, MO: Mosby Elsevier.

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

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

Feagans, L., & Short, E. J. (1984). Developmental differences in the comprehension and production of narratives by reading-disabled and normally achieving children. *Child Development, 55*(5), 1727-1736.

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

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

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

Kaufman, A.S. & Kaufman, N.L. (2004). *Kaufman Brief Intelligence Test-Second Edition.* Circle Pines, MN: AGS Publishing.

Liles, B. Z. (1985). Production and comprehension of narrative discourse in normal and language disordered children. *Journal of Communication Disorders, 18*(6), 409-427.

Merritt, D.D., & Liles, B.Z. (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders, 54,* 429-438.

Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late talkers: Early school age. *Journal of Speech and Hearing Research, 39*(6), 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.

Renfrew, C.E. (1969, 1977). *The Bus Story: A test of continuous speech*. Oxford: Author.

Ripich, D.N., & Griffith, P.L. (1988). Narrative abilities of children with learning disabilities and nondisabled children: Story structure, cohesion and propositions. *Journal of Learning Disabilities, 21(3)*, 165-173.

Roth, F. P., & Spekman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech & Hearing Disorders, 51*(1), 8-23.

Schneider, P. (1996). Effects of pictures versus orally presented stories on story retellings by children with language impairment. *American Journal of Speech-Language Pathology, 5(1),* 86-96.

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

Schneider, P., Dubé, R.V., & Hayward, D. (2004). *The Edmonton Narrative Norms Instrument*. Retrieved August 14, 2008, from University of Alberta: The Edmonton Narrative Norms Instrument. Web site: http://www.rehabmed/ualberta.ca/spa/enni.

Schneider, P., & Hayward, D. (2006). *Who does what to whom: Introduction of referents in children's storytelling from pictures*. Manuscript submitted for publication.

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

Schneider, P., Hayward, D., & Dubé, R.V. (2006b, June). *Using narrative measures to discriminate between children with and without language impairments: More are better*. Poster presented at the Symposium for Research in Child Language Disorders, Madison, WI.

Shipley, K.G., & McAfee, J.G. (2004). *Assessment in Speech-Language Pathology: A Resource Manual, 3rd Edition.* New York: Delmar Learning.

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.

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

Wiig, E.H., Secord, W.A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool-Second Edition*. San Antonio, TX: Harcourt Assessment.

Appendix A

The Story Grammar Model

(Adapted from Stein & Glenn, 1979)

Story Grammar Categories

1) Setting-description of the characters and story context
    a. Major Setting-character introduction
    b. Minor Setting-description of story context
2) Initiating Events-events that provoke a response from the protagonist
3) Internal Responses-reaction of the protagonist to the initiating event which may include goals, thoughts and feelings
4) Plans-the intended actions of the protagonist to deal with the initiating event
5) Attempts-actions of the protagonist to obtain the goal(s)
6) Direct Consequences-the outcome of the protagonist's attempt to attain the goal
7) Reactions-how the character(s) react to the outcome which may include actions or thoughts and feelings