

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Equivalence of Picture Sets for Eliciting Stories from Preschool Children

Kate Dumbrell, Bryn Hornby, Kayla Stephenson

Supervisor: Dr. Phyllis Schneider; Reader: Melissa Skoczylas

Equivalence of Picture Sets for Eliciting Stories

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

ABSTRACT

Narratives are widely used in language assessment and intervention. They are clinically useful, as they require the skills of combining words and sentences into meaningful stories.

Narratives of children with impairments predict later academic achievement, language development, and reading comprehension.

Previous research using narratives looked at development and use of the Edmonton Narrative Norms Instrument (ENNI). The ENNI uses pictures to elicit story generation from children. There are two different story sets that can be used to generate narrative and language samples. The ENNI's Story Grammar norms were based on its existing picture set, A3. A new story, RD, was developed with the goal of having a parallel structure with similar results to A3. The purpose of this study is to investigate whether RD yields similar results to A3. Thirty-six children were given both the RD and A3 form twice, with an average two-week delay between administrations. The stories were scored for Mean Length of Utterance (MLU), Number of Utterances (NUtt), Number of Different Words (NDW), First Mentions (FM) and Story Grammar (SG) information. The scores from each form were compared.

A3 and RD differ in the amount of language they produce and FM of characters. RD produced longer utterances and MLU, higher NDW, and a greater amount of correct FM than A3. However, the forms produced similar results for SG. Although A3 and RD should not be considered as clinically equivalent tools, RD may be used as an alternate tool to analyze and compare SG values to story A3.

INTRODUCTION

Children begin to use narratives in their language at the age of three or four years old (Stadler & Ward, 2005). Narratives are a powerful and revealing tool when collecting a representative sample of a child's language ability. They are a form of oral discourse that people use to communicate personal and created events to others; they are an integral part of culture and a useful way to make sense of our past, evaluate experiences in our present and plan for events in our future (Schick & Melzi, 2010). Narratives have been a part of language development and disorders research for some time. They not only predict academic success, literacy development and written language development, but there has also been evidence to suggest that narratives are related to a child's conceptual development as well (Stadler & Ward, 2005). A child must be able to use complex language that includes the use of explicit vocabulary, appropriate introduction of characters and use of pronouns, cause and effect relationships and temporal connections in their narratives. Narratives are further able to provide a picture of preschoolers' language development beyond single word and sentence acquisition, which includes assessing a child's use of complex, functional and social language (Norbury et al., 2014). Narratives represent language use in everyday social situations and are a substantial part of the school curriculum; they reveal children's linguistic knowledge beyond isolated components of language taken out of context (Hayward & Schneider, 2000). The development of written communication skills and literacy skills is supported partly by oral narrative generation skills (Hayward & Schneider, 2000; Stadler & Ward, 2005). Oral narrative skills are considered a form of literate language that acts as a bridge between oral and written language (Hedberg & Westby, 1993). Further, narratives not only predict later academic success

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

in literacy development, but they also are related to the child's socioemotional development, sociocognitive skills, recall and planning skills, and memory skill (Schick & Melzi, 2010).

Stadler and Ward stated that children begin to tell stories using personal experiences; children's narratives are greatly influenced by their experiences, their environment and their culture (Stadler & Ward, 2005). Children then move on to telling stories during play, retelling stories, and eventually beginning to create their own fictional stories. The authors generated a list of the different levels of stories: labeling, listing, connecting, sequencing and narrating.

Stadler and Ward found that children in the early stages of narrative development often label images in the picture as opposed to telling a story about the picture; listing is where the child lists character attributes or actions in the story; connecting refers to the child having a central topic which connects to the characters' actions; sequencing is where the child starts to demonstrate temporal and cause and effect relations in their narratives; narrating is the final level, where children are able to include all the above levels in addition to a plot that includes planning to attain a goal.

Narratives have been used in a number of different ways to assess and treat child language. However, it is important to note that the way in which stories are elicited from children, and the materials used to elicit the story, will affect the content and quality of the story. Story generation from pictures is arguably more beneficial than story retell in assessing narrative ability because it reveals a child's ability to generate a narrative rather than revealing whether the child can simply recall the story (Schneider & Dube, 2005). Schneider and Dube (2005) further compared the outcomes of a story retell and a story generation task in a group of typically developing kindergarten and Grade 2 children. The results obtained lead to the

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

observation that while story retelling reflects the extent to which the child can use the linguistic input he receives, story generation shows what the child is capable of generating himself, without a sample story. The two story tasks also indirectly reflect a child's cognitive abilities in areas such as sustained auditory attention and verbal working memory. Children require both the skills of retelling and generation for real-life storytelling.

Using picture sets for a story requires children to use their language skills to formulate the story and allows for intervention by identifying the child's strength and weakness patterns in their story productions. However, it is of utmost importance to have a thorough understanding of the complexity of the narrative stimuli used to elicit stories. For instance, although Mercer Meyer's Frog Stories may seem quite similar due to having the same protagonist, being written by the same author and being about the same length, the stories are not very similar in structure. Evidence from Strong (1998) showed that typically-developing children's narratives for three of the Frog Stories varied between 6% and 27% in terms of mean referential cohesion (as cited in Schneider and Hayward, 2010). Thus, one should not assume that different picture sets will elicit the same narrative generation results, even if the sets look very similar on the surface.

Many researchers who study narratives have stated that preschoolers' narrative skills predict future academic and linguistic success; assessing preschoolers' narrative skills provides useful information about their language abilities, the development of which can ultimately affect their success both academically and linguistically (as reviewed by Hayward & Schneider, 2000). Story tasks are an especially useful assessment tool for child language. Story productions can be used as language samples and may be examined closely to determine the child's

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

language abilities. Results taken from the productions allow speech-language pathologists to discriminate between children with language impairments (including expressive and receptive) and those without in a naturalistic language context (Schneider, Hayward & Dube, 2006). This is valuable clinical information as it can aid in the diagnosis and subsequent treatment of language delays in young children. Story generation tasks can also be used as an intervention tool for children with below average language abilities as compared to their peers. Developing an at-risk child's narrative skills through intervention supports literacy development and the prevention of falling behind in reading and writing skill development (Brown et al., 2014). Further, teaching story grammar components has been shown to be beneficial for increasing children's knowledge of story micro- and macrostructures, story generation and narrative retelling (Brown et al., 2014).

Narratives can be analyzed using Stein and Glenn's story grammar model. This model consists of story components that are typically included in a predictable sequence in narratives; the components are related temporally and causally (Stein & Glenn, 1979, as cited in Hayward & Schneider, 2000). This model is used to analyze children's language development and language disorders; children who present with language impairments often use fewer story grammar components, resulting in incomplete stories (Hayward & Schneider, 2000).

To facilitate the assessment of narrative abilities, the Edmonton Narrative Norms Instrument (ENNI) was developed. The ENNI is an assessment tool that was designed for collecting language information from children ages 4 to 9 through storytelling (Schneider, Dubé & Hayward, 2005). This tool is comprised of pictures that portray a story. These pictures are presented to the child, who uses the pictures to generate a story for the examiner. Storytelling

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

norms for preschool to young elementary ages were developed in development of the ENNI. Currently there are two story sets that can be used to generate narrative and language samples. They range in complexity based on the number of characters present and events occurring in the story. The resulting norms of the ENNI for story grammar were based on the first ENNI picture set (A3). The ENNI was carefully developed based on the Story Grammar model of story knowledge and includes information that adults view as being essential in narrative development. The principles of the Story Grammar model describe the content and organization of stories, and how certain “Story Grammar units” (SG units) typically appear in a specific order within a story. SG units are units of information that are characteristic of stories judged by adults and children to be “good” stories (Stein & PolICASTRO, 1984, as cited in Schneider, Dubé & Hayward, 2005). They include characters, settings, initiating events for goal-directed behaviour, attempts to attain the goal, a problem, and a solution to the problem that results in successfully achieving the goal (outcomes). Story Grammar interventions in previous studies have been shown to be advantageous for improving children’s knowledge and complexity of both the narrative structural components and story information (Hayward & Schneider, 2000; Brown et. al., 2014).

PURPOSE

A new story, RD, was developed prior to this study in order to have a story with a parallel structure to the existing complex story of the ENNI, story A3. RD was constructed to have as similar a structure as possible. For example, whereas in story A3 the characters drop a toy plane in a pool, while in the new story the characters lose a balloon in the branches of a tree; in both stories, two adult characters attempt to help them recover their lost objects. This

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

project was a pilot study with the goal of describing the feasibility of the new story, RD, as a new addition to the ENNI clinical tool. This study aims to determine a) whether the two story forms elicited similar language productions from children, and b) whether the forms yielded consistent stories over time (i.e., when told on two occasions with a time gap between retellings). This study received ethics approval from the University of Alberta's Research Ethics Board in the spring of 2015.

METHODS

Participants for this study were recruited by phoning daycares within the Edmonton area. Once the daycares had agreed to participate in the study, the personnel were asked to distribute letters to families of eligible participants. Two letters were given to families, including an information letter about the study and its purpose, and a consent letter. Daycares were followed up with between one and/or two weeks after handing out the forms. The children whose parents completed the consent form and met the study criteria were eligible to participate in our study. The criteria for eligible participants were that they had to be between the ages of four to six years old, speak English as their first language, and could not demonstrate any language or cognitive impairments. A total of 42 children were recruited to participate in the study; valid data was obtained from 36 of these children for a total of 36 subjects. The six children who were excluded from our final data were excluded due to insufficient spoken English (i.e. English was not their first language), lack of cooperation or choosing to decline when the researchers asked for their informed assent. The thirty-six participants (18 males and 18 females) ranged in age from 3 years and 11 months old to 6 years and 5 months old. The mean age was 5 years (60.24 months), with a standard deviation of 8.68

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

months. All children were enrolled in child care programs in Edmonton, Alberta. These daycares were within a 5 km radius of the University of Alberta. One child was recruited via a personal contact of one of the researchers, and was subsequently tested in the child's home. The consent forms had a space where both the mother and father were asked to record the number of years of education they had received. For mothers, the range extended from 10 to 22 years of education in total (including elementary, secondary and post-secondary). The mean number of years for mothers was 16.64, with a standard deviation of 3.37 years. The fathers' education ranged from 9 to 25 years, with a mean of 16.23 and a standard deviation of 3.74 years.

When the researchers arrived at the daycares for testing, they took one child at a time to a quiet space such as a resource room, library, or a table in the hallway, where the staff could still maintain supervision over the child. Then the examiner read out the "Child Assent" form (written in simple language), explaining the study, the risks, and the benefits. The examiner then asked the child if they still wanted to do the study. If the child responded "yes", then either the child was asked to print their name on the form, or, if the child could not write his or her name, the examiner would print the child's name and note that her or she gave verbal consent. If the child said "no", the child was taken back to class and his or her participation in the study was discontinued.

The ENNI was administered in a way that required the child to orally generate a narrative to the examiner, according to ENNI administration procedures (Schneider et al., 2005). The child was informed that the examiner could not see the pictures, and therefore he or she must use good language to tell their story so that the examiner could understand what was happening. To administer the ENNI, a binder containing the stories was held up facing the

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

child so the examiner could not see the pictures. The examiner turned the pages as the child was telling the story, based on non-verbal judgments when the child was ready to move onto the next page. Examples of such judgments were falling intonation at the end of a sentence, a period of silence longer than a few seconds, and/or an expectant look from the child. One of two training stories was administered before the testing stories of the ENNI to ensure that the child was familiar with the storytelling process and to help them get started. One of the training stories was given to half of the children at Time 1, and the second training story was given to the other half of the children at Time 1; children heard the training story that they had not heard before at Time 2. More explicit examiner prompts were allowed during the training stories. For instance, if the child was having trouble starting the story, the examiner might say “would you start with ‘once upon a time’?”. Further explicit prompts were given as necessary.

After the child completed a training story, he or she was asked to generate a narrative for each test story in a previously randomized order (either A3 then RD, or RD then A3). The child was given the following instructions before all stories were administered: “I have some pictures that tell a story. First I’ll show you all the pictures and we’ll go back to the beginning of the story, and then I want you to look at the pictures and tell me the story that you see in the pictures. I won’t be able to see the pictures so you need to tell me the story really well so I can understand it. Okay?”. The child was then instructed to look at all the pictures. The examiner would hold the story and turn the pages at a moderate rate, letting the child look at the pictures and slowing down if asked. Once the end of the story was reached, the examiner would begin at the first page and instruct the child to tell the story. According to the ENNI protocol, examiners gave only general prompts during the test stories. If a child had difficulty,

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

the examiner would say “it’s your story - you can tell it however you want”, or “just try your best”. General supportive comments were often given, such as “you’re doing a great job” or “good story - keep going”. After an approximate two-week period, the same two picture sets (A3 and RD) were given to each child again. The child was asked to produce a story for these picture sets, using the same protocol outlined above. The purpose of this second retelling was to enable comparison of the same stories at different time points. The order of the narratives was counterbalanced.

The narrative productions were recorded on the researchers’ personal cellular phones or iPad devices. If there was a lot of background noise in the room and/or the child’s voice was very quiet and did not consistently improve after reminders to use a “loud voice”, the examiner sometimes had to repeat the child’s utterances for each page so the recording device could accurately record the story. When this happened, the examiner would repeat the child’s utterance verbatim (including any grammatical errors). These repetitions did not seem to distract the children from the story, or change the way they were telling their stories. No identifying information was used in the recordings. The researchers kept track of recordings by labeling them with their own initials (KS, KD or BH), the subject number (01-13), the story title (A3 or RD) and time (1 or 2), and then inputting these into spreadsheets organized by date and daycare location. The researchers then transcribed their recordings using Microsoft Word, again using no identifying information. Each child had a total of four transcripts (two for story A3 and two for story RD). 144 transcripts were transcribed in total (four transcripts each for 36 children).

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

The researchers then checked for inter-rater reliability by each transcribing 10% of one other researcher's audio-recordings. These recordings were selected randomly by the owner and were representative of multiple participants, both stories A3 and RD, and testing at Time 1 and Time 2. The overall inter-rater reliability was 98.4% with the three researcher's individual values being 97.75%, 98.6%, and 98.9%. The Systematic Analysis of Language Transcripts (SALT) program was used to analyze all transcripts. The Standard Measures Analysis tool was used to analyze the total number of words, mean length of utterance (MLU), number of utterances (NUtt), and number of different words (NDW) for every transcript.

The researchers then completed a training session with Dr. Schneider where they learned to score Story Grammar for typically-developing children's ENNI stories. The researchers had the opportunity to compare their scores and ask Dr. Schneider any questions about scoring. They also practiced scoring First Mentions on transcripts taken from the ENNI website. After the training was complete, they each scored their participant's transcripts by hand using the Story Grammar and First Mentions protocols established for the ENNI. The scoring sheets for story A3 were taken directly from the website and can be found on the University of Alberta website at

<https://rehabilitation.ualberta.ca/departments/communication-sciences-and-disorders/resources-for-clinicians-and-researchers/edmonton-narrative-norms-instrument/pictures-analyses-and-comprehension/story-grammar> for Story Grammar and at <https://rehabilitation.ualberta.ca/departments/communication-sciences-and-disorders/resources-for-clinicians-and-researchers/edmonton-narrative-norms-instrument/pictures-analyses-and-comprehension/first-mentions> for First Mentions. The

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

scoring sheets for story RD were modified from the A3 sheets to include the appropriate characters, objects and events. Scoring sheets for RD can be found in Appendices 1 (Story Grammar) and 2 (First Mentions). The researchers then checked for inter-rater reliability by each transcribing 10% of one other researcher's transcripts for Story Grammar, and another 10% of one other researcher's transcripts for First Mentions. These transcripts were selected randomly by the researchers and were representative of multiple participants, both stories A3 and RD, and testing at Time 1 and Time 2. The overall inter-rater reliability for First Mentions was 96.7% with the three researcher's individual values being 100.0%, 96.7% and 93.4%. The overall inter-rater reliability for Story Grammar was 96.9% with the three researcher's individual values being 97.0%, 99.4% and 94.2%.

The SPSS program was used for further statistical analyses on both Story Grammar and First Mentions.

ANALYSIS

Analyses focused on the story information that is typically included in narrative productions according to the story grammar model. The results of the A3 productions and the RD story productions were compared, both at the same time period and across the two time periods, to see if the new story RD produced similar narrative results to story A3. The data was analyzed in SPSS using a within-subjects 2x2 ANOVA, and an alpha level set at 0.05. As researchers were looking for similarities across stories, and thus were anticipating a lack of significant results, beta level was also set at .3; comparisons that fell at or above .3 would be considered to indicate a genuine lack of difference (as recommended by Huberty, 1987). A bivariate correlation was used to evaluate the consistency of story productions over time.

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

RESULTS

Preliminary analyses of three child language variables (i.e., number of different words (NDW), mean length of utterance (MLU), number of utterances (NUtt)) revealed significant results, indicating that the stories produced significantly different amounts of language. For each dependent variable, the stories were analyzed using a 2 (type of story: A3 and RD) x 2 (time: 1st time, 2nd time) within-subjects ANOVA.

Table 1 displays the means and standard deviations for the NUtt, MLU and NDW in the two different stories across the two times. See *Table 1* below.

Table 1

Mean and Standard Deviation (SD) for Type of Story and Time: NUtt, MLU, NDW

Story - NUtt - Mean (SD)		Story - MLU - Mean (SD)		Story - NDW - Mean (SD)		
Time	1	2	1	2	1	2
A3	16.67 (3.521)	16.17 (3.247)	6.4447 (1.05)	6.3008 (1.02)	48.06 (10.02)	46.69 (11.44)
RD	17.86 (4.343)	17.17 (5.085)	6.7817 (1.29)	6.5108 (.92)	53.64 (16.55)	49.22 (13.14)

Note. NUtt = Number of Utterances; MLU = Mean Length of Utterance; NDW = Number of Different Words.

Table 2 displays the correlations for type of story and time: *NUtt, MLU, NDW*. See *Table 2* below.

Table 2

Correlations for type of story and time: NUtt, MLU, NDW

Story and Time	NUtt	MLU	NDW
A3 T1 vs. RD T1	.697* .0000022843	.623* .0000498298	.821* .0000000009

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

A3 T2 vs. RD T2	.741* .0000002412	.792* .0000000089	.794* .0000000077
A3 T1 vs. A3 T2	.605* .0000936714	.714* .0000010183	.738* .000002823
RD T1 vs. RD T2	.462* .0045990257	.794* .0000000074	.722* .0000006659

Note. NUtt = Number of Utterances; MLU = Mean Length of Utterance; NDW = Number of Different Words.

*Correlation is statistically significant.

The descriptive statistics for NUtt displayed the following means and standard deviations (mean (*SD*)) for the different stories: NUtt A31 - 16.67 (3.52), NUtt A32 - 16.17 (3.25), NUtt RD1 - 17.86 (4.34), and NUtt RD2 - 17.17 (5.09). Results for number of utterances indicated that there was a significant main effect of the type of story for NUtt, $F(1, 35) = 6.022$, $p = .019$; partial $\eta^2 = .147$. There was no significant main effect of the time of the story for NUtt, $F(1, 35) = .989$, $p = .327$; partial $\eta^2 = .027$, and there was no significant interaction between the type of the story and the time of the story, NUtt, $F(1, 35) = .091$, $p = .765$; partial $\eta^2 = .003$.

With regard to mean length of utterance, the descriptive statistics displayed the following means and standard deviations (mean (*SD*)) for the different stories: MLU A31 - 6.44 (1.05), MLU A32 - 6.30 (1.02), MLU RD1 - 6.78 (1.29), and MLU RD2 - 6.51 (.92). There was a significant main effect of the type of story for MLU, $F(1, 35) = 5.261$, $p = .028$; partial $\eta^2 = .131$; however, there was no significant main effect of the time of the story for MLU, $F(1, 35) = 3.949$, $p = .055$; partial $\eta^2 = .101$, although the p-value was lower than our beta of .3, indicating that we cannot conclude that time was not a factor. There was no significant interaction between

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

the type of the story and the time of the story for MLU, $F(1, 35) = .635, p = .431$; partial $\eta^2 = .121$.

After analyzing the two stories for number of different words, the descriptive statistics displayed the following means and standard deviations (mean (SD)) for the different stories: NDW A31 - 48.06 (10.02), NDW A32 - 46.69 (11.44), NDW RD1 - 53.64 (16.55), and NDW RD2 - 49.22 (13.14). There was a significant main effect of type of story, $F(1, 35) = 11.54, p = .002$; partial $\eta^2 = .248$; there was a significant main effect of the time of the story on NDW, $F(1, 35) = 4.615, p = .039$; partial $\eta^2 = .116$; and there was no significant interaction between the type of the story and the time of the story, $F(1, 35) = 2.607, p = .115$; $\eta^2 = .069$ (although the p-value was lower than .3 for this result).

Table 3 displays the means and standard deviations for the Story Grammar (SG) scores in the two different stories across the two times. See *Table 3* below.

Table 3

Mean and Standard Deviation for Story Grammar in Each Story

Story	Mean (SD)	
Time	1	2
A3	20.44 (5.02)	20.69 (4.37)
RD	20.39 (4.08)	21.11 (4.02)

Table 4 displays the correlations for Story Grammar (SG) scores in the two different stories across the two times. See *Table 4* below.

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Table 4

Correlations for Story Grammar in Each Story and Time

Story and Time	SG
A3 T1 vs. RD T1	0.629* .0000404262
A3 T2 vs. RD T2	0.788* .0000000113
A3 T1 vs. A3 T2	0.832* .0000000003
RD T1 vs. RD T2	0.586* .0001723925

Note. SG = Story Grammar; T1 = Time 1; T2 = Time 2.

*Correlation is statistically significant.

Story Grammar was analyzed using a 2 (type of story: A3 and RD) x 2 (time: 1st time, 2nd time) within-subjects ANOVA. The descriptive statistics displayed the following means and standard deviations (mean (*SD*)) for the different stories: SG A31 - 20.44 (5.02), SG A32 - 20.69 (4.37), SG RD1 - 20.39 (4.08), SG RD2 - 21.11 (4.02). Results indicated that there was no significant main effect for the two stories' story grammar, $F(1, 35) = .159, p = .692; \eta^2 = .005$, evidencing that the stories produced similar story grammar components. There was no significant main effect of the time of the stories' story grammar, $F(1, 35) = 1.367, p = .250; \eta^2 = .038$ (although once again the p-value was lower than the beta of .3), and there was no significant interaction between the type of story and the time of the stories' story grammar, $F(1, 35) = .450, p = .507; \eta^2 = .013$.

Table 5 displays the means and standard deviations for the First Mentions (FM) scores in the two different stories across the two times. See Table 5 below.

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Table 5

Mean and Standard Deviation for First Mentions in Each Story

Story	Mean (SD)	
	1	2
A3	13.39 (3.38)	14.00 (2.78)
RD	14.28 (2.65)	14.58 (2.59)

Table 6 displays the correlations for First Mentions (FM) scores in the two different stories across the two times. See Table 6 below.

Table 6

Correlations for First Mentions in each story and time.

Story and Time	FM
A3 T1 vs. RD T1	0.683* .0000044040
A3 T2 vs. RD T2	0.806* .0000000029
A3 T1 vs. A3 T2	0.666* .0000093080
RD T1 vs. RD T2	0.718* .0000008421

Note. FM = First Mentions; T1 = Time 1; T2 = Time 2.

*Correlation is significant.

First Mentions (FM) were analyzed using a 2 (type of story: A3 and RD) x 2(time: 1st time, 2nd time) within-subjects ANOVA. The descriptive statistics displayed the following means and standard deviations (mean (*SD*)) for the different stories: FM A31 - 13.39 (3.38), FM A32 - 14.00 (2.78), FM RD1 - 14.28 (2.65), and FM RD2 - 14.58 (2.59). Results indicated that there was a significant main effect of the two stories' first mentions, $F(1, 35) = 8.562, p = .006; \eta^2 = .197,$

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

indicating that the stories produced differing FM. There was no significant main effect of the time of the stories' first mentions, $F(1, 35) = 2.504, p = .123; \eta^2 = .067$ (although note that the p-value was lower than beta), as well as no significant interaction between the type of story and the time of the stories' first mentions, $F(1, 35) = .375, p = .544; \eta^2 = .011$.

DISCUSSION AND LIMITATIONS

The purpose of this study was to examine if the new ENNI story form RD yielded similar results to the original form A3. Results of language, Story Grammar and First Mentions were analyzed to see whether there were differences between the story forms. Within-subject 2x2 ANOVAs revealed that both language measures and First Mentions were significant, indicating that the two story forms produced different results. The measure of Story Grammar was found to be not significant, meaning that the two stories produced similar story grammar components.

The language measure results for "NUtt" (number of utterances) indicated that story RD produced longer utterances on average than story A3 (Mean $A3_1 = 16.67, A3_2 = 16.17$, while Mean $RD_1 = 17.86$ and $RD_2 = 17.17$). Qualitative observations of the stories after the ANOVAs were completed indicated that the children tended to elaborate more on the setting and the characters in story RD, which may have been attributed to the smaller pictures, increased detail in the pictures and/or the main characters appearing quite similar to each other in RD. The MLU value between the stories also differed with RD producing slightly higher values (indicating longer mean length of utterances). This difference is likely attributed to the amount of elaboration provided by children about the characters, setting and events in RD. Further, the NDW language measure revealed that story RD prompted the children to use a larger variety of

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

words in their story description. The correlations between the type of story and time for NUtt, MLU and NDWs were high for all story pairs (A3-RD times 1 and 2), indicating that the patterns remained the same across children. This was interpreted as the children systematically doing something different across the stories, rather than randomly generating different patterns of language each time.

The “First Mentions (FM)” data revealed that story RD consistently contained more appropriate First Mentions (“appropriate” meant that a referent received a full score of 3/3 points). The researchers anticipated that this would be a difference between the two stories given that the children seemed to be inconsistent in their character and object introductions across the two stories. More specifically, in RD children more consistently introduced the objects “balloon” and “ladder” compared to the “airplane” and “net” in form A3. Further, the differences may be attributed to the FM scoring procedure where the term “the lifeguard” (story A3) is scored as a 3, while the term “the policeman” is scored as a 2. The full 3 point value for “the lifeguard” was only given if the child previously mentioned the setting of the story as occurring at the pool; the majority of children did not mention this. If the child failed to mention the pool, they would be given a score of 2/3 for the term “the lifeguard”. This scoring rule was not in place for RD, as the characters and their roles were not contingent upon the setting of the park. As a result, using the term “the” in a first mention of any character in RD resulted in a score of 2, which could have impacted the differing First Mention results between the two stories. However, given that most children did not mention the pool in A3, this factor did not have a large impact on FM in this study.

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

The correlations between the type of story and the time of story were all high (over 0.7), following the same pattern as the language measures correlations. The high correlations indicated that each child consistently and systematically produced significant differences across stories rather than randomly generating their first mentions across the different stories and times.

When interpreting both the language measure scores and the First Mentions scores, one factor to consider is that the setting and characters in story A3 are more cohesive, in that the characters are highly related and belonged to the setting of the pool (e.g., lifeguards and Elephant's mom), while in story RD the characters are not necessarily related to the setting of the park (e.g., policeman and construction worker). These differences may have contributed to the differences in the language measures the children produced, as well as their First Mentions score in both stories. For instance, one qualitative observation that the researchers made was that the children tended to elaborate, clarify or self-correct more in story RD than A3. This may have been due to the fact that the characters were not as closely related to the setting as they were in A3, so the children felt that they needed to explain the action or event in a more detailed manner. This qualitative observation is supported by the quantitative data above for Number of Utterances, Mean Length of Utterances and Number of Different Words which shows that story RD consistently produced higher numbers (meaning more language) in these three areas.

The "Story Grammar (SG)" results were the only results that were not statistically significant, meaning children's Story Grammar components were very similar between the two stories. The data for SG indicated that when story RD was given the second time (i.e., RD-2), it

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

produced slightly more SG grammar components than the other stories and times but this finding was not significant. The other stories and times were more similar in their SG results. Overall, the two stories across the different times produced similar SG components indicating no significant differences. However, although these comparisons were nonsignificant, their *p*-values were lower than the chosen beta value of .3, meaning that the researchers could not conclude that there was truly no difference, just that they did not obtain a difference. The correlations between the two stories and between the times of story were all highly correlated, indicating that each child consistently and systematically produced the same patterns across their stories, rather than randomly generating story grammar components across the different stories and times. These results lead to the conclusion that clinicians may use both story A3 and story RD to attain a baseline measure, compare progress in intervention and conduct post-treatment measures for Story Grammar components.

One important and interesting finding was the direction of differences across Time 1 (T1) and 2 (T2) for both stories. Although most of the time comparisons were not significant, the *p*-values did not reach the chosen beta value of .3, which indicates that it cannot be concluded with confidence that there were no differences. Number of utterances (NUtt), Number of different words (NDW), and MLU were the measures of standard language analyses used in this study. They were all shorter in T2, while the more story-related measures of Story Grammar and First Mentions were longer the second time the children told the story. These results were consistent for both story RD and A3. These findings support the suggestion by Schneider and Hayward (2010) that it can be quite difficult to create two separate picture sets that are truly equivalent across all story component aspects and that produce similar,

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

statistically insignificant story generation results. Story creators need to consider the cohesiveness of the characters and setting, the complete scoring procedures, the detail presented in the pictures and picture size when attempting to create stories that produce similar story generations across children

There were no interactions between the type of story and the time of the story (1st or 2nd) in the language measures, the First Mention measures or the Story Grammar measures. As the children were consistently demonstrating systematic patterns in their story generations, the resulting data is likely of high reliability. Despite the stories having what appeared to be similar structures on the surface level, smaller details such as setting and character cohesiveness, picture depictions and scoring procedures contributed to significant differences between the two stories across the majority of language measures investigated in this study. Time was significant for the NDW language measure; fewer different words were used in T2 for both stories. This may be attributed to a testing effect. The familiarity with the settings and characters may have allowed the participants to provide a more concise summary of their story, leading to fewer words needed in their descriptions of the events.

FURTHER RESEARCH/CONCLUSION

This study extended the research that has been done on the ENNI to date. The ENNI was created as an instrument used to assess storytelling abilities and is available on the University of Alberta website. Professionals and researchers all over the world have used the ENNI. The newly developed picture set (RD) would be recommended as an alternate form to story A3 of the ENNI exclusively for measures of Story Grammar based on this study's results. Both forms A3 and RD yielded clinically similar, statistically nonsignificant results in the amount of Story

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Grammar units that children included. However, it is not recommended that RD be used as an alternate or equivalent form for the three child language measures discussed above (Number of Different Words, Mean Length of Utterances, Number of Utterances) or for First Mentions (FM). RD produced more language and higher FM scores on average compared to A3. As a result, RD cannot be used to assess children's generative story abilities using the ENNI norms for story A3 or to compare treatment efficacy in clinical settings for the above measures as an alternative form to story A3. This study has provided data supporting the use of RD as an alternate assessment form for children's Story Grammar exclusively.

Future research using the ENNI could focus on more explicitly evaluating the factors that may have impacted the children's story generation in this study. Further, improvements could be made in the observable pictorial differences when comparing stories A3 and RD, such as the size of the pictures and level of detail in RD. Improving this detail may help to decrease the level of extra information children tended to provide in form RD (which lead to a greater amount of language). It is recommended that researchers continue investigating, testing and developing alternative forms for the ENNI as it is a highly valuable clinical tool for assessing and intervening in children's language abilities.

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

References

- Brown, J. A., Garzarek, J. E., & Donegan, K. L. (2014). Effects of a narrative intervention on story retelling in at-risk young children. *Topics In Early Childhood Special Education, 34*(3), 154-164.
- Duinmeijer, I., de Jong, J. & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language and Communication Disorders, 47*(5), 542-555.
- Hayward, D., & Schneider, P. (2000). Effectiveness of teaching story grammar knowledge to pre-school children with language impairment. An exploratory study. *Child Language Teaching & Therapy, 16*(3), 255-284.
- Huberty, C. H. (1987). On statistical testing. *Educational Researcher, 16*, 4-9.
- Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*, 485–510.
<http://wom.sagepub.com/login.ezproxy.library.ualberta.ca/content/26/2/1>
- Schick, A., & Melzi, G. (2010). The development of children's oral narratives across contexts. *Early Education & Development, 21*(3), 293-317. doi:10.1080/10409281003680578
- 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. (2005). The Edmonton Narrative Norms Instrument. Retrieved January 2016 from University of Alberta Faculty of Rehabilitation Medicine website: <http://www.rehabresearch.ualberta.ca/enni>.
- Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech & Hearing Services In Schools, 41*(4), 459-473.
- Schneider, P., Hayward, D., & Dubé, R. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal Of Speech-Language Pathology & Audiology, 30*(4), 224-238.
- Stadler, M. A., & Ward, G. C. (2005). Supporting the narrative development of young children. *Early Childhood Education Journal, 33*(2), 73-80. doi:10.1007/s10643-005-0024-4

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Appendices

Appendix 1.

Story Grammar Scoring Sheet for Story RD

Child's Name: _____ Age: _____ Date: _____

Please read the section of the Edmonton Narrative Norms Instrument Manual on scoring SG units before using this sheet.

SG Unit	Acceptable [child need only have one alternative per unit to get credit for that unit]	Score	
Character 1	Dog/ female / girl (or any type of animal such as bunny) [not pronoun]	0	1
Character 2	Rabbit/ male / boy (or any type of animal such as bunny) (not acceptable: pronoun)	0	1
Setting	at park / pulling a wagon/ are playing has/has a balloon/ one asks other to play	0	1
Initiating Event	D pulling wagon/D has a balloon /D shows/gives R his balloon	0	2
Internal Response	R wants / is interested in balloon	0	1
Internal Plan	R decides to take the balloon	0	1
Attempt	R takes balloon / unties the balloon / makes balloon fly / D gives R a turn	0	2
Outcome	balloon flies into tree / R untied balloon or let it go	0	2
Reaction of Dog	D angry/yells/stares at balloon	0	1
Reaction of Rabbit	R feels bad/embarrassed/scared / R stares at balloon/says oops	0	1

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Reaction - both/unknown	“they” are unhappy [code only as replacement for Reaction of Character 1 or 2; there should not be more than 2 reactions total]	0	1
Character 3 (C3)	policeman / other rabbit /other female / her mom / her sister	0	1
Initiating Event	C3 shows up/comes over / D sees C3 / C3 sees balloon in tree / C3 asks what happened	0	2
Internal Response	D/R hopes C3 can help / C3 wants to help	0	1
Internal Plan	D/R decides to ask for help/explains what happened /asks C3 to get balloon / policeman decides to try NOT: D talks to C3 (without specifying what about)	0	1
Attempt	C3 tries to get balloon / reaches for balloon	0	2
Outcome	C3 can’t reach balloon / balloon was too high/stuck	0	2
Reaction C1	R upset / sad / worried / cries / stares at balloon	0	1
Reaction C2	D upset / feels bad / looks sheepish/ sad	0	1
Reaction C3	C3 disappointed / shrugs / says she can’t reach it	0	1
Reaction of both/unknown	they” are disappointed/feels bad [code only as replacement for Reaction of another character; there should not be more than 3 reactions total]	0	1
Character 4 (C4)	other policeman / other rabbit / other male / her father / her brother /other person	0	1
Initiating Event	C4 comes over / has ladder	0	2

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Internal Response	C4 wants to help / knows how to get balloon / offers to help	0	1
Internal Plan	C4 decides to try / has idea / says he will get it D/R/C3 asks C4 to get it	0	1
Attempt* (trying or actually getting balloon)	C4 reaches for balloon/ is going to get it /climbs the ladder to get it/ tries to get it /C4 gets balloon	0	2
Outcome* (him giving the balloon back)	C4 gives balloon to D / D has balloon	0	2
Reaction of Giraffe	D happy / amazed / excited / hugs balloon/ says thanks	0	1
Reaction of Elephant 1	R happy / relieved / feels better / says thanks	0	1
Reaction C4	male worker relieved / pleased	0	1
Reaction of both/unknown	“they” are happy/excited / say thanks [code only as replacement for Reaction of another character; there should not be more than 3 reactions total]	0	1
Total Score:			

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

Appendix 2.

First Mentions Scoring Sheet for Story RD

Child's Name: _____ Age: _____ Date: _____

Please read the section of the Edmonton Narrative Norms Instrument Manual on scoring FM before using this sheet.

Character	Score as 3	Score as 2	Score as 1
Dog – story RD	<p>a/this _____ (e.g., <i>a dog, this bunny</i>)</p> <p>name (e.g., <i>Doug</i>)</p> <p>possessive + noun (e.g., <i>his friend</i> if 'she' already introduced)</p> <p>another animal the other animal (if C mentioned 2 animals and one animal mentioned previously)</p> <p>Pronoun I/me if the child puts him/herself in the story, e.g., "I was at the pool with my friend"</p>	<p>the/that _____ (e.g., <i>the dog, doggie</i>)</p> <p>a [invented word], e.g., <i>a deedee</i></p> <p>someone / somebody</p> <p>possessive + noun (if other character not yet introduced)</p> <p>another/the other _____ (e.g., <i>the other animal</i> if no animal mentioned previously)</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word], e.g., <i>the geegee</i> (an invented name would be scored as 3)</p>
Rabbit – Story RD	<p>a/this _____ (e.g., <i>a rabbit/bunny</i>)</p> <p>name (e.g., <i>Peter</i>)</p> <p>possessive + noun (e.g., <i>her friend</i> if 'he' already introduced)</p>	<p>the/that _____ (e.g., <i>the rabbit/bunny</i>)</p> <p>a [invented word] someone / somebody</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word] (an invented name would be scored as 3)</p>

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

	<p>another _____ (e.g., <i>another animal</i> if other character introduced as animal)</p> <p>the other _____ (e.g., <i>the other animal</i> if C mentioned 2 animals and one animal mentioned individually previously)</p> <p>Pronoun I/me if the child puts him/herself in the story, e.g., "I was at the pool with the giraffe"</p>	<p>possessive + noun (if other character not yet introduced)</p> <p>another/the other _____ (e.g., <i>the other animal</i> if no animal mentioned previously)</p>	
Balloon – Story RD	<p>a/this _____ (e.g., <i>a balloon, a toy</i>)</p> <p>possessive + noun (e.g., <i>her toy, the bunny's balloon</i>)</p>	<p>the/that _____ indefinite vague or empty term, e.g., <i>a thingy/something</i></p> <p>a [invented word]</p>	<p>pronoun (<i>it</i>) definite vague or empty term, e.g., <i>the thingy</i></p> <p>the [invented word]</p>
Policeman – Story RD	<p>a/this _____ (e.g., <i>a policeman, a girl</i>)</p> <p>his/her/their [family member] (e.g., <i>daddy, brother</i> if clear whose family member)</p> <p>name (e.g., <i>Mr./Mrs. Policeman</i>)</p>	<p>the/that _____ (including <i>the policeman</i> e.g., <i>the daddy</i>, unless main characters were introduced as brother and sister)</p> <p>a [invented word]</p> <p>someone / somebody</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word] (an invented name would be scored as 3)</p> <p>the rabbit/bunny (if first rabbit was introduced as a/the rabbit)</p>

EQUIVALENCE OF PICTURE SETS FOR ELICITING STORIES

<p>Man with ladder – Story RD</p>	<p>a/this _____ (e.g., <i>a man, a bunny, a person</i>)</p> <p>another _____ (e.g., <i>policeman, bunny/rabbit or girl</i> if at least one previous character identified with same term)</p> <p>name (e.g., Mr. Ladder Man)</p> <p>someone / somebody</p> <p>the _____ + relative clause (if a plausible role, e.g., <i>the person who climbs</i>)</p>	<p>the/that _____ (e.g., <i>the man, the person who carries the ladder</i>)</p> <p>a [invented word]</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word] (an invented name would be scored as 3)</p> <p>the bunny/rabbit/ the boy bunny/rabbit (if first or second rabbit introduced as a/the [girl] rabbit)</p>
<p>Ladder – Story RD</p>	<p>a/this _____</p> <p>possessive + noun (e.g., <i>his ladder</i>)</p>	<p>the/that _____ indefinite vague or empty term, e.g., <i>a thingy/something</i></p> <p>a [invented word]</p>	<p>pronoun definite vague or empty term, e.g., <i>the thingy</i></p> <p>the [invented word]</p>