GRIFFITH UNIVERSITY

The use of the Renfrew Bus Story with 5-8-year-old Australian children

Author

Westerveld, Marleen, Vidler, Kath

Published

2015

Journal Title

International Journal of Speech-Language Pathology

Version

Accepted Manuscript (AM)

DOI

10.3109/17549507.2015.1024168

Rights statement

© 2015 Taylor & Francis. This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Speech-Language Pathology on Volume 17, 2015 - Issue 3, available online: https://doi.org/10.3109/17549507.2015.1024168

Downloaded from

http://hdl.handle.net/10072/168314

Griffith Research Online

https://research-repository.griffith.edu.au

The use of the Renfrew Bus Story with 5- to 8-year-old Australian children

Marleen F. Westerveld

Kath Vidler

RUNNING HEAD: THE USE OF THE BUS STORY IN AUSTRALIA

Correspondence to

Marleen Westerveld, PhD

Senior Lecturer Speech Pathology

School of Allied Health Sciences

Griffith University

Gold Coast, QLD 4222

Phone: 07 5678 7658

M.westerveld@griffith.edu.au

Keywords: Bus Story, Frog Where Are You, school-age children, story retelling, language sample analysis

This paper has been published:

Westerveld, M. F., & Vidler, K. (2015). The use of the Renfrew Bus Story with 5- to 8-year-old Australian children. *International Journal of Speech-Language Pathology, 17*(3), 304 - 313. doi:10.3109/17549507.2015.1024168

Abstract

Purpose: Research into the language sampling practices of Australian speech pathologists suggests the *Bus Story Test* (Renfrew, 1995) is a frequently used standardised tool for eliciting story retelling samples for screening, diagnosis, goal-setting, and progress monitoring purposes. Because this task has not been normed on an Australian population, this study investigated the usefulness of the *Bus Story* for young school-age Australian children.

Method: 125 Australian primary-school children (ages 5;3 – 8;9) participated in the *Bus Story* task. Children attending Year 2 also retold the story *Frog Where Are You* (Mayer, 1969). Children's performance was analysed for measures of information and sentence length (*Bus Story* only), story length, MLU, number of different words, and clausal density.

Result: Performance on the *Bus Story* improved with year-of-schooling for all measures. Between 21 and 64% of the children performed below expectations on information or length scores when using the published norms. The retell task *Frog Where are You* elicited longer samples, containing a higher number of different words.

Conclusion: Using the published *Bus Story* norms will potentially result in over- identification of language impairment. The retell task *Frog Where are You* may provide a useful alternative for assessing young school-aged children's story retelling ability.

The importance of oral narrative ability to academic success has been well established (see Boudreau, 2008, for a review). Oral narrative competency during the preschool period is an important predictor of future academic performance (Bishop & Edmundson, 1987), and difficulties in oral narrative retelling or comprehension during the school years have been directly linked to reading comprehension deficits (Oakhill & Cain, 2012; Westerveld, Gillon, & Moran, 2008). During the early school years, children are exposed to oral narratives on a daily basis, which is reflected in the current Australian Curriculum (Australian Curriculum Assessment and Reporting Authority [ACARA], 2012). To illustrate, during their Preparatory year (Prep or Year 0), children are expected to provide "a simple, correctly-sequenced retelling of narrative texts" (p. 25); In Year 1, children are asked to "compare different kinds of images in narratives" (p. 29), and in Year 2, children are asked to "explore how language is used to construct characters and settings in narrative and to use grammatical knowledge to predict likely sentence patterns when reading more complex narratives" (p.40). Because children with language / learning difficulties often struggle in comprehending and producing oral narratives (Merritt & Liles, 1987; Pearce, McCormack, & James, 2003; Westerveld et al., 2008), it is not surprising that formal appraisal of these children's oral narrative ability becomes an integral part of the speech pathologist's assessment battery (Westerveld & Claessen, 2014).

A recent survey investigating language sampling practices of 257 Australian speech pathologists working with paediatric populations found that the majority (90.8%) of clinicians routinely collected spontaneous language samples from their clients (Westerveld & Claessen, 2014). Clinicians indicated using a range of contexts (free play, conversation, story re/tell, personal narratives, expository) for eliciting spontaneous language samples, depending on the age of their clients. As expected, for school-age clients, conversation and fictional story re/telling were the most frequently used contexts. Although most respondents reported using informal procedures, such as pictures or objects to collect spontaneous language samples, 62% reported using a standardised test, such as the *Bus Story Test* (Renfrew, 1995), the *Test of Narrative Language* (Gillam & Pearson, 2004), or the *Expression, Reception, and Recall of Narrative Instrument* (Bishop, 2004). Interestingly,

the *Bus Story* was by far the most frequently used standardised test, especially by clinicians working in Victoria and Western Australia (Westerveld & Claessen, 2014). This seemed somewhat surprising, as the *Bus Story* has not been normed on an Australian population. Moreover, specificity of this test in detecting language impairment is weak (Pankratz, Plante, Vance, & Insalaco, 2007). Pankratz et al. (2007) investigated the diagnostic and predictive validity of the *Bus Story* with 64 children (half of whom had an identified language impairment), aged between 4;0 and 5;9 who resided in the Tucson (US) area. Results suggested that use of the *Bus Story* might result in over-identification of language impairment in children with typical development.

Standardisation of the *Bus Story* occurred during 1993-1994 in South East England, based on the performance of 573 children, aged between 3;6 and 8;0, of whom a large proportion lived in "Social Class III areas" (p. 18). It seems that children were originally recruited from one social class area only, and there is no mention of eligibility or exclusion criteria (e.g., identified language impairments, English proficiency, etc). Although English is the main language spoken in both England and Australia, it is not clear if the performance of a group of children from South East England (collected nearly 20 years ago) will be representative of an Australian school-age population. Previous cross-cultural comparisons of narrative language performance has yielded mixed results (Westerveld & Claessen, 2009; Westerveld, Gillon, & Miller, 2004; Westerveld & Heilmann, 2012). To illustrate, Westerveld and Claessen (2009) compared the story retelling performance of 5- and 6-year-old children from New Zealand (NZ) and Western Australia and found significant differences in grammatical accuracy (percentage of grammatically correct utterances) between the groups, with the NZ children showing better performance. There were no differences on measures of story length, semantic diversity, or mean length of utterance. More recently Westerveld and Heilmann (2012) compared story retelling samples produced by 6- and 7-year-old children from NZ and the US. Once again, there were no group differences on measures of story length, semantic diversity, or mean length of utterance. However, there were significant group differences on verbal fluency (percent mazed words) and rate (words per minute), with the NZ children showing better verbal

fluency, but a slower speaking rate compared to their US peers. Taken together these results highlight the need for caution when adopting overseas norms for describing spoken language performance in young school-age children, particularly those measures that may indicate impaired performance in verbal fluency or grammatical accuracy (see also Westerveld, 2011).

Performance on the *Bus Story* during the preschool years has been shown to have good validity in predicting language and academic outcomes three years later, especially in children with language impairment (Bishop & Edmundson, 1987; Pankratz et al., 2007). In a landmark study, Bishop and Edmundson showed that a child's ability at age 4 to relate the main events of a story in the correct sequence is a good prognostic sign (Bishop & Edmundson, 1987, p. 170). In fact, performance on the *Bus Story* Information score at age 4 correctly predicted classification in 83% of the children (as language impaired or non-language impaired) 18 months later. The *Bus Story* Information score is calculated by awarding points to the correct use of 'information points' (or their synonyms) depicted in the model story. This information needs to be provided in the right order, and needs to contain the correct referent (that is, it needs to be clear who or what the 'Agent' is). Points are deducted if the Agent is not correctly specified. Although the Information score is often regarded an indication of the child's semantic abilities (e.g., Bishop & Edmundson, 1987), to obtain high marks, the child also needs to possess adequate syntactic abilities as well as some knowledge about the typical structure of a story. Furthermore, scoring of the information score is not straightforward (Hayward, 2008). The manual provides few examples to guide the marker, and no information is given on inter-rater reliability. To better understand the composition of the *Bus Story* Information score, this study investigated the correlations between the Information score and other more easily generated language measures tapping verbal productivity (number of utterances), semantics (number of different words), and syntax (MLU and clausal density).

Although the *Bus Story* is frequently used by clinicians, it is not the only available narrative retell task. One task that has received much attention in the research literature is *Frog Where Are*

You (Mayer, 1969), a wordless picture book about a boy whose pet frog goes missing one night. Previous research indicates this task is appropriate for a) typically developing children from a wide range of countries, such as New Zealand, the US, Canada, and the UK (Botting, 2002; Heilmann, Miller, Nockerts, & Dunaway, 2010; John, Lui, & Tannock, 2003; Westerveld & Heilmann, 2012); b) adolescents with or without language impairment from Australia (Reed, Patchell, Coggins, & Hand, 2007); as well as c) Spanish-speaking children learning English as a second language (Heilmann et al., 2008). To investigate the usefulness of this task for young school-aged Australian children, the current study administered the task to all Year 2 students and compared their performance on this *Frog story* task to their performance on the *Bus Story*.

The current study

To investigate the usefulness of the *Bus Story* for young school-age Australian children, this study aimed to answer the following questions:

1. How do children attending the first three years of primary school in Australia perform on the *Bus Story*?
2. How does the performance of the Australian children compare to the published *Bus Story* norms?
3. What are the correlations between the *Bus Story* measures (Information and Length) and general language measures of UTT, NDW, MLU, and clausal density?
4. How does performance on the *Bus Story* compare to performance on another story retell task, *Frog Where Are You*, for children attending Year 2?

Methods

Participants

Ethics approval for this project was granted by the Griffith University Human Ethics Committee (PES/31/12/HREC). Approval was also granted by the Department of Education and Training, Queensland Government (550/27/1258). With approval from the district senior speech pathologists,

20 practising speech pathologists agreed to assist in the project. These speech pathologists invited principals of primary schools in their geographical area (from rural, regional, and metropolitan areas) to participate. Of the schools that agreed to participate, teachers were asked to identify children who 1) attended Prep Year, Year 1, or Year 2; 2) spoke English as their first language; 3) were progressing normally at school; and 4) had no history of speech and/ or language impairments. Consent forms were sent home to these children via the teachers, and from the children for whom consent to participate was obtained, participants were randomly selected by the first author (only initials, gender, year of schooling, and ethnicity data were provided by the speech pathologists), making sure there was an equal distribution of girls and boys, and an equal number of participants across the three year-levels.

Although a total of 127 children participated in this study, 125 transcripts were available for analysis (see next section). These children attended Prep (n = 42), Year 1 (41), and Year 2 (n = 42) and were from Australian (85.2%), Aboriginal and Torres Strait Islander (4.0%), Pacific Island (.8%), other (3.2%), or non-specified (6.4%) ethnic backgrounds, as indicated by their parents on the project consent forms. Based on the most recent Census data (Australian Bureau of Statistics, 2010), there may have been an overrepresentation of children whose parents identified their children as *Australian*, although we did not ask for more specific information such as birthplace of the parents. However, children attended schools across Queensland (regional: 54; City: 71); these schools represented the full range of socio-economic areas, as indicated by their decile rankings (1 – 10), based on the schools' postcodes and Socio-economic Indexes for Areas data (SEIFA; Australian Bureau of Statistics, 2008). Distribution of students was as follows: deciles 1 – 3: 37 students; deciles 4 – 7: 56 students; deciles 8 – 10: 32 students). There were 61 girls and 64 boys.

Procedure

All children were seen individually by a certified practising speech pathologist. Assessments lasted approximately 40 minutes and took place in a quiet room at the child's school. All sessions were

audio-recorded using a variety of digital voice recorders. The sessions included the following tasks: a warm-up task, conversation, a story retell (see below), personal narratives, an expository task (year 2 children only), and the *Bus Story* (Renfrew, 1995). For details regarding the warm-up activity, the conversation, and the personal narrative tasks, please see Westerveld et al. (2004). Prep and Year 1 children retold the story *Ana Gets Lost* (Swan, 1992), whereas Year 2 children retold the story *Frog Where Are You* (the Frog story; Mayer, 1969). Administration of the *Bus Story* was counterbalanced to control for practice effect: approximately half the children (n = 55) completed this task after the warm-up, and half the children completed this task as the final activity (n = 70). There were no age-differences between these two groups of children (p = .748).The *Bus Story* and the Frog story retelling task are the focus of the current study.

The *Bus Story* (Renfrew, 1995). In this task, the examiner read the story, while the child followed along with the pictures in a wordless book (four pages containing three pictures each). After listening to the story, the child was asked "Now you tell me the story. Once upon a time, there was a....?" (p. 5). Following the administration guidelines, only minimal or indirect prompts were given, when needed. For example "and then?" or "so...? The model story contained: 15 utterances, MLU: 12.4, number of different words (NDW): 102, and clausal density (CD; total number of clauses divided by the number of utterances): 1.6.

The Frog story: *Frog Where Are You* (Mayer, 1969). The procedures were adopted from Heilmann, Miller, Nockerts, and Dunaway (2010). Following this protocol, children (Year 2 only) were asked to listen to the pre-recorded story using a laptop or desktop computer, while following along with the pictures in a wordless picture book. After listening to the audio, children were asked to retell the story using the book as an aid "Now I would like you to use your own words to tell the story". If the child had difficulty starting the task, examiners were allowed to use the prompt "One day...". Only nonspecific cues were used, if needed, to encourage the child to continue the story

(e.g., *Tell me more; and then?*). The model story contained 55 utterances, MLU: 11.2; NDW: 205, and clausal density (CD): 1.32.

Transcription and coding

All story retelling samples were digitally recorded and sent to the first author's research lab for transcription and coding. All samples were transcribed by trained research assistants using standard coding conventions for Systematic Analysis of Language Transcripts – New Zealand (SALT-NZ) software (Miller, Gillon, & Westerveld, 2012). Following transcription, inspection of the transcripts revealed two instances where the intelligibility of the transcripts was below 75% (two Prep children in the *Bus Story* condition); these transcripts were not included in the analysis. In addition, one Frog story was discarded because of poor (noisy) tape quality. Utterances were segmented into communication units (C-unit), as defined by Loban (1976) as containing a main clause with all its subordinate clauses. However, following standard SALT conventions, elliptical phrases, for example in response to the examiner's prompts were also considered a C-unit. In addition, sentence fragments were counted as separate C-units when the final intonation contour of the utterance indicated that a complete thought has been spoken. Only complete and intelligible (C&I) C-units were used for analysis. All reformulations, repetitions, and disfluencies were placed in parentheses and considered mazes. In addition to transcribing the language samples, the research assistants coded the transcripts for dependent clauses (to calculate clausal density). Only finite clauses (containing a subject and a predicate) were included, following Nippold et al.'s (Nippold, Hesketh, Duthie, & Mansfield, 2005) procedures. After all stories were transcribed the transcripts were analysed using SALT-NZ, which produced a rectangular data file summarising each dependent measure (see section below) for each transcript. *Measures*

The following two measures were calculated for the *Bus Story*, following the guidelines in the test manual (Renfrew, 1995):

Bus Story Information score. Points were awarded when the child included the right information (e.g., *the bus ran away; they made funny faces at each other*), using the correct referent (e.g., *bus, man, train*) so there was no ambiguity (i.e., when there is a change of *actor*, such as the train, bus, man). In addition, the information needed to be mentioned in the right order.

Bus Story (sentence) Length score. There are two sentence-level measures: 1) A5LS: the average number of words in the five longest sentences (excluding and, then, and well); 2) Subordinate clauses (SC): The number of subordinate clauses in the story transcript.

The following measures were calculated for both the Bus stories and the Frog stories:

Clausal Density (CD). Clausal density was calculated by adding the number of identified dependent clauses to the number of utterances to obtain the total number of clauses, and dividing the total number of clauses by the number of utterances.

MLU. Mean length of communication unit (MLU) was computed automatically using SALT-NZ and used as a measure of grammatical ability (Scott & Windsor, 2000; Watkins, Kelly, Harbers, & Hollis, 1995).

UTT. Total number of utterances in communication units was calculated automatically using SALT-NZ to measure verbal productivity (Scott & Windsor, 2000).

NDW. Number of different words was calculated automatically using SALT-NZ as a measure of semantic diversity (Watkins et al., 1995).

Reliability

To ensure accuracy and completeness of transcription and coding, the first author (a) checked all the transcripts for spelling, utterance segmentation, or dependent clause coding errors, (b) listened to the sound files if the transcripts contained unintelligible segments or utterance segmentation issues, (c) made corrections where needed. Any disagreements were resolved through discussion so that

100% agreement was reached for utterance segmentation and dependent clause coding. Next, 10% of the transcripts were randomly selected and scored on the *Bus Story* Information score by a second researcher. Krippendorff alpha coefficients (Krippendorff, 1980) were calculated to document agreement between the two raters. This procedure accounted for chance agreement and the degree of differences between the judgments; for example, interjudge scores of 30 and 31 have a higher level of agreement than scores of 30 and 19. Krippendorff's alpha using ordinal scaling was .94 for the information score, indicating good agreement.

Results

Data were analysed using statistical software SPSS (PASW, 2012). First the results were examined to determine if there were group differences between the children who retold the *Bus Story* at the start of the session and those who completed this task at the end of the session. For each year-level, a series of univariate analysis of variance (ANOVA) tests was completed using each of the language sample measures as the dependent variable and condition (first task versus last task) as the between-subjects variable. Results indicated no significant group differences ($p < .05$) on any of the measures for Prep and year 1. For year 2 children, the group who performed the task last produced significantly longer utterances ($p = .045$) and significantly longer A5LS ($p = .046$).

Performance on the Bus Story by year-level

To answer research question one, we investigated performance at each year-level on both 1) the *Bus Story* measures of Information and Length, and 2) the language sample measures of UTT, NDW, MLU, and CD. To determine the normality of the distribution, several analyses were conducted. First, mean and median, as well as the mean scores achieved by the participants in the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles were calculated based on weighted averages (see Table I). As shown, most mean scores were close to the median scores, indicating distributions around the centre. Skewness statistics (i.e.,tails) and kurtosis (peaked or flat distributions) were also calculated. Values of > 2

standard errors of skewness or kurtosis are considered significant (Tabachnick & Fidell, 1996) and are indicated in Table I. For example, in Prep normal distributions were found for Information, SC, MLU, and NDW. A5LS showed a symmetrical but peaked distribution; UTT showed scores clustering on the lower end; CD showed both a peaked distribution and scores clustering at the low end of the scales (floor effect).

As shown in Tables I and II, there was an improvement in performance by year level on all measures. Multivariate Analyses of Variance (MANOVA) revealed significant effects for Year-level on all measures (*p*'s < .05) with mild to moderate effect sizes. Effect sizes were calculated as partial eta squared (ŋ2) to document the proportion of the variation in each measure that is associated with the children's year of schooling (Lakens, 2013). The amount of variance explained by year level ranged between 4.9% for UTT and 18.8% for NDW. Post hoc analyses (Tukey) indicated a significant improvement (*p* <.05) from Prep to Year 1 on all measures, except the number of utterances, but not from Year 1 to Year 2 (see superscript indicators in Table II).

Insert tables I and II here

Comparing performance to the Bus Story norms

Children's performance on the *Bus Story* Information score and the *Bus Story* (sentence) Length measures (A5LS and SC) was compared to the published *Bus Story* norms. Following the manual's procedures, the participants were divided into age groups using 6-month intervals. Because there were only five participants in the youngest age group (5;0 – 5;5), this group was not considered for the present analysis. First, the participants' mean scores were compared to the means reported in the manual (Renfrew, 1995; pp. 10-11). As shown in Table III, the Australian children scored lower on the *Bus Story* Information score compared to the norms for every age group. Slightly different results were found for performance on the A5LS Sentence measure. Although most age groups performed at a similar level compared to the norms, the oldest age group performed below expectations. When

considering the mean number of subordinate clauses, performance of the Australian children seemed to match the reported norms.

To investigate how many children were considered to score below expectations, the percentage of children scoring below 1SD was calculated for the *Bus Story* Information score. As shown in Table III, between 21% (7;0 – 7;5 years) and 41% (8;0 – 8;5 years) of the Australian children would be considered to score significantly below the norms. For A5LS we calculated the percentage of children scoring below the mean score (as no standard deviations are reported in the manual). Once again, just over 21% of the 7;0 – 7;5 year old children scored below the mean; this percentage increased to 64.7% for the 8-0 – 8;5 year old group. Finally, the mean number of subordinate clauses was considered. Between 17.4% (7;0 – 7;5 years) and 53.8% (6;6 – 6;11) of the children scored below expectations.

Insert Table III here

Correlations between the Bus Story measures and more general language measures

To answer research question number three, bivariate (two-tailed) correlations were calculated between the *Bus Story* measures (Information, A5LS, and SC) and general language measures of UTT, NDW, MLU, and CD. As shown in Table IV, there were significant correlations between all three *Bus Story* measures and more general language measures. In particular, there were large correlations between Information and NDW ($r = .737$, $p < .001$) and between A5LS and MLU ($r = .887$, $p < .001$). As expected the correlation between SC and CD was also large ($r = .925$, $p < .001$).

Insert Table IV here

To investigate if the more general measures of UTT, NDW, and MLU (all of which are computed automatically using SALT), would predict children's performance on the *Bus Story* Information and A5LS measures, regression analyses were performed. When using children's

Information score as the dependent variable and entering UTT, NDW, and MLU as predictors, it was found that the best model F(1,123) = 146.028 *p* <. 001 included only NDW, which was significantly related to Information score (b = .409), t(124) = 12.084, *p* <.001 and explained 54.3% of the variance. When entering A5LS as the dependent variable and entering UTT, NDW, and MLU as predictors, the best model included both MLU and NDW F(2,122) = 324.913, *p* <. 001, with both MLU (b = 1.358), t(124) = 14.367, p <.001 and NDW (b = .053), t(124) = 6.529, *p* <.001 significantly related to A5LS and combined explaining 84.2% of the variance (MLU: 78.7 %; NDW: 5.5%).

Bus Story performance compared to performance on the Frog Where Are You story

Finally, we compared the children's performance (Year 2 of schooling only) on the *Bus Story* to their performance on a different story, *Frog Where Are You* (FWAY; Mayer, 1969) on measures UTT, NDW, MLU, and CD. Although all children participated in this story retelling task, the quality of one recording was too low for accurate transcription. As a result, the performance of 41 children is reported here. As shown in Table V, paired samples t-tests revealed that children produced significantly longer stories, containing a higher NDW in response to the Frog story. There were no differences on measures of MLU or CD. Correlational analyses showed small but significant correlations between children's performance on the two stories on number of UTT (*r* = .323, *p* = .037). Correlations were larger for NDW (*r* = .661, *p* < .001), MLU (*r* = .541, *p* = .002) and CD (*r* = .445, *p* = .003).

Insert Table V here

Discussion

This study investigated the use of the *Bus Story* (Renfrew, 1995) with young school-age children attending primary schools in Queensland, Australia. A total of 127 children participated from Prep (Year 0) to Year 2. All children appeared happy to participate in the task and 125 children produced analysable stories. There was an effect for timing of the task for Year 2 children only on measures of

utterance length, with the group who produced the Bus story last producing significantly longer utterances than the group who was asked to retell the *Bus Story* at the start of the session. This most likely reflects a practice effect and indicates that eight-year-old children with typically developing language may benefit from producing multiple retellings during a single session. Overall, children's performance showed a steady improvement by year group, but this improvement was only statistically significant from Year 0 to Year 1, except for the number of utterances. These results are consistent with those from previous research investigating story retelling performance in young school-aged children (Westerveld et al., 2004), which indicated strong language growth during the early years of schooling. There was variability in performance, however, with the number of utterances produced by the children ranging from 10 to 25 utterances (in Prep), 9 – 23 in Year 1, and 10 to 28 in Year 2. Further inspection of the results showed few outliers. As children with language difficulties were excluded from the study, these results indicate that the range in performance observed most likely reflects normal variability in spontaneous language performance (see also Justice et al., 2006, for a discussion).

To determine the potential of using these results for local norming purposes, distribution statistics were calculated. As shown in Table I, not all measures were normally distributed. This is in line with previous research investigating oral narrative abilities in school-age children (Justice et al., 2006; Westerveld & Gillon, 2010), which showed substantial variability in performance across the language sample measures. This inherent variability in oral narrative performance potentially affects its suitability for norm-referencing. In clinical practice, however, clinicians may use the reported results to determine a child's relative performance compared to his or her peers, provided the child is similar to the group of children who participated in this study. Future research should include a group of children with language impairment to determine the sensitivity of the *Bus Story* task in detecting or confirming language ability status.

When comparing the Australian children's performance to the published *Bus Story* norms on measures of Information and Sentences (A5LS and subordinate clauses), some interesting findings emerged. On the Information score, the Australian children scored consistently lower than the published norms, and between 21% and 41% of the children scored significantly (> 1 SD) below the norms. These findings clearly indicate that using the published norms with Australian children may result in over-identification of children with language difficulties. These results are in line with those by Pankratz et al. (2007), who found that the use of the *Bus Story* resulted in over-identification of children with language impairment (i.e., specificity of 75% for both the Information and the Sentence length scores). Interestingly, these authors used the American version of the *Bus Story*, which was normed on a different sample of 418 children from the US (Cowley & Glasgow, 1994).

In light of the time and effort required in scoring the *Bus Story* transcripts on the Information and A5LS measures, and in interpreting the composition of linguistic skills encompassed in the Information score, we investigated the correlations between the *Bus Story* measures of Information and Sentences (A5LS and subordinate clauses), and the more transparent and easy to calculate measures of story length (UTT), MLU, semantic diversity (NDW), and clausal density (CD). Results revealed significant, large correlations between Information and NDW (.768), between A5LS and MLU (.888), and between SC (number of subordinate clauses) and CD (.922). Furthermore, NDW explained 54.3% of the variance in Information, whereas a combination of MLU (78.8%) and NDW (6.1%) explained 84.9% of the variance in A5LS. Overall, these results have two practical implications. First, calculating NDW and MLU may serve as an adequate proxy of a child's performance on the *Bus Story*. Second, the Information score reflects more than a child's semantic skills. In fact, when semantic diversity, as capsulated by NDW (Watkins et al., 1995), and syntactic ability (MLU) are combined only 54% of the variance is explained. This is important as it is the children's performance on the Information score that has been flagged as an important predictor of later language development and academic performance (Bishop & Edmundson, 1987). It seems reasonable to conclude that the Information score reflects other linguistic skills needed to construct

a story, such as sequencing of events and coherence within and across sentences achieved through appropriate use of pronouns and referents. Future efforts should be directed at creating a more transparent rubric for scoring a child's performance on the *Bus Story* that reflects the current Information score but provides more detailed direction for intervention, if required.

The final question asked if children's performance on the *Bus Story* differed from their performance on a second story retelling task, using the book *Frog Where Are You* (Mayer, 1969). As expected, children produced longer stories, containing a higher NDW in response to the Frog story. In contrast, there were no significant differences on measures of grammatical ability or complexity. This is consistent with previous research investigating the effects of the length of the model story on children's spoken language performance (Holloway, 1986). The Frog story is significantly longer and contains more different words (55 utterances, containing 205 NDW) than the *Bus Story* (UTT: 15; NDW: 102). In addition, the Frog story contains 30 pictures whereas the *Bus Story* only has 12. In contrast, the MLU of the two stories are similar (11.2 and 12.4), with the Frog story containing a lower CD (1.32, compared to 1.6). From a clinical perspective, eliciting a longer sample would be more useful when detailed analysis of a child's linguistic performance is needed for goal setting and intervention purposes. Although only 37% of the children produced at least the 50 recommended utterances for this type of analysis in response to the Frog story (Miller, 1996), only 5% produced less than 28 utterances, which was the maximum number of utterances produced in response to the *Bus Story*.

Limitations

All participants in the current study reside in Queensland and it is not clear if results would generalise to children attending schools in other states or territories in Australia. With the introduction of a common Australian curriculum framework (Australian Curriculum Assessment and Reporting Authority [ACARA], 2012), however, similar levels of performance would be expected of children attending Prep to Year 2 across Australian, regardless of the child's home state or territory.

Another limitation is that a convenience sampling method was used, as opposed to using full population sampling in which no exclusionary criteria are applied to participation (see McFadden, 1996, for a discussion). However, the use of databases containing language samples from typically developing speakers has been endorsed by leading researchers in this field (e.g., Johnston, 2006; Paul & Norbury, 2012). In clinical practice, such databases are useful for describing performance expectations for typical children (Heilmann, Miller, & Nockerts, 2010).

Summary and clinical implications

The results from this study clearly showed that the *Bus Story* was effective in eliciting story retelling samples from young school-age Australian children. All children participated in the task and produced between 9 and 28 analysable utterances. The findings also revealed, however, that caution should be taken when comparing children's performance to the published *Bus Story* norms, as this may result in over-identification of language impairment, especially in the older age groups. Based on the current findings, it may be appropriate to include the *Bus Story* in the speech pathologist's assessment battery to obtain a snapshot of the child's performance in a story retelling context on measures of MLU, NDW, and Information. The results reported in Table I can be used to determine a child's relative performance compared to his or her peers, provided the child is similar to the group of children who participated in this study. In addition, the results can be used as a baseline measure of performance to analyse progress following intervention. However, for a full description of a child's linguistic abilities, analysing a longer spontaneous language sample of at least 50 utterances is recommended, preferably across a range of discourse contexts (Hadley, 1998; Heilmann, Nockerts, & Miller, 2010; Miller, 1996).

Acknowledgements

This project was funded by an Emerging Researcher Grant, awarded to the first author by the School of Arts, Education and Law, Griffith University. The project was also supported by SALT Software LLC, who provided software support, but was not involved in any aspects of the study. In kind assistance

for this project was provided by the Department of Education and Training. The views expressed in this publication do not necessarily present the views of the Department. A sincere thank -you to all the speech pathologists for volunteering their time and assisting in the data collection.

References

Australian Bureau of Statistics. (2008). Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia – Data only, 2006. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012006?OpenDocument

Australian Bureau of Statistics. (2010). Yearbook Australia 2009-2010. Retrieved from http://www.abs.gov.au

Australian Curriculum Assessment and Reporting Authority [ACARA]. (2012). The Australian Curriculum - English. Retrieved from www.australiancurriculum.edu.au

Bishop, D. V. (2004). *Expression, Reception and Recall of Narrative Instrument (ERRNI)*. Oxford, UK: Harcourt Assessment.

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

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching & Therapy, 18*(1), 1-21.

Boudreau, D. (2008). Narrative abilities - Advances in research and implications for clinical practice. *Topics in Language Disorders, 28*(2), 99-114.

Cowley, J., & Glasgow, C. (1994). Renfrew Bus Story test (North American ed.). Centreville, DE: Centreville School.

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

Hadley, P. A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools, 29*, 132-147.

Hayward, D. V., Stewart, G. E., Phillips, L. M., Norris, S. P., & Lovell, M. A. . (2008). Test review: The Renfrew Bus Story. *Language, Phonological Awareness, and Reading Test Directory* (pp. 1-10). Edmonton, AB: Canadian Centre for Research on Literacy.

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28*(2), 178-188.

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41*(1), 84-95.

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme using narrative retells in young school-age children. *American Journal of Speech - Language Pathology, 19*(2), 154-166.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393-404.

Holloway, K. F. C. (1986). The effects of basal readers on oral language structures: A description of complexity. *Journal of Psycholinguistic Research, 15*, 141-151.

John, S. F., Lui, M., & Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Canadian Journal of School Psychology, 18*(1-2), 91-113.

Johnston, J. (2006). *Thinking about child language*. Eau Claire, WI: Thinking Publications.

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology, 15*(2), 177-191.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(November, Article 863), 1-12.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

Mayer, M. (1969). *Frog, where are you?* New York: Penguin Putnam.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension

Journal of Speech and Hearing Research, 30, 539-552.

Miller, J. F. (1996). Progress in assessing, describing, and defining child language disorder. In K. N. Cole, P. S. Dale & D. J. Thal (Eds.), *Assessment of communication and language* (6th ed., pp. 309-324). Baltimore, MD: Paul H. Brookes.

Miller, J. F., Gillon, G., & Westerveld, M. (2012). *Systematic Analysis of Language Transcripts (SALT), New Zealand Version 2012 [computer software]*. Madison, WI: SALT Software LLC.

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research, 48*(5), 1048-1064.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91-121.

Pankratz, M. E., Plante, E., Vance, R., & Insalaco, D. M. (2007). The diagnostic and predictive validity of The Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools, 38*, 390-399.

PASW. (2012). PASW Statistics 20 [Computer Software]. Chicago, IL: IBM SPSS Statistics.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence. Listening, speaking, reading, writing, and communicating* (4th ed.). Baltimore: Mosby Elsevier.

Pearce, W. M., McCormack, P. F., & James, D. G. H. (2003). Exploring the boundaries of SLI: Findings from morphosyntactic and story grammar analyses. *Clinical Linguistics & Phonetics, 17*(4-5), 325-334.

Reed, V. A., Patchell, F. E., Coggins, T. E., & Hand, L. (2007). Informativeness of the spoken narratives of younger and older adolescents with specific language impairment and their counterparts with normal language. *Clinical Linguistics & Phonetics, 21*(11-12), 953-960.

Renfrew, C. E. (1995). *The Bus Story Test: A test of narrative speech* (3rd ed.). Oxford, UK.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339.

Swan, E. (1992). *Ko au na galo (Ana gets lost)*. Wellington, NZ: Learning Media, Ministry of Education.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). NY Harper Collins.

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research, 38*(6), 1349-1355.

Westerveld, M. (2011). Spontaneous language sampling and analysis in children: From research to practice. *ACQuiring Knowledge in Speech, Language and Hearing, 13*(2), 63-67.

Westerveld, M. F., & Claessen, M. (2009, May). *Children's spoken language skills in conversation and narration: A cross -Tasman comparison*. Paper presented at the Speech Pathology Australia Conference: Aspire, Adelaide.

Westerveld, M. F., & Claessen, M. (2014). Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology, 16*(3), 242-249.

Westerveld, M. F., & Gillon, G. T. (2010). Profiling oral narrative ability in young school-aged children. *International Journal of Speech-Language Pathology, 12*(3), 178 - 189.

Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech-Language Pathology, 6*(4), 195-208.

Westerveld, M. F., Gillon, G. T., & Moran, C. (2008). A longitudinal investigation of oral narrative skills in children with mixed reading disability. *International Journal of Speech-Language Pathology, 10*(3), 132-145.

Westerveld, M. F., & Heilmann, J. J. (2012). The effects of geographic location and picture cues on children's story retelling performance. *Asia Pacific Journal of Speech, Language, and Hearing, 15*(2), 129-143.

Table I. Performance on the *Bus Story*: Distribution statistics by year-level of schooling

Age Group	Mean	Median	10%	25%	50%	75%	90%	Skewness	Kurtosis
Prep (n = 42)									
Information	24.95	25.0	14	21	25	30	33	-.642	.119
A5LS	9.6	9.6	7	8.6	9.6	10.6	12.3	-.062	.572*
SC	3.1	2.5	1	1.75	2.5	4.25	6	1.03	1.25
Utterances	16.5	16.5	13	14	16.5	18	20	.776*	.847
MLU	7.2	7.3	6.0	6.4	7.3	8.0	8.5	-.080	.292
NDW	62	61	50	55	61	68	83	.425	.282
CD	1.15	1.14	1.0	1.06	1.14	1.22	1.34	1.496*	2.867*
Year 1 (n = 41)									
Information	29.37	31	20	26	31	34	36	-1.494*	3.954*
A5LS	11.3	11.2	8.8	9.7	11.2	13	14.1	-.262	.754
SC	4.5	5.0	1.2	3	5	6	8	.135	-.407
Utterances	17.3	18	12.2	15	18	19.5	21.8	-4.73*	.068
MLU	7.8	7.8	6.4	7.3	7.8	8.6	9.4	-.758	2.603*
NDW	70.2	70	58	64.5	70	78.5	84.8	-1.159*	2.679*
CD	1.23	1.22	1.07	1.11	1.22	1.33	1.43	.903*	.886
Year 2 (n = 41)									
Information	30.8	32	20	26.5	32	35	40.5	-.622	.153
A5LS	11.9	11.7	9	10.2	11.7	13.8	14.4	.017	-.366
SC	5.1	5	1.3	3	5	7	8.7	.783*	1.238*
Utterances	18.3	18	13.6	16	18	20	22.7	.421	.954
MLU	8.3	8.2	7.1	7.6	8.2	9.0	10.0	.362	.008
NDW	75.6	77	57.6	68	77	83.3	90	-.119	.916
CD	1.25	1.25	1.07	1.12	1.25	1.35	1.46	.630	-.103

Note: A5LS= average number of words of the 5 longest sentences; SC = number of subordinate clauses; MLU = mean length of utterance in words; NDW = number of different words; CD = clausal density. * distribution shows significant level of skewness / kurtosis (i.e., >2 standard errors of skewness or kurtosis)

Table II. Mean performance on the *Bus Story* by Year level (with SD) and range.

Measures	**Prep**	**Year 1**	**Year 2**	**Effect size***
	n = 42	**n = 41**	**n = 42**	**η^2**
Age (years;months)	6;0 (0;3)	6;11 (0;4)	7;11 (0;4)	.855
	5;3 – 6;5	6;4 – 7;7	7;4 – 8;9	
Bus Story Information	25 (6.7)[ab]	29.4 (7.0)[a]	30.8 (7.4)[b]	.116
	8 - 36	3 – 41	11 – 44	
Bus Story A5LS	9.6 (1.9)[cd]	11.3 (2.1)[c]	11.9 (2.2)[d]	.180
	5.0 – 14.2	5.0 – 16.4	7.0 – 17.0	
Bus Story SC	3.1 (2.2)[ef]	4.5 (2.4)[e]	5.1 (2.8)[f]	.108
	0 - 10	0 - 10	1 – 14	
Utterances	16.5 (2.9)[g]	17.3 (3.4)	18.3 (3.6)[g]	.049
	11 - 25	9 – 23	10 – 28	
MLU	7.2 (1.0)[hi]	7.8(1.1)[h]	8.3 (1.1)[i]	.152
	4.5 – 9.7	3.9 – 10.0	6.0 – 10.8	
NDW	61.6 (11.4)[jk]	70.2 (11.5)[jl]	75.6 (13.3)[kl]	.188
	36 – 86	29 – 87	38 - 108	
CD	1.15 (.13)[mn]	1.24 (.15)[m]	1.25 (.15)[n]	.080
	1.0 – 1.59	1.0 – 1.67	1.05 – 1.65	

Note: A5LS = Average of the 5 longest sentences; SC = number of subordinate clauses; MLU = mean length of utterance; NDW = number of different words; CD = clausal density. Measures with the same superscript letter are statistically significant ($p < .05$). For example MLU in Prep is significantly shorter than MLU in Year 1 ([h]), and MLU in Year 2 ([i]). The difference in MLU between Year 1 and Year 2 is not significant.*Effect sizes were calculated as partial eta squared.

Table III. Comparing the participants' performance to the *Bus Story* norms, including the percentage of participants scoring below expectations

Age group		Bus Story Information Score			Bus Story A5LS			Sub Clauses Mean		
	n	Australia	BUS Norms	% below 1SD	Australia	BUS Norms *	% below M	Australia	BUS Norms*	% below M
5;6 – 5;11	16	26.0 (7.38)	26.79 (6.48)	18.8	9.72 (2.11)	9	< 9	3.56 (2.68)	3	43.8%
		8 - 35			5 – 12.8		25.0	0 - 10		
6:0 – 6;5	30	26.13 (5.97)	28.66 (6.77)	26.7	9.74 (1.74)	9 - 10	< 9	3.1 (1.73)	3	43.3%
		12 - 36			5.8 – 14.2)		23.3	0 – 8		
6;6 – 6;11	13	27 (9.13)	30.42 (6.66)	23.1	10.72 (2.96)	10 - 11	< 10	3.8 (2.73)	3- 4	53.8%
		3 - 38			5 – 16.4		46.2	0 – 9		
7;0 – 7;5	23	31.04 (5.55)	32.6 (6.87)	21.7	12.15 (1.43)	11 – 12	< 11	5.4 (2.234)	4	17.4%
		20 – 41			8.8 – 14.2		21.7	0 – 10		
7;6 – 7;11	22	30.5 (7.02)	35.33 (7.31)	31.8	11.79 (1.97)	12 – 13	< 12	4.68 (2.41)	4 - 5	45.5%
		17 - 43			7 – 14.2		45.5	1 - 9		
8;0 – 8;5	17	30.65 (8.69)	37.2 (5.92)	41.2	12.07 (2.61)	13 – 14 - 15	< 13	5.41 (3.34)	5 - 6	52.9%
		11 – 44			8.8 - 17		64.7	1 – 14		

Note: Means (with SD in brackets) and range of scores are reported for the Australian participants. * Only the mean scores are provided in the *Bus Story* manual.

Table IV. Correlations between the *Bus Story* measures and more general language measures (n = 125)

	Info	**A5LS**	**SC**	**UTT**	**NDW**	**MLU**	**CD**
Info	---	.595**	.582**	.461**	.737**	.535**	.451**
A5LS		---	.793**	.171	.758**	.887**	.744**
SC			---	.101	.653**	.747**	.925**
UTT				---	.629**	-.005	-.161
NDW					---	.654**	.473**
MLU						---	.764**
CD							---

Note: Info = *Bus Story* information score; A5LS = Average of 5 longest sentences; SC = number of subordinate clauses; UTT = number of utterances; NDW = number of different words; MLU = mean length of utterance; CD = clausal density. * p = .005, ** p < .001

Table V. Group performance (Year 2 only, n = 41) on the *Bus Story* and the Frog story

	Bus Story **Mean (SD)**	**Frog story** **Mean (SD)**	**Stats**
UTT	18.3 (3.6)	46.0 (10.8)	t (1,40 = -18.55, p < .001)
NDW	75.6 (13.3)	128.1 (26.6)	t (1,40 = -17.19, p < .001)
MLU	8.4 (1.0)	8.5 (1.2)	p = .539
CD	1.25 (.15)	1.22 (.13)	p = .112

Note: UTT= number of utterances; NDW = number of different words; MLU = mean length of utterance; CD = clausal density.