Evaluating the Clinical Utility of the Profile of Oral Narrative Ability for 4-Year-Old

Children

Marleen F. Westerveld

Gail T. Gillon

Lynda Boyd

College of Education

University of Canterbury

Christchurch, New Zealand

Correspondence to:

Marleen F. Westerveld, PhD

School of Literacies and Arts in Education

College of Education

University of Canterbury

Private Bag 4800

Christchurch 8140

New Zealand

Email: m.westerveld@gmail.com

Phone: + 64 3 343 7771

Abstract

This study investigated if the story retelling and comprehension task *Ana Gets Lost*, that is frequently used with school-aged children, has clinical utility with a preschool population. The study also assessed the task's concurrent and predictive validity with norm-referenced tests of language performance. A total of 92 typically developing 4-year-old children participated. After 12 months, 57 children were available for a follow-up session. At each session, children listened twice to the story while looking at the pictures and then retold the story without the use of pictures. After the first exposure the children were asked comprehension questions to assess their oral narrative comprehension. Children's performance was analysed on measures of comprehension, narrative quality, semantics, morphosyntax, and verbal productivity to provide a Profile of Oral Narrative Ability (PONA). Results showed normal distribution of some of the measures and acceptable concurrent and predictive correlations with two norm-referenced measures of language ability. Although the results indicate the potential usefulness of this tool with preschool children, further research should investigate its potential as a screening measure of oral narrative performance.

Introduction

Oral language proficiency in preschoolers is a well-known predictor of future academic success, especially reading comprehension (e.g., Catts, Fey, Zhang, & Tomblin, 2001; Storch & Whitehurst, 2002). One way to investigate preschool language performance is by eliciting an oral narrative language sample. Assessing children's oral language skills in a narrative context allows insight into the child's ability to use language in a decontextualised manner, a skill that is vital for accessing the curriculum and participating in classroom interactions (Milosky, 1987). Story retelling ability in particular has been found to predict later language development and reading comprehension (Bishop & Adams, 1990; Bishop & Edmundson, 1987). It is therefore not surprising that research efforts have focused on finding ways in which to effectively and efficiently assess oral narrative ability, using a story retelling task, in preschool-aged children (Justice, Bowles, Pence, & Gosse, 2010; Pankratz, Plante, Vance, & Insalaco, 2007). The current study aimed to add to this growing body of research by investigating if a story retelling and comprehension task that is frequently used with school-aged children in New Zealand (Westerveld & Gillon, 2010) has clinical utility with preschool-age children.

An oral narrative can be defined as a monologue describing an experience or events that are chronologically sequenced (e.g., Engel, 1995). Oral narrative ability should be assessed in two modalities, comprehension and production (Boudreau, 2008; Skarakis-Doyle & Dempsey, 2008), and analysed at two levels: 1) macrostructure level, which considers the overall quality of the oral narrative, and 2) microstructure level, which takes into account the narrative's more 'local' features, including measures of grammar and vocabulary (Hughes, McGillivray, & Schmidek, 1997). From an educational perspective children need to develop adequate skills in all these aspects to communicate effectively in their preschool environment, as outlined in the early childhood curriculum (Ministry of Education, 1996,

2007). Furthermore, theoretical models of the links between oral narrative ability and reading comprehension support the importance of evaluating oral narrative ability across the modalities of comprehension and production (Graesser, Singer, & Trabasso, 1994). As Zwaan and Graesser (1998, p. 196) aptly explained, "adequate comprehension is tantamount to the construction of a coherent and appropriate model of the state of affairs denoted in the text [..]; this model is called a *mental model*". When reading or listening to narratives (or stories), good comprehenders will use or activate these mental models to explain actions of the characters as well as the sequence of events in the story.

Existing methods to assess story retelling ability in preschoolers include *The Bus Story* (Renfrew, 1995) , the *Narrative Assessment Protocol* (Justice et al., 2010), and the story retelling tasks used to create the *SALT Narrative Story Retell Database* (http://www.saltsoftware.com/salt/downloads/NarStoryRetellRDBDoc.pdf). However, these assessments do not examine oral narrative ability across the modalities of comprehension and production. To the authors' knowledge, the only exceptions are the *Narrative Comprehension of Picture Books Task* (NC task; Paris & Paris, 2003), the *Edmonton Narrative Norms Instrument* (ENNI; Schneider & Dubé, 2005), and the *Expression, Reception and Recall of Narrative Instrument* (ERRNI; Bishop, 2004). Although the NC task (Paris & Paris, 2003) considers the child's comprehension of a story, it only appraises the child's story retelling performance at macrostructure level; while the ENNI (Schneider & Dubé, 2005) investigates the child's story retelling ability at both micro- and macrostructure level and includes a comprehension component, no normative data are available for this comprehension measure; the ERRNI (Bishop, 2004) provides norms for 4- and 5-year-old children, but floor effects are seen for this age-group, which makes the task more suitable for children aged 6 years and over. The current study therefore uses the *Profile of Oral Narrative Ability* (PONA; Westerveld & Gillon, 2010). The PONA derives its macrostructure and microstructure

measures of oral narrative production performance from a story retelling task and also involves a comprehension component. Using the PONA will allow us to extend previous research into the clinical utility of this tool (Westerveld & Gillon, 2010) to a younger age-group and help determine if the PONA can be used successfully across the age range from preschool (i.e., 4- year-old children) to school-age children.

When investigating the clinical utility of an assessment tool, several factors need to be considered. First, the time required to elicit and analyse the story retelling sample is important. Traditionally, oral language samples have been audio-taped and transcribed, often using transcription software, such as *Systematic Analysis of Language Transcripts* (SALT; Miller, Gillon, & Westerveld, 2008). Although some authors argue that this is too time-consuming, and therefore favour online analysis of oral narrative ability by viewing video recordings of the child (Justice et al., 2010), no study has directly compared the time (including the amount of training required to accurately analyse the oral narratives, either online or offline) and/or ease of access to equipment (video vs audio) needed for these two coding processes. It is therefore not clear if online analysis has a real time-advantage over offline transcription. A second consideration is the availability of norm-referenced data to allow the clinician to determine a child's performance relative to age-matched peers. Apart from the ENNI (Schneider & Dubé, 2005), the *SALT Narrative Story Retell Reference Database* (http://www.saltsoftware.com/salt/downloads/NarStoryRetellRDBDoc.pdf), and *The Bus Story Test* (Renfrew, 1995), very few norms exist against which to compare preschool-age children's oral narrative perfomance in a story retelling context.

To further investigate the clinical usefulness of an assessment tool, both reliability and validity must be evaluated. Reliability includes interrater reliability, which measures the amount of variance between two or more raters when scoring the performance of the same children. Results from previous research into the clinical utility of the PONA with school-

aged children suggested acceptable interrater reliability levels for the macrostructure analysis of the story retellings (i.e., α = .96; Westerveld & Gillon, 2010). "Measurement validity concerns the extent to which an instrument measures what it is intended to measure" (Portney & Watkins, 2000, p. 79). The face validity (i.e., does the task assess story comprehension and retelling ability) and content-description validity (the rationale for the way in which the story retelling task is administered) of the PONA have been described in detail (Westerveld & Gillon, 2010). However, one aspect of measurement validity that has not yet been established is criterion-related validity; concurrent validity checks the performance on aspects of the PONA against a criterion test, administered at the same time; predictive validity investigates if performance on the PONA is a valid predictor of a future criterion score. Considering the hypothesised significance of the child's story comprehension and retelling performance as an indicator of current and future language ability, this aspect of measurement validity becomes important.

The current study

The current study aimed to extend the research by Westerveld and Gillon (2010) into the clinical utility of the PONA by investigating its usefulness for preschool-age children. The story retelling task was administered to 92 four-year-old children who showed typical development. Fifty-seven (62%) of these children were re-assessed after 12 months. We were particularly interested in the distribution statistics as well as concurrent and predictive validity. More specifically, the following questions were addressed:

1. How do 4-year-old children perform on the PONA?

2. Is performance on the PONA sensitive to age?

3. Is performance on the PONA at age 4 related to the child's more general language ability as measured by the *Peabody Picture Vocabulary Test – Fourth Edition* (PPVT-4; Dunn & Dunn, 2007)?

4. Is performance on the PONA at age 4 or age 5 related to performance on a standardised test of listening comprehension at age 5?

5. Is performance on the PONA at age 4 predictive of performance at age 5?

Method

Participants

The data presented are part of a longitudinal study investigating the literacy development of children attending kindergarten in New Zealand (NZ). The participants were recruited from seven kindergartens in suburban Christchurch, NZ. Teachers were asked to hand out information sheets to parents of children who met the following criteria: 1) age between 4;0 – 4;11 years, 2) spoke NZ English as their first language, 3) had no history of speech and language difficulties, and 4) were currently not receiving specialist services. A total of 92 children (54 girls, 38 boys) were referred and participated in the story retelling task. These children were from NZ European (82.6%), Maori (12%), Pasifika (2.2%), and 'other' (3.3%) ethnic backgrounds. Based on the 2006 Census data, this ethnic distribution is representative of the Canterbury region (Statistics New Zealand, 2009), except for an under-representation of Asian ethnicity (expected 5%).

All 92 children produced answers to the comprehension component of the task of sufficient detail to be analysed (see below); 73 children provided stories that could be analysed (6 transcripts were discarded because of examiner error; 5 children refused to retell the story; 6 children only provided the title and/or too little information to score the story; the intelligibility of two samples was below 80%). These 73 children came from low-mid to high socio-economic backgrounds as indicated by the Ministry of Education school ranking system. There were 43 girls (59%) and 29 boys (41%) from NZ European (86.5%), Maori (8%), Pasifika (1.5%), and 'other' (4%) ethnic backgrounds.

After approximately one year, all 92 families were contacted and invited to participate in stage 2 of the study. A total of 57 parents agreed (62% of the original sample); 34 girls and 23 boys of NZ European (89.5%), Maori (8.8%), and 'other' (1.8%) ethnic backgrounds. By that stage, all children had made the transition from kindergarten to primary school. In New Zealand, children typically start their first year of formal schooling on their fifth birthday. To determine if the children who were available for follow-up at age 5 ($n = 57$) were representative of the whole group ($n = 92$) at age 4, t-tests for independent samples were used to compare the 4-year-old children available for testing at age 5 ($n = 57$) to the 4-year-olds who were unavailable ($n = 35$). There were no significant differences ($p > .05$) on any of the measures (including age, PPVT score, oral narrative measures) and we therefore assumed that this continuing group was representative of the original 4-year-old group.

Procedure and Materials

Using a longitudinal repeated measures design, the children were seen on two occasions (age 4 and age 5), approximately 12 months (M = 12, range = 11 − 13 months) apart. Assessment sessions lasted 45 - 60 min. The children were seen individually by trained research assistants in a quiet room at their kindergarten or at their school. All sessions were recorded, using digital voice recorders (Olympus DM1). The following assessments were administered:

Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4; Dunn & Dunn, 2007). This standardised test was administered at age 4 to determine the children's receptive vocabulary skills. The PPVT-4 has been normed for children and adults from the age of 2 years 6 months through 90 years and older. It has excellent reliability (test-retest = .93; split-half = .94). As reported in the manual, the PPVT-4 has an average concurrent correlation of .82 with the *Expressive Vocabulary Test, Second Edition* (EVT-2; Williams, 2007). For 3- to 5-year-old children, it shows moderate concurrent correlations with the following

lexical/semantic subtests from the *Comprehensive Assessment of Spoken Language* (CASL; Carrow-Woolfolk, 1999): *Basic Concepts*: $r = .50$; *Sentence Completion*: $r = .54$, and *Antonyms*: $r = .41$.

Story Retelling. This task was administered at age 4 and age 5. The children were asked to listen twice to an audio-recording of an unfamiliar story, while looking at the pictures of the story book on a computer screen (using Microsoft Office Power Point™). The story was an English translation of *Ko au Na Galo* [*Ana Gets Lost*] (Swan, 1992). The story is about a Pasifika girl named Ana, who gets lost in the city while looking for her mother and father. The book is a 10-page reader of the kind that is typically used in NZ classrooms, with coloured pictures and Tokelauan (language spoken on Tokelau, territory of NZ in the South Pacific Ocean; population approximately 1400) text. Following the first exposure to the story, the children were asked eight questions, yielding a comprehension score. Following the second exposure to the story, the children were asked to retell the story without the use of pictures, using the prompt "*This time I would like <u>you</u> to tell the story into my digital recorder so that other children can listen to <u>your</u> story next time*". For a full description of the task, including the prompts, the model story, and the comprehension questions, please visit http://www.saltsoftware.com/salt/downloads/NewZealand.cfm.

Listening Comprehension. The *Understanding Spoken Paragraphs* subtest from the *Clinical Evaluations of Language Fundamentals – Fourth Edition* (Semel, Wiig, & Secord, 2006) was used at age 5 to assess the children's ability to answer questions after listening to four spoken paragraphs of increasing length and complexity. The factual and inferential questions target the main idea, the sequence, some details, and predictive information. The CELF-4 has been normed for ages 5 to 21 years. The subtest is supplemental for children between the ages of 5;0 and 8;11 and was standardised on 825 Australian children. It shows excellent internal consistency for the 5;0 – 5;11 age group ($r = .81$) and moderate correlations

with all other subtests of the CELF-4 that tap receptive language skills ($r > .42$) (Semel et al., 2006).

Transcription and Analysis

The digital sound files were transferred to a computer using Olympus DSS Player Pro Dictation Module version 4.4.0 (Olympus, 2006) and were transcribed while using headphones and Olympus RS28 foot pedals. The sound files were transcribed and coded by a researcher experienced in language sample analysis, using standard Systematic Analysis of Language Transcripts (NZ version) software conventions (SALT- NZ; Miller et al., 2008). Utterance segmentation was based on communication units (C-units), defined as one main clause with all its subordinate clauses (Loban, 1976). However, following standard SALT conventions, elliptical responses (phrases) 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 (see http://www.saltsoftware.com/training/transcription/Cunits.pdf). Only complete and intelligible (C&I) C-units were used for analysis. All reformulations, repetitions, and disfluencies were placed in parentheses and considered mazes. Finally, all child utterances that were unrelated to the story (i.e., comments, questions) were excluded from the analysis. Formal beginnings (title) and formal endings (e.g., "The end") were also excluded from the analysis.

Oral Narrative Measures

Oral narrative comprehension (ONC). The children were asked eight questions following the first exposure to the story (e.g., *"Who is the story about? Why did Ana have to stay at home? Who found Ana? Why were Ana's parents happy to see her?"*). To ensure all children had access to the same information prior to listening to the story for a second time, children were given the correct information if they did not respond to the questions or if their

answers were clearly incorrect. A scoring guide was used to determine if the child's answers were considered correct or incorrect and if further prompts were allowed. ONC was calculated as the number of questions answered correctly.

Oral narrative quality (ONQ). All transcripts were analysed at macrostructure level by scoring the stories on a story quality rubric. The rubric was adapted from Jones and Lodholz (1999); it covers six text structure elements: introduction, main character/s, supporting character/s, conflict, resolution, and conclusion as well as a measure of holistic coherence and a measure of 'theme'. The rubric is available for download at http://www.saltsoftware.com/salt/downloads/NewZealand.cfm. The child was awarded points for each characteristic: 5 points for proficient inclusion, 3 points for emerging proficiency, and 1 point for minimal or no evidence of inclusion. The scores were totalled to yield an ONQ score (minimum score 8, maximum score 40).

Microstructure measures of oral narrative performance. Only measures that are known to be sensitive to language development as well as language ability were considered when developing the PONA (Westerveld & Gillon, 2010). The following microstructure measures were used: Number of Different Words (NDW) to reflect semantic diversity (Watkins, Kelly, Harbers, & Hollis, 1995), total number of utterances (UTT) to measure verbal productivity (Scott & Windsor, 2000), mean length of utterance in morphemes (MLU-M) to measure grammatical ability (Scott & Windsor, 2000; Watkins et al., 1995), and grammatical accuracy (GA) in percent grammatically correct utterances to measure grammatical competency (see Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004).

Reliability

To ensure accuracy and completeness of transcription and coding, the first author listened to all the story sound files and checked the transcripts for spelling, error coding (GA), and

utterance segmentation errors. All errors were corrected. For the Oral Narrative Quality (ONQ) coding, 20% of the transcripts were randomly selected and rescored by an independent examiner (experienced with language sample analysis), who was trained in scoring the stories on the ONQ rubric. Krippendorff alpha coefficients (Krippendorff, 1980) were calculated to document agreement between scorers. This procedure accounted for chance agreement and the degree of differences between the judgments; interjudge scores of 30 and 31 have a higher level of agreement than scores of 30 and 19. Krippendorff stated that alpha values below .67 are unreliable, alphas between .67 - .80 are adequate for making tentative conclusions, and alphas at or above .80 reflect good agreement. Krippendorff's alpha using ordinal scaling was .91 for the ONQ scoring, indicating a good agreement between the two independent scorers.

Results

Internal consistency

The results were analysed using statistical software SPSS (PASW, 2008). First we wanted to investigate the internal consistency of the measures. Correlation coefficients between the oral narrative measures at age 4 were calculated and are reported in Table I. As expected, some oral narrative measures were significantly correlated. For example Oral Narrative Quality (ONQ) showed significant correlations with Oral Narrative Comprehension (ONC), length of the narrative in number of utterances (UTT), number of different words (NDW), and MLU-M. In contrast, there were no significant correlations between ONC and the number of utterances used to retell the story (UTT), or between ONQ and Grammatical Accuracy (GA), reflecting the multidimensional nature of the oral narrative measures. We also found that NDW and UTT were highly correlated ($r = .91$, see Table I). Both measures were retained, however, as they reflect conceptually different language abilities.

Table I. Correlations between the oral narrative measures at age 4 (*n* = 73)

Measures	ONC	ONQ	NDW	UTT	MLU-M	GA
ONC	---	.48**	.21	.07	.35**	.24*
ONQ		---	.71**	.55**	.52**	.01
NDW			---	.91**	.64**	-.24*
UTT				---	.42**	-.29*
MLU-M					---	-.21
GA						---

Note: ONC = Oral narrative comprehension, maximum score is 8; ONQ = Oral narrative quality, range of scores 8 - 40; NDW = Number of different words; UTT = total number of utterances; MLU-M = Mean length of utterance in morphemes; GA = grammatical accuracy in percent grammatically correct utterances. * $p < .05$; ** $p < .001$

Distribution statistics

We investigated the normality of the sample's distribution at age 4 to determine if the data can potentially be used for normative purposes. First, mean, median, and percentile scores were calculated. As shown in Table II, most mean scores were slightly above the median score (except for ONC and GA), indicating off-centre distributions. Next, skewness (investigates the degree of symmetry of the distribution) and kurtosis (looks at the peakedness/flatness of the distribution) statistics were calculated. Values of > 2 standard errors of skewness or kurtosis are considered significant (Tabachnick & Fidell, 1996). As shown in Table II, the measures showing a normal distribution were ONQ, ONC, and MLU-M. Both GA and UTT showed significantly peaked distributions with GA scores clustering at the high end of the scale (ceiling effect) and UTT clustering at the low end of the scale (floor effect).

Table II. Distributional performance and statistics on the oral narrative measures (*n* = 73)

Measures	M	Mdn	SD	5%	10%	25%	50%	75%	90%	95%	Skewness	Kurtosis
ONC (n = 92)	4.6	5	1.6	2	2.3	4	5	6	6.7	7	-.45	.004
ONQ	18.8	18	5.4	10	12	16	18	22	28	28	.36	-.32
NDW	28	25	15.2	9	12	16	25	38	49	61	.97*	.86
UTT	8.1	7	4.6	2.7	3	5	7	11	13	18.6	1.43*	2.84*
MLU-M	5.9	5.7	1.6	3.1	3.8	4.9	5.7	7.0	7.9	8.9	.20	.02
GA	76.6	80	23	31	44	64	80	100	100	100	-1.06*	1.13*

Note: * indicates that performance on these measures show an abnormal level of skewness/kurtosis. ONC = Oral narrative comprehension, maximum score is 8; ONQ = Oral narrative quality, range of scores 8 - 40; NDW = Number of different words; UTT = total number of utterances; MLU-M = Mean length of utterance in morphemes; GA = grammatical accuracy in percent grammatically correct utterances.

Sensitivity to age

To determine if performance on the task was sensitive to development (age), children's performance on each of the oral narrative measures at age 5 was compared to their performance at age 4, using paired *t*-tests (*n* = 57). There was a significant improvement in performance with age on all measures: ONC $t(1,56) = -6.05$, $p < .001$; ONQ $t(1,37) = -3.42$, $p < .05$; UTT $t(1,37) = -2.25$, $p < .05$; MLU-M $t(1,37) = -2.77$, $p < .05$; NDW $t(1,37) = -3.85$, $p <.001$; and GA $t(1,37) = -2.12$, $p < .05$. Effect sizes were calculated as *d* (Cohen, 1988), where small *d* = .20, medium effect size *d* = .50, large effect size *d* = .80. Effect sizes were medium to large as shown in Table III, indicating a marked improvement on all measures of oral narrative ability with age.

Criterion-prediction validity

We also wanted to investigate if performance on the oral narrative task was related to well-constructed norm-referenced measures of spoken language performance. Two measures of concurrent validity were calculated. First, we examined the relationship between the oral narrative measures and performance on the PPVT-4 (Dunn & Dunn, 2007) for all children who participated at age 4. The PPVT-4 is a widely-used, psychometrically sound measure of receptive vocabulary as well as a screening measure of verbal ability. As shown in Table IV, performance on the PPVT-4 showed significant moderate to large positive correlations with measures of ONQ (*r* = .44) and ONC (*r* = .55). Correlations between performance on the PPVT-4 with microstructure measures of oral narrative performance were not significant, however.

Table III. Group performance on the oral narrative measures and the standardised language measures (with SD and range) at age 4 and age 5

Measures	n	4-year-old	n	5-year-old	Effect size *d**
Age in months	92	55 (3.3) 48 - 59	57	66.6 (3.5) 60 - 72	
PPVT-4	92	108 (12.1) 82 - 134		---	
USP#		---	57	8.1 (2.4) 3 - 13	
ONC	92	4.6 (1.6) 0 - 8	57	6.2 (1.3) 2 - 8	1.13
ONQ	73	18.8 (5.4) 8 - 32	51	22.7 (6.1) 10 - 36	.79
NDW	73	28 (15) 5 - 76	51	36 (13) 14 - 72	.88
UTT	73	8.0 (4.6) 2 – 25	51	9.3 (3.4) 3 - 20	.51
MLU-M	73	5.9 (1.6) 2.5 – 10.4	51	6.8 (1.4) 3.9 – 11.8	.63
GA	73	76.6 (23) 0 - 100	51	87.3 (13.8) (50 – 100)	.49

Note: # USP = *Understanding Spoken Paragraphs* (CELF-4), raw scores are presented. *effect sizes are reported for all measures, indicating a significant improvement in performance with age. ONC = Oral narrative comprehension, maximum score is 8; ONQ = Oral narrative quality, range of scores 8 - 40; NDW = Number of different words; UTT = total number of utterances; MLU-M = Mean length of utterance in morphemes; GA = grammatical accuracy in percent grammatically correct utterances.

Second, we investigated the concurrent relationship between the oral narrative measures at age 5 and performance on the *Understanding Spoken Paragraphs* (USP) subtest of the CELF-4 (Semel et al., 2006). As a group (*n* = 57), the children performed in the normal range on the USP (M = 10.68, SD = 2.6), although individual standard scores ranged between SS 5 and SS 20. Performance on the USP (using raw scores) showed significant moderate

positive correlations with measures of ONQ ($r = .39$), ONC ($r = .39$), NDW ($r = .36$), and

MLU-M ($r = .30$). Table IV shows the results.

Table IV. Concurrent and predictive correlations between standardised language measures and performance on aspects of the PONA

	PPVT-4 Age 4	USP Age 5
Measures at age 4		
ONC	.54**	.44**
ONQ	.44**	.35*
NDW	.13	.43**
UTT	.02	.38*
MLU-M	.21	.28
GA	.11	-.07
Measures at age 5		
ONC		.39**
ONQ		.39**
NDW		.36**
UTT		.25
MLU-M		.30*
GA		.18

Note: PPVT-4 = *Peabody Picture Vocabulary Test*, *Fourth Edition*; USP = *Understanding Spoken Paragraphs* (CELF-4); ONC = Oral narrative comprehension, maximum score is 8; ONQ = Oral narrative quality, range of scores 8 - 40; NDW = Number of different words; UTT = total number of utterances; MLU-M = Mean length of utterance in morphemes; GA = grammatical accuracy in percent grammatically correct utterances.

To investigate predictive validity, we calculated correlations between performance on all oral narrative measures of the PONA at age 4 and performance on the USP at age 5 (see Table IV, $n = 57$). There were significant moderate to large positive correlations between performance at age 4 on the PONA and performance on the USP at age 5 on ONC ($r = .44$), NDW ($r = .43$), UTT ($r = .38$), and ONQ ($r = .35$). Correlations between the grammatical measures of MLU-M and GA at age 4 and USP at age 5 were not significant.

Predictive validity

Finally, we considered the correlations between performance on individual oral narrative measures contained in the PONA at age 4 and performance on these same measures at age 5 for the children who participated in both sessions. The only significant ($p < .05$) positive correlations were seen between measures of ONC at age 4 and age 5 ($r = .28$), and NDW at age 4 and age 5 ($r = .39$). All other correlations were non-significant: ONQ ($r = .29$), UTT ($r = .20$), MLU-M ($r = .17$), GA ($r = .30$).

Table V. Correlations between language measures at age 4 and age 5 ($n = 57$)

Measures	Age 5					
	ONC	ONQ	NDW	UTT	MLU-M	GA
Age 4						
ONC	.28*	.09	.19	.01	.30*	.35*
ONQ	.11	.29	.35*	.17	.33*	.07
NDW	.25	.27	.39*	.17	.36*	.07
UTT	.26	.30	.40*	.20	.42**	.01
MLU-M	.24	-.01	.27	.11	.17	.03
GA	.18	-.01	-.27	-.25	-.01	.30

Note: ONC = Oral narrative comprehension, maximum score is 8; ONQ = Oral narrative quality, range of scores 8 - 40; NDW = Number of different words; UTT = total number of utterances; MLU-M = Mean length of utterance in morphemes; GA = grammatical accuracy in percent grammatically correct utterances. * $p < .05$; ** $p < .001$

Discussion

The aim of this study was to investigate if the *Profile of Oral Narrative Ability* (PONA; Westerveld & Gillon, 2010) has clinical utility with preschool-age children. The PONA comprises a range of oral narrative measures derived from a story retelling task (*Ana gets Lost*; Swan, 1992). Measures include oral narrative production at macrostructure (oral narrative quality) and microstructure levels (semantic and grammatical ability, verbal productivity), as well as oral narrative comprehension. The task was field-tested with 92 four-year-old children. Approximately 60% these children (*n* = 57) were seen for follow-up assessment one year later at age 5.

The type of story (unfamiliar story but set in a culturally appropriate context) and task protocol (e.g., two listenings to the story interspersed with comprehension questions) proved an effective method to engage 4-year-old children with typical development in retelling a story. The careful analysis of the examiner protocols also revealed useful clinical information. Unfortunately, six of the collected story retelling samples had to be discarded because the examiner did not follow the protocol, despite training in administering the task. Although the elicitation protocol explicitly stated the importance of using neutral prompts to encourage the child to continue his or her story, closer inspection of these six story retelling transcripts revealed that the examiner's use of specific prompts may have inflated these children's performance on measures of oral narrative quality (ONQ), number of utterances (UTT), and the number of different words (NDW). These samples were therefore deleted from the data analyses. From a reliability perspective, this has important practical implications, especially when a spontaneous language task is used for assessment purposes and administered by professionals who may not be familiar with language sample analysis procedures or with administering standardised assessments. Thus, it is important for speech-language pathologists or teachers administrating spontaneous language elicitation tasks to not

only listen to the child's recorded responses, but to carefully evaluate their own prompts and verbal cues provided during the elicitation task.

In addition to the six discarded transcripts due to examiner error, at age 4 a total of 11 children (12%) either refused to retell the story or provided too few utterances for reliable analysis. Interestingly, at age 5 a total of 6 children (10%) provided story retellings that were too short for analysis. One plausible explanation is that some children considered the task too daunting as they were not allowed to refer to the pictures when retelling the story, thus stressing their memory capacity. Unfortunately, previous research by Justice et al. (2010), who investigated oral narrative skills in preschoolers using a task in which children were allowed to refer to pictures when telling the story, did not report the number of children refusing to participate, so no direct comparisons are possible. Further research investigating whether story retelling elicitation rates improve when allowing 4-year-old children to refer back to the pictures would be useful.

Another explanation is that some children were unable to retell the story because of subtle oral language difficulties (i.e., difficulties not sufficiently obvious as to have caused parent or teacher concern since all participants included had no reported history of difficulty and were not receiving any specialist intervention). Closer inspection of the children's oral narrative comprehension (ONC) as well as their performance on the PPVT-4 showed that, although there was a trend for the group of children who did not provide a story retelling to score lower on these measures than the group of children who did provide a story of sufficient length to be analysed, this was not statistically significant. Future research should consider the use of a warm-up or practice task to clarify the issue of language competence versus performance in this age group.

A third plausible explanation is noncompliance. No data were collected on the children's typical behavioural patterns when engaging with adults or their personality

characteristics in terms of self-confidence or shyness. Preschool children are less "practised" in responding in more formal ways to adult requests compared to school-aged children and a 12% noncompliant rate may indeed be an expected level for 4-year-old children. Further research in the use of language sampling within this young age group is necessary to provide insight into these issues.

Descriptive analyses

Consistent with previous research into children's spontaneous language performance, descriptive analyses revealed irregular distributions of some of the oral narrative measures (Justice et al., 2006; Liles, Duffy, Merritt, & Purcell, 1995; Westerveld & Gillon, 2010). The number of utterances (UTT) and grammatical accuracy (GA) in particular showed peaked distributions; performance on UTT and NDW clustered at the low end of the spectrum and GA showed a ceiling effect (31.5% of the children achieved a GA of 100%). In contrast, normal distributions were found on measures of ONC, ONQ, and MLU-M. These normal distributions not only show that the task is particularly suitable for capturing preschool children's performance on these measures, they also indicate that the sample size of children who participated in the study is adequate for norming purposes. Furthermore, these findings confirm our previous hypothesis that the ONC measure showed a ceiling effect in the 5- to 7-year-age groups (Westerveld & Gillon, 2010) and that the task is thus more suitable for 4-year-olds. Finally, the results indicate that the task may be too simple to adequately capture grammatical accuracy in children with typical development. Future research should now investigate the performance of children with language impairment to determine the sensitivity of the task to language ability status and its potential to be used as a screening measure.

Next, we determined the task's sensitivity to age. From a theoretical point of view, oral narrative ability develops with age (see Hughes et al., 1997, for an overview). Not only do children's semantic and syntactic skills improve during the preschool years, children also

develop and refine their *mental models* for fictional stories, which in turn improves their comprehension and production of stories (Hudson & Shapiro, 1991). As a result, an improvement in performance on all measures of the PONA was expected between age 4 and age 5. The results confirmed this hypothesis and showed a significant improvement (with medium to large effect sizes) on all oral narrative ability measures when comparing children's performance at age 5 to their performance at age 4. Although a practice effect could have inflated the children's performance at age 5, this seems unlikely considering the assessment sessions were 12 months apart. Moreover, performance levels on all oral narrative measures at age 5 in the current study are similar to those of the group of 5-year-olds we reported on in previous research (Westerveld & Gillon, 2010).

Criterion-related validity

Story retelling ability in preschool-age children is considered an important predictor of future language performance and academic attainment (e.g., Bishop & Edmundson, 1987); assessment of this skill should thus be included in the speech-language pathologist's standard assessment battery for preschoolers with suspected language impairment. To determine a child's relative performance compared to his or her peers, normative data need to be consulted. Although several norm-referenced story retelling assessment tasks are available (Justice et al., 2010; Renfrew, 1995; Schneider & Dubé, 2005), research into the criterion-related validity of these tasks is scarce. The current study therefore investigated whether performance on the PONA was correlated to performance on standardised measures of language performance: a) the PPVT-4 (Dunn & Dunn, 2007) and the *Understanding Spoken Paragraphs* (USP) subtest of the CELF-4 Australian version (Semel et al., 2006). At age 4, results indicated significant moderate concurrent correlations between performance on two aspects of the PONA, oral narrative comprehension (ONC) and oral narrative quality (ONQ), and performance on the PPVT-4 (.54 and .44, respectively). At age 5 there were significant

moderate concurrent correlations between most aspects of the PONA and performance on the *Understanding Spoken Paragraphs* task (ranging from .30 to .44). These correlations are consistent with those from previous studies investigating concurrent correlations between oral narrative measures and performance on a standardised language test in preschool-age children (Justice et al., 2010). To illustrate, Justice and colleagues found correlations of .34 and .35 between children's performance on the *Narrative Assessment Profile* (NAP) and their performance on the core language composite of the *CELF Preschool-2* (Wiig, Secord, & Semel, 2004). The NAP, however, is a composite score of children's performance on (mainly) microstructure measures, so direct comparisons between the studies are not possible.

It was somewhat surprising that the correlation between children's performance on the PPVT-4 was not significantly correlated to the semantic diversity measure of NDW (see Table IV). Previous research addressed the criterion validity of the PPVT-III (Dunn & Dunn, 1997) by investigating concurrent correlations between performance on the PPVT-III and spontaneous language sample measures in 3- to 6-year-old children (Ukrainetz & Blomquist, 2002). As one would expect, no significant correlations were found between performance on the PPVT-III and the two non-semantic measures of MLU and total number of words. However, a significant correlation was found between performance on the PPVT-III and the expressive semantic measure NDW ($r = .36$, $p = .05$). The most obvious explanation for the apparent differences in findings between Ukrainetz and Blomquist's (2002) research and the current study relates to the way the language samples were collected. Ukrainetz and Blomquist elicited 150 utterances across conversation, narration, and exposition, whereas the current study used a narrative task yielding, on average, 8 utterances. It is generally recommended that at least 50 utterances should be elicited for full analysis of a child's strengths and weaknesses in spoken language performance (Miller, 1996). In addition, some elicitation contexts yield higher semantic diversity, especially in older children (Westerveld,

Gillon, & Miller, 2004). Future research should consider the predictive validity of performance on a story retelling task, such as the one used in the current study, and performance on a spontaneous language task that is likely to yield at least 50 utterances.

The study also considered the predictive correlations between children's performance on aspects of the PONA at age 4 and the USP subtest of the CELF-4 at age 5. Once again, significant moderate correlations were found with measures of ONC, ONQ, NDW, and MLU-M (ranging from .30 to .39). These findings compare favorably to previous investigations into the predictive correlations between language sample measures and standardised test performance (Pankratz et al., 2007). Results from the current study add to our knowledge base by highlighting the importance of sampling children's oral narrative retelling performance across the modalities (comprehension and expression) and at both microstructure and macrostructure levels.

When investigating the correlations between children's performance at age 4 to their performance at age 5, an interesting finding emerged. The only measures at age 4 that were significantly correlated to performance at age 5 were ONC and NDW. These findings suggest that some of the narrative measures derived from the story retelling samples, including those tapping grammatical ability (GA and MLU-M) and story quality (ONQ) may, by themselves, not be indicative of future performance in children with typical development. Unfortunately, few longitudinal studies used the same narrative assessment task over time, and most of these did not report predictive validity (Fey et al., 2004; Paul, Hernandez, Taylor, & Johnson, 1996). One notable exception is the study by Bishop and Edmundson (1987) in which a cohort of more than 80 children, who were referred for speech pathology assessment around the age of 4, participated. To investigate how many of these children would overcome their language difficulties, all children were seen for follow-up assessments at age 4½ and 5½. Bishop and Edmundson asked children to retell *The Bus Story* (Renfrew, 1969) and assessed

their ability to express semantic relationships (with the aid of pictures) as captured in the *Bus Story Information* score. Significant correlations (ranging between .76 and .797) were found between the *Bus Story Information* test scores obtained by the children over time. Furthermore, performance on the *Bus Story Information* subtest was one of the four top variables to predict language status outcome at age 5½ from 4-year scores. When investigating patterns of impairment, Bishop and Edmundson (1987) concluded that children most at risk of ongoing language difficulties were those who, at age 4, showed: a) expressive language difficulties across the domains of phonology, syntax and morphology, and semantics; or b) verbal comprehension as well as expressive language difficulties. Although direct comparisons between Bishop and Edmundson's study (1987) and the current investigation are not possible, we tentatively conclude that 4-year-old children who score poorly on ONC and/or NDW measures of the PONA may display weaknesses in oral narrative ability that warrant further investigation.

Summary of findings

In summary the following main positive findings were identified:

- The story retelling task was effective in eliciting adequate story retelling samples for meaningful analyses in most 4- year- old children tested;

- Data showed normal distributions amongst the 4-year -old population on measures of ONC, ONQ, and MLU-M;

- The story retelling task used was sensitive to age with significant improvement shown on all measures from 4 to 5 years of age;

- Moderate concurrent correlations between performance on oral narrative comprehension (ONC) and oral narrative quality (ONQ), and performance on the PPVT-4 were evident and between most aspects of the PONA and performance on the *Understanding Spoken Paragraphs* task from the CELF-4.

Future research should investigate the performance of children with language impairment to determine the sensitivity of the task to language ability status and to examine its potential to be used as a screening measure of oral language performance.

Clinical utility

The results from the current study extend previous research into the clinical utility of the PONA with school-aged children (Westerveld & Gillon, 2010) by demonstrating the clinical potential of the PONA in 4-year-old children. The story retelling task is quick to administer (10 mins) and easy to score. It yields important information about the child's ability to a) understand a story, b) retell a well-sequenced story containing important story elements, c) combine words into sentences, d) produce a number of sentences to retell a story, e) produce grammatically correct sentences, and d) use a variety of words to make up the story. One potential drawback for widespread use of this tool may be the need for transcription (Justice et al., 2010), which is considered time-consuming. As seen in Table II, however, children produced on average 8 utterances, which should not take long to transcribe, either by hand or by using more formal transcription software such as SALT-NZ (Miller et al., 2008).

The PONA needs to be used as part of an overall broader assessment battery for preschool children suspected of language impairment. As mentioned in the introduction, it is now well accepted that at least 50 utterances are needed for reliable analysis of a child's strengths and weaknesses in spoken language performance (Miller, 1996). The current study's findings, however, suggest that the PONA may be a clinically useful screening measure of spontaneous language performance.

Acknowledgments

References

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

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry, 31*, 1027-1050.

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.

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

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language (CASL)*. Circle Pines, MN: American Guidance Service.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*(1), 38-50.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test - III*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test - 4*. Circle Pines, MN: American Guidance Service.

Engel, S. (1995). *The stories children tell: Making sense of narratives in childhood*. NY: W.H. Freeman.

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

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395.

Hudson, J. A., & Shapiro, L. R. (1991). From knowing to telling: The development of children's scripts, stories, and personal narratives. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 89-136). Hillsdale, NJ: Lawrence Erlbaum.

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

Jones, M. A., & Lodholz, C. (1999). *Story Grammar Rubric*. Unpublished document, Madison Metropolitan School District and the Language Analysis Lab, Waisman Center, University of Wisconsin, Madison.

Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly, 25*, 218-234.

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.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative

    discourse ability in children with language disorders. *Journal of Speech and Hearing*

    Research, 38, 415-425.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL:

    National Council of Teachers of English.

Miller, J., Gillon, G., & Westerveld, M. (2008). *SALT NZ: Systematic Analysis of Language*

    Transcripts - New Zealand Version. Madison, WI: SALT Software.

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.

Milosky, L. M. (1987). Narratives in the classroom. *Seminars in Speech and Language, 8*(4),

    329-343.

Ministry of Education. (1996). *Te Whariki. Early Childhood Curriculum*. Wellington, NZ:

    Learning Media, Retrieved from

    http://www.educate.ece.govt.nz/learning/curriculumAndLearning/TeWhariki.aspx.

Ministry of Education. (2007). *The strands of Te Whariki: Communication. Assessment for*

    Learning: Early Childhood Exemplars. Wellington, NZ: Learning Media. Retrieved

    from

    http://www.educate.ece.govt.nz/learning/curriculumAndLearning/Assessmentforlearn

    ing/KeiTuaotePae/Book14.aspx.

Olympus. (2006). DSS Player Pro Dictation Module (version 4.4.0) [Computer Software].

    Melville, NY.

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.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*(1), 36-76.

PASW. (2008). PASW Statistics 18 [Computer Software]. Chicago, IL: IBM SPSS Statistics.

Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late talkers: Early school age. *Journal of Speech and Hearing Research, 39*(6), 1295-1303.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

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

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

Schneider, P., & Dubé, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology, 14*(1), 52-60.

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.

Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals - Fourth Edition - Australian Standardised Edition*. Marrickville: Harcourt Assessment.

Skarakis-Doyle, E., & Dempsey, L. (2008). Assessing story comprehension in preschool children. *Topics in Language Disorders, 28*(2), 131-148.

Statistics New Zealand. (2009). Retrieved from http://www.stats.govt.nz/Census/2006CensusHomePage.aspx

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

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.

Ukrainetz, T., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy, 18*, 59-78.

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. 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.

Wiig, E. H., Secord, W., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool - Second Edition (CELF preschool-2)*. Toronto, Canada: The Psychological Corporation.

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition*. Bloomington, MN: Pearson Assessments.

Zwaan, R. A., & Graesser, A. C. (1998). Introduction to special issue of SSR: Constructing meaning during reading. *Scientific Studies of Reading, 2*(3), 195-198.