Should we use sentence- or text-level tasks to measure oral language proficiency in year-one students following whole-class intervention?

Maria Lennox^{1,2}

Marleen Westerveld^{2,3}

David Trembath³

Address correspondence to:

Maria Lennox

Principal Education Officer: Student Services

Metropolitan Region

Department of Education and Training

Queensland

Phone: (07) 3028 8211

Email: maria.lennox@det.qld.gov.au

¹Department of Education and Training, Queensland

²Griffith Institute for Educational Research, Griffith University, Australia

³ Menzies Health Institute Queensland, Australia

This paper was published as:

Lennox M, Westerveld M, F, Trembath D, Should We Use Sentence- or Text-Level Tasks to Measure Oral Language Proficiency in Year-One Students following Whole-Class Intervention? Folia Phoniatr Logop 2017;69:169-179. doi: 10.1159/000485974

Acknowledgements

The authors wish to acknowledge the significant contributions of Tania Kelly and Janice Zee for their instrumental role in the collection of the data used in this study. The authors wish to thank the schools, parents, teachers and teacher aides for their commitment and support of this project. Sincere thanks to the Griffith University Master of Speech Pathology students for generously volunteering their time to support this study. Finally, thank you to the students for their willing participation. The views expressed in this publication do not necessarily present the views of the Department of Education and Training, Queensland.

Declaration of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

Abstract

Aims: To compare students' oral language proficiency on sentence- versus text-level tasks at school-entry and following tier 1 intervention in their first year of formal schooling.

Methods: 104 students participated in this study. Participants were part of a broader longitudinal study and were enrolled at three, low socio-economic, linguistically diverse Australian primary schools. Tasks were administered to all students at the beginning and end of the school year. Performance on the sentence-level task, the *Renfrew Action Picture Test* (RAPT), was analysed for Information and Grammar as per the test manual. Performance on the text-level task, the Profile of Oral Narrative Ability was analysed for measures of story length, mean length of utterance, grammatical accuracy, number of different words, and story quality.

Results: Results showed that both tasks are sensitive to measure progress following tier 1 intervention. However, RAPT concern status was not related to oral narrative concern status. Furthermore, if only the RAPT task had been used, between 11% and 21% of students performing below expectations in oral narrative would not have been identified.

Conclusion: It is recommended that the assessment of oral language proficiency of students from culturally diverse, low socio-economic backgrounds goes beyond the sentence-level and includes an oral narrative retell task.

Key Words: *Renfrew Action Picture Test*, oral narrative skills, school-age children, assessment, culturally and linguistically diverse backgrounds, response to intervention.

Introduction

Considering the importance of oral language competence for academic success, efficient approaches to assess the oral language skills of at-risk students are important to speech-language pathologists (SLPs) and educators [1-3]. Two types of tasks are frequently used to measure oral language proficiency in school-based settings: sentence- and text-level tasks. Often, time and resource restrictions prohibit the use of both tasks within an assessment battery. Consequently, it is important to establish which task provides the most meaningful information related to student oral language proficiency within the school setting. The current study will investigate this issue in relation to Australian students who attended schools in an area known to be home to a high percentage of children from culturally and linguistically diverse (CALD) backgrounds. Of these students, it is estimated that only one in five have sufficient language skills to efficiently access and successfully participate in education [4]. The results from this study will assist SLPs and educators to make informed decisions regarding the use of assessment tasks for measuring oral language progress of Australian students from low socio-economic backgrounds during their first year of formal schooling.

Assessment within the Response to Intervention framework

Without early identification and support, at-risk students are likely to fall further behind their same-aged peers [see 5]. This poses unique challenges for educators resulting in many schools, both nationally and internationally, adopting a response to intervention (RtI) framework [6].

RtI is a framework in which student learning needs are met through differentiated levels of support and high quality teaching [6, 7]. This framework comprises three tiers of support: tier 1, whole class instruction based on evidence informed curriculum; tier 2, small

group support to address gaps in student learning; and tier 3, more intensive support and frequent monitoring for students who demonstrate a lack of progress [8, 9]. School-wide screening data are used to determine baseline-level functioning for all children and identify students who may benefit from further support [10]. Then, progress monitoring is regularly conducted to evaluate whether or not instructional support is effective at each of the three levels [11]. Speech language pathologists (SLPs) may play a key role in this process. For example, at the end of the school year following explicit, evidence-based, classroom intervention (tier 1), oral language skills are re-evaluated to determine if further intensified intervention is required (tier 2/tier 3) according to student need [8, 12].

The RtI framework relies on data-driven, evidence based practice which necessitates valid, reliable tasks that are administered by adequately skilled professionals [5]. The interpretation of these tasks guides the selection of intervention approaches [13]. However, screening of oral language to inform intervention approaches is time-intensive and thus expensive, underlining the importance of selecting tools that are reliable to administer and sensitive in detecting oral language difficulties, or change in oral language performance following intervention [10, 14].

The importance of oral language skills for academic success

At school, students require strong oral language skills across word- (vocabulary), sentence- (syntax), and text- (discourse) levels to access curriculum content [15]. To illustrate, in the Foundation year (i.e., the first year of formal schooling in Australia), students are expected to “bring vocabulary from personal experiences, relating this to new experiences and building a vocabulary for thinking and talking about school” (p. 21). At the sentence-level, students are expected to “recognise that sentences are key units for expressing ideas (ACELA1435, p.21). At the text-level, students are expected to “sequence ideas in spoken

texts, retelling well known stories, retelling stories with picture cues, retelling information using story maps” (p. 22). Furthermore, the concurrent and predictive relationship between oral language performance and reading success has been well documented [16-18]. A large body of research proposes that early weakness in vocabulary, syntax/grammar, and discourse (such as oral narrative skills) predicts future reading difficulties [19, 20].

Factors relating to language and literacy ability in school-aged children

Socio-economic status is among the most prominent factors influencing early language and literacy delays and consequently future reading ability [21-24]. As a group, students from low socio-economic areas are at increased risk of academic failure [5, 25-28]. The Australian Early Development Census (AEDC) is a population measure conducted on commencement of the first year of formal schooling and collects information on individual students across five domains of childhood development [29]. The AEDC reported that children from socially disadvantaged areas were 4.1 times more likely to be developmentally vulnerable, relative to children in less disadvantaged areas [30, 31]. Low socio economic status is one factor that influences the development of oral language and literacy skills prior to school entry. Other factors that impact the literacy development of Australian students include attendance at pre-primary education, exposure to oral narratives, and speaking a primary language other than English [27, 32-34].

Although bilingualism in itself is not a risk factor, oral language proficiency has been found to be a predictor of success for bilingual students [35]. In fact, the assessment of oral language proficiency in 4 – 5 year-old students at school-entry has been shown to predict language and literacy ability at the end of primary school for Australian bilingual children [35]. The findings of this longitudinal Australian study by Dennaoui et al. [35] suggests that bilingual students who commence formal education with limited English proficiency are at

risk compared to their proficient bilingual peers, even after six years of exposure to the academic language environment [35]. In this study teachers utilised a three point rating scale at school entry to classify English proficiency [35]. This study highlights the importance of early identification of oral language challenges to ensure targeted educational support for students who are at risk of falling behind their same aged peers.

Assessment of oral language proficiency

High quality data collection informs decision making within an RtI framework and is achieved through universal progress monitoring [5]. It is widely accepted within the literature that a range of valid, reliable, and sensitive tools, with normative data are needed in order to assess early oral language skills [9, 34, 35]. Traditionally, school-based SLPs have been recommended to utilise a variety of data gathering and assessment practices [12]. In Australia, standardised assessments are typically employed to determine eligibility for state-funded services, but may also be used to evaluate baseline function and set intervention goals [11]. Although standardised or norm-referenced assessments allow clinicians to evaluate student performance based on a standardised sample (e.g., *Clinical Evaluation of Language Fundamentals – fourth Edition* [36]), these types of tests are not always suitable in clinical practice. For example, appropriate norm-referenced assessment tools are not readily available for Australian populations of students who identify as CALD, due to the heterogeneity of languages other than English spoken within Australian homes [37]. For this reason, clinicians may opt for more informal measures, including locally developed tasks. Other reasons for selecting informal measures may be due to convenience and cost [38, 39]. Informal measures reduce the bias of standardised norm-referenced assessments that over-identify students from culturally diverse or low socioeconomic backgrounds [40, 41]. For example, there is a growing body of evidence supporting the use of oral narrative tasks to appraise student

language abilities in both monolingual and bilingual students [42, 43]. The findings from these studies indicate student performance on oral narrative tasks was similar for both groups of students, suggesting bilingual students were not disadvantaged by this measure [42, 43].

These challenges in choosing appropriate assessment tasks have prompted further investigation into the types of oral language tasks that are more suitable within an RtI framework [5]. Two commonly assessed areas of language proficiency are syntax (i.e., ability to produce grammatical sentences), and oral narratives (ability to produce well-sequenced sentences at text- or discourse-level). One frequently administered sentence-level task is the *Renfrew Action Picture Test* (RAPT) [44] that is claimed to be useful in assessing spoken language produced by children aged between 3 and 8 years. However, Jacobs (2014) reported on the use of the RAPT to examine the cross-cultural performance of young school-age students [45, 46]. These results suggested that this measure may place students at risk of being over-identified as having difficulties in language. Sentence-level tasks by nature only capture a student's ability to formulate grammatically correct utterances, using appropriate vocabulary, and do not reveal if students have challenges sequencing utterances to produce, for example, a coherent narrative or story. Considering the importance of oral narrative ability for classroom participation and academic success [47], a task tapping students' oral language proficiency at text-level may be more informative than sentence-level assessment tasks.

Numerous studies internationally have utilised text-level assessment tasks, such as oral narrative retell or generation tasks to appraise students' oral language proficiency at school-entry [40, 48, 49]. Indeed, several oral narrative measures are easily accessible and commonly used by Australian SLPs [38]. Examples include *The Bus Story Test* [50] [e.g., 26, 51] and the Profile of Oral Narrative Ability (PONA), a story retelling task that has been used frequently in Australia and New Zealand with children ages 4;0 to 7;6 years [2, 34, 52].

However, despite the recognised importance of text-level skills for accessing the curriculum [53], few community samples have been used to investigate the comparative merit of utilising either sentence- or text- level assessment tasks to collect information regarding student oral language proficiency at school entry or following tier 1 intervention.

The current study

To assist educators and SLPs to obtain information about student performance and progress and make decisions regarding intervention approaches, it is important to establish which oral language measures provide the most relevant and functional information within an educational context. Information related to the tasks to elicit language and monitor student progress across the school year, the capacity to measure student progress over time, the relationship between the measures and the performance of students on these measures compared to local and published norms are all elements that are considered when evaluating these measures. This study aimed to address some of these issues and evaluated the performance of 104 students from low socioeconomic areas. Student performance was evaluated at the start and the end of their Foundation year, using a sentence-level task (the RAPT) and a text-level task (the PONA). A detailed description of the intervention falls outside of the scope of the current article, but is described in detail in Lennox et al. [54]. In summary, the 24-week, tier 1, book-based intervention targeted oral language (i.e., vocabulary, narrative, shared reading) and emergent literacy skills (i.e., phonological awareness). Intervention was delivered by trained teachers and teacher aides for one hour a day, four days a week. Each sixty-minute session was comprised of a thirty minute-whole class session led by the classroom teacher and a thirty-minute small group session led by the classroom teacher or teacher aide. The following questions were asked:

1. How do Foundation year students perform on the RAPT and the PONA tasks at the start and end of the first formal year of schooling?
2. What are the correlations between student performance on the RAPT and language measures derived from the PONA at the start and at the end of the school year?
3. Do students demonstrate progress on RAPT and PONA in response to tier 1 intervention?
4. Are the RAPT and PONA tasks comparable in determining at-risk status at the end of the first year of school?

Based on previous studies, we hypothesized that both tasks would be effective in describing and measuring progress related to the oral language proficiency of foundation year students [44, 46, 52]. We also expected a strong relationship between the RAPT and PONA tasks, as both tasks measure oral language domains [54]. Consequently, it was hypothesized that all students would demonstrate progress on these measures [2, 44]. Past studies suggest that oral narrative tasks require children to draw upon a number of foundational oral language skills and working memory skills [52, 55]. Based on this knowledge, it was anticipated that the oral narrative task would be more sensitive in identifying oral language difficulties than a sentence-level task.

Method

Participants

Ethics approval was sought and granted by the Griffith University Human Ethics Committee and the Department of Education and Training, Queensland. The current study is part of a larger project investigating the effectiveness of a school-based emergent literacy program on the language and literacy skills of students in their first year of schooling [54]. Consent forms were provided on enrolment or sent home to parents of all foundation year

students enrolled across the three participating schools. Only students for whom consent was provided, were included in the data collection. In addition, students' verbal consent to participate was gained prior to administering the assessment battery.

The assessment data from 104 foundation year students (45 male, 59 female) across eight classes were available for analysis. These students were aged between 4;7 and 5;7 (mean 5; 0) at the start of the school year. The percentage of students who identified as CALD, 66% was compared to the percentage not CALD, 38%. Information obtained from the Australian Early Development Census (AEDC) [56] and the MySchool website [15] underlines the cultural diversity of the geographical location in which the three schools were located. Individual schools collected data on ethnicity as per their local school process, however, specific details of exposure and use of other languages, dominant language/s, and parental birth place was not sought.

Procedure

Volunteer Griffith University Master of Speech Pathology students were trained in the administration of the assessment battery under the supervision of three, school-based, certified practising SLPs. Each participating student was seen individually for approximately 45 mins and assessed on a battery of tasks. Included in this battery were the Profile of Oral Narrative Ability [52], using the *Ana Gets Lost* story [57] and the *Renfrew Action Picture Test* [44]. Both assessment tasks were conducted at school-entry (time 1) and repeated at the end of the school year (time 2).

Assessment tasks

The *RAPT* [44]. Each student is shown 10 coloured picture cards and asked to respond to a short question (e.g., what is the girl doing?). Only minimal, indirect prompting is allowed

with some exceptions, as outlined in the manual. In accordance with the RAPT manual, the following two scores are calculated:

- *RAPT Information*: students are awarded points if the correct words (e.g., nouns, verbs and prepositions) are used to convey information within the pictures (e.g., cuddle/hold/hug, teddy bear/bear/teddy).
- *RAPT Grammar*: students are awarded points if the grammatical structures (e.g., tense, sentence structure and plurality) are used correctly (e.g., subject-verb-object, -ing).

Total raw scores were calculated for both Information and Grammar. To answer research question four, student performance was compared to published RAPT norms.

Profile of Oral Narrative Ability [52]. This oral narrative task was administered as per the protocol outlined by Westerveld and Gillon Westerveld and Gillon [52]. It involved students listening to the pre-recorded story *Ana Gets Lost* [57] on a computer screen. Immediately following the first exposure to the story, the students were asked eight comprehension questions. Another task was completed as a distractor, before students were asked to listen to the story a second time. Following this second exposure, students were asked to retell the story without any picture prompts. Students were provided with a verbal prompt if they had difficulty starting the story. Indirect, unspecific prompts were the only other prompts provided to students, if needed.

The following measures were calculated, based on their sensitivity to age and language ability [52]:

- *Oral narrative quality* (ONQ) was used to evaluate the students' ability to retell a coherent story and was scored using a story quality rubric. Student responses were evaluated across eight story grammar elements: introduction, theme, main character/s, supporting character/s, conflict, coherence, resolution and conclusion. For each of the story grammar elements, students are awarded a score of: one point if element is not

included, three points if emerging ability is demonstrated, or five points for proficient skills demonstrated. The minimum score that can be obtained is 8 and maximum score is 40.

- *Total number of utterances* (UTT) in communication units was used to measure story length [58].
- *Mean Length of Utterance* (MLU) in morphemes was calculated as a measure of morphosyntactic development [59].
- *Number of Different Words* (NDW) was used to measure semantic diversity [58].
- *Grammatical Accuracy* (GA) was calculated automatically, using SALT-NZ and represents the percentage of utterances that are grammatically accurate [60].

Transcription and Reliability

All oral narrative samples were digitally recorded, transcribed, and coded for quality by a trained research assistant in accordance with the standard coding conventions for SALT NZ software [61]. Following transcription, the first author carefully examined all oral narrative retellings and made appropriate changes where required. Furthermore, 20% of the transcripts were randomly selected using SPSS and scored independently by the first author on the ONQ rubric. As per the guidelines outlined by Krippendorff [62], a good agreement (Krippendorff's alpha 0.824) was obtained.

Results

Data were analysed using the statistical software SPSS. To answer research question one, descriptive statistics were used to examine student performance on all measures at time 1 and time 2 (Table I). As shown in Table I, there was a wide range of performance on both

tasks at both time points. This cannot be accounted for by outliers, as very few were observed. Skewness and kurtosis z-scores were calculated by dividing the score by its standard error [63]. Scores that fell outside the ± 2.58 range were not considered normally distributed [64]. At time 1, all measures were normally distributed with the exception of NDW which showed positive skewness and kurtosis scores, indicating floor effects. Number of utterances (UTT) also had a positive kurtosis score at time 1. Similarly, at time 2, all scores again showed normal distribution with the exception of UTT which continued to show positive skewness.

Insert Table I here

To answer research question two, bivariate correlations were performed (for time 1 and time 2) between the two RAPT measures Information and Grammar and the PONA measures: Utterances (UTT), number of different words (NDW), MLU, grammatical accuracy (GA), and oral narrative quality (ONQ) and comprehension (ONC). Table II shows the results. There were significant positive correlations between RAPT Information and RAPT Grammar ($r = 0.790$ at time 1, and $r = .620$ at time 2, $p < 0.001$). There were also significant, strong positive correlations between UTT and NDW ($r = .892$ at time 1, and $r = .875$, $p < 0.001$), indicating that an increase in the number of utterances used to tell a story is accompanied by an increase of the number of different words used.

As expected, significant positive, mild to moderate correlations were found between most of the RAPT measures and the PONA measures at both times, with some exceptions.

First, at time 1, UTT did not correlate significantly with the RAPT measures Information ($r = .089$) and Grammar ($r = .154$). In addition, MLU did not correlate significantly with measures of Information ($r = .171$) and ONQ ($r = .124$). Finally, GA did not correlate significantly with UTT ($r = -.021$), MLU ($r = -.048$), or NDW ($r = .066$).

At time 2, several measures that were initially not significantly correlated at time 1, became significantly positively correlated. These included UTT and Information ($r = .315$), ONQ, and MLU ($r = .446$).

Insert Table II here

Research question three was investigated by completing repeated measures (RM) ANOVAs as well as computing change scores for each variable (i.e. time 2 score minus time 1 score). Effect sizes were calculated as partial eta squared (η^2) to determine how reliably the independent variable explains the dependent variable [65]. Partial eta squared effect sizes can be interpreted as small = 0.009, medium = 0.0588, large = 0.1379 [66]. All students showed improvement on all measures between time 1 and time 2, as indicated by the medium to large effect sizes ranging from 0.091 to .756 (see Table I).

Finally, research question four examined whether poor performance on the RAPT corresponded to poor performance on aspects of the PONA and vice versa. To investigate this we chose measures that were conceptually linked and significantly correlated at time 1 and time 2: RAPT Information and ONQ; RAPT Information and NDW; and RAPT Grammar and GA. Student performance on the PONA was compared to local norms reported in Westerveld and Vidler [34]; student performance on the RAPT was compared to the published norms [44]. Crosstabs were created by generating concern/no concern categories based on 1SD below the means for both of the measures (see Table III, IV and V). An odds

ratio was then calculated to investigate the association between at-risk status on the above mentioned measures scores. The odds ratio of students identified as ‘of concern’ on ONQ and NDW versus RAPT Information is 2.92, 95% confidence interval 0.78 -11.0 and 2.93 (95% CI 0.85 - 10.13), respectively. This suggests that students are approximately three times more likely to be identified of concern on ONQ and NDW than the RAPT Information measure. Similarly, the odds ratio for GA and RAPT Grammar is 5.18, (95% CI 2.03 - 13.20). Finally, as shown in Table III, IV and V approximately 8.7% of students are ‘of concern’ on the RAPT measures Information and Grammar, but are not ‘of concern’ on the PONA measures NDW, ONQ and GA.

Insert Table III

Insert Table IV

Insert Table V

Discussion

This study sought to examine the merits of two oral language tasks in appraising the oral language performance of a sample of Australian students from low socio-economic and linguistically diverse backgrounds at school-entry and at the end of their first school year following tier 1 intervention, with the ultimate aim of informing high quality data-driven practice. Two tasks were used: *The Renfrew Action Picture Test* [44], a sentence-level task, and the Profile of Oral Narrative Ability [52], an oral narrative retell task that assessed text-level skills. Results from 104 students were available for analysis.

The first aim of this study was to investigate how a student sample of CALD students perform prior to and following tier 1 intervention by assessing performance at the start and the end of their first formal year on these two oral language measures. As shown in Table I, a wide range of performance was noticed on both tasks. However, few outliers were observed

and most of the measures were normally distributed, indicating there were no floor or ceiling effects. Taken together, it is apparent that both tasks have clinical utility in eliciting and describing the spoken language skills in a group of young school-age students from low socio economic, culturally diverse backgrounds.

Research question two evaluated the relationship between the two oral language tasks. Significant, moderate to strong correlations were found between RAPT Information and Grammar scores at both time points, which is consistent with the manual reporting that the Grammar score is around 75% to 82% of the Information score, dependent on the child's age [44]. This highlights the importance of clinicians realising the RAPT may provide an overall indication of oral language proficiency at sentence-level, rather than two separate oral language domains (i.e., semantics and syntax/morphology). This confirms the statement made in the manual that 'the Grammar score is, to some extent, dependent on the amount of Information given' p. 18 [44].

The mild to moderate significant correlations across the RAPT and the PONA tasks at both time points indicate that both tests tap the same underlying construct, oral language proficiency, as correlations that are significant are unlikely to occur in a population by chance. The correlations we found between the PONA measures (GA and UTT; ONQ and MLU; GA and MLU; GA and NDW) are consistent with those from previous studies into the oral narrative abilities of typically developing children from monolingual households [52] and extend those findings to a sample of children who are from low socio-economic and CALD backgrounds.

Contrary to expectation, there were differences in concurrent correlations between all the language measures from time 1 to time 2. The most likely explanation for these differences in correlations at time 1 and time 2 is the growth in verbal productivity as shown by much longer narratives at the end of the school year (Mean 14.2 utterances, Table I)

compared to school-entry (Mean 8.1 utterances, Table I), most likely due to exposure to stories as part of the tier 1 intervention [54]. These findings are consistent with previous research that suggests students from low SES, culturally diverse backgrounds may start school with limited exposure to narratives [40], which may have affected the length of their story retellings at time 1.

Research question three investigated whether the two language tasks were sensitive to progress following tier 1 intervention. As shown in Table I, the results clearly indicated that students made significant progress across all measures, regardless of the task. Overall, the medium to large effect sizes suggest that the intervention effect would have been noticeable in clinical practice. Interestingly, improvement in MLU was small (mean 0.66), despite a medium effect size. This finding, however, is consistent with previous literature that reports MLU for school-aged children often shows very little variation [52, 60]. Collectively, these observations highlight the sensitivity of the two language tasks when evaluating student performance following tier 1 intervention.

Finally, we were interested in identifying whether sentence- or text-level tasks were comparable in identifying which students may need tier 2 intervention by determining at-risk status at the end of the first year of school. Our results indicated that there was an increase in the odds ratio across all cross tabulations. Significant associations were found between the RAPT Information score and the measures of NDW and ONQ (see Table III and IV). Results showed that when using the RAPT Information score to determine students ‘of concern’ at time 2, this measure does not guarantee that the students will be able to retell coherent stories with diverse vocabulary. Our data suggest that when using the RAPT Information cut-off to determine if tier 2 intervention is warranted, between 11.5 to 15.4% of students with oral narrative difficulties would not be identified. Considering the significant importance of vocabulary and oral narrative proficiency to future reading ability [67], these findings should

not be underestimated and clearly demonstrate the need to use assessment tasks that go beyond the sentence-level.

As shown in Table V, using GA (derived from the narrative task), 21.2% of students were identified ‘of concern’ when compared to 8.7% identified ‘of concern’ using the RAPT Grammar measure. These results clearly demonstrate that the RAPT task is not sensitive in detecting or confirming student performance on grammatical ability at text –level. The most likely explanation is that performance on the RAPT task relies on the student’s ability to use morphological endings (e.g. past tense -ed). Therefore, the RAPT Grammar score should not be used in isolation as a way to describe student performance in grammar.

Interestingly, 8.7% (n = 9) of students were identified as ‘of concern’ on the RAPT but scored within acceptable limits on at least one of the PONA measures NDW, GA, and/or ONQ. Closer inspection of these students’ performance on the RAPT was conducted. The findings suggested that these students had difficulty producing specific syntactic structures in response to the RAPT (picture) prompts. However, despite these difficulties, they were able to retell a short, well-sequenced, and accurate story. We therefore reiterate our opinion that a text-level task such as the PONA is a more ecologically valid way of appraising oral language performance in young school-age students [43].

Limitations

Students who participated in this study were all attending Queensland primary schools. Although it is unclear whether the results of this study will generalise to a broader population across Australia, all states and territories now adhere to one common English Curriculum [15], suggesting the performance of the students in the current study may reflect those from disadvantaged areas in other parts of Australia. Although it is well accepted that the oral language performance of bilingual children cannot be adequately appraised using one

language, such assessments are not readily available [68]. Furthermore, the advantages of using dynamic assessment over static assessment for students from CALD backgrounds are acknowledged, however, this was not feasible in the context of the current study. Detailed information regarding ethnicity, specifically the language/s spoken within the home and the dominant language requires further evaluation and was not recorded consistently across the school sites. This information would further assist in making comment around home language exposure and use.

The RAPT manual provides limited details regarding the ethnicity of the standardisation sample, nor does it provide adequate information regarding reliability (e.g., test-retest, inter-rater) and validity (e.g., concurrent, predictive, internal consistency) [12]. Furthermore, standardisation of this test was completed in 1987/1988. Taken together, it is unclear how suitable the RAPT is for use with an Australian population.

The administration of the test battery was conducted in the same order for all students, with the oral narrative and comprehension task first and the RAPT last. Performance on the RAPT task may have potentially been inflated as a result of a practice effect or a familiarity effect with the examiner. This may have underestimated the identification of students at risk of future difficulties. Future research should address this issue by randomising the administration of the tasks.

Conclusion

This study compared student performance on a sentence- and text-level task to determine whether these tasks are suitable when used at school-entry, and at the end of the school year following tier 1 intervention for Foundation year students from low socio-economic, linguistically diverse backgrounds. The findings suggest that the RAPT seems to evaluate one single language construct as demonstrated by the moderate-strong correlations

between Grammar and Information. Second, although both tasks seemed suitable to elicit scoreable responses from this group of students and were sensitive to progress over time, our results indicated that reliance on the sentence-level task would potentially result in under-identification of up to 15% of students who might be considered at-risk moving into their second year of schooling.

During the early school years, children are exposed to oral and written narratives on a daily basis and as such students' ability to respond to, comprehend and retell stories is of utmost importance for accessing the curriculum [15]. The results from this study suggest that assessment beyond the sentence-level skills assessed using the RAPT are required for this student population. The findings of this study illustrated that profiling students' oral narrative ability may provide a more robust view of oral language competence for academic success.

References

1. Justice LM, Bowles R, Pence K, Gosse C. A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly*. 2010;25:218-34. doi: 10.1016/j.ecresq.2009.11.002.
2. Westerveld MF, Gillon GT, Boyd L. Evaluating the clinical utility of the Profile of Oral Narrative Ability in 4-year-old children. *International Journal of Speech-Language Pathology*. 2012;14(2):130-40. doi: 10.3109/17549507.2011.632025.
3. Dennaoui K, Nicholls RJ, O'Connor M, Tarasuik J, Kvalsvig A, Goldfeld S. The English proficiency and academic language skills of Australian bilingual children during the primary school years. *International journal of speech-language pathology*. 2016;18(2):157-65. doi: 10.3109/17549507.2015.1060526.
4. Wake M, Levickis P, Tobin S, Gold L, Ukoumunne OC, Goldfeld S, et al. Two-Year Outcomes of a Population-Based Intervention for Preschool Language Delay: An RCT. *Pediatrics*. 2015;136(4):e838-e47. doi: 10.1542/peds.2015-1337. PubMed PMID: 26347428.
5. Ball CR, Trammell BA. Response-to-intervention in high-risk preschools: Critical issues for implementation. *Psychology in the Schools*. 2011;48(5):502-12. doi: 10.1002/pits.20572.
6. Barnett DW, Daly EJ, Jones KM, Lentz FE. Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education*. 2004;38(2):66-79. doi: 10.1177/00224669040380020101.
7. Division for Early Childhood of the Council for Exceptional Children, National Association for the Education of Young Children, National Head Start Association.

Frameworks for response to intervention in early childhood: Description and Implications 2013 29/12/2016.

8. Greenwood CR, Bradfield T, Kaminski R, Linas M, Carta JJ, Nylander D. The Response to Intervention (RTI) approach in early childhood. *Focus on Exceptional Children*. 2011;43(9):1.
9. Kaminski RA, Abbott M, Bravo Aguayo K, Latimer R, Good RH. The preschool early literacy indicators: Validity and benchmark goals. *Topics in Early Childhood Special Education*. 2014;34(2):71-82. doi: 10.1177/0271121414527003.
10. Hempenstall K. Research-driven reading assessment: Drilling to the core. *Australian Journal of Learning Difficulties*. 2009;14(1):17-52. doi: 10.1080/19404150902783419.
11. Justice LM. Evidence-Based Practice, Response to Intervention, and the Prevention of Reading Difficulties. *Language, Speech, and Hearing Services in Schools*. 2006;37(4):284. doi: 10.1044/0161-1461(2006/033).
12. Paul R, Norbury C. *Language disorders from infancy through adolescence: listening, speaking, reading, writing, and communicating*. St. Louis, Mo: Elsevier Mosby; 2012.
13. Carta JJ, Greenwood CR, Atwater J, McConnell SR, Goldstein H, Kaminski RA. Identifying preschool children for higher tiers of language and early literacy instruction within a response to intervention framework. *Journal of Early Intervention*. 2014;36(4):281-91. doi: 10.1177/1053815115579937.
14. Reilly S, Wake M, Ukoumunne OC, Bavin E, Prior M, Cini E, et al. Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics*. 2010;126(6):e1530-E7. doi: 10.1542/peds.2010-0254.
15. Australian Curriculum Assessment and Reporting Authority [ACARA]. *The Australian Curriculum - English 2016*. Available from: <http://www.australiancurriculum.edu.au/>

16. Bishop DVM, Adams C. A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry*. 1990;31:1027-50. doi: 10.1111/j.1469-7610.1990.tb00844.x/pdf.
17. Oakhill JV, Cain K. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*. 2012;16(2):91-121.
18. Westerveld MF, Moran CA. Expository language skills of young school-age children. *Language, Speech, and Hearing Services in Schools*. 2011;42(2):182-91. doi: 10.1044/0161-1461(2010/10-0044).
19. Catts HW, Fey ME, Zhang X, Tomblin JB. Language basis of reading and reading disabilities: Evidence from a longitudinal investigation *Scientific Studies of Reading*. 1999;3(4):331-61. doi: 10.1207/s1532799xssr0304_2.
20. Language Reading Research Consortium. The Dimensionality of Language Ability in Young Children. *Child Development*. 2015;86(6):1948-65. doi: 10.1111/cdev.12450.
21. Sénéchal M, LeFevre J-A. Parental Involvement in the Development of Children's Reading Skill: A Five-Year Longitudinal Study. *Child development*. 2002;73(2):445-60. doi: 10.1111/1467-8624.00417.
22. Neumann MM. A socioeconomic comparison of emergent literacy and home literacy in Australian preschoolers. *European Early Childhood Education Research Journal*. 2016;24(4):555-66. doi: 10.1080/1350293X.2016.1189722.
23. Snow CE, Burns MS, Griffin P. Preventing reading difficulties in young children. Washington, DC: National Academy Press; 1998.
24. OECD. PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I): Programme for International Student Assessment; 2014.

25. Mullis IVS, Martin MO, Foy P, Drucker KT. PIRLS 2011 International Results in Reading. Chestnut Hill, MA, USA: TIMMS & PIRLS International Study Centre; 2012.
26. Snow PC, Eadie PA, Connell J, Dalheim B, McCusker HJ, Munro JK. Oral language supports early literacy: A pilot cluster randomized trial in disadvantaged schools. *International Journal of Speech-Language Pathology*. 2014;16(5):495-506. doi: 10.3109/17549507.2013.845691.
27. Hay I, Fielding-Barnsley R. Competencies that underpin children's transition into early literacy. *Australian Journal of Language and Literacy*. 2009;32(2):148-62.
28. Wheldall R, Glenn K, Arakelian S, Madelaine A, Reynolds M, Wheldall K. Efficacy of an evidence-based literacy preparation program for young children beginning school. *Australian Journal of Learning Difficulties*. 2016;21(1):21-39. doi: 10.1080/19404158.2016.1189443.
29. Australian Government. Australian Early Development Census: About the AEDC data collection. In: Census AED, editor. Canberra, Australia: Department of Education and Training; 2016.
30. Australian Government. Australian Early Development Census National Report 2015: A snapshot of early childhood development in Australia. In: Census AED, editor. Canberra, Australia: Department of Education and Training; 2016.
31. McIntosh B, Crosbie S, Holm A, Dodd B, Thomas S. Enhancing the phonological awareness and language skills of socially disadvantaged preschoolers: An interdisciplinary programme. *Child Language Teaching and Therapy*. 2007;23(3):267-86. doi: 10.1177/0265659007080678.
32. Dowsett CJ, Huston AC, Imes AE, Gennetian L. Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly*. 2008;23(1):69-93. doi: 10.1016/j.ecresq.2007.06.003.

33. Assel MA, Landry SH, Swank PR, Gunnewig S. An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing*. 2007;20(5):463-94. doi: 10.1007/s11145-006-9039-5.
34. Westerveld MF, Vidler K. Spoken language samples of Australian children in conversation, narration and exposition. *International Journal of Speech-Language Pathology*. 2016;18(3):288-98. doi: 10.3109/17549507.2016.1159332.
35. Dennaoui K, Nicholls RJ, O'Connor M, Tarasuik J, Kvalsvig A, Goldfeld S. The English proficiency and academic language skills of Australian bilingual children during the primary school years. *International Journal of Speech-Language Pathology*. 2016;18(2):157-65. doi: 10.3109/17549507.2015.1060526.
36. Semel E, Wiig E, Secord W. *Clinical evaluation of language fundamentals - 4th edn - Australian standardised edition*. Marrickville: Harcourt; 2006.
37. McLeod S. Resourcing speech-language pathologists to work with multilingual children. *International Journal of Speech-Language Pathology*. 2014;16(3):208-18. doi: 10.3109/17549507.2013.876666.
38. Westerveld MF, Claessen M. Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology*. 2014;16(3):242-9. doi: 10.3109/17549507.2013.871336.
39. Caesar LG, Kohler PD. Tools Clinicians Use: A Survey of Language Assessment Procedures Used by School-Based Speech-Language Pathologists. *Communication Disorders Quarterly*. 2009;30(4):226-36. doi: 10.1177/1525740108326334.
40. Spencer TD, Petersen DB, Slocum TA, Allen MM. Large group narrative intervention in Head Start preschools: Implications for response to intervention. *Journal of Early Childhood Research*. 2015;13(2):196-217. doi: 10.1177/1476718X13515419.

41. Abel AD, Schuele CM, Arndt KB, Lee MW, Blankenship KG. Another look at the influence of maternal education on preschoolers performance on two norm-referenced measures. *Communication Disorders Quarterly*. 2016. doi: 10.1177/1525740116679886.
42. Ebert KD, Pham G. Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language Speech and Hearing Services in Schools*. 2017;48(1):42-55. doi: 10.1044/2016_LSHSS-16-0007.
43. Boerma T, Leseman P, Timmermeister M, Wijnen F, Blom E. Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *International Journal of Language & Communication Disorders*. 2016;51(6):626-38. doi: 10.1111/1460-6984.12234.
44. Renfrew C. The Renfrew language scales - Action picture test (Revised ed.). Milton Keynes United Kingdom: Speechmark Publishing; 2010.
45. Ward E, Fisher J. A picture-elicited screening procedure. *Child Language Teaching and Therapy*. 1990;6(2):147-59. doi: 10.1177/026565909000600203.
46. Jacobs D. Renfrew language scales: Applicability to a Victorian population. *Speech Pathology Australia National Conference Adelaide, South Australia* 2014.
47. Spencer TD, Slocum TA. The effect of a narrative intervention on story retelling and personal story generation skills of preschoolers with risk factors and narrative language delays. *Journal of Early Intervention*. 2010;32(3):178-99. doi: 10.1177/1053815110379124.
48. Petersen DB, Gillam SL, Gillam RB. Emerging Procedures in Narrative Assessment: The Index of Narrative Complexity. *Topics in Language Disorders*. 2008;28(2):115-30. doi: 10.1097/01.TLD.0000318933.46925.86.

49. Boudreau D. Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders*. 2008;28(2):99-114. doi: 10.1097/01.TLD.0000318932.08807.da.
50. Renfrew CE. *The Bus Story Test: A test of narrative speech*. 3rd ed. Oxford, UK 1995.
51. Westerveld MF, Vidler K. The use of the Renfrew Bus Story with 5-8-year-old Australian children. *International Journal of Speech-Language Pathology*. 2015;17(3):304-13.
52. Westerveld MF, Gillon GT. Profiling oral narrative ability in young school-aged children. *International Journal of Speech-Language Pathology*. 2010;12(3):178-89. doi: 10.3109/17549500903194125. PubMed PMID: WOS:000277984200002.
53. Dickinson DK, Tabors PO. *Beginning literacy with language: young children learning at home and school*. Baltimore, Md: Paul H. Brookes Pub. Co; 2001.
54. Lennox M, Westerveld MF, Trembath D. Evaluating the effectiveness of PrepSTART for promoting oral language and emergent literacy skills in disadvantaged preparatory students. *International Journal of Speech-Language Pathology*. 2016:1-11. doi: 10.1080/17549507.2016.1229030.
55. Florit E, Roch M, Levorato MC. The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Applied Psycholinguistics*. 2013;34(2):1-21. doi: 10.1017/S0142716411000749.
56. Australian Government. *Australian Early Development Census [AEDC] Canberra 2012* [cited 2014 20/11/2014].
57. Swan E. *Ko au na galo (Ana gets lost)*. Wellington New Zealand: Learning Media, Ministry of Education; 1992.

58. Scott CM, Windsor J. 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*. 2000;43(2):324-39.
59. Watkins RV, Kelly DJ, Harbers HM, Hollis W. Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*. 1995;38(6):1349-55. PubMed PMID: ISI:A1995TJ24600014.
60. Fey ME, Catts HW, Proctor-Williams K, Tomblin JB, Zhang X. Oral and Written Story Composition Skills of Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*. 2004;47(6):1301-18. doi: 10.1044/1092-4388(2004/098).
61. Miller JF, Gillon GT, Westerveld MF. Systematic Analysis of Language Transcripts (SALT), New Zealand Version 2012 [computer software]. Madison, WI: SALT Software LLC; 2012.
62. Krippendorff K. Content analysis: An introduction to its methodology. Newbury Park, CA: Sage; 1980.
63. Peat JK, Barton B. Medical statistics: a guide to SPSS, data analysis, and critical appraisal. Hoboken, NJ;Chichester, West Sussex;: John Wiley & Sons Inc; 2014.
64. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*. 2012;10(2):486.
65. Vacha-Haase T, Thompson B. How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*. 2004;51(4):473-81. doi: 10.1037/0022-0167.51.4.473.
66. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*. 2011;6(2):135-47. doi: 10.1016/j.edurev.2010.12.001.

67. Nation K, Snowling MJ. Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*. 2004;27(4):342-56.
PubMed PMID: ISI:000225473700002.
68. Westerveld MF. Emergent literacy performance across two languages: Assessing four-year-old bilingual children. *International Journal of Bilingual Education and Bilingualism*. 2014;17(5):526-43. doi: 10.1080/13670050.2013.835302.