

Retelling stories: The validity of an online oral narrative task

Child Language Teaching and Therapy
2023, Vol. 39(2) 150–174
© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions
DOI: 10.1177/02656590231155861
journals.sagepub.com/home/ctt



Gail Gillon

Child Well-being Research Institute, University of Canterbury, New Zealand
Better Start National Science Challenge, Liggins Institute, University of Auckland,
New Zealand

Brigid McNeill

Child Well-being Research Institute, University of Canterbury, New Zealand
School of Teacher Education, University of Canterbury, New Zealand
Better Start National Science Challenge, Liggins Institute, University of Auckland,
New Zealand

Amy Scott

Child Well-being Research Institute, University of Canterbury, New Zealand
Better Start National Science Challenge, Liggins Institute, University of Auckland,
New Zealand

Megan Gath

Child Well-being Research Institute, University of Canterbury, New Zealand

Marleen Westerveld

Griffith Institute for Educational Research, Griffith University, Australia

Abstract

This study examined the validity of data collected from a novel online story retell task. The task was specifically designed for use by junior school teachers with the support of speech–language therapists or literacy specialists. The assessment task was developed to monitor children’s oral language progress in their first year at school as part of the Better Start Literacy Approach for early literacy teaching. Teachers administered the task to 303 5-year-olds in New Zealand at school entry and after 20 weeks and 12 months of schooling. The children listened to a story with pictures via iPad presentation and were then prompted to retell the story. The children’s spontaneous language used in their story retell was captured and uploaded digitally via iPad

Corresponding author:

Gail Gillon, Child Well-being Research Institute, University of Canterbury, Antigua street, Private Bag 4800, Christchurch 8140, New Zealand.

Email: gail.gillon@canterbury.ac.nz

audio recording and analyzed using semi-automated speech recognition and computer software. Their responses to factual and inferential story comprehension questions were also analyzed. The data suggested that the task has good criterion validity. Significant correlations between story retell measures and a standardized measure of children's oral language were found. The Better Start Literacy Approach story retell task, which took approximately 6 min for teachers to administer, accurately identified children with low oral language ability 81% of the time. Growth curve analysis revealed that the task was useful for monitoring oral language development, including for English as second language learners. Boys showed a slower story comprehension growth trajectory than girls. The Better Start Literacy Approach story retell task shows promise in providing valid data to support teacher judgement of children's oral language development.

Keywords

oral narrative, oral language, assessment, language sampling, story retell

I Introduction

It is well-established that children's oral narrative production and comprehension abilities are important for reading comprehension and social interaction, underpinning educational success, and socio-emotional wellbeing (Nation et al., 2010; Portilla et al., 2021; Suggate et al., 2018; Westerveld et al., 2008). Children's oral narratives provide insights into their oral language skills at word, sentence, and text-level. Information related to children's semantic, syntactic, morphological, and phonological knowledge, as well as children's knowledge of discourse structure can be gleaned even from a relatively short oral narrative (Heilmann et al., 2010b; Murphy et al., 2022). Children's narrative comprehension demonstrates their ability to understand discourse-level language, including factual and inferential comprehension, and is a useful measure of young children's oral language comprehension (Paris and Paris, 2003).

Developing children's oral narrative skills is a common and important goal in early schooling. However, it is challenging for teachers to regularly monitor children's oral narrative development in valid and reliable ways, particularly given the time typically needed for detailed language sample analysis (Westerveld and Claessen, 2014). Increased global migration rates (McAuliffe and Triandafyllidou, 2021) and improved understanding of the cultural and social importance of bilingualism (Ramírez-Esparza et al., 2020) and indigenous languages (Gaffney et al., 2021), necessitate that teachers have access to narrative assessments that are valid for linguistically diverse learners. This study investigated the validity of a novel online oral narrative production and comprehension task. The assessment was specifically designed for implementation during regular teaching practice within English medium teaching contexts. It was also designed to provide data to support teachers' observations in ways that would foster collaborative practices between speech–language therapists (SLTs), literacy specialists and class teachers in supporting children with greater oral language and early literacy needs. The task incorporates a number of features to reduce teacher workload, improve data usability to guide teaching practice, and provide consistent task presentation to engage children in the activity.

I Oral Narrative Production and Comprehension

Children develop their oral narrative skills during the preschool and early school years. From a story structure perspective, oral narratives consist of a sequence of goal-directed attempts or

actions that serve to solve a problem (Trabasso and Nickels, 1992). Analyzing American English speakers' narrations of the picture story "Frog, Where are You" Trabasso and Nickels (1992) found that children made significant developmental gains in their oral narrative ability between the ages of 3 and 6 years. They reported that 3-year-old children produced descriptive statements, 4-year-old children included some goal-directed statements in their oral narratives, and 5-year-old children demonstrated a shift to organizing their story around a goal. To construct an narrative, however, children need to draw upon oral language skills across the domains of vocabulary, syntax, morphology, and phonology (see Hughes et al., 1997). Previous studies have found a clear developmental progression in children's oral narratives on measures of vocabulary (number of different words), mean length of utterance, and clausal density (Heilmann et al., 2010a; Justice et al., 2006; Westerveld et al., 2004).

The oral narrative skills of bilingual learners generally progress in a consistent way across languages, but may be influenced by bilingual factors such as learning languages simultaneously or sequentially and the educational context for language learning (Pesco and Kay-Raining Bird, 2016). Huang et al. (2022) examined oral narrative skills from a story retell task for children for whom Spanish was their home language but who spoke both Spanish and English. The 95 participants were taught in Grade 1–3 dual Spanish and English learning programs within a southwestern city in the USA. The researchers found the children showed similar performance across most oral narrative measures when retelling a story in Spanish compared to when telling the story in English. The exception was the higher number of different words used when retelling the story in their home language (Spanish). To investigate oral narrative development across languages and with bilingual learners a pan European research group (Narrative and Discourse Working Group of COST Action IS0804) developed a novel oral narrative assessment task specifically for young children from linguistically and culturally diverse backgrounds: Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012). A series of studies tested this oral narrative assessment with children aged 3–9 years who are speakers of one or more of the following languages: English, Finnish, German, Greek, Hebrew, Italian, Russian, Slovak, Swedish, and Turkish (Gagarina et al., 2016). Pesco and Kay-Raining Bird's (2016) summary of generalizations that can be drawn from these studies include: developmental changes in children's oral narrative abilities were particularly evident between 4 and 7 years; children performed similarly across languages on measures of story complexity and in some cases in their story structure knowledge; and when differences were observed across languages, the children's performance was almost always superior in their majority language for sequential bilingual learners.

In addition to considering children's production of oral narratives, their comprehension of narratives also requires evaluation. To fully comprehend a narrative, both factual and inferential comprehension are essential. Factual comprehension demonstrates understanding of information that is explicitly stated, such as correctly answering who, what, and where questions. Inferential comprehension demonstrates understanding of information that is not explicitly stated, but is inferred by the ability to draw on relevant prior knowledge and clues in the text or illustrations. For example, children need to understand causal relationships between the characters' mental states, goals, and actions and integrate background information to understand why certain situations occur. Research suggests that oral narrative comprehension improves with age (e.g. Westerveld and Gillon, 2010), with significant development observed in inferential comprehension skills from ages 3- to 6-years (Filiatrault-Veilleux et al., 2016).

2 Importance of oral language monitoring

The ability of teachers to accurately monitor children's progress in oral language development is critically important. In addition to the primary role of oral language to children's communication

and social development, oral language skills form the basis of children's written language development (see Snowling and Hulme, 2021, for a review). From a theoretical perspective, it has long been recognized that skilled readers draw upon both word recognition and oral language comprehension abilities to comprehend written text. Tunmer and Hoover (2019) described a Cognitive Foundations Framework based on the Simple View of Reading (Gough and Tunmer, 1986) to highlight the cognitive skills children need for reading success. Within this framework, alphabet knowledge, phonological awareness and phonological decoding ability are central to word recognition. Semantic, syntactic knowledge and the ability to integrate relevant background knowledge and inferencing skills are all central for language comprehension. Oral language comprehension ability together with word recognition skills can account for the majority of the variance in children's reading comprehension performance (Hoover and Tunmer, 2018). Teaching approaches that support children in their first year at school to develop their phoneme awareness, phonic knowledge and their ability to use these skills in decoding written words have consistently proven effective in advancing children's early reading and spelling abilities (see Gillon, 2018 for a review). However, developing children's skills in other areas of oral language is also important. Children's semantic (e.g. vocabulary) knowledge contributes directly to reading comprehension and indirectly to word recognition (Tunmer and Chapman, 2012). Both receptive and expressive vocabulary skills are important to reading, with explicit and systematic vocabulary instruction to improve reading comprehension abilities having long been advocated (see Manyak et al., 2021; Wright and Cervetti, 2017 for reviews). Furthermore, Metsala et al. (2021) found that children's syntactic and morphological knowledge makes unique and individual contributions as children's reading comprehension performance advances. Finally, background knowledge (Smith et al., 2021) and higher-level oral language skills of inferencing, comprehension monitoring and story structure knowledge (Oakhill and Cain, 2012) have all been found to be distinct predictors of more advanced reading comprehension abilities.

Young children who enter school with lower levels of oral language are at risk for persistent reading difficulties which may have a long-term negative impact on both their educational and health outcomes (McLeod et al., 2019). It is important, therefore, that teachers' assessment and instructional practices are appropriate to support individual children's needs to advance their oral language production and comprehension skills. Monitoring young children's oral narrative development in response to quality teaching may provide insights with regard to those who require additional support to prevent later reading comprehension difficulties. Petersen et al. (2020) implemented a pilot study using parallel forms of reading and listening comprehension for children in second and third grades. Children were selected from dual language immersion teaching classes focusing on either English and Spanish ($n = 78$) or English and Navajo (Dine) ($n = 32$). Listening and reading English subtests from the Narrative Language Measures (NLM) of the CUBED assessment (Petersen and Spencer, 2016) were administered. The results showed a large correlation ($r = 0.75$ controlling for reading fluency) between oral language comprehension and reading comprehension measures. Furthermore, factor analysis supported a unidimensional model of comprehension highlighting the relationship between oral and reading comprehension. Monitoring data that can support teacher practice of strengthening children's oral language comprehension in their first year at school is therefore likely to be associated with later reading comprehension success.

Monitoring oral narrative skill development may also highlight for class teachers' areas of relative strength for children. Some cultures have a strong history of oral narrative and storytelling passed down through the generations which may positively influence children's oral narrative development. For example, Gardner-Neblett and Iruka (2015) found, through analyzing data from the Early Childhood Longitudinal Study in the USA, that preschool children's oral narrative ability was more closely linked to early literacy skills for African American children than children

from other ethnic groups. The researchers hypothesized that this may reflect a historical African American cultural preference for oral communication and the emergence over time of a diverse range of oral narrative styles among many African American children. This oral narrative ability may help foster these children's ability to transfer knowledge of oral language to early literacy contexts.

Monitoring oral narrative development in boys during the preschool and early school years may be particularly important (Fey et al., 2004; Gardner-Neblett and Sideris, 2018). The Progress in International Reading Literacy Study (PIRLS) (Mullis et al., 2017) consistently demonstrated that boys' reading comprehension is lower than girls in the fourth grade at school. In the 2016 PIRLS data, girls outperformed boys in comprehending narrative fiction in 48 of the 50 participating countries and in comprehending factual information in 38 of the 50 participating countries. Teachers' ability to evaluate children's oral narrative comprehension in their first year at school in valid and reliable ways will provide insights as to whether boys are developing these foundational skills for later reading comprehension in similar ways to girls, particularly in relation to comprehending fictional texts.

3 Teacher practices of monitoring oral narrative development

Teachers routinely assess children's phonic, phoneme awareness, word decoding or word recognition skills in ways that provide quantitative or objective data to support their observations of children's learning. This combination of teacher judgment and objective testing on phonic and phoneme awareness tasks may be particularly helpful in accurately identifying children who are at risk for persistent reading difficulties (Snowling et al., 2011). However, teachers' routine assessment of other aspects of children's oral language production and comprehension in ways that provide objective data to support teachers' observations is less common. Rather, teacher judgement based on observation and listening to children's oral language during class activities is often used.

In New Zealand, the country for the current study, Cameron et al. (2019) analyzed the responses of 745 junior school teachers from a national survey. (These teachers taught children aged 5 and 6 years). The researchers were seeking to understand how teachers assess young children's oral language and early literacy skills. Teachers acknowledged that oral language assessment was an area that required further attention and the majority of the teachers (53%) indicated the need for new oral language assessments that are culturally relevant, related to current research, and easy to administer. All teachers who answered the question related to assessment type they currently use ($n = 684$) reported using a "running record," with 85% undertaking a running record within 10 weeks of a child starting school. This involves children reading aloud a children's "reader" (typically reading a text they have read before) and teachers noting the child's reading errors. Following this oral reading activity, 75% of teachers surveyed indicated that they then asked the children to retell the story and 81% of respondents indicated that they asked children some questions about the story. Teachers indicated that they conducted this type of assessment on a regular basis to monitor children's language progress (e.g. once every couple of months). Often, the language comprehension questions asked after the story retell were teacher-generated questions (76%). The respondents used predominantly teacher judgement (94% of teachers), to evaluate the quality of the children's story retell.

Contrary findings related to the accuracy of teacher judgement of school-aged children's oral language abilities are reported in the literature. While some studies have found teacher judgement of young children's need for specialist services such as speech-language pathology to be reasonably accurate (Williams, 2006), others have reported that teacher ratings of children's oral communication skills show poor sensitivity (Antoniazzi et al., 2010). Teacher judgement of students' abilities

can be biased by a number of factors such as the average ability of the children in the teacher's class or how children perform on other proximal skills. For example, teacher judgement of children's reading ability may be biased by their knowledge of children's spelling performance (Schmitterer and Brod, 2021). The ability for teachers to plan and identify specific oral language teaching goals or to accurately monitor a child's oral language progress based only on teacher judgement is challenging. Recent data suggest that Year 1 class teachers have a bias to overestimate children's language and early literacy abilities and are less likely to see the extent of diversity in children's language learning (Sanrey et al., 2021). Furthermore, children's story retell quality may vary depending on the quality, relevance, length, use of picture prompts, cultural relevance and familiarity of the story children are being asked to retell (Boudreau, 2008). Judgements about children's comprehension through asking questions at the end of a story is influenced by factors related to the test items, text genre and children's reading ability (Kim and Petscher, 2020). If variables such as the type of story and types of questions asked change at each story retell assessment, then teachers may be unable to use these data to monitor growth in oral narrative abilities over time. Teachers need more detailed and valid data around individual children's oral language performance in order to help inform their teaching decisions and to corroborate their judgements based on observations.

Although a variety of valid and reliable standardized, norm-referenced oral narrative and oral language comprehension assessments are commercially available (e.g. Gillam and Pearson, 2017), teachers' regular use of such assessments is often limited (Cameron et al., 2019). Challenges related to the cultural appropriateness of a standardized test for a specific population, the time involved in test administration and analysis, and the relevance of the data (i.e. standardized scores) to guide teaching decisions may all pose barriers for teachers to use these assessments in everyday class practice. One less formal method to assess children's oral narrative abilities is through the use of "language sampling" techniques. This method involves recording children's spontaneous language during an oral narrative activity, such as telling or retelling a story, or relating a personal experience, transcribing verbatim the children's utterances and then analyzing the language. Analyzing children's spontaneous language is considered a more authentic and less biased way to observe children's oral language abilities, particularly for children who come from linguistically diverse backgrounds (Wood et al., 2018a).

SLTs have long used language sampling techniques as part of best practice in the assessment of children's oral narrative development (Kemp and Klee, 1997), although inconsistency in language sample data collection methods and methods of data analysis is evident (Voniati et al., 2021; Westerveld and Claessen, 2014). Children's oral narrative language samples are typically analyzed at the microstructure level (e.g. identifying the number of different words used, mean length of utterances, types of words used, word level errors) as well as at the macrostructure level (e.g. analyzing children's use and understanding of the goal-directed nature and structure of a narrative). Few class teachers use language sample analysis techniques to examine children's spontaneous language as part of their regular monitoring of children's oral narrative development (Cameron et al., 2019). An obvious barrier to their use is the time involved in transcription as well as the need for increased teacher professional learning in oral language assessment (Malec et al., 2017; Voniati et al., 2021).

Justice et al. (2010) and Bowles et al. (2020) examined the validity of the Narrative Assessment Protocol (NAP & NAP-2) designed for preschool teachers' use. The NAP was specifically designed to bypass the need for verbatim transcription of a child's oral narrative to reduce administration time through the use of a scoring protocol sheet. In examining task validity and reliability, research assistants (e.g. graduate students in speech-language pathology or education) made judgements related to both microstructure and macrostructure elements from listening to a video recording of

3–6-year-old children's story telling attempts. The occurrence of language elements (e.g. use of elaborated noun phrases: zero, one, two, or three or more occurrences) was recorded against pre-determined items listed on the protocol. Good psychometric properties were reported from the task, which took about 20 min to administer. Although NAP-2 holds promise for larger scale use by educators, some limitations of the scoring protocol are evident as acknowledged by the researchers. In contrast to previous research indicating the multidimensional nature of narrative ability (e.g. Justice et al., 2006; Westerveld and Gillon, 2010), the items contained in the NAP-2 loaded onto a single factor of narrative ability, namely language complexity, and did not include measures of productivity such as the number of different words or the total number of words.

4 Current study

This current study extended the work of Bowles et al. (2020) through investigating oral narrative assessment tasks for junior school class teachers' use as part of their regular teaching practice. We examined the usefulness of oral language sample analysis, using semi-automated speech recognition from an online narrative story retelling task followed by comprehension questions. Previous research demonstrated that asking young children to retell a story they have heard is a valid and useful context to gather a spontaneous sample of children's oral language for analysis (Westerveld et al., 2012). The current assessment was specifically designed for teachers to implement and to use the data gained to inform their teaching practice. With advancements in online technologies, automated speech recognition, and digital device use in classrooms, opportunities for teachers to access more detailed information about children's oral language in more natural settings is feasible through reduced time involved in administration and analysis. Automatic presentation of a story to ensure consistency of story presentation across time and across children, guidance on prompts used during the story retell and consistency of comprehension questions may enhance the usefulness of story retell tasks.

Given the importance of microstructural aspects of children's narrative samples to understanding and predicting longer-term language ability (e.g. Murphy et al., 2020), the development of psychometrically valid and reliable measurements of young children's narrative skills is critical. The current study examined quantitative data generated from a novel online assessment task related to the microstructural elements of children's story retelling narrative and children's factual and inferential oral language comprehension to validate this novel assessment task.

We focus on class teachers' assessment of children's story retelling and story comprehension in their first year at school, since this is an area less researched within the educational domain. Malec et al. (2017) undertook a systematic review of research related to oral language assessment use with young school-aged children. Of the 201 studies that were relevant to their initial search criteria, most were published in speech–language pathology or special education journals. Nine of the ten final articles that met criteria for their in-depth review involved speech–language pathologists or research assistants implementing the assessments. The researchers emphasized that research related to oral narrative assessments is important for teachers, but the relevance of this research to class teaching practice needs to be a focus for future research. Teachers interviewed as part of Malec et al.'s study highlighted the need for further guidance on oral language assessments (such as story retelling) that are suitable for implementation in a class context and that can inform their teaching practice.

5 Study aims

This study aimed to establish the validity of a novel online oral narrative production (story retell) and comprehension task that has been specifically designed for class teachers to use within their

regular teaching practice. The validity of the task to measure 5-year-old children's oral language skills at school entry and across their first year at school was investigated. The study also aimed to identify growth patterns in the children's oral narrative production and comprehension measures across the first year at school.

The following research questions were asked:

1. Are oral narrative story retell and narrative comprehension scores on the novel assessment task concurrently associated with language abilities measured with an established measure of oral language (scores on the Clinical Evaluation of Language Fundamentals Preschool—CELF-P2; Wiig et al., 2006)? (Criterion validity).
2. Are oral narrative story retell and narrative comprehension scores concurrently associated with reading ability (comprehension, accuracy, and rate, as assessed on the Neale Analysis of Reading Ability, Neale, 1999)? (Predictive validity).
3. Are the oral narrative story retell and narrative comprehension scores sensitive to growth over the first year of school?
4. Do patterns of growth in oral narrative story retell and narrative comprehension over the first year of school differ based on sex or whether children speak English as a second language?

II Method

2 Participants

The participants were drawn from a larger study investigating the efficacy of the *Better Start Literacy Approach (BSLA)* in New Zealand new entrant and Year 1 classes (Gillon et al., 2019, 2020, 2022). At study commencement, there were 303 children (51% male, 49% female) with returned parent consent forms to participate in the study, who also met our eligibility criteria of being in their first 20 weeks at school and aged between 5 years 0 months to 5 years 5 months (mean age 62.6 months, $SD = 1.61$). The cohort included the following ethnic groups: 60% NZ European, 22% Asian, 10% Māori, 9% Pasifika, and 9% other ethnicities (note that multiple ethnic affiliations were allowed). Within this cohort, 12% of the children spoke English as a second language. Approximately half of the children (48.4%) were from schools in low- to middle-income neighborhoods with the remainder in middle to high-income neighborhoods as indicated by the school's address against the New Zealand Deprivation Index. There were no significant differences between the number of males and females with regard to their school socioeconomic areas [$F(1, 301) = 0.28, p = 0.60$].

Children entered the study in two cohorts: Cohort 1 children ($n = 149$) completed assessments at three time points across their first year at school: baseline (Time 1), 20 weeks post-baseline (Time 2), and then approximately 12 months post-baseline (Time 3). At time 3, 134 children from cohort 1 (90%) remained in the study. From the second cohort of children to enter the study ($n = 154$), 151 children (98%) completed Time 1 and Time 2 assessments. Children from this cohort only completed the third assessment if they were identified as requiring Tier 2 small group teaching support (based on the design of the larger study; see Gillon et al., 2022 for more information). A total of 42 children from this group were selected to complete the Time 3 assessment. The results reported below relevant to the Time 3 assessment therefore include 176 children (134 children from Cohort 1 and 42 children with greater learning needs from Cohort 2 (based on lower phonological awareness and non-word reading scores).

3 Assessments

Class teachers administered the BSLA online story retell and narrative comprehension task to the children in their class at the three assessment points over a 12-month period. Research assistants (qualified teachers, SLTs, or SLT senior students) also helped some teachers with administering these assessments, particularly at the Time 3 assessment point due to time constraints of the study.

a Online story retell task development. Using a purpose-built website for the BSLA online story retell task, analysis and reporting features were developed through an iterative process. In the following section, the current state of these online technologies is described (continued development work is in progress to further advance reporting efficiencies and speech automation features, Scott et al., 2022). Both research evidence and practical implications guided the development of the task in order to be suitable for teachers to use in their everyday class practice. These aspects are as follows:

1. Teacher professional learning: Teachers completed an online learning module (approximately 3 h of learning) to strengthen their knowledge of English phonology, semantics, syntax, and morphology at the commencement of the study. Teachers could revisit this learning content at any time during the course of the study. In addition, teachers also completed phoneme awareness and oral narrative assessment and teaching modules. These modules were designed to be delivered online, but for the purposes of this study, members of the research team (PhD qualified SLTs and literacy specialists) provided this professional learning and development to the teachers in two face-to-face workshop presentations (approximately 3.5 h for each workshop) and through coaching during the assessment implementation.
2. Selection of story retell task. A child's ability to generate their own story (story telling) or to retell a story they have heard (story retelling) are both stable constructs of young children's oral narrative ability over time (Pinto et al., 2018). The researchers selected a story retell task for use in this study for several reasons: it is a common activity within the school curriculum; it can easily be presented in an online digital format; it provides insight into children's language use, as well as knowledge of story structure; it allows for easier automated speech transcription through artificial intelligence expecting what a child might say given the constrained story stimulus; it provides the opportunity for an oral language comprehension activity, and previous research has suggested that oral narrative retells are generally longer and contain more story components than story generation contexts (Merritt and Liles, 1989).
3. Cultural relevance and story type. Within culturally responsive teaching frameworks (e.g. Gillon and Macfarlane, 2017; Ratima et al., 2020) it is important to ensure that the story to be retold is relevant to the children's cultural context (McConnell and Loeb, 2021). The research team wrote a story called "*Tama and the Playground*" (Gillon et al., 2019b) specifically for this purpose, ensuring that the characters, illustrations and story reflected the New Zealand cultural context and would be familiar to young school-aged children in New Zealand (see Supplemental Appendix A). The structure and length of the story was based on a previous story that had proven useful for story retell elicitation in 4- to 7-year-old children (Westerveld and Gillon, 2010).
4. Digital presentation of the task. The consistency of story retell task presentation is critical if teachers are to access valid data related to children's oral narrative development (Boudreau, 2008). An online presentation ensures consistency of presentation and removes potential

bias that might influence a child's response (e.g. bias of one teacher being much more animated in the story telling than another, or a teacher's speech clarity or accent influencing what a child hears, or teachers providing additional prompts or clarification when telling the story or when asking comprehension questions). The BSLA story retell task in this study used a purposely built assessment website which the teachers accessed via their computer or iPad to present the task to the children. Task instructions for the child to follow were presented on the screen for the teacher to read aloud to the child at the beginning of the task (see Supplemental Appendix A). The child then listened to the story with accompanying pictures on the screen (pictures only with no written text to control for children's reading ability). A professionally trained male speaker with a New Zealand accent recorded the story. The story encompassed five digital pages of pictures, and the pre-recorded audio of the story matched each picture. The children could move through the story in their own time—they were prompted to click to go to the next page by a "ding" sound. After clicking to move to the next page, the story automatically began. Pause and replay buttons were available on each screen page of the story should an interruption have occurred during the telling of the story.

b Story retell. Following the presentation of the story, teachers were instructed to ensure the iPad or computer recording device was switched on and then children were prompted to retell the story in their own words, using the online slide prompts read to them by the teacher (see Supplemental Appendix A). Children then manually progressed through the same series of pictures presented on the screen, while retelling their story. Encouraging but non-specific prompts were also provided to teachers to use if required (e.g. "what happened at the beginning?"; "and then what happened?").

c Story comprehension. Following the retell, children were asked five comprehension questions which were presented on the screen for the teacher to read aloud. Children's responses to the questions were also recorded. Three of the questions were factual questions tapping understanding of character identities and actions. Two were inferential questions tapping comprehension of story context and motivation for character's actions (see Supplemental Appendix A for details). Following earlier pilot work and feedback from teachers, five questions was determined as sufficient to capture students' factual and inferential understanding of the story, while not adding unnecessary time to the task. A scoring rubric was developed based on the pilot trials which identified children's most likely responses to these questions. Each response was scored as 2 for correct, 1 for partially correct, or 0 for incorrect. If the child did not respond or replied "I don't know," they were given a score of 0. Current assessment developments allow for teachers to score online in real time with the children's scores being automatically calculated and provided in a report format for the teacher.

5. Story recording and analysis. The latest software developments of the task included the following features¹: through the use of an iPad or computer laptop, the child's story retell was captured. This audio file of the child's story retell was uploaded from the recording device to the student's profile on the assessment platform and was immediately available for listening. Once the teacher had completed the task, a flag was generated in the assessment system, indicating to the research assistant (SLT) that an audio file was ready for transcription. These audio files were downloaded in bulk, transcribed automatically using speech-to-text software, checked for accuracy and then coded for analysis using Systematic Analysis of Language Transcripts (SALT) software (Miller et al., 2017). Following this analysis, pdfs of the language transcripts and analysis were uploaded in bulk back into the assessment

website, populating the children's individual assessment profile for teachers to access. This semi-automated transcription process and analysis took about 12 min per child at the time of the study (ongoing software development has reduced this time to less than 7 min per child, see Scott et al., 2022). This analysis was available for the teachers to access within 48 h of the audio file being uploaded (current developments being piloted remove the need for the research assistant in this process unless the teacher requests assistance).

The SALT analysis can generate a wide range of outputs. However, to keep the coding process manageable, we limited the data to areas particularly useful to guide teaching practice. In addition, given the short length of the language sample, measures of language productivity, semantic diversity and utterance length are more reliable measures (Heilmann et al., 2010). For each child's transcript of their story retell, we generated the following quantitative data automatically using SALT (see Supplemental Appendix A for analysis details):

- (a) Number of utterances/sentences (c-units)
- (b) Number of words
- (c) Number of different words
- (d) Number of nouns
- (e) Number of verbs
- (f) Number of adjectives
- (g) Number of adverbs
- (h) Mean length of utterance/sentence (average number of words per utterance)
- (i) % of intelligible utterances
- (j) Word level errors (e.g. pronoun errors)
- (k) Omitted bound morphemes (e.g. omission of regular -ed tense ending)

6. Task efficiency to monitor growth. The complete task of children listening to the *Tama and the Playground* story, retelling the story, and answering five comprehension question on average took 6 min to complete per child. The digital presentation with onscreen administration instructions and prompt questions also allowed for a teacher assistant to present the task to the child. The assessment website automatically generated reports for teachers and for children's parents, which teachers could adapt for their own context. The audio file of the child's oral narrative was in a format that teachers could easily share digitally with the child's family, SLT or literacy specialist who may be involved in supporting the child's learning.

d Clinical Evaluation of Language Fundamentals Preschool. Children's oral language abilities were assessed at the baseline assessment point (school entry) using the Clinical Evaluation of Language Fundamentals Preschool—Second Edition-Australian and New Zealand Edition (CELF-P2; Wiig et al., 2006). The CELF-P2 is a widely used measure with validity demonstrated using the standardization samples (Wiig et al., 2006) and has been previously shown to have good internal consistency (range: .73–.96 across subtests) and test–retest reliability (.78–.94; Black et al., 2020; Denman et al., 2017).

Core language scores were derived from a combined score on the three subscales of sentence structure, word structure, and expressive vocabulary. The combined cohort's mean standard score was 95.3 ($SD = 17.32$). A one way analysis of variance (ANOVA) indicated that girls ($M = 99.07$, $SD = 15.50$) scored significantly higher than boys ($M = 91.70$, $SD = 18.25$) on the CELF at baseline [$F(1300) = 14.26$, $p < 0.001$].

e *Neale Analysis of Reading Ability*. Reading accuracy, comprehension, and rate were assessed at the Time 3 assessment point using the Neale Analysis of Reading Ability (NARA; Neale, 1999). The NARA is a widely used test in New Zealand of children's reading ability. Concurrent validity has been established and parallel forms reliability is very high (.91–.96 across forms; McKay, 1996).

Raw scores for accuracy, comprehension and rate were converted to percentile ranks for descriptive purposes. Mean percentile ranks were as follows: accuracy mean = 47.51 ($SD = 33.08$), comprehension mean = 39.85 ($SD = 32.50$), rate mean = 42.85 ($SD = 27.03$).

3 Oral narrative data reliability

For the purposes of this study, 20% of the children's audio files were randomly selected from each assessment time point for reliability scoring. A single assessor re-transcribed these audio files and analyzed these new transcripts using SALT coding. Reliability was calculated as the intra-class correlation (ICC) between the two sets of scores. ICCs for the 11 variables derived from SALT coding ranged from 0.811 to 0.998 across time points.

To assess the interrater reliability of the narrative comprehension scoring, audio recordings of 76 children (25%) answering the comprehension questions were re-scored by an independent second assessor. Reliability was calculated as the ICC between the two sets of scores. The ICC of 0.884 indicated high interrater reliability.

III Results

Prior to undertaking analyses related to task validity, we examined the associations between the oral narrative scores generated from the analyses of the children's story retells and their underlying factor structure. Narrative comprehension was positively correlated with all oral narrative production variables other than omitted bound morphemes and word-level errors. Aside from these two error variables (which were correlated with each other), all remaining nine oral narrative variables were significantly positively correlated with each other (r^2 's > 0.22, all $p < 0.001$; see Supplemental Appendix B, Table B1).

The distributions of scores at baseline on narrative comprehension and number of different words in children's story retell are presented in Figures 1 and 2, respectively. The skewness and kurtosis were both well-within the acceptable ranges across all time points, indicating normally distributed data (narrative comprehension: skewness range -0.08 to -0.59 , kurtosis range 0.23 to -0.50 ; number of different words: skewness range 0.14 – 0.19 , kurtosis range 0.22 to -0.36 across assessment points). This suggests that the difficulty levels of these tasks are appropriate for 5–6-year-old children. Descriptive statistics, including skewness and kurtosis, for all oral narrative variables are provided in Supplemental Appendix B, Table B2.

A principal component analysis (PCA) was used to examine the structure of the oral narrative production scores for potential dimension reduction. All eleven oral narrative production scores were entered into the analysis, and a rotated solution was produced through Promax rotation. Narrative comprehension was excluded, as it was measured through a separate task and was considered to be conceptually distinct. The rotated loading plot, shown in Figure 3, was indicative of a two-factor solution, with the two error variables (omitted morphemes and word level errors) loading on a separate component factor. The extraction of two eigenvalues with values greater than 1 confirmed the two-factor solution (accounting for a combined 64.85% of the variance in scores). Thus, the results support the creation of two composite scores of oral narrative production skills.

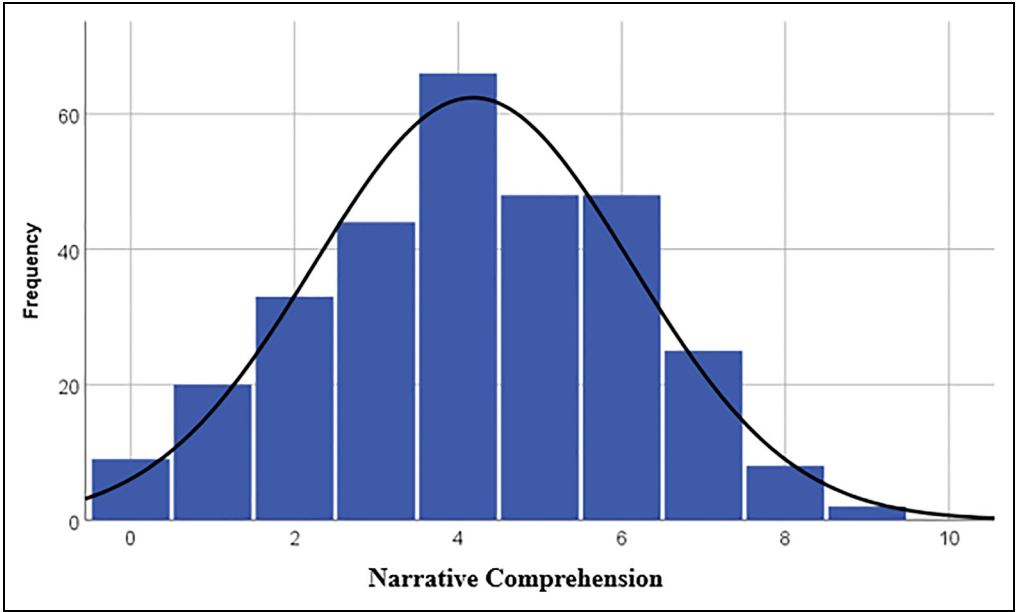


Figure 1. Distribution of narrative comprehension scores at baseline.

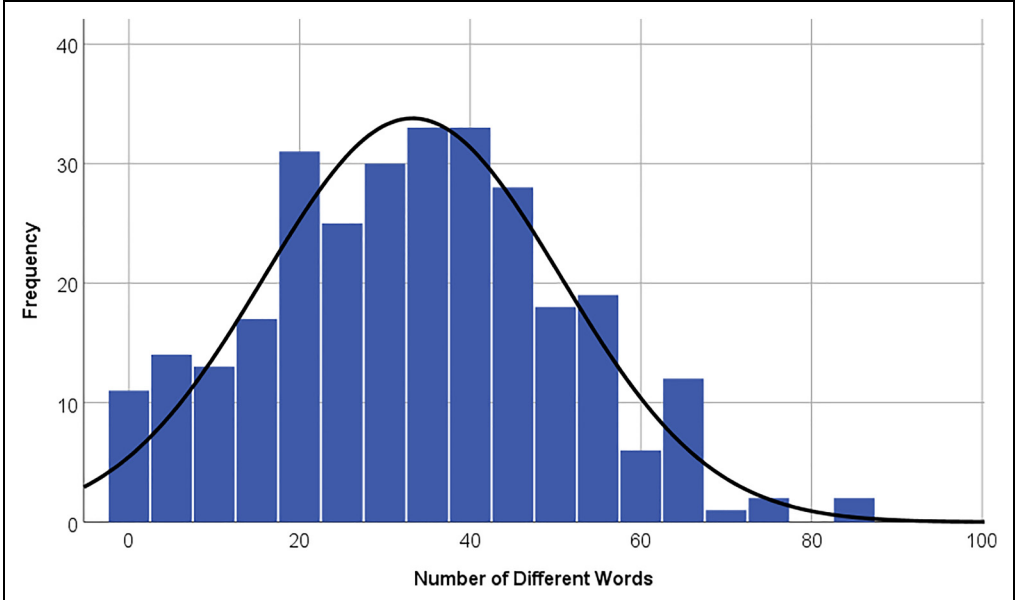


Figure 2. Distribution of number of different words at baseline.

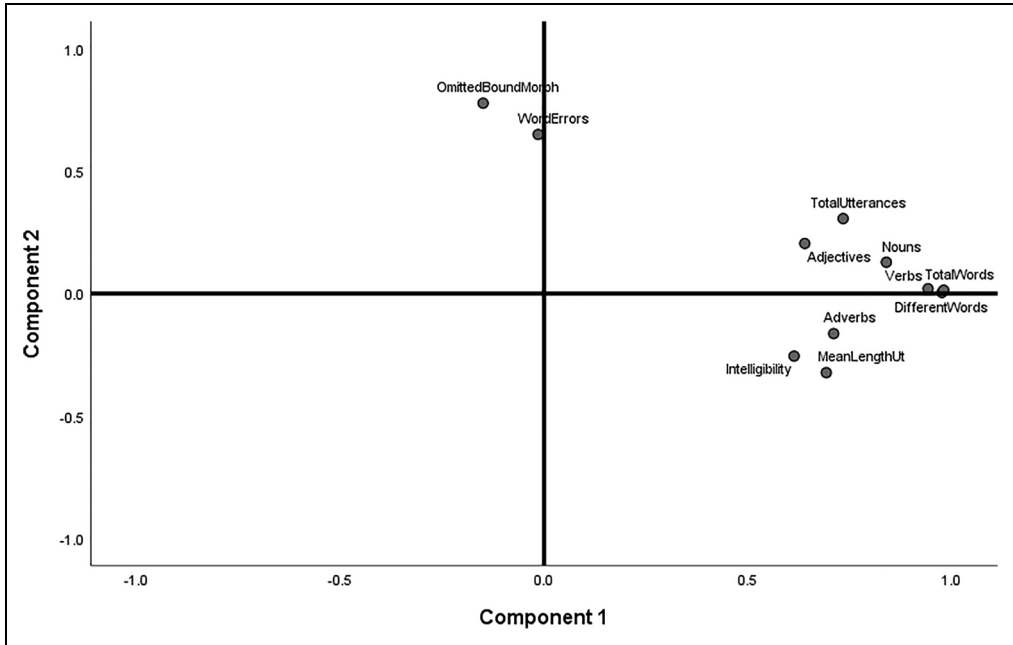


Figure 3. Rotated component loading plot for oral narrative production variables.

Table 1. Concurrent associations at school entry between Clinical Evaluation of Language Fundamentals Preschool (CELF-P2) core language scores, comprehension, oral narrative production composite score, and oral narrative error composite score.

Oral narrative variable	Correlation with CELF
Narrative comprehension	$r = 0.38, p < 0.001$
Narrative productivity composite	$r = 0.32, p < 0.001$
Narrative production error composite	$r = -0.25, p < 0.001$

n = 303. CELF scores are CELF-P2 core language scores

I Concurrent validity at school entry

The first research question aimed to establish the validity of the story retell and comprehension tasks as measures of oral language skills at school entry. We first examined concurrent associations of these measures with children’s CELF Core Language scores. Table 1 presents the correlations between CELF scores and scores on the narrative production and comprehension task, all assessed at school entry.

As seen in the table, CELF scores were significantly positively correlated with both comprehension and oral narrative production measures. CELF scores were negatively correlated with the oral narrative error composite score.

a Identification of children with low language. We next used discriminant function analysis to determine the ability of the comprehension, oral narrative productivity score, and oral narrative error

score to identify children with low oral language skills. We first identified children with low language skills as those scoring 77 or below on the CELF ($n = 41$; 13.9%), representing 1.5 standard deviations below the standardized CELF mean. A discriminant function analysis predicting a low CELF score from narrative comprehension, the oral narrative production composite, and the error composite assessed at school entry confirmed the discriminative ability of the comprehension and oral narrative production measures (Wilks' Lambda = 0.83, $\chi^2(3) = 40.67$, $p < 0.001$). Children with low language skills on the CELF were correctly classified by their performance in story retell production and comprehension questions 81.3% of the time.

2 Predictive validity at the end of the first year of school

The second research question examined the correlation between comprehension and oral narrative production skills measured at the end of the first year of school and children's concurrent scores on the NARA (raw scores). Narrative comprehension scores were positively correlated with NARA on accuracy ($r = 0.20$, $p = 0.009$), comprehension ($r = 0.22$, $p < 0.001$), and rate ($r = 0.17$, $p = 0.02$). The error composite was negatively correlated with concurrent scores on NARA accuracy ($r = -0.27$, $p = .006$), comprehension ($r = -0.35$, $p < 0.001$), and rate ($r = -0.22$, $p = 0.03$). The oral narrative production composite was not significantly correlated with any of the NARA scores: accuracy ($r = 0.11$, $p = 0.17$), comprehension ($r = 0.14$, $p = 0.06$), and rate ($r = 0.08$, $p = 0.30$).

3 Growth curve models

To examine the change in narrative comprehension and narrative production skills over the three assessment points, we used linear mixed effects growth curve modeling. Growth curve models allow for the modeling of individual trajectories while also providing insights on the average change over time and factors associated with change (Hedeker and Gibbons, 2006). Trajectories of comprehension and oral narrative skills over time were modelled as a function of sex (male, female) and English as a Second Language (ESL) status (yes, no), including both fixed and random effects.

Growth curve models provide a more robust and flexible means for assessing growth in longitudinal data than the typical repeated measures analysis of variance, particularly for datasets that contain missing data. Our approach to model building was as follows: Model 1 examined trajectories as a function of time, Model 2 added in the effects of sex and age, and Model 3 added interaction terms to the model to test whether trajectories differed between groups. Log-likelihood ratio tests were used to compare each subsequent growth curve model to the previous one. Final model statistics are presented in Supplemental Appendix B (Table B3) and we outline the key findings for each dependent variable below.

a Narrative comprehension. Analysis indicated a significant effect of time, with comprehension scores increasing by 0.87 points on average at each subsequent time point (95% CIs: 0.73, 1.01). Furthermore, comprehension performance differed significantly based on both sex and ESL status, with females scoring 0.54 points higher (out of 10) on average than males (95% CIs: 0.22, 0.86) and ESL speakers scoring on average 0.62 points lower than English as a first language speakers (95% CIs: -1.02, -0.23).

A log-likelihood ratio test indicated that model fit was significantly improved by adding the sex by time interaction term [$\chi^2(1) = 4.60$, $p < 0.03$] but not when including the non-significant ESL by time interaction. In other words, growth trajectories did not differ based on whether children were ESL learners, but did differ based on sex. Females showed more growth in narrative comprehension

over time than males, with the difference between males and females increasing by 0.31 points at each subsequent assessment point (95% CIs: 0.03, 0.59). Figure 4 shows the mean narrative comprehension scores of males and females at each assessment point.

b Oral narrative production composite. There was a significant effect of time on the oral narrative production composite, with scores increasing by 0.06 points (6%) on average at each subsequent time point (95% CIs: 0.05, 0.07). Across assessment points, oral narrative production scores differed significantly based on sex but not ESL status, with females scoring 0.05 points (5%) higher on average than males [95% CIs: 0.02, 0.07]. The investigation of interaction effects indicated no significant differences in growth trajectories based on sex or ESL status.

c Oral narrative error composite. Error composite scores differed significantly by time, with error scores decreasing by 0.014 points (23% of the mean error rate) on average at each subsequent time point (95% CIs: -0.02 , -0.008). Across assessment points, error scores differed significantly based on ESL status but not sex, with ESL children showing more word errors at 0.025 points higher (increase of 2.5% for words with errors) on average than English as a first language speakers (95% CIs: 0.009, 0.04). The investigation of interaction effects indicated no significant differences in growth trajectories (i.e., reduction in errors) based on sex or ESL status.

4 Comprehension of factual and inferred content

To determine whether children's comprehension differed based on the type of information required by the question, we classified the comprehension into questions that were factual (questions 1, 4, and 5) and those that were inferential (questions 2 and 3; comprehension questions are provided in Supplemental Appendix A). Mean scores on factual and inferential questions over time by sex are presented in Supplemental Appendix B, Table B4. Figure 5 shows that children consistently scored higher on factual questions than inferential questions over time.

We next examined whether growth over time differed by sex. At baseline, there were no significant differences between males and females on factual or inferential questions.

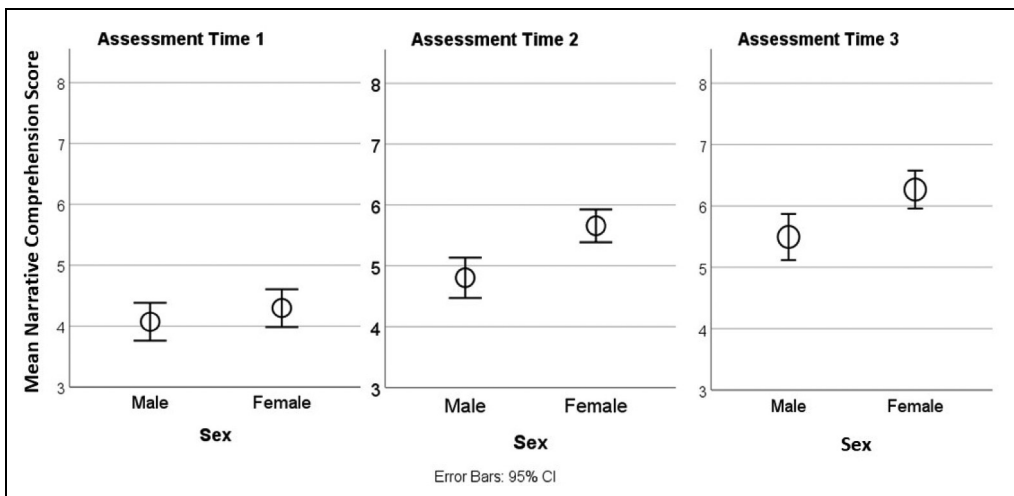


Figure 4. Oral narrative comprehension of males and females across the three assessment points.

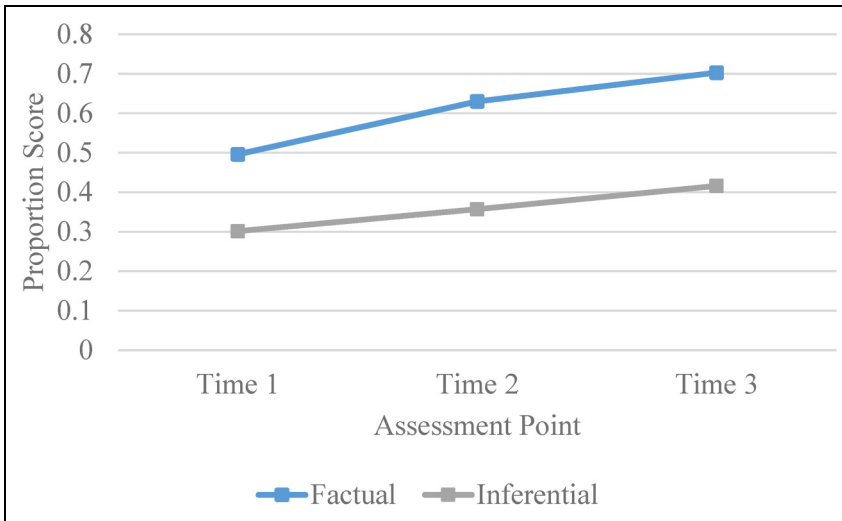


Figure 5. Factual and inferential scores over time.

When examining change over time, males and females showed a similar growth pattern for factual questions [$F(1171)=0.16$, $p=0.85$]. However, females showed significantly greater growth over time on inferential questions, $F(1.9,327.6)=2.94$, $p=0.05$. The mean scores at Time 2 and Time 3 differed significantly between males and females (see Figure 6).

IV Discussion

This study investigated the validity of data gathered from a novel, digitally presented oral narrative (story retell) and narrative comprehension assessment. The task was specifically designed for teachers' use in monitoring the oral language development of children in their first year at school as part of the Better Start Literacy Approach (BSLA) (Gillon et al., 2022). The teachers of new entrant and year 1 classes in New Zealand involved in this study received relevant professional learning prior to task implementation and were supported by SLTs or literacy specialists. The presentation of the task via iPad using a recorded voice with accompanying pictures provided consistency of presentation for all children who completed the task. Both factual and inferential knowledge were assessed through five comprehension questions the teachers asked the children after their attempt to retell the story.

Data were analyzed from teachers' implementation of the BSLA story retell task to 303 5-year-old children at school entry. Follow-up assessments using the same task were administered following 20 weeks of teaching and again 12 months after the initial assessment. Analyses revealed that a short story retell, followed by comprehension questions that took approximately 6 min for teachers to administer to a child using an iPad presentation, is potentially a useful task to monitor aspects of children's oral language development. First, the story retell task showed good internal structure with moderate correlations between oral narrative production and comprehension measures. Second, the quantitative measures in the story retell selected for data reporting loaded onto two factors. This suggests that the BSLA story retell task is useful for teachers to gain insight into two distinct aspects of children's oral language. One factor grouped together measures that teachers could use to describe a child's oral language productivity in retelling a story (i.e.

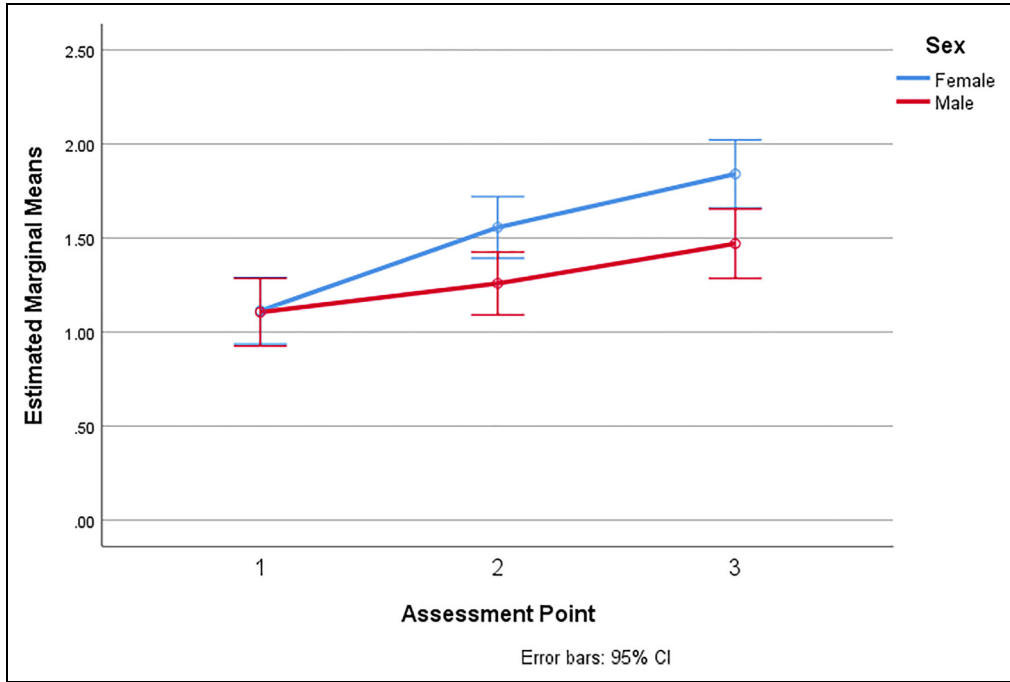


Figure 6. Scores on inferential comprehension questions over time by sex.

measures related to the number, length, and intelligibility of children’s utterances when retelling the story as well as the number and type of different words used). The other factor grouped together the two measures teachers could use to describe children’s grammatical knowledge through their word level errors and omitted bound morphemes.

Criterion task validity was demonstrated through examination of associations between quantitative measures of children’s story retell narratives and a standardized assessment of their oral language development (CELF-P2). Previous research has established that language performance on measures such as the CELF is associated with microstructural aspects of children’s narratives (Murphy et al., 2020). In the present study, the narrative productivity and error composite scores, as well as the children’s comprehension scores, were moderately correlated with children’s performance on the CELF-P2 score (in line with the correlations reported by Murphy et al., 2020), and the task predicted 81% of the children with low levels of oral language as measured by the CELF-P2. This is impressive considering the ease with which teachers can implement the story retell task using a digital presentation as part of their regular class assessments. We anticipate that in addition to informing their teaching practice, teachers will share the data from the child’s BSLA story retell with SLTs or literacy specialists to support children whose speech, oral language or early literacy development they are particularly concerned about. Continued research investigating the relationship between measures of microstructural elements of oral narratives and standardized language assessments is warranted, particularly through longitudinal designs with large samples of children (Murphy et al., 2020).

Predictive task validity in relation to connected text reading and comprehension was demonstrated to some extent. Analysis of the concurrent association between oral narrative and reading performance at the end of the first year of school showed that oral narrative comprehension and

the oral narrative error composite (omitted morphemes and word level errors) were significantly correlated with reading accuracy, reading comprehension and reading rate. The oral narrative production composite, however, was not correlated with reading performance. These results are consistent with previous research showing the connection between oral narrative and reading ability (Snowling and Hulme, 2021). The relatively modest nature of the correlation between variables (r values between 0.17 and 0.35) is likely due to the young age of our study participants. At earlier stages of literacy development, the primary task is learning to decode text efficiently and broader language skills (such as oral narrative) explain relatively less variance in early reading outcomes (Snowling and Hulme, 2021). Similarly, Gardner-Neblett and Iruka (2015) reported that oral narrative skills in 4-year-old children did not mediate the relationship with a literacy composite at 5-years with the exception of African American children. The authors also attributed the relatively weak relationship between oral narrative and reading variables to the young age of their cohort.

The finding regarding the lack of association between the production composite and reading skills, but a significant correlation for the error composite score suggests that particular features of microstructural analysis in oral narrative may be more strongly associated with literacy outcomes. Wellman et al. (2011) examined the utility of oral narrative ability at 3–6 years as a predictor of reading outcomes at 8–12 years in children with typical language, speech difficulties, and children with language difficulties. The microstructural analysis included the number of t -units, number of words and different words, and mean length of t -unit (i.e. similar to the production composite used in the current study). They found the only reading outcome predicted by microstructural analysis was non-word reading. However, an error composite score was not included in their microstructure measures. Our findings suggest that including a measure of children's grammatical knowledge (i.e. error analysis) from the children's story retelling attempts, in addition to measures of productivity, may be useful to support teaching practice.

Analysis of the distribution of narrative comprehension scores revealed that the scores were normally distributed and therefore at a suitable level of difficulty for this age group. Consistent with other studies (Westerveld et al., 2021), comprehension of factual information was easier than understanding inferred meaning for 5-year-old children. These data have practical implications for teaching practice. Teachers frequently ask children questions about a story as part of large group teaching activity in junior classes. If they are aware of how individual children in their class respond to different types of questions, they can more easily scaffold the task to suit a child's learning needs, such as asking a child with a high narrative comprehension score a more complex question that requires inferencing skills and asking a child with a low comprehension score a simple question that requires factual knowledge.

Growth curve modelling demonstrated that this task is useful to monitor children's oral language production and comprehension over the course of a child's first year at school. Significant growth in the children's oral narrative production and comprehension scores at each assessment point was demonstrated, and a significant reduction in the number of word level errors in the children's story retelling attempts over time was evident. Similar growth patterns over time were shown for children for whom English was a second or other language, suggesting that the task is useful for monitoring these children's English oral language development. Interestingly, boys showed less growth in their first year at school than girls in narrative inferential comprehension. This is a novel finding in the research literature. On average, boys entered school with lower oral language skills than girls, but they showed the same growth trajectories as girls for oral narrative production. That is, they were not catching up to girls nor falling further behind after 12 months at school. In contrast, for narrative comprehension, particularly in comprehending inferred meaning, their scores demonstrated a pattern of falling further behind the performance of girls in a 12-month period. Understanding how to mitigate this pattern and to accelerate the oral narrative comprehension

skills of boys in their first year at school warrants further research. Such research may help to change the consistent global finding that boys' reading comprehension of fictional text lags behind girls' performance in many countries around the world (Mullis et al., 2017).

1 Collaborative practices

The importance of collaboration between SLTs and teachers to support the oral language and literacy outcomes of learners is well documented in the literature (e.g. Archibald, 2017). There are, however, multiple barriers to effective inter-professional collaboration including lack of time to develop meaningful inter-professional relationships and limited preparation regarding effective collaboration practices in preservice programs (Pfeiffer et al., 2019). The implementation of the BSLA story retell task trialled in the current approach can act as a facilitator of collaborative practice, particularly in regard to universal (Tier 1) curriculum and supporting children with language difficulties in the classroom context (Tiers 2 and 3). Key aspects from the task implementation model used in the current study that may enable collaborative practice include: the provision of professional learning and development regarding oral language for teachers, administration of the task by teachers which somewhat shifts the balance of power regarding oral language, the inclusion of metrics in the tool that both teachers and SLTs find useful and can guide teaching practice, efficiency of administering and scoring the oral narrative, including the ability for the sound file of the child's story retell to be electronically shared between teachers and SLTs, and developing a shared language around oral language assessment across groups.

The validity of the BSLA story retell task for tracking gains in the English language skills of children with English as a second language has important practical implications. Wood et al. (2018b) also reported that an oral narrative task was sensitive to change in the microstructural language ability of dual language learners ($n = 74$) over 8 months from Kindergarten to Grade 1. The narrative task in the Wood et al., (2018b) study, however, was administered by research assistants and transcribed manually which limits its application to classroom practice. Given the increasing linguistic diversity of learners and the importance of objectively monitoring learners' growth in English (and their other language/s where possible), the tool trialled in the current study will also enhance assessment practices for English as a second language learners.

2 Study limitations

This study is one of the first studies to focus on a range of microstructure elements of children's story retell through the use of digital presentation, semi-automated speech recognition and computerized scoring within regular class teaching contexts. A limitation of this study, however, is that we did not investigate children's oral narrative production at the macrostructure level. Macrostructural elements related to the quality of children's oral narrative and children's knowledge of story structure are also important to consider (Suggate et al., 2018; Wellman et al., 2011), particularly in relation to children's written text production (Zanchi et al., 2020). Continued research into methods that enhance practitioners' scoring reliability for macrostructure elements of children's story retells is necessary (Karusoo-Musumeci et al., 2022). Establishing reliability of both microstructure and macrostructure analyses in ways that are time efficient and easy to undertake for teachers will support the wider spread use of oral narrative assessments from research and controlled clinical contexts to regular classroom use.

The children in this study heard the same story three times over a course of a year. Story familiarity can influence story retelling performance (Boudreau, 2008), but the time frame between the assessment points (a minimum of 20 weeks) and the use of a novel story used only in this

assessment task helped to mitigate the influence of story familiarity. Future studies could explore the use of different, but quantitatively comparable stories to mitigate any influence of story familiarity. This study did not gather data related to the teachers' experiences and perceptions of administering the BSLA story retell task or their experiences in using the data automatically generated for each child within their class teaching practice. This is an area for future research. Children's perceptions and motivation to engage in online oral narrative story retelling tasks is also worthy of investigation.

In considering teacher time constraints to administer the task, children were only asked five narrative comprehension questions following their retelling of the story. This limited the data available regarding performance on factual versus inferred meaning. Given the novel finding relating to gender difference in the growth trajectory of inferential oral narrative comprehension, further experimental research to understand this finding is necessary, including demonstration of the validity of these particular items in assessing inferential abilities.

3 Technological development

The technological development behind the automation of this oral narrative story retell task has been substantial. Scott et al. (2022) presented a detailed description of this task's development relating to software, automatic speech recognition and user interface. Speech models have been developed based on thousands of children retells of the *Tama and the Playground* (Gillon et al., 2019b) story used in this study, which now allow for transcripts to be automatically transcribed for 5- to 6-year old children to ~80% accuracy. The current iteration of the task uses automatic transcription of the language sample, with minimal support from research assistants to check accuracy, complete coding and SALT analysis. The current system in place takes ~7 min to check transcripts and code for SALT analysis. Current development work includes further training of the speech model to increase the accuracy and embed basic language analysis into the tool. The short-term aim is to have transcription accuracy to a suitably high level that teachers can complete the checking of transcripts themselves, removing the need for research assistants.

4 Summary

There is a need for teachers to be able to monitor children's oral language development in their first year at school using valid assessment tasks that are practical to implement within the class context. The novel online oral narrative production and comprehension task examined in this study (the BSLA story retell task) was specifically designed for use by junior school class teachers. The task proved to have good internal structure, be of a suitable level of difficulty for 5-year-old children and have good criterion validity. The task proved useful to monitor children's development over time, including for children who are English language learners. Analyses highlighted the need to focus further attention on the narrative comprehension development of boys. The task holds much promise to support teacher judgment and observation of children's oral narrative and comprehension development in their first year at school.

Acknowledgements

We are very grateful to the children, families and teachers who participated in this project. We also acknowledge the work of our senior research assistants, Dr Amanda Denston and Dr Jo Walker who supported our teachers in their implementation of our assessment task. We acknowledge the work of Global Office Limited, Christchurch, NZ in designing the online capabilities and speech automation aspects of the assessment task.






Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Funding was provided by Ministry of Business Innovation and Employment New Zealand (grant number 15-02688) and the Ministry of Education Foundational Learning Grant. This project received approval from the University of Canterbury Human Ethics Committee Ref: 2019/23/ERHEC.

ORCID iDs

Gail Gillon  <https://orcid.org/0000-0001-9501-8147>
Brigid McNeill  <https://orcid.org/0000-0002-7298-880X>
Amy Scott  <https://orcid.org/0000-0002-6737-9815>
Megan Gath  <https://orcid.org/0000-0002-8647-1262>
Marleen Westerveld  <https://orcid.org/0000-0002-5194-2335>

Supplemental material

Supplemental material for this article is available online.

Note

- 1 At the time of the study implementation, not all of these automated reporting features were available and teachers in this study were emailed back the language transcripts and analyses rather than the current format which enables teachers to access their children's transcripts directly from an assessment website.

References

- Antoniazzi D, Snow P, and Dickson-Swift V (2010) Teacher identification of children at risk for language impairment in the first year of school. *International Journal of Speech Language Pathology* 12(3): 244–252.
- Archibald LMD (2017) SLP-educator classroom collaboration: a review to inform reason-based practice. *Autism and Developmental Language Impairments* 2: 1–17.
- Black C, DeFelice M, and Plant R (2020) Clinical evaluation of language fundamentals: preschool, second edition (CELF preschool-2). In: Volkmar FR (ed) *Encyclopedia of Autism Spectrum Disorders*. Springer New York, pp. 1–3. https://doi.org/10.1007/978-1-4614-6435-8_102527-1.
- Boudreau D (2008) Narrative abilities: advances in research and implications for clinical practice. *Topics in Language Disorders* 28(2): 99–114.
- Bowles RP, Justice LM, Khan KS, et al. (2020) Development of the narrative assessment protocol-2: a tool for examining young children's narrative skill. *Language, Speech & Hearing Services in Schools* 51(2): 390–404.
- Cameron TA, Carroll JLD, Taumoepeau M, et al. (2019) How do New Zealand teachers assess children's oral language and literacy skills at school entry? *New Zealand Journal of Educational Studies* 54(1): 69–97.
- Denman D, Speyer R, Munro N, et al. (2017) Psychometric properties of language assessments for children aged 4–12 years: a systematic review. *Frontiers in Psychology* 8: 1515–1515.
- Fey ME, Catts HW, Proctor-Williams K, et al. (2004) Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research* 47(6): 1301–1318.
- Filiatrault-Veilleux P, Bouchard C, Trudeau N, et al. (2016) Comprehension of inferences in a narrative in 3- to 6-year-old children. *Journal of Speech, Language, and Hearing Research* 59(5): 1099–1110.
- Gagarina N, Klop D, Kunnari S, et al. (2012) MAIN—multilingual assessment instrument for narratives. *ZAS Papers in Linguistics* 56: 1–36.

- Gagarina N, Klop D, Tsimpli IM, et al. (2016) Narrative abilities in bilingual children. *Applied Psycholinguistics* 37(1): 11–17.
- Gaffney JS, Ritchie J, Skerrett M, et al. (2021) Imagining possible worlds with young children, families, and teachers: sustaining indigenous languages and family pedagogies. *Kotuitui* 16(2): 234–249.
- Gardner-Neblett N and Iruka IU. (2015) Oral narrative skills: explaining the language-emergent literacy link by race/ethnicity and SES. *Developmental Psychology* 51(7): 889–904.
- Gardner-Neblett N and Sideris J (2018) Different tales: the role of gender in the oral narrative–reading link among African American children. *Child Development* 89(4): 1328–1342.
- Gillam R and Pearson N (2017) Test of narrative language – 2: Pro-Ed.
- Gillon G and Macfarlane AH (2017) A culturally responsive framework for enhancing phonological awareness development in children with speech and language impairment. *Speech, Language and Hearing* 20(3): 163–173.
- Gillon G, McNeill B, Denston A, et al. (2020) Evidence-based class literacy instruction for children with speech and language difficulties. *Topics in Language Disorders* 40(4): 357–374.
- Gillon G, McNeill B, Scott A, et al. (2022) A better start literacy approach: effectiveness of Tier 1 and Tier 2 support within a response to teaching framework. *Reading and Writing*: 1–34. DOI: 10.1007/s11145-022-10303-4.
- Gillon G, McNeill B, Scott A, et al. (2019) A better start to literacy learning: findings from a teacher-implemented intervention in children’s first year at school. *Reading and Writing* 32(8): 1989–2012.
- Gillon G, McNeill B, and Scott A (2019b) Tama and the Playground. University of Canterbury Child Wellbeing Research Institute: Christchurch, NZ.
- Gillon GT (2018) *Phonological Awareness: From Research to Practice* (2nd ed.). New York: The Guilford Press.
- Gough PB and Tunmer WE (1986) Decoding, reading, and reading disability. *Remedial and Special Education* 7(1): 6–10.
- Hedeker D and Gibbons RD (2006) *Longitudinal Data Analysis*. Wiley-Interscience. <https://doi.org/10.1002/0470036486>.
- Heilmann J, Miller JF, Nockerts A, et al. (2010a) Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology* 19(2): 154–166.
- Heilmann J, Nockerts A, and Miller JF (2010b) Language sampling: does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools* 41(4): 393–404.
- Hoover WA and Tunmer WE (2018) The simple view of reading: three assessments of its adequacy. *Remedial and Special Education* 39(5): 304–312.
- Huang BH, Bedore LM, Ramírez R, et al. (2022) Contributions of oral narrative skills to English reading in Spanish–English Latino/a dual language learners. *Journal of Speech, Language, and Hearing Research* 65(2): 653–671.
- Hughes DL, McGillivray L, and Schmidek M (1997) *Guide to Narrative Language: Procedures for Assessment*. Eau Claire, WI: Thinking Publications. <https://go.exlibris.link/ZYGMQxQ9>.
- Justice LM, Bowles R, Pence K, et al. (2010) A scalable tool for assessing children’s language abilities within a narrative context: the NAP (narrative assessment protocol). *Early Childhood Research Quarterly* 25(2): 218–234.
- Justice LM, Bowles RP, Kaderavek JN, et al. (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.
- Karusoo-Musumeci A, Pearce WM, and Donaghy M (2022) The effect of workshop training on rater variability in children’s oral narrative assessment. *Child Language Teaching and Therapy* 38(1): 8–21.
- Kemp K and Klee T (1997) Clinical language sampling practices: results of a survey of speech-language pathologists in the United States. *Child Language Teaching and Therapy* 13(2): 161–176.
- Kim Y-SG and Petscher Y (2020) Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening comprehension). *Annals of Dyslexia* 71(2): 218–237.
- McAuliffe M and Triandafyllidou A (eds) (2021) *World Migration Report 2022*. Geneva: International Organization for Migration (IOM).

- Malec A, Peterson SS, and Elshereif H (2017) Assessing young children's oral language: recommendations for classroom practice and policy. *Canadian Journal of Education* 40(3): 362–392.
- Manyak PC, Manyak AM, and Kappus EM (2021) Lessons from a decade of research on multifaceted vocabulary instruction. *The Reading Teacher* 75(1): 27–39.
- McConnell GE and Loeb D (2021) Production of narratives by at-risk American Indian children in the midwest. *Topics in Language Disorders* 41(2): 153–168.
- McKay M (1996) The Neale analysis of reading ability revised - systematically biased? *British Journal of Educational Psychology* 66(2): 259–266.
- McLeod S, Harrison LJ, and Wang C (2019) A longitudinal population study of literacy and numeracy outcomes for children identified with speech, language, and communication needs in early childhood. *Early Childhood Research Quarterly* 47: 507–517.
- Merritt DD and Liles BZ (1989) Narrative analysis: clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders* 54(3): 438–447.
- Metsala JL, Sparks E, David M, et al. (2021) What is the best way to characterise the contributions of oral language to reading comprehension: listening comprehension or individual oral language skills? *Journal of Research in Reading* 44(3): 675–694.
- Miller JF, Gillon GT, and Westerveld MF (2017) Systematic Analysis of Language Transcripts (SALT), New Zealand / Australia Version 18 [Computer Software]. SALT Software LLC.
- Mullis IVS, Martin M, Foy P, et al. (2017) Progress in international reading literacy study, PIRLS 2016 International Results in Reading.
- Murphy KA, Springle AP, Sultani MJ, et al. (2022) Predicting language performance from narrative language samples. *Journal of Speech, Language, and Hearing Research* 65(2): 775–784.
- Neale M (1999). *Neale Analysis of Reading Ability* (3rd ed.). Melbourne: Australian Council for Educational Research.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010) *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC: Authors.
- Nation K, Cocksey J, Taylor JSH, et al. (2010) A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry* 51(9): 1031–1039.
- Oakhill JV and Cain K (2012) The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading* 16(2): 91–121.
- Paris AH and Paris SG (2003) Assessing narrative comprehension in young children. *Reading Research Quarterly* 38(1): 36–76.
- Pesco D and Kay-Raining Bird E (2016) Perspectives on bilingual children's narratives elicited with the multilingual assessment instrument for narratives. *Applied Psycholinguistics* 37(1): 1–9.
- Petersen DB, Spencer TD, Konishi A, et al. (2020) Using parallel, narrative-based measures to examine the relationship between listening and reading comprehension: a pilot study. *Language, Speech, and Hearing Services in Schools* 51(4): 1097–1111.
- Petersen DB and Spencer TD (2016) CUBED. Language Dynamics Group.
- Pfeiffer DL, Pavelko SL, Hahs-Vaughn DL, et al. (2019) A national survey of speech-language pathologists' engagement in interprofessional collaborative practice in schools: identifying predictive factors and barriers to implementation. *Language, Speech, and Hearing Services in Schools* 50(4): 639–655.
- Pinto G, Tarchi C, and Accorti Gamannossi B (2018) Kindergarteners' narrative competence across tasks and time. *The Journal of Genetic Psychology* 179(3): 143–155.
- Portilla AY, Almanza V, Castillo AD, et al. (2021) The development of narrative skills in children: a systematic review. *Revista de Investigación en Logopedia* 11(2): e67607.
- Ramírez-Esparza N, García-Sierra A, and Jiang S (2020) The current standing of bilingualism in today's globalized world: a socio-ecological perspective. *Current Opinion in Psychology* 32: 124–128.
- Ratima MT, Smith JP, Macfarlane AH, et al. (2020) *The Hikairo Schema for Primary: Culturally Responsive Teaching and Learning*. Wellington: NZCER Press. <https://go.exlibris.link/psdCjBDg>.

- Sanrey C, Bressoux P, Lima L, et al. (2021) A new method for studying the halo effect in teachers' judgement and its antecedents: bringing out the role of certainty. *British Journal of Educational Psychology* 91(2): 658–675.
- Schmitterer AMA and Brod G (2021) Which data do elementary school teachers use to determine reading difficulties in their students? *Journal of Learning Disabilities* 54(5): 349–364.
- Scott A, Gillon G, McNeill B, et al. (2022) The evolution of an innovative online task to monitor children's oral narrative development. *Frontiers in Psychology* 13. DOI: 10.3389/fpsyg.2022.903124.
- Smith R, Snow P, Serry T, et al. (2021) The role of background knowledge in reading comprehension: a critical review. *Reading Psychology* 42(3): 214–240.
- Snowling MJ, Duff F, Petrou A, et al. (2011) Identification of children at risk of dyslexia: the validity of teacher judgements using 'Phonic Phases'. *Journal of Research in Reading* 34(2): 157–170.
- Snowling MJ and Hulme C (2021) Annual research review: reading disorders revisited – the critical importance of oral language. *Journal of Child Psychology and Psychiatry* 62(5): 635–653.
- Suggate S, Schaughency E, McAnally H, et al. (2018) From infancy to adolescence: the longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development* 47: 82–95.
- Trabasso T and Nickels M (1992) The development of goal plans of action in the narration of a picture story. *Discourse Processes* 15(3): 249–275.
- Tunmer WE and Chapman JW (2012) Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading* 16(2): 122–140.
- Tunmer WE and Hoover WA (2019) The cognitive foundations of learning to read: a framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties* 24(1): 75–93.
- Voniati L, Armostis S, and Tafiadis D (2021) Language sampling practices: a review for clinicians. *Evidence-Based Communication Assessment and Intervention* 15(1): 24–45.
- Wellman RL, Lewis BA, Freebairn LA, et al. (2011) Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech, and Hearing Services in Schools* 42(4): 561–579.
- Westerveld MF and Claessen M (2014) Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology* 16(3): 242–249.
- Westerveld MF, Filiatrault-Veillex P, and Paynter J (2021) Inferential narrative comprehension ability of young school-age children on the autism spectrum. *Autism & Developmental Language Impairments* 6: 239694152110356.
- Westerveld MF and Gillon GT (2010) Profiling oral narrative ability in young school-aged children. *International Journal of Speech-Language Pathology* 12(3): 178–189.
- Westerveld MF, Gillon GT, and Boyd L (2012) Evaluating the clinical utility of the profile of oral narrative ability for 4-year-old children. *International Journal of Speech-Language Pathology* 14(2): 130–140.
- Westerveld MF, Gillon GT, and Miller JF (2004) Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech Language Pathology* 6(4): 195–208.
- Westerveld MF, Gillon GT, and Moran C (2008) A longitudinal investigation of oral narrative skills in children with mixed reading disability. *International Journal of Speech-Language Pathology* 10(3): 132–145.
- Wiig EH, Secord WA, and Semel E (2006) *Clinical Evaluation of Language Fundamentals-Preschool: Australian Standardised Edition*, 2nd ed. Sydney: Harcourt Assessment.
- Williams C (2006) Teacher judgements of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy* 22(2): 135–154.
- Wood C, Wofford MC, and Schatschneider C (2018a) Relationship between performance on oral narrative retells and vocabulary assessments for Spanish-English speaking children. *Communication Disorders Quarterly* 39(3): 402–414.
- Wood C, Wofford MC, Gabas C, et al. (2018b) English Narrative language growth across the school year: young Spanish–English dual language learners. *Communication Disorders Quarterly* 40(1): 28–39.
- Wright TS and Cervetti GN (2017) A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly* 52(2): 203–226.
- Zanchi P, Zampini L, and Fasolo M (2020) Oral narrative competence and literacy skills. *Early Child Development and Care* 190(14): 2309–2320.