2016

Predicting Second Grade Listening Comprehension Using Prekindergarten Measures

Crystle N. Alonzo

Gloria Yeomans-Maldonado

Kimberly A. Murphy
Old Dominion University, kamurphy@odu.edu

Beau Bevens

Language and Reading Research Consortium (LARRC)

Original Publication Citation

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. Topics in Language Disorders, 36(4), 312-333. doi:10.1097/TLD.0000000000000102

Top Lang Disorders
Vol. 36, No. 4, pp. 312–333

Predicting Second Grade Listening Comprehension Using Prekindergarten Measures

Crystle N. Alonzo, Gloria Yeomans-Maldonado, Kimberly A. Murphy, Beau Bevens, and Language and Reading Research Consortium (LARRC)

Purpose: The purpose of this study was to determine prekindergarten predictors of listening comprehension in second grade. **Methods:** Within a large, 5-year longitudinal study, children progressing from prekindergarten to second grade were administered a comprehensive set of prekindergarten measures of foundational language skills (vocabulary and grammar), higher-level

Author Affiliations: *The School of Health and Rehabilitation Sciences, MGH Institute of Health Professions, Boston, Massachusetts (Ms Alonzo and Mr Bevens); Crane Center for Early Childhood Research and Policy, College of Education and Human Ecology, The Ohio State University, Columbus, Ohio (Ms Yeomans-Maldonado); and Communication Disorders and Special Education, Old Dominion University, Norfolk, Virginia (Dr Murphy).*

This article was prepared by a Task Force of the Language and Reading Research Consortium (LARRC) consisting of Tiffany Hogan (Convener), Crystle Alonzo, Gloria Yeomans-Maldonado, Kimberly Murphy, Beau Bevens, Kate Cain, and Hugh Catts. LARRC project sites and investigators are as follows:

Ohio State University (Columbus, OH): Laura M. Justice (Site PI), Richard Lomax, Ann O'Connell, Jill Pentimonti[1], Stephen A. Petrill[2], Shayne B. Piasta.

Arizona State University (Tempe, AZ): Shelley Gray (Site PI), Maria Adelaida Restrepo.

Lancaster University (Lancaster, UK): Kate Cain (Site PI).

University of Kansas (Lawrence, KS): Hugh Catts[3] (Site PI), Mindy Bridges, Diane Nielsen.

University of Nebraska-Lincoln (Lincoln, NE): Tiffany Hogan (Site PI), Jim Bovaird, J. Ron Nelson.[4]

MGH Institute of Health Professions (Boston, MA): Tiffany Hogan (Site PI).

[1]Jill Pentimonti is now at American Institutes for Research.

[2]Stephen A. Petrill was a LARRC coinvestigator from 2010 to 2013.

[3]Hugh Catts is now at Florida State University.

[4]J. Ron Nelson was a LARRC coinvestigator from 2010 to 2012.

This work was supported by grant # R305F100002 of the Institute of Education Sciences' Reading for Understanding Initiative. We are deeply grateful to the numerous staff, research associates, school administrators, teachers, children, and families who participated. Key personnel at study sites include Lisa Baldwin-Skinner, Lauren Barnes, Garey Berry, Jennifer Bostic, Shara Brinkley, Janet Capps, Tracy Centanni, Beth Chandler, Lori Chleborad, Willa Cree, Dawn Davis, Kelsey Dickerhoof, Jaclyn Dynia, Michel Eltschinger, Kelly Farquharson, Yvonne Fraser, Abraham Aldaco Gastelum, Tamarine Foreman, Rashaun Geter, Sara Gilliam, Cindy Honnens, Miki Herman, Hui Jiang, Elaine Joy, Jaime Kubik, Trudy Kuo, Gustavo Lujan, Chi Luu, Junko Maekawa, Carol Mesa, Denise Meyer, Maria Moratto, Marcie Mutters, Amy Pratt, Trevor Rey, Amber Sherman, Shannon Tierney, and Stephanie Williams.

The views presented in this work do not represent those of the federal government, nor do they endorse any products or findings presented herein.

The authors have indicated that they have no financial and no nonfinancial relationships to disclose.

Corresponding Author: *Tiffany P. Hogan, PhD, Professor, Communication Sciences and Disorders, School of Health and Rehabilitation Sciences, MGH Institute of Health Professions, 36 1st Ave, Boston, MA 02129 (thogan@mghihp.edu).*

DOI: 10.1097/TLD.0000000000000102

language skills (inferencing, comprehension monitoring, and text structure knowledge), listening comprehension, working memory, and nonverbal processing, as well as second grade measures of listening comprehension. **Results:** A prekindergarten measure of listening comprehension—the *Test of Narrative Language*—and a prekindergarten measure of foundational language skills and working memory—the *Clinical Evaluation of Language Fundamentals-4* Recalling Sentences—were significant predictors of second grade listening comprehension. **Conclusions:** Our findings show that a quick, reliable measure of sentence imitation and/or listening comprehension, administered in prekindergarten, provides insight into a child's second grade listening comprehension. Knowing who is at risk for comprehension failure will allow educators to make informed, evidence-based decisions on the need for further in-depth assessment and language-intensive instruction to stave off future reading disabilities. **Key words:** *comprehension*, *language*, *listening comprehension*, *prediction*, *prekindergarten*, *sentence imitation*, *working memory*

NUMEROUS studies have focused on the creation of screening instruments for diagnosing word reading difficulties to provide early intervention (e.g., Bridges & Catts, 2011). Successful reading comprehension involves, however, more than just accurate word reading; it also requires proficient listening comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). In fact, a significant number of children who develop adequate word reading skills still have poor reading comprehension because of inadequate listening comprehension (Catts, Adlof, & Weismer, 2006). Despite the importance of listening comprehension for reading comprehension, there have been only a few investigations into the precursors of listening comprehension in young children. The purpose of this study is to determine prekindergarten predictors of listening comprehension in second grade. We aimed to provide educators and clinicians a set of readily available, early screening measures to identify children at risk for comprehension difficulties who would then require further in-depth assessment.

LISTENING COMPREHENSION

Importance of listening comprehension

According to the widely known "Simple View of Reading" model, reading comprehension is the product of accurate word recognition and proficient listening comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). Skilled word recognition is the ability to translate printed text into pronounceable words, whereas listening comprehension is the ability to process, integrate, and understand the meaning of text when it is heard instead of read (Hogan, Adlof, & Alonzo, 2014; Molloy, 1997). Of course, comprehension is no simple task; listening comprehension is a dynamic process that has both a language and cognitive basis. Some key language influences on comprehension are vocabulary, inferencing, and background knowledge (see Cain, Oakhill, Barnes, & Bryant, 2001; Elleman, Lindo, Morphy, & Compton, 2009; Hogan et al., 2014); whereas some key cognitive influences include working memory and attention (see Daneman & Merikle, 1996; Lorch et al., 2000).

Over time, the influence of listening comprehension on children's reading comprehension grows immensely. For example, one large, longitudinal study showed that, from second to eighth grade, listening comprehension became more important to reading comprehension than word reading, such that by eighth grade individual differences in listening comprehension accounted fully for individual differences in reading comprehension (Adlof, Catts, & Little, 2006; also see Language and Reading Research Consortium [LARRC], 2015a). Beyond contributions to reading comprehension, listening comprehension is an important standalone skill for everyday functioning at home and in the classroom, for example, to understand orally presented stories and complex instructions (Hogan, Bridges, Justice, & Cain, 2011).

Predictors of listening comprehension

Although multiple studies have identified early predictors of word recognition, substantially fewer have investigated early predictors of listening comprehension. Educators currently work most often within a Response-To-Intervention (RTI) model (Fuchs, Mock, Morgan, & Young, 2003; Jimerson, Burns, & VanDerHeyden, 2007) to determine who is at risk for reading disability. Within most RTI models, word reading is the focus of early assessment and intervention (Ukrainetz, 2006). Therefore, interventions in the early grades target the precursors to proficient word reading (e.g., letter-sound correspondence, phonemic awareness, and basic word decoding skills). These RTI models often lack assessment and treatment of listening comprehension. The focus on word reading alone may result in underidentification and, thus, limited instruction for children with deficient listening comprehension. A recent study by Allen, Ukrainetz, and Carswell (2012), showed that some children benefited less from word reading instruction within current models of RTI because their primary weakness was in listening comprehension, not word decoding. They found that first graders termed "early responders" within RTI had word reading fluency difficulties that resolved after a few months of additional word reading instruction. Those in the study who did not respond to more word reading instruction were those with deficits in listening comprehension, which was not a focus of remedial intervention. Knowing the early predictors of listening comprehension would provide a more comprehensive assessment of children's risk for future reading disabilities within an RTI framework. Moreover, identifying a child's core weakness—word reading or listening comprehension—would guide targeted intervention to stave off *all* future reading disabilities.

Several studies have found that school age children with poor reading comprehension, despite good word reading, had low language skills as early as kindergarten (Catts et al., 2006; Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Nation, Cocksey, Taylor, & Bishop, 2010). A recent study by Justice, Mashburn, and Petscher (2013) showed that children as young as 15 months old with poor language had later reading comprehension difficulties in fifth grade. Because language skills develop from an early age and are independent of word reading, these findings highlight that language skills likely serve as the foundation for competent listening comprehension. In addition to language skills, listening comprehension likely draws on other early developing skills known to influence language processing, including working memory and vocabulary knowledge.

Surprisingly few studies have investigated predictors of listening comprehension, independent of reading comprehension (e.g., Elwér et al., 2013; Kim, 2015; 2016; Potocki, Ecalle, & Magnan, 2013; Tighe, Spencer, & Schatschneider, 2015). Furthermore, only three of these included children in the prereader stage of reading development (Florit, Roch, Altoe, & Levorato, 2009; Florit, Roch, & Levorato, 2014; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012). Moreover, these studies account for only a limited amount of variance in listening comprehension. In one of the studies of older children, third grade verbal and nonverbal reasoning, word reading fluency, and working memory significantly predicted 33%–42% of the variance in future listening comprehension in seventh and tenth grades (Tighe et al., 2015). In a concurrent prediction study, Florit et al. (2009) showed that working memory measures were predictors of listening comprehension in prekindergarten and kindergarten, which accounted for 6% and 4% of the variance, respectively, whereas vocabulary knowledge explained the most variance (22%) in listening comprehension. Florit et al. (2014) in a follow-up study found that prekindergarten vocabulary and higher-level language measures predicted approximately 50% of the variance in later listening comprehension in kindergarten. In the following sections, we describe in more detail research on early predictors of listening comprehension.

Foundational language predictors

Language skills can be categorized into foundational language skills and higher-level language skills (see Hogan et al., 2011, for a review). Both levels contribute to listening comprehension. Foundational language skills, including vocabulary and grammatical knowledge, are those that develop relatively easily and quickly for most children during the course of early childhood and provide the foundation for higher-level language skills. Higher-level language skills build on foundational skills for constructing mental models of a text's meaning and include many subcomponents such as knowledge of complex grammatical structures, cohesive language and other elements of academic discourse, and sensitivity to pragmatic elements (e.g., author style and intention). Receptive and expressive word knowledge consistently predict individual differences in listening comprehension (Florit et al., 2009; 2014; Kim, 2015; Lepola et al., 2012; Tighe et al., 2015). Furthermore, researchers have found that children with specific language impairment, who have deficits in vocabulary, also tend to have poorer listening comprehension skills (Kelso, Fletcher, & Lee, 2007; Vandewalle, Boets, Boons, Ghesquière, & Zink, 2012).

Few studies have examined the predictive power of grammatical knowledge on listening comprehension, and those that have demonstrate little consensus. Florit, Roch, and Levorato (2013) found that grammatical knowledge did not play a specific role in listening comprehension, whereas Kim (2015, 2016) found that grammatical knowledge both directly and indirectly, via comprehension monitoring, predicted listening comprehension. Despite equivocal findings, it is reasonable to consider grammatical knowledge in the discussion of possible predictors of listening comprehension because listening comprehension involves understanding and integrating words and phrases with important information and relations indicated through grammatical conventions. Of note, recent studies have found that sentence recall, which is arguably a measure of grammatical knowledge, is a strong predictor of reading comprehension (Adlof, Catts, & Lee 2010; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015).

Higher-level language predictors

Higher-level language skills are those that integrate words, phrases, and sentences to build a mental model of a text and its meaning and include the ability to draw inferences, monitor comprehension, and identify text structures (Oakhill & Cain, 2012). Higher-level language skills predict individual differences in listening comprehension in 4- to 6-year-olds after controlling for foundational language skills (Florit et al., 2014; Kim, 2015; 2016; Lepola et al., 2012). The most commonly identified higher-level language contributor to listening comprehension is inferencing ability both directly (Kim, 2016; Lepola, et al., 2012) and indirectly (Florit et al., 2014). Inferencing is one's ability to fill in the gaps in text and move past the literal meaning of words to create a comprehensive mental model (Cain & Oakhill, 2014).

Comprehension monitoring and text structure knowledge are two additional higher-level language skills. Comprehension monitoring requires significant cognitive resources, such as working memory and attention, to reflect on one's background knowledge to detect inconsistencies or violations within a text (Oakhill & Cain, 2012). Through experience with spoken narratives, children as early as 30 to 39 months develop monitoring skills prior to learning to read words (Skarakis-Doyle & Dempsey, 2008). Two studies found that comprehension monitoring directly (Kim, 2015) and indirectly (Kim, 2016) predicted listening comprehension in kindergarten children in South Korea. Text structure knowledge is the ability to recognize relationships within and between texts to aid comprehension (Oakhill & Cain, 2012). Both comprehension monitoring and text structure knowledge are weak in children with poor comprehension (Cain, 1996; Oakhill, Hartt, & Samols, 2005; Yuill

& Oakhill, 1991), which provides further evidence that both could reasonably serve as early predictors of later individual differences in listening comprehension.

One could argue that a measure of listening comprehension is itself a measure of higher-level language. Indeed, some measures of higher-level language skills are similar in format to measures of listening comprehension. For example, inferencing tasks commonly require one to read short passages and answer questions that assess inferencing skills. Similarly, listening comprehension measures require one to listen to passages, of varied lengths, and answer questions, some of which may assess inferencing. Because listening comprehension can be measured reliably in prereaders (Leslie & Caldwell, 2010), it is reasonable to hypothesize that early listening comprehension measures may serve as significant predictors of later listening comprehension measures.

Working memory predictors

Working memory is the mental workspace where we simultaneously store and manipulate incoming information (Baddeley, 1986). During a listening task, including listening to written text, we hold information as we adapt our mental model to integrate new information. As such, working memory is a potential predictor of listening comprehension because of the demands placed on memory resources during listening comprehension tasks (Florit et al., 2009; Tighe et al., 2015). As noted earlier, Florit et al. (2009) found that working memory predicted listening comprehension in 4- to 6-year-olds even after accounting for individual differences in verbal language abilities. In addition, they found that the predictive power of working memory on listening comprehension was stable from prekindergarten to kindergarten. Further evidence of the relationship between working memory and listening comprehension is supported by that fact that children and adults with poor listening comprehension consistently have low working memory abilities (McInnes, Humphries, Hogg-Johnson, & Tannock, 2003).

Other potential predictors

It is reasonable to consider other potential predictors of listening comprehension aside from language and working memory. For example, nonverbal intelligence and mother's education have been associated with early reading comprehension (Catts, Fey, Zhang, & Tomblin, 2001). These metrics may quantify individual differences in processing speed and literacy experience, respectively, both of which could affect listening comprehension. Finally, chronological age, which may index literacy experience and overall mental maturity, is likely to be an additional predictor of listening comprehension.

The present study

In the present study, we add to the extant literature by using data from a 5-year longitudinal study to predict second grade listening comprehension from a broad set of prekindergarten predictors, including multiple measures of foundational language (vocabulary and grammar), higher-level language (inferencing, comprehension monitoring, and text structure knowledge), working memory, and nonverbal processing, as well as pertinent demographic factors such as age and mother's education. We intentionally gave preference to those measures that are available to educators and clinicians while also excluding word-reading-based measures, such as letter identification, because our focus was on predicting listening comprehension. Our goal is to provide a clinically useful compilation of prekindergarten measures that are readily available for educators to predict children's future listening comprehension abilities for use in RTI frameworks. Based on past studies, we hypothesized that we would be able to explain a substantial amount of variance in second grade listening comprehension because of our comprehensive set of potential prekindergarten predictors. We hypothesized that a foundational language measure such as vocabulary knowledge would emerge as one of our best predictors of future listening comprehension based on its prevalence as a top predictor in past studies (Florit et al., 2014; Kim, 2015). We also hypothesized that an early measure of listening comprehension would best predict future listening comprehension because commonly the best predictor of one's current ability in a domain is one's prior ability in that domain. Furthermore, listening comprehension can be measured reliably in young children and likely draws on the same set of skills required for future listening comprehension (Florit et al., 2013). Finally, we considered the possibility that a measure that requires multiple component skills, such as sentence imitation, which is arguably reliant on both foundational language skills and working memory (Catts, Nielsen, Bridges, & Liu, 2014; Kidd, 2013), would be a good predictor of future listening comprehension; indeed, it is consistently a strong predictor of reading comprehension (e.g., Adlof et al., 2010; Alloway & Gathercole, 2005; Badian, 1982; Hulme et al., 2015; Scarborough, 1998).

METHODS

Participants

The participants were children whose parents gave consent for them to take part in a 5-year longitudinal study conducted by LARRC, which enrolled 420 prekindergarten children in year 1 at four university sites (Arizona State University, the University of Kansas, the Ohio State University, and the University of Nebraska-Lincoln). The purpose of LARRC was to examine reading and listening comprehension in children prekindergarten to third grade. The current sample included all participants who began in the prekindergarten cohort during the initial year of the LARRC's study. Children were recruited through information packets, containing flyers and consent forms that were sent home by their classroom teachers. Participants were tested from January to May of each academic year from prekindergarten to third grade. By year 4 of the study (second grade), 328 of the original 420 prekindergarten children remained in the study. For the present study, we used data on 318 second grade children who had complete data on the three second grade listening comprehension measures that represented our outcome of interest.

Table 1 shows the demographic information for child participants in prekindergarten during year 1. Overall, our sample was predominantly White and non-Hispanic. Median family income was \$60,000 to \$85,000 per year, and median mother's education level was the attainment of 2- to 4-year degrees.

Table 1. Selected baseline child characteristics

Characteristic	Prekindergarten
N	318
Age in months, *M* (*SD*)	61 (3.81)
Female (%)	43.08
Individualized Education Program (%)	13.52
White (%)	93.08
Hispanic (%)	8.18
English home language (%)	91.19
No response (%)	6.92
Free/reduced price lunch (%)	10.69
No response (%)	7.55
Family income (%)	
≤\$30,000	9.43
\$30,001–\$60,000	21.38
\$60,001–\$85,000	22.33
≥\$85,001	38.36
No response	8.49
Mother's highest level of education (%)	
No high school diploma	.63
High school but no college	9.43
Some college no degree	19.5
2- or 4-year degree	38.68
Graduate degree	24.21
No response	7.55

Note. For age, standard deviation is in parentheses.

Procedures

Every school year during the 5-year study, measures were administered in several sessions within the 5-month testing period (January through May). Assessments were grouped into blocks to make each testing session a reasonable length ($\leq$60 min). The full battery of assessments included measures of listening and reading comprehension, language, memory, and word recognition, which required a total of 5–6 hours to administer. All assessments were administered by trained research staff in the child's school, local university site, community center, or home. Assessors underwent comprehensive training, which included the completion of online training modules (with quizzes) and in-laboratory fidelity checks by trained supervisors to ensure reliable measurement administration and fidelity across sites.

Measures

The measures presented in this study were taken from the larger test battery described earlier. Raw scores (total correct) were used in analyses except where indicated. For published measures, standard test procedures were followed (including basal and ceiling rules) except for the *Clinical Evaluation of Language Fundamentals-4* (CELF-4; Semel, Wiig, & Secord, 2003). The CELF-4 was developed and normed for children ages 5 and up, but in the present study, we administered it to our prekindergarten sample because of the longitudinal nature of our study. To adequately measure development over time, although some tests were not normed for prekindergarten children, some subtests or test procedures were modified to better suit the prekindergarten children. In these cases, we describe our modifications next.

Listening comprehension measures

Three measures of listening comprehension were administered. The first was a modified version of the Understanding Spoken Paragraphs (USP) subtest of the CELF-4. This task assesses children's ability to comprehend narratives of increasing length and complexity (e.g., syntactic and lexical). Modifications included administering two paragraphs instead of three, and preparation of new paragraphs for prekindergarten children. These paragraphs were written by experienced research staff, and were based on lexile and readability measures to ensure that they were suitable for prekindergarten children. Children listened to both paragraphs and then answered five questions pertaining to each. Questions tapped skills such as accurate memory of information presented, general knowledge relevant to the story, and inferencing ability. Children's responses were recorded and later postscored in the research laboratory. Interrater reliability, estimated using the intraclass correlation coefficient (ICC), was .99. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .71$).

The second measure of listening comprehension was the receptive portion of the *Test of Narrative Language* (TNL; Gillam & Pearson, 2004), which assessed children's ability to comprehend narratives. Children listened to three stories read by the examiner and then answered 40 open-ended questions about the stories. A slight modification was made from the standardized procedures: in the LARRC TNL protocol, children were asked to retell the third story before they answered comprehension questions. Children's responses were recorded and postscored in the research laboratory (ICC = .97). Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .87$).

Our final measure was the Listening Comprehension Measure (LCM), an experimenter-designed measure that was adapted in part from the *Qualitative Reading Inventory-5* (Leslie & Caldwell, 2010). For the prekindergarten measure, children listened to one narrative and two expository passages read by the examiner while looking at picture supports and then answered 15 open-ended questions total with no pictures present. For the second grade measure, children listened to two narrative and two expository passages read by the examiner, with no pictures present and then answered 29 open-ended questions total. The questions required recall of information that was provided either explicitly or implicitly in the passages. Children's responses were recorded and postscored in the research laboratory (ICC = .96). Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .78$).

Foundational language measures

Two measures of vocabulary, one receptive and one expressive, were administered. The *Peabody Picture Vocabulary Test-4* (PPVT-4; Dunn & Dunn, 2007), Form A, was administered to assess the breadth of children's receptive vocabulary. Children selected one of four pictures that corresponded to the target word spoken by the examiner. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .96$). The *Expressive Vocabulary Test-2* (EVT-2; Williams, 2007), Form A, was administered to assess breadth of expressive vocabulary. Children were shown a picture and asked to provide either a label or a synonym. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .94$).

Five measures of grammatical knowledge were administered. The first two were subtests of the CELF-4. The first was the Word Classes (WC) subtest, which measures children's ability to understand and express relationships between words that are related by semantic class features. For the receptive portion of the test, the examiner showed a set of three to four pictures to the children, named each picture, and asked the children which two words go together best. For the expressive portion, children were then asked to explain how the two words go together. These responses were scored offline; thus, they were audio recorded and postscored by trained personnel in the research laboratory (ICC = .99). The sum of the receptive and expressive raw scores was used in analyses. Internal consistency reliability coefficients for our sample were all acceptable: $\alpha = .88$ for receptive, $\alpha = .85$ for expressive, and $\alpha = .92$ for the combined score.

The second subtest of the CELF-4 was Word Structure (WS), which assessed children's ability to apply word structure rules to mark inflections, derivations, and comparisons, and to select and use appropriate pronouns to refer to people, objects, and possessive relationships. Children listened to a model sentence and then produced a similar sentence using appropriate inflectional morphology. To adapt this measure for prekindergarteners, we implemented a discontinue rule of eight incorrect responses. For the most part, this measure was scored onsite; only questionable items were postscored offsite with 100% agreement among raters. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .83$).

The final two administered measures of grammar were two probes from the *Rice/Wexler Test of Early Grammatical Impairment* (TEGI; Rice & Wexler, 2001). The Past Tense (TEGI-T) probe assessed children's production of regular and irregular past tense verbs. The Third Person Singular probe (TEGI-S) assessed children's use of the third person singular morpheme (/s/ or /z/) in a picture elicitation task. Internal consistency reliability coefficients for our sample were acceptable for both the Past Tense probe ($\alpha = .86$) and the Third Person Singular probe ($\alpha = .85$).

The *Test for Reception of Grammar-2* (TROG-2; Bishop, 2003) was the final measure of grammar and assessed children's comprehension of grammar. Each of the 20 grammatical contrasts, marked by inflections, function words, and word order, was assessed in a block of four items. One point was awarded for a block if all four items were correct. The score used in analyses was the number of blocks correct. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .84$).

Higher-level language measures

Three experimenter-designed measures of higher-level language were administered. The first, the Inference Task (IT; based on the work of Cain & Oakhill, 1999; Oakhill & Cain 2012), assessed children's ability to generate inferences from narrative texts that were read to them. Children listened to two stories and were asked eight questions about each story. This measure was recorded and postscored in the research laboratory (ICC = .86). Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .78$).

The second, Comprehension Monitoring–Knowledge Violations Task (KVT; based on Baker, 1984), assessed children's ability to monitor their comprehension of short stories, some of which included inconsistent information. Five out of the seven test stories included inconsistent information. Children listened to each story and were asked whether it made sense. If they replied that it did not make sense (indicating comprehension of the inconsistency), they were asked to tell what was wrong with the story. Children received one point for each inconsistent story for which they correctly identified the inconsistency (possible range: 0–5). Internal consistency reliability coefficient for our sample was just below an acceptable standard ($\alpha = .69$).

Our final higher-level language measure was the Picture Arrangement Task (PAT), adapted from the Picture Arrangement Test of the *Wechsler Intelligence Scale for Children-3* (WISC-III UK edition; Wechsler, 1992), which assessed children's knowledge of narrative text structure, specifically their ability to sequence a series of picture cards into a causally and temporally coherent story. The total number of correct stories was tallied for the raw score, which was used in analyses (possible range: 0–12). Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .85$).

Working memory measures

Four measures of working memory were administered. Memory Updating (MU) is a researcher-designed task based on the work of Belacchi, Carretti, and Cornoldi (2010) that assessed children's ability to regulate and modify the contents of working memory, using comparison of objects. Words were limited to one to two syllables. They were also controlled for overlapping initial sounds and word frequency within each trial (e.g., "rabbit" and "rug" could not occur in the same trial and the average word frequency for each trial was three; Storkel & Hoover, 2010). During test administration, children listened to a series of words presented by the examiner and were asked to tell which one/two/three/four/five things were the smallest in lists that varied from 2 to 12 words. For example, one test item asked children to recall the four smallest things out of a list of 10 words. Testing was discontinued if children incorrectly recalled words in both items within a level. The raw score used in analyses was the total number of words answered correctly across all levels administered. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .74$).

The Auditory Working Memory (AM) subtest of the *Woodcock Johnson III Tests of Cognitive Abilities: Normative Update* (WJ-III; Woodcock, McGrew, & Mather, 2001) assessed children's ability to temporarily store and recode orally presented information. Children listened to an audio-recorded series of words and digits, ranging in length from two to eight items total. They were then asked to repeat the series of words in sequential order, followed by the series of digits in sequential order. Two points were awarded if both the words and digits were correctly repeated, and one point was awarded if either the words or digits were correctly repeated. Testing was discontinued after three consecutive incorrect items within a set. The internal consistency reliability coefficient for our sample was acceptable ($\alpha = .78$).

Another experimental measure, the Non-Word Repetition Task (NWRT), assessed children's phonological short-term memory; specifically, the ability to repeat 16 nonwords that varied in two characteristics known to influence performance difficulty: (a) length (two, three, four, or five syllables) and (b) phonological complexity (high or low phonotactic probability). For analyses, the average percentage of consonants correct for all words was used. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .89$).

A modified version of the Recalling Sentences (RS) subtest of the CELF-4 assessed children's ability to repeat sentences of increasing length and complexity. To make this subtest more appropriate for prekindergarten children, the first two items from the Recalling Sentences subtest of the *CELF-Preschool-2* (Wiig, Secord, & Semel, 2004) were inserted as the first two items on this version of the test. Children's responses were recorded and later postscored in the research laboratory (ICC = .99). Sentences received a score of 0 to 3, depending on the number of words incorrectly recalled (0 = ≥4 errors; 1 = 2 or 3 errors; 2 = 1 error; 3 = no errors). Testing was discontinued after five consecutive scores of 0. The internal consistency reliability coefficient for our sample was acceptable ($\alpha = .92$). Of note, this Recalling Sentences task has been purported to assess both grammatical knowledge and working memory (Catts et al., 2014; Kidd, 2013), which is a point of consideration in our analytical model.

Nonverbal intelligence measures

To measure nonverbal intelligence, we administered the *Kaufman Brief Intelligence Test-2* (KBIT-2; Kaufman & Kaufman, 2004) nonverbal matrices. The KBIT-2 nonverbal matrices measure a child's ability to solve problems by assessing the child's ability to perceive relationships and complete visual analogies using pictures or abstract designs instead of words. Internal consistency reliability coefficient for our sample was acceptable ($\alpha = .79$).

Analytical strategy

In the present study, we used a data-driven approach to select prekindergarten measures prior to determining whether the selected measures predict listening comprehension in second grade. This allowed us to run a more parsimonious analysis of the potential predictors by reducing our measures to only those that had the highest loadings on our three constructs of interest: foundational language, higher-level language, and working memory. We also gave preference to measures that were readily available to educators and clinicians because our primary aim was to impact clinical practice. This two-step strategy is described in detail here.

Step 1

In Step 1, we used a confirmatory factor analysis to select the highest loading indicators of each of our three prekindergarten constructs: foundational language, higher-level language, and working memory. We used a cut-point of .70 or higher to select prekindergarten indicators as potential predictors of second grade listening comprehension because this criterion has been recommended as a cut-off point for obtaining a reliable factor/construct (Comrey & Lee, 1992). Although we recognize that this cut-point is arbitrary, we chose it to guide us in the preliminary step needed to reduce the number of predictors. Figure 1 includes a graphical representation of our three-factor model. The confirmatory factor analysis was conducted with MPlus 6.12 software, with a robust maximum likelihood estimator (Muthén & Muthén, 2011).

Note that because our three constructs—foundational language, higher-level language, and working memory—have been found to be highly correlated at early stages of language development (LARRC, 2015b), these three factors were allowed to correlate in our model. Also, the residuals of the PPVT-4 and EVT-2 as well as the two probes of the TEGI (TEGI-S and TEGI-T) were allowed to correlate, respectively. The CELF-4 Recalling Sentences measure was allowed to be an indicator of the foundational language factor and the working memory factor because this measure involves both processes (Catts et al., 2014; Kidd, 2013). Model fit comparisons provided evidence that when CELF-4 Recalling Sentences was part of the working memory factor, there was a slightly better model fit than when it was part of the foundational language factor. The loading for CELF-4

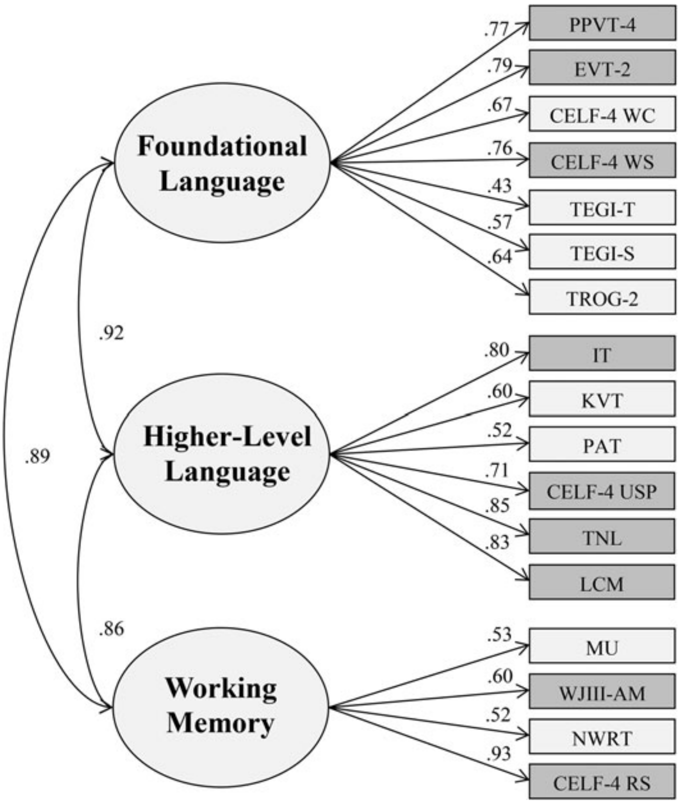

Figure 1. Confirmatory factor analysis of the prekindergarten measures used to select predictors of listening comprehension in second grade. CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; CELF-4 USP = Clinical Evaluation of Language Fundamentals-4—Understanding Spoken Paragraphs; CELF-4 WC = Clinical Evaluation of Language Fundamentals-4—Word Classes; CELF-4 WS = Clinical Evaluation of Language Fundamentals-4—Word Structure; EVT-2 = Expressive Vocabulary Test-2; KVT = Knowledge Violation Task; LCM = Listening Comprehension Measure; MU = Memory Updating Task; NWRT = Non-Word Repetition Task; PAT = Picture Arrangement Task; PPVT-4 = Peabody Picture Vocabulary Test-4; TEGI-S = Rice/Wexler Test of Early Grammatical Impairment-Past Tense; TEGI-T = Rice/Wexler Test of Early Grammatical Impairment-Third Person Singular; TNL = Test of Narrative Language; TROG-2 = Test for Reception of Grammar-2; IT = Inference Task; WJ-III AM = Woodcock Johnson III Tests of Cognitive Abilities: Normative Update—Auditory Working Memory.

Recalling Sentences was higher than .70 on both the foundational language and working memory constructs, and thus, for the purpose of the variable selection process, this measure would have been included as part of Step 2 regardless of which factor it loaded on. Next, we review all indicators that met the .70 criterion for each factor. An exception to this rule was made for WJ-III Auditory Working Memory, whose standardized loading was .60. We decided to include this measure in the model because it is a commonly used and accepted measure of working memory in clinical settings and was the highest loading traditional measure of working memory on our working memory factor.

Step 2

In the second step, a linear multiple regression was employed using our prekindergarten measures (determined through confirmatory factor analysis—see Step 1) to predict listening comprehension skills in second grade. Our dependent variable was defined as the second grade listening comprehension factor score extracted from a latent variable using the CELF-4 Understanding Spoken Paragraphs, the Test of Narrative Language, and the Listening Comprehension Measure. In addition, all models included mother's education, age in months, and a measure of nonverbal intelligence.

The linear multiple regression was followed by relative importance analyses of the measures with the objective of understanding the unique variance contribution of each of the components in our models, as described in Grömping (2006), "relative importance refers to the quantification of an individual regressor's contribution to a multiple regression model" (p. 1). Note that if we sum the proportionate contribution of the variance for each of the predictors used in each of the models, we arrive at the total R^2 for the model. As such, we can interpret this decomposition in R^2 as the relative importance that each of our predictors has on the overall model. The R package relaimpo (Grömping, 2006) was used to calculate the relative importance of each of the coefficients.

Missing Data

Different levels of missing data were present in our sample. Complete case analyses would have considerably reduced our sample size; thus, multiple imputation was used to conduct all subsequent analyses. For the confirmatory factor analysis described in Step 1, the estimation method took care of any missing data; however, the methods used in Step 2 do not rely on maximum likelihood estimation, requiring a different strategy to handle missing data.

The percentage of missingness for all prekindergarten measures used in the analyses ranged from 1% to 23% (M = 5%, SD = 0.06%). Because of assessor error, the measure with the highest percentage of missingness was the Inference Task (23%). Without the Inference Task influencing the missing percentage, the mean percentage of missingness was 4% (SD = 0.03%, minimum = 1%, maximum = 8%).

According to Little's MCAR test (Little, 1988), data were missing completely at random χ^2 (2261) = 2238.02, p = .630. Thus, SAS 9.4 (SAS Institute, 2014) was used to impute a total of 10 data sets using all measures listed in the present study as part of the imputation model. In addition, because we had nested data but were not interested in contextual effects, we applied an inflation factor to the standard errors in the multiple regression stage to protect against Type I error (see Moulton, 1986, p. 387).

RESULTS

Table 2 includes the descriptive statistics of prekindergarten measures as well as our second grade listening comprehension measures used to calculate the listening comprehension factor score. A wide range of individual differences was apparent on all measures. Table 3 reports the correlations of all prekindergarten and second grade measures used for analyses.

Table 4 presents the results of the three-factor confirmatory factor analysis described as part of the first step in selecting the best measures from each construct that were used to predict listening comprehension in second grade. To examine the model fit of the confirmatory factor analysis, the following indices were used: comparative fit index (CFI; Bentler, 1990), root mean squared error of approximation (RMSEA; Steiger, 1990), and standardized root mean square residual (SRMR; Hu & Bentler, 1998). CFI is considered adequate when it exceeds .95 (Hu & Bentler, 1999), RMSEA when it is below .08 (and good fit when below .05; Browne & Cudeck, 1993), and SRMR when below .05 (Hu & Bentler, 1999). Model fit was good with RMSEA = .04, CFA = .97, and SRMR = .04.

Of note, the results of the confirmatory factor analysis indicated that the three factors were highly correlated: the correlation between higher-level language and foundational language was r = .92, higher-level language and working memory was r = .86, and foundational language and working memory was r = .89. Although the correlations among factors were relatively high, concerns were alleviated when a one-factor model was run and model

Table 2. Descriptive statistics of measures used in analyses

Variable	*N*	Mean	*SD*	Observed Range	Theoretical Range
Prekindergarten					
CELF-4 USP	306	6.18	2.12	0 to 10	0 to 10
TNL	304	16.34	6.77	1 to 32	0 to 40
LCM	313	7.38	3.08	0 to 13	0 to 15
PPVT-4	316	95.66	18.37	33 to 142	0 to 228
EVT-2	318	71.01	13.62	24 to 107	0 to 190
CELF-4 WC	303	22.88	8.30	0 to 39	0 to 48
CELF-4 WS	308	16.06	5.55	0 to 29	0 to 32
TEGI-T	295	8.79	4.22	0 to 17	0 to 18
TEGI-S	298	7.18	2.78	0 to 10	0 to 10
TROG-2	316	6.56	3.73	0 to 18	0 to 20
IT	245	0.87	0.40	0 to 2	0 to 2
KVT	316	2.08	1.59	0 to 5	0 to 5
PAT	298	2.86	2.97	0 to 12	0 to 12
MU	317	4.39	2.71	0 to 16	0 to 30
WJ-III AM	311	6.22	4.59	0 to 18	0 to 21
NWRT	296	0.51	0.15	0 to .86	0 to 1
CELF-4 RS	302	33.13	14.15	0 to 76	0 to 102
KBIT-2	317	15.57	3.94	0 to 29	0 to 46
Second grade					
CELF-4 USP	317	6.72	1.76	0 to 10	0 to 10
TNL	313	29.48	4.37	12 to 38	0 to 40
LCM	313	19.18	4.74	1 to 27	0 to 29
LC Factor	318	0.00	3.29	−14.76 to 5.49	—

Note. Raw scores are reported.
CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; CELF-4 WC = Clinical Evaluation of Language Fundamentals-4—Word Classes; CELF-4 WS = Clinical Evaluation of Language Fundamentals-4—Word Structure; CELF-4 USP = Clinical Evaluation of Language Fundamentals-4—Understanding Spoken Paragraphs; EVT-2 = Expressive Vocabulary Test-2; IT = Inference Task; KBIT-2 = Kaufman Brief Intelligence Test-2; KVT = Knowledge Violation Task; LC = Listening Comprehension; LCM = Listening Comprehension Measure; MU = Memory Updating Task; NWRT = Non-Word Repetition Task; PAT = Picture Arrangement Task; PPVT-4 = Peabody Picture Vocabulary Test-4; SD = Standard Deviation; TEGI-S = Rice/Wexler Test of Early Grammatical Impairment-Past Tense; TEGI-T = Rice/Wexler Test of Early Grammatical Impairment-Third Person Singular; TNL = Test of Narrative Language; TROG-2 = Test for Reception of Grammar-2; WJ-III AM = Woodcock Johnson III Tests of Cognitive Abilities: Normative Update—Auditory Working Memory.

fit was compared between the two models. Specifically, a chi-square difference test indicated that the three-factor model was a superior fit ($\chi^2(3) = 59.77$, $p < .0001$) to the one-factor model. In addition, because the goal was to predict a second grade outcome and there is evidence of a language factor bifurcation by second grade (LARRC, 2015b), defining the three-factor model in prekindergarten was justified.

As previously noted, with the exception of WJ-III Auditory Working Memory, measures with standardized loadings higher than .70 were selected in Step 1 to be later used as predictors of second grade listening comprehension (see Table 4 or Figure 1). The nine measures that were selected were WJ-III Auditory Working Memory (loading = .60), CELF-4 Word Structure (loading = .76), Peabody Picture Vocabulary Test (loading = .77), Expressive Vocabulary Test (loading = .79), the Inference measure (loading = .80), CELF-4 Understanding Spoken Paragraphs (loading = .71), Test of Narrative

Table 3. Correlations of all behavioral measures used in the study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Prekindergarten																						
1. KBIT-2	1.00																					
2. CELF-4 WS	0.28	1.00																				
3. TEGI-S	0.17	0.52	1.00																			
4. TEGI-T	*0.10*	0.36	0.42	1.00																		
5. TROG-2	0.35	0.47	0.32	0.31	1.00																	
6. PPVT-4	0.42	0.58	0.43	0.34	0.54	1.00																
7. EVT-2	0.38	0.56	0.41	0.36	0.51	0.71	1.00															
8. CELF-4 WC	0.38	0.46	0.39	0.22	0.48	0.49	0.55	1.00														
9. IT	0.35	0.52	0.40	0.33	0.44	0.60	0.60	0.50	1.00													
10. KVT	0.28	0.50	0.31	0.23	0.43	0.50	0.50	0.39	0.44	1.00												
11. PAT	0.48	0.36	0.19	*0.08*	0.43	0.39	0.41	0.40	0.42	0.32	1.00											
12. CELF-4 USP	0.31	0.51	0.29	0.23	0.39	0.47	0.50	0.48	0.58	0.43	0.41	1.00										
13. TNL	0.32	0.61	0.47	0.34	0.46	0.62	0.62	0.55	0.70	0.51	0.42	0.60	1.00									
14. LCM	0.42	0.58	0.42	0.28	0.45	0.60	0.59	0.54	0.69	0.45	0.38	0.60	0.72	1.00								
15. NWRT	0.21	0.37	0.28	0.26	0.23	0.28	0.33	0.24	0.24	0.22	0.24	0.31	0.40	0.32	1.00							
16. WJ-III AM	0.31	0.47	0.31	0.22	0.41	0.35	0.39	0.36	0.42	0.28	0.45	0.42	0.46	0.43	0.36	1.00						
17. MU	0.28	0.39	0.26	0.13	0.25	0.38	0.40	0.28	0.34	0.33	0.36	0.31	0.38	0.39	0.25	0.35	1.00					
18. CELF-4 RS	0.36	0.64	0.50	0.38	0.52	0.60	0.66	0.54	0.60	0.47	0.40	0.55	0.69	0.67	0.48	0.54	0.47	1.00				
Second grade																						
19. CELF-4 USP	0.15	0.31	0.17	0.25	0.28	0.24	0.28	0.32	0.36	0.28	0.14	0.40	0.42	0.43	0.24	0.28	0.21	0.35	1.00			
20. TNL	0.33	0.48	0.32	0.25	0.42	0.51	0.53	0.44	0.53	0.40	0.30	0.56	0.58	0.56	0.32	0.35	0.32	0.55	0.48	1.00		
21. LCM	0.25	0.46	0.31	0.24	0.38	0.51	0.54	0.40	0.51	0.44	0.31	0.51	0.56	0.56	0.27	0.35	0.26	0.54	0.43	0.63	1.00	
22. LC Factor	0.33	0.52	0.35	0.28	0.45	0.54	0.58	0.48	0.59	0.46	0.33	0.60	0.64	0.63	0.34	0.39	0.33	0.61	0.63	0.94	0.84	1.00

Note. All bivariate correlations were significant except for those in italics.
CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; CELF-4 USP = Clinical Evaluation of Language Fundamentals-4—Understanding Spoken Paragraphs; CELF-4 WC = Clinical Evaluation of Language Fundamentals-4—Word Classes; CELF-4 WS = Clinical Evaluation of Language Fundamentals-4—Word Structure; EVT-2 = Expressive Vocabulary Test-2; IT = Inference Task; KBIT-2 = Kaufman Brief Intelligence Test-2; KVT = Knowledge Violation Task; LC = Listening Comprehension; LCM = Listening Comprehension Measure; MU = Memory Updating Task; NWRT = Non-Word Repetition Task; PAT = Picture Arrangement Task; PPVT-4 = Peabody Picture Vocabulary Test-4; TEGI-S = Rice/Wexler Test of Early Grammatical Impairment-Past Tense; TEGI-T = Rice/Wexler Test of Early Grammatical Impairment-Third Person Singular; TNL = Test of Narrative Language; TROG-2 = Test for Reception of Grammar-2; WJ-III AM = Woodcock Johnson III Tests of Cognitive Abilities-Normative Update—Auditory Working Memory..

Table 4. Standardized factor loadings for prekindergarten

Measures	Estimate	*SE*	Estimate/*SE*	*p*
Foundational language construct				
PPVT-4	0.77	0.04	22.08	<.0001
EVT-2	0.79	0.02	32.53	<.0001
CELF-4 WC	0.67	0.04	18.76	<.0001
CELF-4 WS	0.76	0.03	29.67	<.0001
TEGI-T	0.43	0.05	8.31	<.0001
TEGI-S	0.57	0.05	12.26	<.0001
TROG-2	0.64	0.04	15.88	<.0001
Higher-level language construct				
IT	0.80	0.03	31.39	<.0001
KVT	0.60	0.04	15.41	<.0001
PAT	0.52	0.05	11.19	<.0001
CELF-4 USP	0.71	0.03	22.86	<.0001
TNL	0.85	0.02	44.08	<.0001
LCM	0.83	0.02	40.42	<.0001
Working memory construct				
MU	0.53	0.04	12.22	<.0001
WJ-III AM	0.60	0.04	15.98	<.0001
NWRT	0.52	0.05	11.04	<.0001
CELF-4 RS	0.93	0.02	47.59	<.0001

Note. Italicized measures represent those that were subsequently used for analyses.
CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; CELF-4 USP = Clinical Evaluation of Language Fundamentals-4—Understanding Spoken Paragraphs; CELF-4 WC = Clinical Evaluation of Language Fundamentals-4—Word Classes; CELF-4 WS = Clinical Evaluation of Language Fundamentals-4—Word Structure; EVT-2 = Expressive Vocabulary Test-2; IT = Inference Task; KBIT-2 = Kaufman Brief Intelligence Test-2; KVT = Knowledge Violation Task; LC = Listening Comprehension; LCM = Listening Comprehension Measure; MU = Memory Updating Task; NWRT = Non-Word Repetition Task; PAT = Picture Arrangement Task; PPVT-4 = Peabody Picture Vocabulary Test-4; SE = Standard Error; TEGI-S = Rice/Wexler Test of Early Grammatical Impairment-Past Tense; TEGI-T = Rice/Wexler Test of Early Grammatical Impairment-Third Person Singular; TNL = Test of Narrative Language; TROG-2 = Test for Reception of Grammar-2; WJ-III AM = Woodcock Johnson III Tests of Cognitive Abilities: Normative Update—Auditory Working Memory.

Language (loading = .85), the Listening Comprehension Measure (loading = .83), and CELF-4 Recalling Sentences (loading = .93). All of the loadings were significant at $\alpha = .05$.

A summary of the results of the four multiple regressions that were conducted to assess which of the nine prekindergarten predictors were associated with the second grade listening comprehension factor score can be found in Tables 5 and 6. All models controlled for mother's education, age of children in months, and nonverbal intelligence.

Model 1

The first model included all nine prekindergarten predictors (Model 1, Table 5). For this model, only the three prekindergarten listening comprehension tasks significantly and positively predicted second grade listening comprehension. Because the second grade listening comprehension factor score (i.e., the outcome of interest) was based on a latent variable defined using similar versions of the three listening comprehension measures used in prekindergarten (i.e., CELF-4 Understanding Spoken Paragraphs, Test of Narrative Language, and the Listening Comprehension

Table 5. Model results for prekindergarten measures predicting second grade listening comprehension

	Model 1			Model 2		
	Estimate	***p***	**Relative Importance (Variance Decomposition)**	**Estimate**	***p***	**Relative Importance (Variance Decomposition)**
Intercept	− 5.861	.021*	—	− 5.618	.035*	—
Mother's education	0.034	.754	0.011	0.053	.639	0.014
Age in months	− 0.059	.153	0.004	− 0.056	.195	0.004
KBIT-2	0.014	.747	0.014	0.040	.369	0.018
TNL	0.350	.000*	0.083	0.127	.001*	0.108
CELF-4 USP	0.081	.037*	0.079	—	—	—
LCM	0.188	.023*	0.080	—	—	—
CELF-4 WS	0.029	.457	0.043	0.0505	.220	0.055
CELF-4 RS	0.026	.148	0.066	0.0364	.047*	0.085
PPVT-4	0.010	.418	0.044	0.0111	.409	0.055
EVT-2	0.025	.145	0.055	0.0293	.107	0.069
IT	0.074	.914	0.049	0.9374	.153	0.069
WJ-III AM	0.003	.942	0.020	0.0209	.617	0.026
R^2			0.548			0.505

*$p < .05$.

Note. CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; CELF-4 USP = Clinical Evaluation of Language Fundamentals-4—Understanding Spoken Paragraphs; CELF-4 WS = Clinical Evaluation of Language Fundamentals-4—Word Structure; EVT-2 = Expressive Vocabulary Test-2; IT = Inference Task; LCM = Listening Comprehension Measure; KBIT-2 = Kaufman Brief Intelligence Test-2; PPVT-4 = Peabody Picture Vocabulary Test-4; TNL = Test of Narrative Language; WJ-III AM = Woodcock Johnson III Tests of Cognitive Abilities: Normative Update—Auditory Working Memory.

Measure), it was hypothesized that they would be predictive of second grade listening comprehension. In terms of variance decomposition, relative importance analysis suggested that each of the three prekindergarten listening comprehension tasks explained between 7.9% and 8.3% of the variance of second grade listening comprehension. Together, they were responsible for 24.2% of the variance. Overall, this model explained 54.8% of variance of second grade listening comprehension.

Model 2

To determine whether additional prekindergarten measures (besides listening comprehension) were significant predictors of second grade listening comprehension, a model with only a single measure of prekindergarten listening comprehension was run. Given its availability among service providers, the Test of Narrative Language was chosen as the only prekindergarten listening comprehension measure in this model. Results of this model are presented under Model 2, Table 5. Specifically, the Test of Narrative Language and CELF-4 Recalling Sentences were positive and significant prekindergarten predictors of second grade listening comprehension. In terms of relative importance, the Test of Narrative Language explained 10.8% of the total variance, whereas the CELF-4 Recalling Sentences was responsible for 8.5% of the variance. Overall, this model explained 50.5% of the variance of second grade listening comprehension.

Model 3

Given that only the Test of Narrative Language and the CELF-4 Recalling Sentences were significant in Model 2, we set out to further understand the variance decomposition of the significant prekindergarten predictors by running a model with only the Test of Narrative Language and the CELF-4 Recalling Sentences as prekindergarten predictors (Model 3, Table 6). Similar to Model 2, this model indicated that both the Test of Narrative Language and the CELF-4 Recalling Sentences were positive and significant predictors of the outcome. However, the variance decomposition indicated that the Test of Narrative Language was responsible for 21.7% of the variance in second grade listening comprehension, whereas the CELF-4 Recalling Sentences was responsible for 18.3% of this variance. Overall, this model explained 47.3% of the variance of second grade listening comprehension.

Model 4

We ran a model with only the CELF-4 Recalling Sentences (Model 4, Table 6) to understand its independent influence on second grade listening comprehension. This model showed that both nonverbal intelligence and the CELF-4 Recalling Sentences were positive and significant predictors of second grade listening comprehension. In addition, variance decomposition indicated that the CELF-4 Recalling Sentences was responsible for 29.6% of the variance of second grade listening comprehension, whereas nonverbal intelligence was responsible for 5.1%. Overall, this model explained 39.5% of the variance of second grade listening comprehension.

To summarize, the largest amount of variance predicted by our models with multiple measures was 54.8%. Our results indicated that when including all three prekindergarten listening comprehension measures, they each significantly predicted second grade listening comprehension. We chose to examine the predictive power of including only one measure of listening comprehension—the Test of Narrative Language—because it is readily available for purchase by educators and clinicians. In doing so, we found that the prekindergarten Test of Narrative Language was responsible for approximately 8.3%–21.7% of the variance, depending on the

Table 6. Reduced model results for prekindergarten measures predicting second grade listening comprehension

	Model 3			Model 4		
	Estimate	***p***	**Relative Importance (Variance Decomposition)**	**Estimate**	***p***	**Relative Importance (Variance Decomposition)**
Intercept	−5.421	.015*	—	−5.940	.013*	—
Mother's education	0.126	.204	0.031	0.189	.074	0.042
Age in months	−0.028	.446	0.005	−0.010	.797	0.006
KBIT-2	0.072	.057	0.038	0.091	.024*	0.051
TNL	0.193	<.0001*	0.217	—	—	—
CELF-4 RS	0.067	<.0001*	0.183	0.120	<.0001*	0.296
R^2			0.473			0.395

*$p < .05$.

Note. CELF-4 RS = Clinical Evaluation of Language Fundamentals-4—Recalling Sentences; KBIT-2 = Kaufman Brief Intelligence Test-2; TNL = Test of Narrative Language.

inclusion of other predictors in the model. Another significant language predictor was a measure of sentence imitation. Depending on other predictors in the model, the CELF-4 Recalling Sentences was responsible for explaining 6.6%–29.6% of the variance in second grade listening comprehension.

DISCUSSION

Although many screening instruments are available to predict risk for word reading disabilities, there is a paucity of early screening measures for detecting risk for difficulties in listening comprehension. Using a large, 5-year longitudinal data set, the purpose of this study was to identify prekindergarten predictors of second grade listening comprehension that could serve as quick reliable screening tools for identifying children at risk for later listening comprehension difficulties. These identified children would then require a more comprehensive assessment of language to confirm whether they indeed have a listening comprehension deficit. With a focus on immediate application, we gave preference to readily available, and relatively quick, measures to determine whether those measures could significantly predict listening comprehension.

As we hypothesized, we were able to predict a significant amount of variance in second grade listening comprehension based on prekindergarten measures administered 4 years prior. In our best statistical model, we accounted for approximately 55% of the variance in individual differences in listening comprehension. Past studies accounted for at most 50%. It is likely we were able to account for a slightly larger portion of the variance in later listening comprehension compared with past studies because of our comprehensive array of potential predictors. Although even at 55%, there is room for substantial error when predicting an individual child's future listening comprehension. So, in practice, these measures should be used as part of a screening protocol to determine whether a child requires further assessment. To that end, educators may consider administering these measures, singly or in a bundle, within an early screening or RTI framework.

We hypothesized that a measure of vocabulary would significantly predict future listening comprehension because vocabulary is a robust predictor of listening comprehension (Florit et al., 2014; Kim, 2015). However, our results did not support this conclusion. Moreover, our standardized working memory measure did not predict listening comprehension even though past studies by Florit and colleagues (2009) have found that working memory was a significant, albeit small, predictor of concurrent and future listening comprehension in young children. It is likely that vocabulary and working memory were not significant in our predictive models because of the inclusion of other measures that required similar skills as those used in second grade listening comprehension tasks. For example, as hypothesized, our earlier measures of listening comprehension significantly predicted future listening comprehension. Often one's past ability can predict one's current ability. Listening comprehension was no exception to this rule. It seems a child's ability to comprehend aurally in prekindergarten is similar to his ability to do the same in second grade, 4 years later. Based on our findings, listening comprehension may be a stable skill across the early grades.

Our other significant predictor of second grade listening comprehension was a measure of one's ability to repeat sentences accurately in prekindergarten. This finding is in line with other studies showing that an early measure of sentence imitation is predictive of later reading comprehension (e.g., Adlof et al., 2010; Alloway & Gathercole, 2005; Badian, 1982; Scarborough, 1998). Sentence imitation tasks tap into children's grammatical knowledge and working memory (Catts et al., 2014; Kidd, 2013). Indeed, there is an active debate on the relative importance of those skills—grammatical knowledge versus working memory—for explaining individual differences in sentence imitation (see Klem et al., 2015). We found that our measure of sentence imitation loaded equally on both foundational language and working memory constructs, which leads us to conclude that one's ability to repeat sentences draws on both grammatical knowledge and working memory. It is this task's multifactorial nature that likely makes sentence imitation a strong predictor of listening comprehension, a skill that also draws on grammatical knowledge to make sense of multiple clauses and working memory to hold those clauses while adapting to new information with each new clause.

Limitations

Although this study is unique in its size and scope with a strong analytical design, there are limitations worth noting. First, we needed to reduce the number of predictors in our regression models because inclusion of all possible predictors from our comprehensive assessment battery would result in a number of regression models that would have made accurate interpretation untenable. We used research-supported cut points to determine our predictors within a strong theoretical framework. We also showed preference to those measures that are available to educators and clinicians. Moreover, we did not include word-reading-based indicators, such as letter identification, because we chose to predict listening comprehension, not word reading. Those decisions, taken together, were primarily clinically driven, and, as such, could have influenced which measures we found to be statistically significant predictors of listening comprehension. Second, we did not provide specific cut points for educators who choose to use these measures for screening purposes to determine who is at risk versus who is not at risk for future listening comprehension difficulties and may need further in-depth diagnostic assessment. We acknowledge that our longitudinal sample may not represent children in more diverse classrooms. Furthermore, although our study accounted for a larger amount of variance in predicting listening comprehension than other studies, that variance is not large enough to generate an accurate formula to determine which individual child will have future difficulty with listening comprehension based on prekindergarten measures. Instead, our findings provide guidance on which measures educators and clinicians may choose to use to screen to determine who requires additional in-depth, comprehensive assessment. We suggest that appropriate cut points be determined based on local distribution of test scores and based on local funding allocations for early intervention assessment and treatment.

CONCLUSIONS

This study expanded on previous research by determining prekindergarten predictors of second grade listening comprehension using data from a large, longitudinal sample of children. Our findings show that a quick, reliable, readily available measure of sentence imitation and/or listening comprehension, administered in prekindergarten, can be an important screening measure to determine who is at risk for deficient second grade listening comprehension. Current literacy screening measures do not typically include predictors of listening comprehension, a critical skill for reading comprehension. Screening for risk for later poor listening comprehension will allow educators and clinicians to make informed, evidence-based decisions about who requires further in-depth testing to determine which children would benefit from language-intensive instruction to stave off future reading disabilities of *all* types, including those involving comprehension deficits.

REFERENCES

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. doi:10.1177/0022219410369067

Adlof, S. M., Catts, H. W., & Little, T. D. (2006).

Should the simple view of reading include a fluency component? *Reading and Writing*, *19*(9), 933–958. doi:10.1007/s11145-006-9024-z

Allen, M. M., Ukrainetz, T. A., & Carswell, A. L. (2012). The narrative language performance of three types of at-risk first-grade readers. *Language, Speech, and Hearing Services in Schools*, *43*(2), 205–221. doi:10.1044/0161-1461(2011/11-0024)

Alloway, T. P., & Gathercole, S. E. (2005). The role of sentence recall in reading and language skills of children with learning difficulties. *Learning and Individual Differences*, *15*(4), 271–282. doi:10.1016/j.lindif.2005.05.001

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Badian, N. A. (1982). The prediction of good and poor reading before kindergarten entry: A 4-year follow-up. *The Journal of Special Education*, *16*(3), 309–318. doi:10.1177/002246698201600306

Baker, L. (1984). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology, 76*, 588–597. doi:10.1037//0022-0663.76.4.588

Belacchi, C., Carretti, B., & Cornoldi, C. (2010). The role of working memory and updating in Coloured Raven Matrices performance in typically developing children. *European Journal of Cognitive Psychology*, *22*(7), 1010–1020. doi:10.1080/09541440903184617

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. doi:10.1037/0033-2909.107.2.238

Bishop, D. V. M. (2003). *Test for reception of grammar* (2nd ed.). Minneapolis, MN: Pearson.

Bridges, M. S., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*(4). doi:10.1177/0022219411407863

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136–136. doi:10.1177/0049124192021002005

Cain, K. (1996). Story knowledge and comprehension skill. In Cornoldi C., & Oakhill J. (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 167–192). Mahwah, NJ: Lawrence Erlbaum.

Cain, K., & Oakhill, J. (1999). Inference making and its relation to comprehension failure. *Reading and Writing, 11*, 489–503. doi:10.1023/a:1008084120205

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique, 114*, 647–662. doi:10.4074/s0003503314004035

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & cognition, 29*(6):850–859.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278–293. doi:10.1044/1092-4388(2006/023)

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32*(1), 38–50. doi:10.1044/0161-1461(2001/004)

Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y.-S. (2014). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities, 49*(5), 451–465. doi:10.1177/0022219414556121

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearson.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1–44.

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology, 115*, 497–516. doi:10.1016/j.jecp.2012.12.001

Florit, E., Roch, M., Altoe, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, *27*(4), 935–951. doi:10.1348/026151008×397189

Florit, E., Roch, M., & Levorato, M. C. (2013). The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Applied Psycholinguistics, 34*(2), 395–415.

Florit, E., Roch, M., & Levorato, M. C. (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. *Reading and Writing*, *27*(5), 793–817. doi:10.1007/s11145-013-9464-1

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, *18*(3), 157–171. doi:10.1111/1540-5826.00072

Gillam, R. B., & Pearson, N. A. (2004). *Test of narrative language*. Austin, TX: Pro-Ed.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. doi:10.1177/074193258600700104

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, *17*(1), 1–27.

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, *16*(3), 199–207. doi:10.3109/17549507.2014.904441

Hogan, T. P., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children, 44*(3), 1–18.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. doi:10.1007/bf00401799

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424. doi:10.1037//1082-989x.3.4.424

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*(12), 1877–1886. doi:10.1177/0956797615603702

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). *Handbook of response to intervention*. New York, NY: Springer. doi:10.1007/978-0-387-49053-3

Justice, L., Mashburn, A., & Petscher, Y. (2013). Very early language skills of fifth-grade poor comprehenders. *Journal of Research in Reading*, *36*(2), 172–185. doi:10.1111/j.1467-9817.2011.01498.x

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Brief Intelligence Test, Second Edition. Minneapolis, MN: Pearson, Inc.

Kelso, K., Fletcher, J., & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language & Communication Disorders*, *42*(1), 39–57. doi:10.1080/13682820600693013

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, *33*(3), 208–223. doi:10.1097/TLD.0b013e31829d623e

Kim, Y.-S. G. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, *86*(1), 128–144. doi:10.1111/cdev.12293

Kim, Y.-S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, *141*, 101–120. doi:10.1016/j.jecp.2015.08.003

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. doi:10.1111/desc.12202

Language and Reading Research Consortium. (2015a). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*, 151–169. doi:10.1002/rrq.99

Language and Reading Research Consortium. (2015b). The dimensionality of language ability in young children. *Child Development, 86*(6), 1948–1965. doi:10.1111/cdev.12450

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly*, *47*(3), 259–282. doi:10.1002/RRQ.020

Leslie, L., & Caldwell, J. (2010). *Qualitative reading inventory* (5th ed.). Bloomington, MN: Pearson Assessments.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. doi:10.1080/01621459.1988.10478722

Lorch, E. P., Milich, R., Sanchez, R. P., van den Broek, P., Baer, S., Hooks. K., et al. (2000). Comprehension of televised stories in boys with attention deficit/hyperactivity disorder and nonreferred boys. *Journal of abnormal psychology, 109*(2), 321–330.

McInnes, A., Humphries, T., Hogg-Johnson, S., & Tannock, R. (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of Abnormal Child Psychology*, *31*(4), 427–443. doi:10.1023/A:1023895602957

Molloy, P. (1997). *The role of individual differences in working memory in reading and listening comprehension in intermediate grade students* (Doctoral dissertation), *58*(6), 2072-A-2073-A (UMI No. AAMNN19626).

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics, 32*, 385–397. doi:10.1016/0304-4076(86)90021-7

Muthén, L. K., & Muthén, B. O. (2011). Mplus (version 6.12) [computer software], Los Angeles.

Nation, K., Cocksey, J., Taylor, J. S. H., & Bishop, D. V. M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*, 1031–1039. doi:10.1111/j.1469-7610.2010.02254.x

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91–121. doi:10.1080/10888438.2010.529219

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*(7–9), 657–686. doi:10.1007/s11145-005-3355-z

Potocki, A., Ecalle, J., & Magnan, A. (2013). Narrative comprehension skills in 5-year-old children:

Correlational analysis and comprehender profiles. *The Journal of Educational Research*, *106*(1), 14–26. doi:10.1080/00220671.2012.667013

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler test of early grammatical impairment*. San Antonio, TX: The Psychological Corporation.

SAS Institute. (2014). *Base SAS 9.4 procedures guide: Statistical procedures*. Cary, NC: Author.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, *48*(1), 115–136. doi:10.1007/s11881-998-0006-5

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). Bloomington, MN: Pearson.

Skarakis-Doyle, E., & Dempsey, L. (2008). The detection and monitoring of comprehension errors by preschool children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1227–1243. doi:10.1044/1092-4388(2008/07-0136)

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. doi:10.1207/s15327906mbr2502_4

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*(2), 497–506.

Tighe, E. L., Spencer, M., & Schatschneider, C. (2015). Investigating predictors of listening comprehension in third-, seventh-, and tenth-grade students: a dominance analysis approach. *Reading Psychology*, *36*(8), 700–740. doi:10.1080/02702711.2014.963270

Ukrainetz, T. A. (2006). The implications of RTI and EBP for SLPs: Commentary on LM Justice. *Language, Speech, and Hearing Services in Schools*, *37*(4), 298–303.

Vandewalle, E., Boets, B., Boons, T., Ghesquière, P., & Zink, I. (2012). Oral language and narrative skills in children with specific language impairment with and without literacy delay: A three-year longitudinal study. *Research in Developmental Disabilities*, *33*(6), 1857–1870. doi:10.1016/j.ridd.2012.05.004

Wechsler, D. (1992). *Wechsler intelligence scale for children* (3rd ed.). New York, NY: The Psychological Corporation.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). *Clinical evaluation of language fundamentals-preschool* (2nd ed.). San Antonio, TX: Psychological Corporation.

Williams, K. T. (2007). *Expressive vocabulary test* (2nd ed.). Bloomington, MN: Pearson.

Woodcock, R.W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, United Kingdom: Cambridge University Press.