Journal of Educational Psychology

Reading Prosody Unpacked: A Longitudinal Investigation of Its Dimensionality and Relation With Word Reading and Listening Comprehension for Children in Primary Grades

Young-Suk Grace Kim, Jamie M. Quinn, and Yaacov Petscher

Online First Publication, April 6, 2020. http://dx.doi.org/10.1037/edu0000480

American Psychological Association

Journal of Educational Psychology

ISSN: 0022-0663

http://dx.doi.org/10.1037/edu0000480

Reading Prosody Unpacked: A Longitudinal Investigation of Its Dimensionality and Relation With Word Reading and Listening Comprehension for Children in Primary Grades

Young-Suk Grace Kim
University of California, Irvine

Jamie M. Quinn and Yaacov Petscher
Florida State University

We investigated the dimensionality of various indicators of reading prosody, and the relations of word reading and listening comprehension to the identified dimension(s) of reading prosody, using longitudinal data from Grades 1 to 3. A total of 371 English-speaking children were assessed on oral text reading, word reading, and listening comprehension in the fall and spring of each year (i.e., 6 waves of data). From oral text reading, reading prosody was evaluated on pause structures (pause duration, pause frequency) and pitch (intonation contour, F_0 change) using spectrographic analysis, and on expressiveness, smoothness, phrasing, and pacing using the Multi-Dimensional Fluency Scale (MFS). A bifactor structure described the data best across the 6 waves, composed of (a) a ratings and pause *general* factor, which captured common variance among MFS, pause frequency, and pause duration; (b) ratings (MFS) and pause *specific* factors, which captured variance over and above the ratings and pause general factor; and (c) a separate pitch factor, which captured variance in intonation contour and F_0 change. Word reading and listening comprehension were related to the identified dimensions of reading prosody, but when they were in a model together, word reading, not listening comprehension, was uniquely related to reading prosody across the six waves. These results indicate that reading prosody is multidimensional and that a pitch factor is a dissociable skill from the general ratings and pause prosody. Furthermore, word reading is the primary driver for the development of various dimensions of reading prosody, at least for children in primary grades.

Educational Impact and Implications Statement

Reading prosody, reading texts with appropriate expression, has been widely considered an important feature of text reading fluency. We found that multiple aspects and indicators of reading prosody are best described as a multidimensional construct composed of a pause and ratings dimension and a pitch dimension. Children's word reading and listening comprehension skills were both related to these dimensions of reading prosody, but word reading had a consistent and independent relation. These results indicate the importance of word reading development for expressive reading for children in primary grades.

Keywords: listening comprehension, Multi-Dimensional Fluency Scale, reading prosody, spectrograph, word reading

Supplemental materials: http://dx.doi.org/10.1037/edu0000480.supp

Fluent reading, widely known as text or oral reading fluency, has garnered substantial attention in research and educational settings. Definitions of text reading fluency vary (Kuhn, Schwanenflugel, Meisinger, Levy, & Rasinski, 2010; National Institute of Child Health and Human Development, 2000; Wolf & Katzir-Cohen, 2001), but they typically include the following three

Editor's Note. Kathy S. Binder served as the action editor for this article.—SG

Young-Suk Grace Kim, School of Education, University of California, Irvine; Jamie M. Quinn, Florida Center for Reading Research, Florida State University; Yaacov Petscher, College of Social Work, Florida State University.

We thank participating schools, teachers, and children. This research was supported by Grants R305A120147 and R305A180055 by the Institute of Education Sciences, US Department of Education. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency.

Correspondence concerning this article should be addressed to Young-Suk Grace Kim, School of Education, University of California, Irvine, 3455 Education Building, Irvine, CA 92697. E-mail: Youngsk7@uci.edu or young.kim@uci.edu

aspects—text reading accuracy, text reading speed, and expression (i.e., reading prosody). Clear in this is a recognition of reading prosody as one of the defining features of fluent reading (Dowhower, 1991; Kuhn & Stahl, 2003; Schreiber, 1991; Wolf & Katzir-Cohen, 2001). In fact, reading prosody is hypothesized to be "at the heart of the development of reading skill" (Kuhn et al., 2010, p. 239) and is related to reading comprehension (e.g., Arcand et al., 2014; Binder et al., 2013; Calet, Gutierrez-Palma, & Defior, 2015; Groen, Veenendaal, & Verhoeven, 2019; Klauda & Guthrie, 2008; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel, Hamilton, Wisenbaker, Kuhn, & Stahl, 2004; Veenendaal, Groen, & Verheoeven, 2014). However, most previous studies on text reading fluency have focused on the accuracy and speed aspects of connected text reading (text reading *efficiency* to be precise; Baker et al., 2008; Baker, Park, & Baker, 2012; Daane, Campbell, Grigg, Goodman, & Oranje, 2005; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Kim, 2015; Kim & Wagner, 2015; Kim, Park, & Wagner, 2014; Kim, Petscher, Schatschneider, & Foorman, 2010; Kim, Wagner, & Foster, 2011; Kim, Wagner, & Lopez, 2012; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Silverman, Speece, Harring, & Ritchey, 2013; Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009) with considerably less empirical attention to reading prosody.

To address this gap in the literature, our goals in the present study were to investigate (a) the dimensionality of various indicators of reading prosody and (b) the relations of word reading and oral language comprehension (i.e., listening comprehension) to the identified dimension(s) of reading prosody, using longitudinal data from children in primary grades in elementary school (from Grade 1 to Grade 3). Note that in the present study, we focus on prosody in reading *connected* texts that is part of the definition of text reading fluency, not on prosodic sensitivity in isolated words, known as word prosody or prosodic sensitivity (see Calet, Gutierrez-Palma, Simpson, Gonzalez-Trujillo, & Defior, 2015; Kim & Petscher, 2016; Schwanenflugel & Benjamin, 2017; Whalley & Hansen, 2006; Wood, Wade-Woolley, & Holliman, 2009). Also note that we use and differentiate the terms text reading *fluency* and text reading *efficiency*. Although the definition of text reading fluency includes efficiency (accuracy and speed) *and* prosody of reading connected texts, most studies have operationalized text reading fluency as text reading efficiency, excluding reading prosody. In a similar vein, *text* reading fluency is used over the widely used broad term, reading fluency, because theoretically and empirically *text* reading fluency is a differentiated construct from *word* reading fluency (Jenkins et al., 2003; Kim, 2015; Kim & Wagner, 2015).

Reading Prosody

Reading prosody refers to prosodic rendering of the written text when reading aloud. Prosody concerns suprasegmental rhythmic and melodic features of speech, including pitch (intonation), stress (loudness), and duration (Dowhower, 1991). Pitch changes at the end of a sentence—typically declining pitch at the end of a declarative sentence, and rising pitch at the end of a yes–no question (e.g., Schwanenflugel et al., 2004). Pauses are expected in meaningful semantic units (e.g., phrasal unit) as well as between sentences. Intrasentential pauses are usually shorter than intersentential pauses (Cooper & Paccia-Cooper, 1980; Schwanenflugel et al., 2004). Individuals with skilled reading prosody read texts with appropriate raising and lowering of pitch, phrasing or grouping of words into meaningful units, lengthening of certain vowels, and duration of pauses (Binder et al., 2013; Dowhower, 1991; Groen et al., 2019; Miller & Schwanenflugel, 2006; Schwanenflugel et al., 2004).

Clear in this brief review is that there are multiple aspects of reading prosody (e.g., pitch, pause structure, phrasing, stress,). These different aspects of reading prosody have been measured using rating scales and spectrographic analyses. In a rating scale, students' oral reading is evaluated against a priori established criteria on overall expressiveness or on multiple aspects (e.g., phrasing). For example, the National Assessment of Education Progress (NAEP) Oral Reading Fluency Scale evaluates students' overall reading expressiveness primarily based on one's ability to read in meaningful phrase units, and this scale was shown to be both reliable and related to students' reading skills (Pinnell et al., 1995; Sabatini, Wang, & O'Reilly, 2019). Other rating scales evaluate multiple aspects of prosody. In the Comprehensive Oral Reading Fluency Scale (Benjamin et al., 2013), students' oral reading is evaluated on rate and accuracy (on a scale of 1 to 8), expressive intonation (on a scale of 1 to 4), and natural pausing (on a scale of 1 to 4). The most widely used rating scale to date is the Multi-Dimensional Fluency Scale (Rasinski, 2004; Rasinski, Rikli, & Johnston, 2009; Zutell & Rasinski, 1991). The Multi-Dimensional Fluency Scale evaluates reading prosody on a scale of 1 to 4 in the following four aspects: (a) expression and volume, (b) phrasing, (c) smoothness, and (d) pace. The expression and volume aspect focuses on sounding or reading like natural language, phrasing evaluates choppiness and intonation, smoothness evaluates pauses and smooth rhythm, and pace evaluates conversational pace. The Multi-Dimensional Fluency Scale has been used widely in previous studies and has been shown to be reliable and valid in several languages, including English (Paige, Rasinski, & Magpuri-Lavell, 2012), Spanish (González-Trujillo, Calet, Defior, & Gutiérrez-Palma, 2014), Dutch (Veenendaal, Groen, & Verhoeven, 2014, 2015), and Turkish (Yildiz et al., 2014).

Reading prosody has been also examined using spectrographic analysis. Spectrographic analysis allows researchers to measure specific aspects of reading prosody with precision. For example, pitch can be measured by marking the difference between fundamental frequency (F_0) in the last peak and F_0 at the end of the sentence in hertz (Hz). Duration of intersentential and intrasentential pauses can be measured in milliseconds, and intensity (or loudness as a measure of stress) can be measured in decibel (dB). A series of studies on reading prosody using spectrographic measurements have been conducted by Schwanenflugel and her colleagues. For example, they measured reading prosody in terms of duration (intersentential pause duration, intrasentential pause duration) and pitch (sentence-final F_0 change, and child-adult F_0 match), and examined the relations of reading prosody to reading skills (Schwanenflugel et al., 2004). Other studies revealed that more advanced readers in second grade have shorter pauses and fewer ungrammatical pauses (which captures the phrasing aspect of prosody in studies using a rating scale; Valle, Binder, Walsh, Nemier, & Bangs, 2013); that children marked direct quote, contrastive words, and exclamations with a higher pitch than when in an unmarked context (Schwanenflugel, Westmoreland, & Benjamin, 2015); that reading prosody values vary depending on text complexity (Benjamin & Schwanenflugel, 2010); and that various indicators such as inappropriate (or ungrammatical) pauses, adult-like intonation contour (i.e., vocalic nuclei), and pitch (F_0) change are related to reading comprehension (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel et al., 2004, 2015). Álvarez-Cañizo, Suárez-Coalla, and Cuetos (2015) also examined similar indicators using spectrographic measurements in Spanish and found that children with poor reading comprehension skills had poor reading prosody.

In the present study, we expand our understanding of reading prosody by examining the dimensionality of multiple, widely used reading prosody indicators (e.g., intonation, pitch, pauses, phrasing, smoothness, pace), using both spectrographic analysis and a rating scale. Although these multiple aspects of reading prosody have been examined in previous studies, to our knowledge, few studies have explicitly investigated the dimensionality of these diverse indicators. One exception is Benjamin et al.'s (2013) study in which reading prosody was measured by ungrammatical pauses and sentence-final pitch indicators using spectrographic analysis, and a principal component exploratory factor analysis revealed two factors: a pitch factor and a pause factor. Other studies also suggest that pitch variables may be capturing a somewhat different dimension than pause structure variables because pause structure variables have weak relations with pitch variables (Benjamin & Schwanenflugel, 2010; Binder et al., 2013; Schwanenflugel et al., 2004, 2015). Given that various aspects (whether measured by rating scale or spectrogram approaches) are purported to capture the construct of reading prosody, it is important to examine whether multiple indicators of reading prosody are best described as having a single dimension, multiple dissociable but related dimensions, or a bifactor structure with a general factor that captures common variance across all the indicators and with specific factors (those over and above the general factor; see the Data Analytic Strategy below). In the present study, we measured multiple aspects of students' reading prosody using spectrographic measurements (intonation contour, sentence-final F_0 change, intersentential pause duration, frequency of ungrammatical pauses, and total pause frequency) and the Multi-Dimensional Fluency Scale ratings (expression and volume, phrasing, smoothness, and pace; Rasinski, Rikli, & Johnston, 2009; Zutell & Rasinski, 1991).

Predictors of Reading Prosody

If reading prosody is an important part of text reading fluency, then what are the contributing skills to reading prosody? According to the automaticity theory (LaBerge & Samuels, 1974) and the verbal efficiency theory (Perfetti, 1992), one apparent skill that is necessary for reading prosody development is decoding or word reading skill (e.g., Chall, 1996; Kuhn & Stahl, 2003; Schwanenflugel et al., 2004) because automaticity in word reading or decoding allows cognitive resources (e.g., working memory and attention) to be available for additional processes such as semantic processing and prosodic reading. Extant evidence indeed supports a relation of word reading to reading prosody. For example, word reading efficiency (accuracy and rate) was very strongly and negatively related to intrasentential pause and moderately related to pitch change for third graders (Schwanenflugel et al., 2015; also see Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2008; and Schwanenflugel et al., 2004) and adults with low literacy skills (Binder et al., 2013). Word reading skill is expected to be particularly strongly related to reading prosody in the beginning phase of reading development because of its large constraining role of reading development. For example, reading prosody measured by the Multi-Dimensional Fluency Scale was strongly related with nonword reading and text reading efficiency for second graders (.72–.78), whereas it was weakly to moderately related for fourth graders (.24–.44) learning to read in Spanish (Calet et al., 2015).

Another potential skill that might contribute to reading prosody development is oral language skill because prosodic reading involves semantic processing (Kuhn et al., 2010). Prosodic reading—reading with appropriate raising and lowering of pitch, appropriate grouping of words in meaningful units, and reading with adult-like prosodic contour—would be facilitated by semantic processing or meaning construction and integration (comprehension; Fodor, 1998; Frazier, Carlson, & Clifton, 2006; Webman-Shafran, 2018). Therefore, children's ability in listening comprehension, which involves semantic processing (Adlof, Catts, & Little, 2006; Florit & Cain, 2011; Gough & Tunmer, 1986; Kim, 2016, 2017, 2020), might be related to reading prosody. This relation is likely to emerge after children have developed a certain level of word reading skill as children's word reading skill has a large constraining role in comprehension processes during the beginning phase of reading development (e.g., Adlof et al., 2006; Kim & Wagner, 2015). Studies have reported a weak to moderate relation of reading prosody to vocabulary (Arcand et al., 2014; Groen et al., 2019; Paige, Rasinski, Magpuri-Lavell, & Smith, 2014; Ravid & Mashraki, 2007; Veenendaal et al., 2015) and to syntactic knowledge (Veenendaal et al., 2015). To our knowledge, no prior work has investigated the relation of listening comprehension to reading prosody.

The Present Study

The goals of the present study were to address several gaps in the literature on reading prosody—its dimensionality, its predictors (the relations of word reading and listening comprehension to reading prosody), and the potentially changing nature of these dimensions and predictors with development, using longitudinal data from primary grade children in elementary school. The theory and evidence reviewed above suggest that word reading plays a large constraining role in reading development. Then, it is reasonable to speculate that the relations among identified dimensions of reading prosody (e.g., pitch and pause structure) and the relations of word reading and listening comprehension to reading prosody may change with development. The following were two research questions that guided the present study: (a) What is the dimensionality of various reading prosody indicators (intonation contour, sentence-final F_0 change, intersentential pause duration, frequency of ungrammatical pauses, and total pause frequency that are measured by spectrographic analysis; and expression and volume, phrasing, smoothness, and pace that are measured by the Multi-Dimensional Fluency Scale ratings) for children in the lower grades of elementary school (i.e., Grades 1 to 3)? Do the relations among identified dimensions change with development?; (b) How are children's word reading and listening comprehension skills related to the identified dimensions of reading prosody? Do the relations change from Grade 1 to Grade 3? These questions were addressed by using longitudinal data from Grade 1 to Grade 3.

The first research question on dimensionality was addressed by fitting a series of alternative models shown in Figure 1 (see Data Analytic Strategy section below for details). We hypothesized that the reading prosody variables would be either dissociable but related factors (Figure 1b) or have a bifactor structure (Figures 1c–1e; see Data Analytic Strategy section below for details). We also expected that the relations between various dimensions of reading prosody (e.g., pitch and pause structure) would become stronger with the development of children's word reading skills—that is, as the constraining role of word reading decreases, cognitive resources will be increasingly available for reading prosody, which in turn will allow stronger relations between prosody variables. The second research question was addressed by including word reading and listening comprehension as predictors of the dimensions of reading prosody identified in the first research question (see Figure 3). We posited that word reading would be strongly related to reading prosody across the grades, particularly to pause-related prosody (e.g., ungrammatical pause frequency;

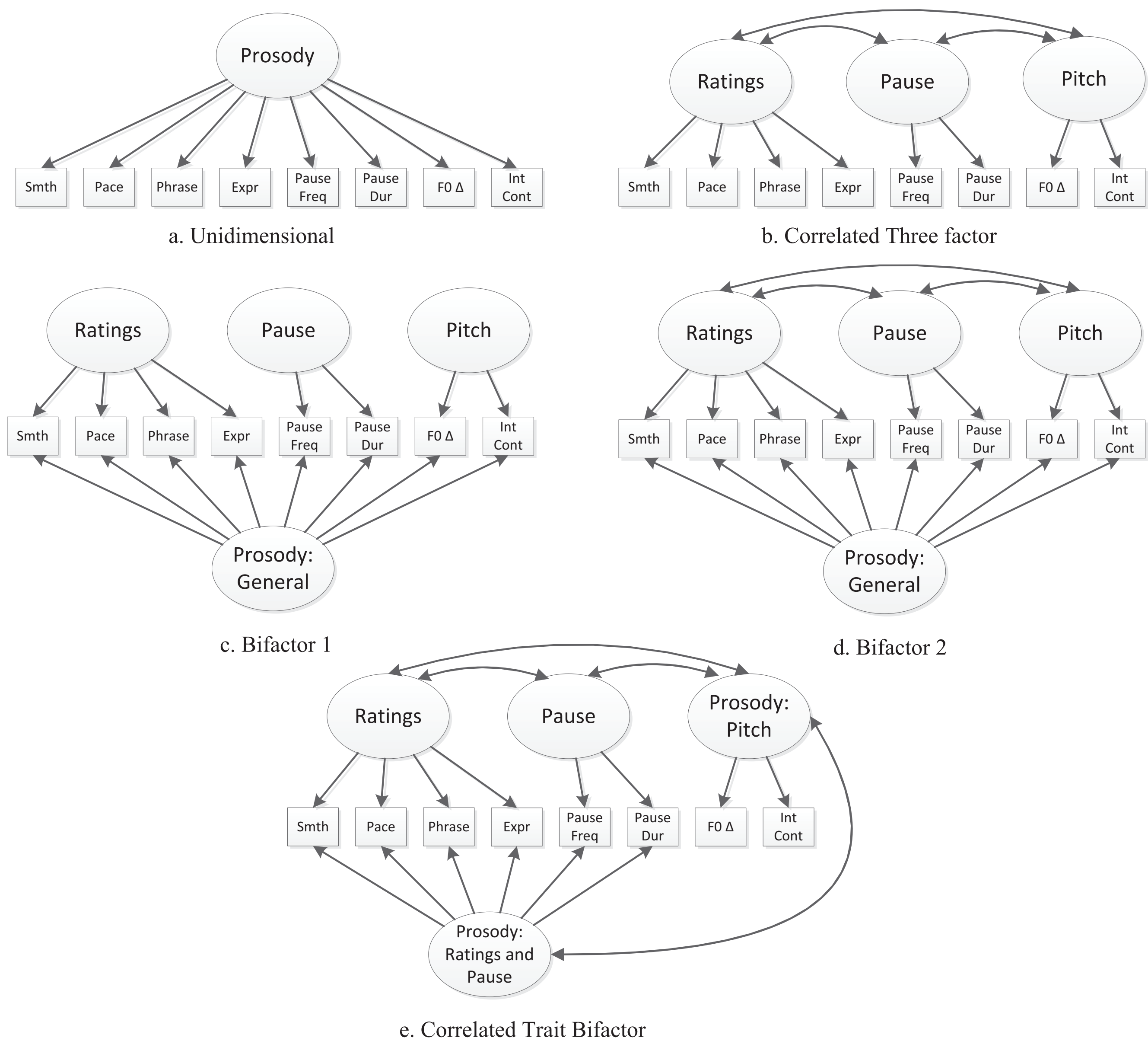

Figure 1. Dimensionality models fit to the prosody data. Residual variances for all indicators were estimated, but not shown for figure brevity. All pathways were estimated, and factor variances were fixed at 1 for model identification purposes. Smth = smoothness; Pace = pacing; Phrase = phrasing; Expr = expression and volume; Pause Freq = pause frequencies; Pause Dur = pause durations; F_0 Δ = fundamental frequency change; Int Cont = intonation contour.

Binder et al., 2013; Schwanenflugel et al., 2004, 2015), whereas listening comprehension may be related to reading prosody after the very initial phase of reading development (e.g., Grade 3).

Method

Participants

The sample students were 371 English-speaking children who participated in a 3-year longitudinal study from Grade 1 to Grade 3 in a Southeastern state of the United States. Grades 1 to 3 are an important period when students are rapidly developing their decoding skills and associated reading prosody as well as language skills (e.g., listening comprehension). Data were collected in the fall and spring of each year, totaling six waves. The average age of students was 6.36 years ($SD = 0.53$), 7.33 years ($SD = 0.52$), and 8.34 years ($SD = 0.54$) in the fall of Grades 1, 2, and 3, respectively. Fifty-two percent of students in Grade 1 ($n = 192$), 46% of students in Grade 2 ($n = 172$), and 39% of students in Grade 3 ($n = 146$) qualified for free or reduced lunch, a proxy variable for poverty status. The racial/ethnic breakdown was as follows in Grade 1: 59.8% White, 25.9% Black, 5.9% Hispanic, 2.4% Asian/Pacific Islander, 0.3% American Indian/Alaska Native, and 5.9% identified as two or more races/ethnicities. Slightly less than half of the sample was female ($n = 180$; 48.5%). Human subjects approval was obtained from the Florida State University (HSC No. 2015.16488).

Measures

Reading prosody. Children were presented with three grade-level passages in each wave and were asked to read each passage aloud. After each passage, a simple literal comprehension question (e.g., name of a main character in the story) was asked to ensure that children read for meaning. Passages were normed in the state where the study was conducted prior to the study, and they were composed of 155 to 198 words in Grade 1, 187 to 200 words in Grade 2, and 200 to 307 words in Grade 3. One passage in each wave was used as a linking passage between waves (e.g., one passage between Waves 1 and 2, another passage between Waves 2 and 3, etc.). Students' oral reading was digitally recorded (i.e., saved as a *.wav file).

Reading prosody was measured by spectrographic analysis and a rating scale. Spectrographic analyses were informed by previous studies (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel et al., 2004) and included the following five indicators: (a) vocalic nucleus, (b) sentence-final change in F_0, (c) intersentential pause duration (in ms), (d) frequency of ungrammatical pauses, and (e) total pause frequency. Vocalic nucleus is a measure of intonation contour in hertz (pattern of pitch changes in the voice; Miller & Schwanenflugel, 2008). Sentence-final change in F_0 is the difference in hertz from the final pitch peak to final F_0. There were three interrogative sentences (i.e., sentence 1 for passage 3 in Wave 1 [Have you seen a rainbow?], sentence 2 for passage 3 in Wave 5 [Do you know what that means?], and sentence 1 for passage 3 in Wave 6 [Are you ready for a float trip in a canoe?]), which had positive F_0 change. The positive F_0 change values from these sentences were multiplied by -1 to follow the same distributions as F_0 change values from declarative sentences where F_0 decreases (i.e., intonation goes down). Following previous work, durations longer than 100 ms between words or phrases were considered pauses and were measured by visually marking the spectrograph because one hundred milliseconds is considered the minimum pause length that can be reliably measured (Arcand et al., 2014; Miller & Schwanenflugel, 2006, 2008). Ungrammatical pauses were inappropriate pauses that did not fit into major syntactic boundaries (e.g., clause boundaries) or reasonable phrasal boundaries where a pause would be expected (see Benjamin & Schwanenflugel, 2010). Praat software (Version 5.4; Boersma & Weenink, 2015) was used to measure each of these indicators. For the spectrographic analysis, the first three sentences in the oral reading of each passage were used because of the resource-intensive nature of the coding and the large amount of data (see Miller & Schwanenflugel, 2006; Schwanenflugel et al., 2004, for a similar approach). One graduate student and three undergraduate students in a speech and language pathology program underwent rigorous training and coded data using Praat. Similarity reliability coefficients (Shrout & Fleiss, 1979), which indicated the proximity of the coder's score for each variable to that of the primary coder, ranged from .90 to .99, using 78 cases.

Students' reading prosody was also rated by a widely used scale, the Multi-Dimensional Fluency Scale (MFS; Rasinski, 2004; Rasinski, Homan, & Biggs, 2009; Zutell & Rasinski, 1991). The MFS assesses reading prosody in four areas: (a) expression and volume, (b) phrasing, (c) smoothness, and (d) pace. Expression evaluates the extent to which students' reading is similar to natural language with adequate expression and volume. Phrasing is related to marking clause and sentence units. Smoothness is the extent to which students easily resolve word and structure difficulties. Pace rates conversational pace, whereby a high rating indicates that the students read neither too fast nor too slow (see Rasinski, 2004; Zutell & Rasinski, 1991, for the scale). Each aspect was rated on a scale of 1 to 4 (1 for *not fluent reading* to 4 for *fluent reading*). Three individuals (one doctoral student in education, one individual with a master's degree in education, and one individual with a bachelor's degree in speech and language pathology) were rigorously trained with exact percent agreement ranging from .80 to .90, using 72 cases.

Listening comprehension. Listening comprehension was measured by two normed tasks, the Oral Comprehension test of the Woodcock-Johnson III Tests of Achievement (henceforth WJOC; Woodcock, McGrew, & Mather, 2001) and the Listening Comprehension subscale of the Oral and Written Language Scales, second edition (OWLS henceforth; Carrow-Woolfolk, 2011). In WJOC, the child heard a sentence or sentences and was asked to supply a missing word such as nouns, adjectives, and verbs (e.g., People sit in _____.; Eduardo is very attentive to the flower seeds that he plants in his garden every year. He tries to make sure all of the conditions are right so that everything will _______.). In OWLS, an assessor read increasingly difficult words, phrases, and sentences aloud, and the student responded by indicating which one of four pictures best depicted the stimuli (e.g., The sheep eats the grass; Although Bill Smith had said, "I won't take any dogs with long ears," the opposite situation actually occurred.). Cronbach's alpha estimates ranged from .74 to .79 in WJOC and .91 to .93 in OWLS.

Word reading. Word reading was assessed with three tasks: the Letter-Word Identification (LWID) task of the Woodcock-Johnson Tests of Achievement (Woodcock et al., 2001), the Word Reading subtask of the Wechsler Individual Achievement Test, third edition (WIAT; Wechsler, 2009), and the Sight Word Efficiency (SWE) subtask of the Test of Word Reading Efficiency, second edition (Torgesen, Wagner, & Rashotte, 2012). In LWID and WIAT, the child was asked to read aloud words of increasing difficulty. SWE measured a student's ability to read as many sight words of increasing difficulty as accurately as possible within a short time limit (45 s). Cronbach's alpha estimates were .91 to .92 in LWID and .95 in WIAT at each wave of data. Test–retest reliability for SWE was reported to be .93 (Torgesen et al., 2012).

Procedure

Students were individually assessed in several sessions of 30–40 min per session by rigorously trained research assistants in quiet spaces in participating schools.

Data Analytic Strategy

Confirmatory factor analysis (CFA) and structural equation modeling were primary analytic strategies, using Mplus Version 8.3 (Muthén & Muthén, 1998–2018). Because of the censored nature of the pause variables (both duration and frequency were censored from below, i.e., had values close to zero), all models were fit using maximum likelihood estimation with robust standard errors (MLR) in Mplus.

Research question 1: Dimensionality models of reading prosody. The dimensionality of the prosody indicators was determined using CFA. Performances on pause frequency and ungrammatical pause frequency were combined and averaged because of extremely high correlations (see the Results section). Thus, students' performance on four prosody indicators from spectrogram data (intonation contour [i.e., vocalic nuclei], F_0 change, pause frequency, and pause duration) and four indicators from the rating scale (expression and volume, phrasing, pacing, and smoothness) were averaged across passages per wave. These eight indicators were used in five CFA models that were fit to the data per wave. The five models were informed by theory and prior evidence. For example, a unidimensional model was tested as the baseline model where reading prosody is described as a single construct that is measured with various specific indicators or aspects (see Figure 1a). Alternatively, previous studies suggested that pitch and pause structure may be related but differentiated constructs (see Benjamin et al., 2013, and the literature review above). However, the factor structure of the four aspects in the Multi-Dimensional Fluency Scale, and their relations with other prosody indicators are unknown (see Figure 1b). It is also possible that various aspects of reading prosody have a bifactor structure with a general dimension or factor that captures commonality across all prosody indicators and residual specific factors (e.g., pitch, pause, or ratings; see Figure 1c). Variations of these models are also plausible (see Figure 1d and 1e). Below are detailed descriptions of each model.

The first model was a unidimensional model (see Figure 1a), where a reading prosody factor was indicated by the four variables from the spectrograph data (pause frequency, pause duration, F_0 change, and intonation contour) and the four variables from the rating scale (expression and volume, phrasing, pacing, and smoothness). The second model was a correlated three-factor model (see Figure 1b), where a ratings factor was indicated by expression and volume, pacing, phrasing, and smoothness; a second pause factor was indicated by pause frequency and pause duration; and a third pitch factor was indicated by intonation contour and F_0 change. Third, a bifactor model was estimated (see Figure 1c, bifactor 1), where the factors from the correlated three-factor model (ratings, pause, and pitch, which are called specific factors in the bifactor model) and a general factor orthogonal to these factors (i.e., the general factor was not allowed to covary with the specific factors) were indicated by the eight variables (i.e., the four variables for spectrograph and the four variables for the rating scale), and the specific factors were not allowed to be correlated. In a bifactor model (Gibbons & Hedeker, 1992), the general factor (in this case, reading prosody) captures common variance across all the manifest variables or indicators, and thus, it theoretically captures the most reliable portion of the variance for each of the indicators. The specific factors, orthogonal to the general factor and to each other, help to explain item response variance (item residual variance) that is not captured by the general factor (e.g., method variance). Bifactor 2 (see Figure 1d) was another bifactor model, which was identical to the bifactor 1 model, but the specific factors (i.e., ratings, pause, and pitch) were allowed to correlate.

Finally, we fit a correlated trait bifactor model (see Figure 1e). In this model, there was a Prosody: Ratings and Pause general factor indicated by the four rating scale indicators (i.e., smoothness, pacing, phrasing, expression and volume) and two pause variables (pause frequency and pause duration). In addition, there was a separate Prosody: Pitch factor, indicated by F_0 change and intonation contour. The Prosody: Ratings and Pause general factor remained orthogonal to the specific factors of Ratings and Pause, but the correlation between Ratings and Pause was allowed to be estimated. Prosody: Pitch was allowed to correlate with the general factor, Prosody: Ratings and Pause, and with the Ratings and Pause specific factors. This alternative model was informed by previous evidence and preliminary analysis in the present study. Pitch and pause variables measured by spectrographic analysis were related but different factors when using exploratory factor analysis (Benjamin et al., 2013), and pitch variables have weak relations with pause variables (Binder et al., 2013; Schwanenflugel et al., 2004). Furthermore, our preliminary analysis revealed that the pitch variables, intonation contour and F_0 change, had little to no loadings on the general prosody factor in the Figures 1c and 1d models in all the waves.

Research question 2: The relations of word reading and listening comprehension to reading prosody. After establishing the dimensionality of the reading prosody variables, the relations of word reading and listening comprehension to the reading prosody factors were investigated using the structural equation model in Figure 3. A latent variable for word reading was created from the three normed word reading measures (i.e., LWID, WIAT, and SWE), and a latent variable for listening comprehension was created from the two normed measures (i.e., WJOC and OWLS). Children's demographic variables were not included in the analysis because none of them were statistically significant after accounting for word reading and listening comprehension.

Multiple criteria were used to determine model fit. The confirmatory fit index (CFI) and the Tucker-Lewis index (TLI) were used, whereby values above .90 are considered adequate and values above .95 are considered excellent (Hu & Bentler, 1999). We also used the root mean-squared error of approximation (RMSEA) and its associated confidence interval, where values under .08 are preferred (Kline, 2016). For determining the best fitting nested models, the difference in the Satorra-Bentler chi-square tests of model fit were used, whereby the preferred model was one in which the difference in Satorra-Bentler chi-square estimates was significantly better, or significantly closer to zero (Muthén & Muthén, 1998–2018). According to the nested model test available in Mplus 8.3 (Muthén & Muthén, 1998–2018) and using the criteria outlined by Asparouhov and Muthén (2019), the five tested models were nested such that the unidimensional model (H_0) was nested in the correlated three-factor model (H_1); the correlated three-factor model was nested in the bifactor Model 1; the bifactor Model 1 was nested in the bifactor Model 2; and the correlated trait bifactor model was nested in bifactor Model 2. In certain cases, if the assumption of nested models was not met (e.g., if the nested model [H_0] has degrees of freedom fewer than or equivalent to those of the comparison model [H_1]; see Results section for examples of such models), we looked at the differences in the sample-size adjusted Bayesian information criterion (*n*BIC) between compared models. Models with *n*BIC values closer to negative infinity were preferred, where a difference of 5 was considered *strong* evidence for a better fitting model and a difference of 10 was considered *very strong* evidence for a better fitting model (Raftery, 1995).

Results

Descriptive Statistics

The descriptive statistics for each reading prosody indicator separated by wave are reported in Table 1 (descriptive statistics by passage across waves are found in the online supplemental materials), and the descriptive statistics for word reading and listening comprehension variables by wave are presented in Table 2. Sample sizes were $n = 350$ in Wave 1, $n = 366$ in Wave 2, $n = 326$ in Wave 3, $n = 329$ in Wave 4, $n = 311$ in Wave 5, and $n = 300$ in Wave 6. The number of students with missing data in each wave were as follows: $n = 21$ in Wave 1, $n = 5$ in Wave 2, $n = 45$ in Wave 3, $n = 42$ in Wave 4, $n = 60$ in Wave 5, and $n = 71$ in Wave 6. As shown in Table 1, there were differing rates of missingness for each of the measured variables. The differing rates of missingness were mostly due to technical difficulties (e.g., digital recorder malfunction). For example, most missing data occurred in estimating pause durations between sentences and measuring F_0 change across sentences, where these values were more likely to be missing for passage 2 and passage 3 for Wave 1 and Wave 2 (see Table S1 in the online supplemental materials for more information). In addition, some students had creaky voice (or glottal fry), and they were not included in the reading prosody coding. Based on chi-square difference tests, there were no differences on any of the demographic variables between students who had missing data and those who did not ($ps > .1$), with one exception in Wave 1: Students who qualified for the free and reduced lunch program

Table 1
Descriptive Statistics for the Prosody Variables (Spectrograph and Rating Scale) by Wave, Averaged Across Passages

Variable	Wave 1					Wave 2					Wave 3					Wave 4					Wave 5					Wave 6				
	n	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*
Spectrograph																														
F_0 change	259	−158.10	−9.30	−62.97	24.55	333	−141.40	−19.63	−61.83	21.99	319	−127.90	−1.65	−63.58	21.87	317	−111.60	−9.30	−53.96	21.12	302	−131.50	−15.20	−60.33	23.71	284	−102.90	−3.75	−44.51	19.22
Intonation contour	263	199.40	319.80	258.5	27.66	337	183.60	331.040	258.19	29.42	320	190.90	321.490	254.61	28.81	318	191.280	320.610	251.15	27.79	302	180.310	316.140	247.2	30.17	286	185.20	308.150	245.07	26.9
Pause duration	258	0.140	10.80	1.370	1.29	335	0.210	3.060	0.86	0.46	319	0.160	2.080	0.7	0.33	316	0.130	1.840	0.67	0.28	300	0.210	1.620	0.58	0.22	284	0.080	1.330	0.58	0.21
Pause frequency (Avg.)	264	1	31	12.83	7.86	335	1	21	9.51	4.86	320	1.330	26	11.54	6.35	317	0	20	9.35	4.79	301	0.330	13.250	5.11	3.18	285	0	13.50	4.45	3.14
Rating scale																														
Smoothness	259	1	3.670	1.73	0.79	327	1	4	1.8	0.68	307	1	4	1.96	0.74	312	1	4	2.34	0.77	276	1	4	2.73	0.84	270	1	4	2.89	0.73
Pacing	258	1	4	1.73	0.79	327	1	4	1.77	0.73	307	1	4	1.98	0.83	312	1	4	2.39	0.9	276	1	4	2.79	0.82	270	1	4	3.04	0.75
Phrasing	259	1	4	1.78	0.79	327	1	4	1.85	0.73	307	1	4	2.08	0.81	312	1	4	2.49	0.8	276	1	4	2.93	0.76	270	1	4	3.08	0.7
Expression and volume	259	1	4	1.83	0.78	328	1	4	2.05	0.73	307	1	4	2.32	0.74	312	1	4	2.68	0.8	276	1	4	2.97	0.67	270	1	4	3.1	0.57

Note. Total sample sizes per wave were as follows: $n = 350$ in Wave 1, $n = 366$ in Wave 2, $n = 326$ in Wave 3, $n = 329$ in Wave 4, $n = 311$ in Wave 5, and $n = 300$ in Wave 6. F_0 = fundamental frequency. Pause duration is reported in seconds.

Table 2
Descriptive Statistics for the Word Reading and Listening Comprehension Variables by Wave

Variable	Wave 1				Wave 2				Wave 3				Wave 4				Wave 5				Wave 6			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
Listening comprehension																								
WJOC – Raw	2	21	13.03	3.64	8	24	16.18	3.67	6	26	16.31	3.55	9	28	19.16	3.69	10	28	18.89	3.77	13	29	21.68	3.60
WJOC – SS	56	134	103.79	13.09	73	148	112.26	13.40	71	140	106.02	12.67	71	146	113.25	12.69	58	153	106.23	13.53	71	136	113.24	12.03
OWLS – Raw	40	95	69.80	12.44	49	105	78.40	12.68	46	111	83.81	11.60	56	111	89.47	11.18	67	116	92.80	11.19	65	117	96.91	10.17
OWLS – SS	50	135	100.64	14.08	62	133	105.39	13.82	65	135	106.66	13.06	55	137	109.33	13.70	57	135	108.77	13.41	57	138	110.21	13.67
Word reading																								
LWID – Raw	17	57	32.49	6.75	25	53	38.46	6.39	24	62	41.84	6.60	30	61	45.59	6.47	33	63	48.21	6.73	34	65	50.67	6.44
LWID – SS	68	147	110.99	13.54	75	142	112.47	12.89	73	135	108.48	12.07	69	133	108.24	11.85	49	133	105.54	12.14	59	135	105.69	11.67
WIAT – Raw	1	40	14.10	9.42	2	45	22.91	9.93	5	56	28.00	10.51	8	63	33.40	9.84	13	60	37.41	10.70	17	63	41.40	10.67
WIAT – SS	60	147	100.15	16.10	65	148	104.80	16.04	63	143	103.27	15.33	63	141	105.15	14.33	59	138	103.70	14.90	62	139	104.30	14.88
SWE – Raw	4	67	31.72	14.98	11	73	43.69	14.25	15	80	51.68	13.44	21	83	58.50	12.03	16	86	61.69	11.82	27	90	66.21	11.53
SWE – SS	55	144	100.60	17.59	55	144	105.01	16.79	55	137	103.86	16.57	55	138	104.74	16.24	55	136	101.15	15.93	55	134	101.41	15.49

Note. Total sample sizes per wave were as follows: n = 350 in Wave 1, n = 366 in Wave 2, n = 326 in Wave 3, n = 329 in Wave 4, n = 311 in Wave 5, and n = 300 in Wave 6. Min = minimum; Max = maximum; Raw = raw score; SS = standard score; WJOC = Woodcock-Johnson III oral comprehension; OWLS = Oral and Writing Language Scales; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Tests; SWE = Sight Word Efficiency.

were more likely to be missing than students who did not qualify for free and reduced lunch, $\chi^2(1) = 5.03$, $p = .025$, Cramer's $V = .12$. Complete information on these chi-square tests can be found in the online supplemental materials (Table S2). Little's test of missing data, which tests whether data can be considered missing completely at random (MCAR), failed to establish MCAR within the data for Waves 1 to 4 (Wave 1, $p < .001$; Wave 2, $p < .001$; Wave 3, $p = .035$; Wave 4, $p = .007$), but the data could be considered MCAR at Wave 5, $\chi^2(3750) = 3719.23$, $p = .636$, and at Wave 6, $\chi^2(3224) = 3206.64$, $p = .583$. Missing data in Waves 1 to 4 were mostly attributable to issues with the spectrographic measurement as mentioned above, so we considered these missing data to be missing at random (MAR).

Beginning with the spectrogram data, intonation contour remained similar across the waves although average intonation contour was highest in Wave 1 (258.5) and lowest in Wave 6 (245.07). At all waves, the means for F_0 change were negative, indicating that students, on average, decreased their pitch from the peak to the end of the sentence. The frequency of pauses within a sentence and the duration of pauses between sentences decreased over time. Pause frequencies and durations decreased from an average of 12.8 pauses within a sentence and 1.37 s between sentences in Wave 1 to an average of 4.45 pauses within a sentence and 0.58 s between sentences in Wave 6. Total pause frequencies and ungrammatical pause frequencies were nearly perfectly correlated ($.98 \leq r\text{s} \leq .99$); therefore, we averaged the frequencies of ungrammatical pauses and total pauses across the three passages per wave to get an average pause frequency metric seen in Table 1. Average scores on the rating scale increased over time, such that students read with better pacing, smoothness, phrasing, and expression and volume from the fall of Grade 1 (Wave 1: average rating 1.73–1.83) to the spring of Grade 3 (Wave 6: average rating 2.89–3.10).

Students' mean performances on word reading and listening comprehension were in the average and somewhat high average ranges. For listening comprehension tasks, mean standard scores ranged from 100.64 in OWLS at Wave 1 to 113.25 in WJOC at Wave 4. For the word reading tasks, mean standard scores ranged from 100.15 in WIAT at Wave 1 to 112.47 in LWID at Wave 2.

Because of space constraints, the 13×13 correlation matrices per wave across the included tasks are presented in tables in Appendix A. Among the spectrogram variables, intonation contour was significantly and negatively related to F_0 change ($-.46 \leq r\text{s} \leq -.35$) such that those with greater intonation contour made greater changes in F_0. F_0 change was weakly related with pause frequency ($.10 \leq r\text{s} \leq .27$). There were practically no relations of intonation contour with pause frequency ($-.11 \leq r\text{s} \leq .01$) and pause duration ($.00 \leq r\text{s} \leq -.04$). There were moderate correlations between pause duration and pause frequency, which increased over time ($r = .32$ in Wave 1 to $r = .51$ in Wave 6). Among the rating scale indicators, there were strong correlations in each wave ($.73 \leq r\text{s} \leq .92$). F_0 change, pause frequency, and pause duration were negatively and weakly to strongly related with rating scale variables ($-.79 \leq r\text{s} \leq -.10$), whereas intonation contour tended to have no relation or positive but weak relations with rating scale variables ($-.01 \leq r\text{s} \leq .27$).

There were negative correlations between the spectrogram measures and the word reading ($-.79 \leq r\text{s} \leq -.18$) and oral language measures ($-.42 \leq r\text{s} \leq -.04$), with the exception of intonation contour, which was not related to any language or word reading variables at any wave ($-.03 \leq r\text{s} \leq .11$). There were moderate to strong correlations between the rating scale variables and the word reading measures ($.55 \leq r\text{s} \leq .77$), and weak to moderate correlations between the rating scale variables and the oral language measures ($.23 \leq r\text{s} \leq .43$). As expected, correlations among the word reading measures were very strong ($.73 \leq r\text{s} \leq .91$), and correlations between the oral language measures were moderate to strong ($.57 \leq r\text{s} \leq .64$).

Research Question 1: Dimensionality of Reading Prosody Variables

As described above, five alternative models were fit within each wave to determine the dimensionality of the reading prosody indicators (see Figure 1). Note that for ease of interpretation, F_0 change was reverse coded for the dimensionality and predictive models below so that positive F_0 change values indicate greater F_0 change. Table 3 presents model fit indices and model comparisons. Overall, the correlated trait bifactor model (Figure 1e) was selected as the final model for the following reasons. First, for nested models, model fit differences were examined by the difference in the Satorra-Bentler chi-square test. As shown in the last two columns of Table 3, the correlated three-factor model (Figure 1b) was preferred to the unidimensional model (Figure 1a) in all waves. The bifactor 1 model (Figure 1c) was preferred to the correlated three-factor model in Waves 2, 3, and 4. The bifactor 2 model (Figure 1d) was preferred to the bifactor 1 model in Wave 4. The bifactor 2 model was preferred to the correlated three-factor model in Waves 1, 5, and 6. The correlated trait bifactor model (Figure 1e) was preferred to the bifactor 1 model in Wave 3 and to the bifactor 2 model in Wave 6. In Wave 5, the correlated trait bifactor model did not fit significantly differently compared with the bifactor 2 model.

Second, for the models that did not meet the assumption of nested models (i.e., model fit comparisons of the correlated trait bifactor model with the other models in Waves 1, 2, and 4), comparisons were conducted using *n*BIC (see the Data Analytic

Table 3
Model Fit Statistics for the Dimensionality Analyses

Model	χ^2	*df*	SCF	*p*	CFI	TLI	RMSEA	90% CI	*n*BIC	Comparison	Model test
Wave 1											
1. Unidimensional	58.49	20	1.06	<.001	.97	.96	.08	[.06, .11]	8498.07		
2. Correlated three-factor	31.17	18	.97	.03	.99	.98	.05	[.02, .08]	8471.18	2 vs. 1	$\Delta\chi^2(2) = 16.99^{***}$
3. Bifactor 1	27.93	15	.88	.02	.99	.98	.06	[.02, .08]	8472.66	3 vs. 2	$\Delta\chi^2(3) = 3.98^{ns}$
4. Bifactor 2	10.93	11	.85	.45	1.00	1.00	.00	[.00, .06]	8467.16	4 vs. 2	$\Delta\chi^2(7) = 19.78^{**}$
5. Correlated trait bifactor	***13.12***	***11***	***.67***	***.29***	***1.00***	***1.00***	***.03***	***[.00, .07]***	***8466.69***	***5 vs. 4***	***ΔnBIC = .47***[a]
Wave 2											
1. Unidimensional	90.41	20	.97	<.001	.96	.94	.10	[.81, .12]	9770.32		
2. Correlated three-factor	41.45	18	.97	.001	.99	.98	.06	[.04, .09]	9728.12	2 vs. 1	$\Delta\chi^2(2) = 48.96^{***}$
3. Bifactor 1	27.15	15	.94	.03	.99	.99	.05	[.02, .08]	9721.51	3 vs. 2	$\Delta\chi^2(3) = 13.11^{**}$
4. Bifactor 2	42.69	11	.65	<.001	.98	.95	.09	[.06, .12]	9734.42	4 vs. 3	$\Delta\chi^2(4) = -1.23^{ns}$
5. Correlated trait bifactor	***21.97***	***10***	***.81***	***.02***	***.99***	***.98***	***.06***	***[.03, .09]***	***9726.91***	***5 vs. 3***	***ΔnBIC = 5.4***[a]
Wave 3											
1. Unidimensional	127.30	20	1.02	<.001	.94	.91	.13	[.11, .15]	9386.87		
2. Correlated three-factor	69.34	18	1.01	<.001	.97	.95	.09	[.07, .12]	9332.35	2 vs. 1	$\Delta\chi^2(2) = 53.86^{***}$
3. Bifactor 1	55.56	15	.98	<.001	.98	.96	.09	[.07, .12]	9324.41	3 vs. 2	$\Delta\chi^2(3) = 13.44^{**}$
4. Bifactor 2	48.74	11	.90	<.001	.98	.94	.10	[.08, .13]	9324.54	4 vs. 3	$\Delta\chi^2(4) = 8.82^{ns}$
5. Correlated trait bifactor	***19.65***	***12***	***.97***	***.07***	***1.00***	***.99***	***.04***	***[.00, .08]***	***9297.10***	***5 vs. 3***	$\boldsymbol{\Delta\chi^2(3) = 14.26^{**}}$
Wave 4											
1. Unidimensional	124.74	20	.98	<.001	.94	.92	.13	[.11, .15]	9057.20		
2. Correlated three-factor	68.80	18	.99	<.001	.97	.96	.09	[.07, .12]	9008.14	2 vs. 1	$\Delta\chi^2(2) = 60.82^{***}$
3. Bifactor 1	54.11	15	.95	<.001	.98	.96	.09	[.07, .12]	8999.55	3 vs. 2	$\Delta\chi^2(3) = 14.04^{**}$
4. Bifactor 2	36.61	11	.82	<.001	.99	.96	.09	[.06, .12]	8988.51	4 vs. 3	$\Delta\chi^2(4) = 16.36^{**}$
5. Correlated trait bifactor	***28.98***	***10***	***.99***	***.001***	***.99***	***.97***	***.08***	***[.05, .11]***	***8989.65***	***5 vs. 4***	***ΔnBIC = 1.14***[a]
Wave 5											
1. Unidimensional	147.44	20	1.03	<.001	.92	.88	.15	[.12, .17]	8309.38		
2. Correlated three-factor	72.01	18	1.05	<.001	.96	.95	.10	[.08, .12]	8238.00	2 vs. 1	$\Delta\chi^2(2) = 89.71^{***}$
3. Bifactor 1	67.31	15	1.00	<.001	.97	.94	.11	[.08, .13]	8237.09	3 vs. 2	$\Delta\chi^2(3) = 6.38^{ns}$
4. Bifactor 2	19.96	11	.98	.046	.99	.99	.05	[.01, .09]	8199.40	4 vs. 2	$\Delta\chi^2(7) = 48.32^{***}$
5. Correlated trait bifactor	***23.72***	***12***	***1.04***	***.02***	***.99***	***.98***	***.06***	***[.02, .09]***	***8201.88***	***5 vs. 4***	$\boldsymbol{\Delta\chi^2(1) = 3.00^{ns}}$
Wave 6											
1. Unidimensional	130.18	20	1.06	<.001	.92	.89	.14	[.12, .16]	7488.27		
2. Correlated three-factor	63.04	18	1.04	<.001	.97	.95	.09	[.07, .12]	7420.91	2 vs. 1	$\Delta\chi^2(2) = 58.41^{***}$
3. Bifactor 1	65.70	15	.95	<.001	.96	.93	.11	[.08, .14]	7425.63	3 vs. 2	$\Delta\chi^2(3) = 2.11^{ns}$
4. Bifactor 2	31.55	11	1.06	<.001	.99	.96	.08	[.05, .11]	7406.33	4 vs. 2	$\Delta\chi^2(7) = 31.85^{***}$
5. Correlated trait bifactor	***8.42***	***12***	***.97***	***.75***	***1.00***	***1.00***	***.00***	***[.00, .04]***	***7378.58***	***5 vs. 4***	$\boldsymbol{\Delta\chi^2(1) = 7.10^{**}}$

Note. SCF = scaling correction factor used for Satorra-Bentler chi-square tests; CFI = confirmatory fit index; TLI = Tucker-Lewis index; RMSEA = root mean squared error of approximation; *n*BIC = sample size–adjusted Bayesian information criterion; ns = not significant. Bold and italicized model is the best fitting model in each wave.
[a] Not a practically important difference.
** $p < .01$. *** $p < .001$.

Plan above). The correlated trait bifactor model did not fit practically differently compared with the bifactor 2 model in Wave 1 (ΔnBIC = 0.47) and Wave 4 (ΔnBIC = 1.14), and the correlated trait bifactor model did not fit practically differently compared to the bifactor 1 model in Wave 2 (ΔnBIC = 5.4).

Finally, both F_0 change and intonation contour did not load significantly onto the Prosody factor in the unidimensional model (Figure 1a) nor onto the general prosody factor in bifactor models in the bifactor 1 model (Figure 1c) and the bifactor 2 model (Figure 1d). Based on the model fit comparisons and the results of loadings, the correlated trait bifactor model (Figure 1e) was selected as the final model. Figure 2 shows the final model results for the six waves, which are detailed below.

Beginning with the Prosody: Ratings and Pause general factor, the loadings of the rating scale indicators were large and positive ($.85 \leq \lambda \leq .96$), the loadings for pause frequency were large and negative ($-.85 \leq \lambda \leq -.68$), and the loadings for pause duration were negative and moderate in strength ($-.54 \leq \lambda \leq -.42$). The negative loadings of the pause frequency and pause duration variables indicate that the Prosody: Ratings and Pause general factor captures higher scores in the rating scale variables and lower frequencies and shorter durations of pause. To estimate reliability, factor reliability was calculated using McDonald's coefficient omega (ω; McDonald, 1999), which is shown to be an appropriate measure of factor reliability when data are modeled in a bifactor structure (Reise, 2012). Prosody: Ratings and Pause was a reliable factor in all waves (McDonald's ω range = .75–.78). The pattern of consistent loadings and high reliability indicates that the Prosody: Ratings and Pause general factor captured variance common to both the rating scale indicators and pause indicators.

The Prosody: Pitch factor had similar patterns of loadings across all waves. F_0 change loaded positively and strongly ($.72 \leq \lambda \leq .97$), and intonation contour loaded positively and moderately ($.37 \leq \lambda \leq .50$). Although the pattern of loadings was consistent, this factor was not consistently reliable (range across waves: McDonald's ω = .54–.72; Wave 5 was the only wave to reach a McDonald's omega criterion of at least .70). The correlation between Prosody: Pitch and the Prosody: Ratings and Pause general factor was positive, and the magnitude increased from weak to moderate over time: .16 at Wave 1 ($p = .282$), .27 at Wave 2 ($p < .001$), .30 at Wave 3 ($p < .001$), .29 at Wave 4 ($p < .001$), .32 at Wave 5 ($p < .001$), and .37 at Wave 6 ($p < .001$). The correlation between Prosody: Pitch and the Ratings specific factor was significant and positive in Waves 3 through 6 ($.16 \leq rs \leq .91$). Prosody: Pitch was significantly and positively related to the Pause specific factor in only Wave 3, $r = .18$, $p = .023$.

Turning to loadings for specific factors, the patterns of loadings for the Ratings specific factor was not consistent (i.e., the loading pattern was not consistent over time). The only consistent loading was expression and volume, which loaded significantly at every wave except for Wave 2 (Wave 1 $\lambda = -.38$, $p = .001$; Wave 3 $\lambda = .46$, $p < .001$; Wave 4 $\lambda = .47$, $p < .001$; Wave 5 $\lambda = .13$, $p = .02$; Wave 6 $\lambda = .47$, $p < .001$). There were no remaining consistent indicators of the Ratings specific factor. The Ratings specific factor was also unreliable at all six waves (range: McDonald's ω = .00–.21). The pattern of inconsistent loadings and unreliability in the Ratings specific factor across the six waves of data indicates that there was no additional construct-relevant variance that could be accounted for beyond what was captured in the Prosody: Ratings and Pause general factor.

The Pause specific factor was not a reliable factor in any wave (range: McDonald's ω = .04–.47). Pause duration was not significant in Wave 1 ($p = .19$), loaded significantly but weakly in Wave 2 ($\lambda = .28$, $p = .036$), and loaded strongly and significantly in Waves 3 to 6 ($.85 \leq \lambda \leq .88$, $ps < .001$). The correlation between the Pause- and Ratings specific factors was not estimable in Wave 1 because of a model nonconvergence issue and was fixed to zero; the correlation was significant and negative in Wave 3, $r = -.17$, $p < .01$ and was not significantly different from zero in the remaining waves ($ps > .05$).

In summary, the correlated trait bifactor model, where the ratings and pause indicators had a bifactor structure and the pitch indicators were captured by another factor (Figure 1e), described the data best across all the waves. As stated above, the general factor, Prosody: Ratings and Pause, was the most reliable factor and captured common variance across the six ratings and pause indicators (expression, phrasing, smoothness, pace, pause frequency, pause duration). There was a trend that the Pause specific factor accounted for an additional portion of variance in the pause indicators over and above what was captured in the Prosody: Ratings and Pause general factor, but the Pause specific factor was not reliable in any wave. The Ratings specific factor was also unreliable across the waves over and above the Prosody: Ratings and Pause general factor.

Research Question 2: The Relations of Word Reading and Listening Comprehension to the Identified Dimensions of Reading Prosody

For the CFA models including listening comprehension and word reading (see Appendices B, C, and D for factor loadings, model fit statistics, and factor correlations, respectively), the loadings were significant and strong to very strong for the word reading measures ($.77 \leq \lambda s \leq .98$, $ps < .001$) and were strong to very strong for the listening comprehension measures ($.69 \leq \lambda s \leq .84$, $ps < .001$). The correlations between word reading and listening comprehension were moderate ($.52 \leq rs \leq .67$, $ps < .001$). Word reading was positively and strongly correlated with the Prosody: Ratings and Pause general factor ($.80 \leq rs \leq .87$, $ps < .001$), was positively and weakly to moderately correlated with the Prosody: Pitch factor ($.19 \leq rs \leq .34$, $ps < .001$), and was negatively and weakly to moderately correlated with the Pause specific factor ($-.35 \leq rs \leq -.22$, $ps < .003$). Word reading was not related to the Ratings specific factor ($ps > .27$). Listening comprehension was moderately and positively correlated with the Prosody: Ratings and Pause general factor ($.46 \leq rs \leq .55$, $ps < .001$). Listening comprehension was not related to the Ratings specific factor ($ps > .11$) or Pause specific factor ($ps > .05$). Listening comprehension was positively and weakly related to the Prosody: Pitch factor in Wave 3, Wave 5, and Wave 6 ($.15 \leq rs \leq .19$, $ps < .05$).

The structural equation model shown in Figure 3 was modified by removing the pathways from word reading and listening comprehension to the specific Ratings factor and the specific Pause factor. The specific Ratings factor was not reliable, nor did it have any significant relations with word reading and listening comprehension factors in bivariate correlations (see the preceding para-

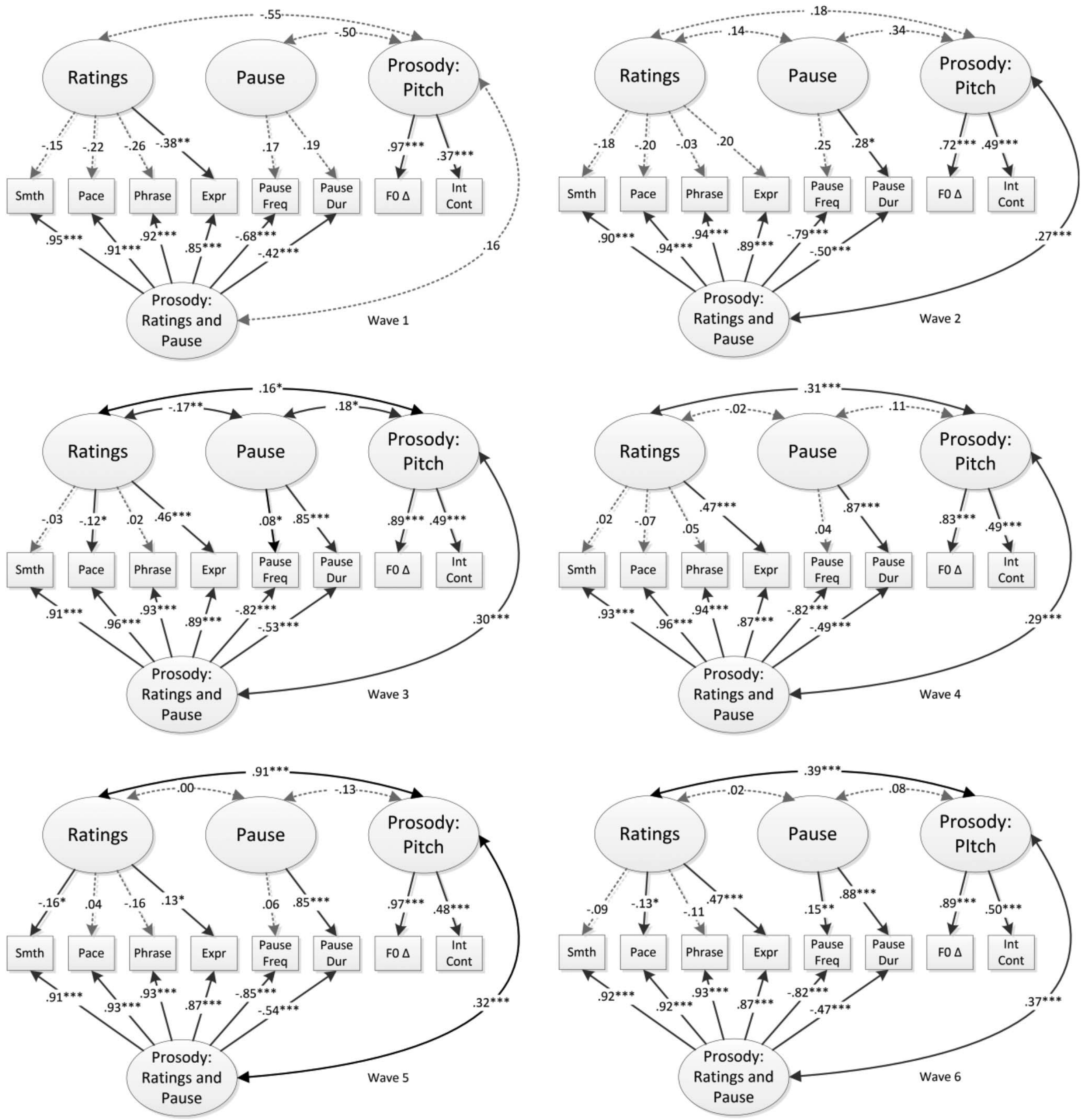

Figure 2. Dimensionality of prosody by wave. Residual variances were estimated, but not shown for figure brevity. Factor variances were fixed at 1 for model identification purposes. Gray, dashed pathways were not statistically significant ($p > .05$). Smth = smoothness; Pace = pacing; Phrase = phrasing; Expr = expression and volume; Pause Freq = pause frequencies; Pause Dur = pause durations; F_0 Δ = fundamental frequency change; Int Cont = intonation contour. * $p < .05$. ** $p < .01$. *** $p < .001$.

graph). The specific Pause factor was related to word reading, but this factor was not reliable. Below, we report results from the modified Figure 3 model, but the results of the original Figure 3 model with all pathways is presented in Appendix E.

The modified Figure 3 model was fitted across each of the six waves of data, using the Prosody: Ratings and Pause factor and the Prosody: Pitch factor as the outcomes. Results using standardized regression weights are presented in Figure 4 (factor loadings from

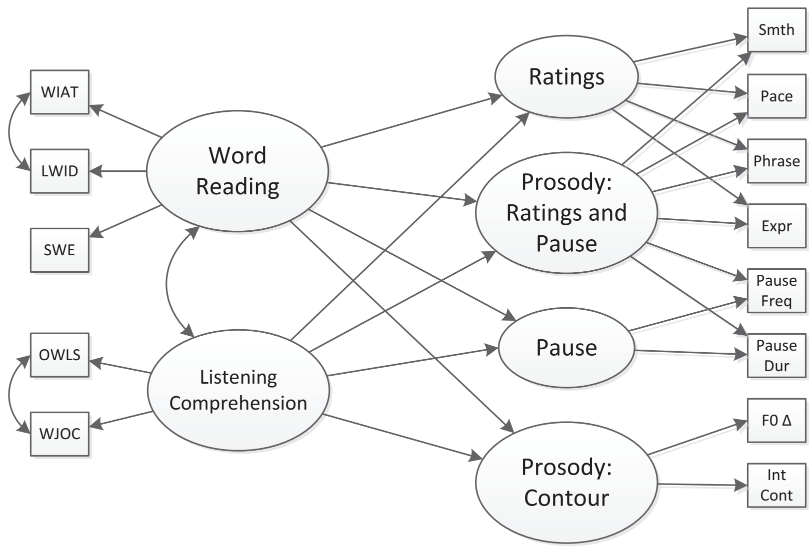

Figure 3. Structural equation model whereby word reading and listening comprehension predict reading prosody general and specific factors. All pathways were estimated, and factor variances were fixed at 1 for model identification purposes. Residual variances for all indicators are not shown for figure brevity. WIAT = Wechsler Individual Achievement Test; LWID = Letter-Word Identification; SWE = Sight Word Efficiency; OWLS = Oral and Written Language Scales; WJOC = Woodcock-Johnson III oral comprehension; Smth = MFS (Multi-Dimensional Fluency Scale) smoothness; Pace = MFS pacing; Phrase = MFS phrasing; Expr = MFS expression and volume; Pause Freq = pause frequencies; Pause Dur = pause durations; F_0 Δ = fundamental frequency change; Int Cont = intonation contour.

the estimated model are presented separately in Table 4). Results at Wave 1, the fall of Grade 1, were as follows. Word reading was strongly related to the Prosody: Ratings and Pause general factor ($\gamma = .83$, $p < .001$). After accounting for word reading, listening comprehension did not explain any additional variance in the Prosody: Ratings and Pause general factor ($\gamma = .03$, $p = .62$). A similar pattern was found for the Prosody: Pitch factor such that word reading moderately and independently predicted Prosody: Pitch ($\gamma = .32$, $p < .001$), but listening comprehension did not independently predict Prosody: Pitch ($\gamma = -.01$, $p = .88$). The correlation between listening comprehension and word reading was moderate, $r = .55$, $p < .001$.

Results in subsequent waves were highly similar to those in Wave 1 (see Figure 4 and Table 4). Word reading was related to the Prosody: Ratings and Pause general factor ($.78 \leq \gamma\text{s} \leq .95$, $p\text{s} < .001$) and the Prosody: Pitch factor ($.21 \leq \gamma\text{s} \leq .38$, $p\text{s} < .01$), whereas listening comprehension was not ($-.10 \leq \gamma\text{s} \leq .08$, $p\text{s} \geq .05$) after accounting for word reading. Across all six waves, 68.3 to 80.6% of the variance was explained in the Prosody: Ratings and Pause general factor (Waves 1 to 6, respectively: 72.0%, 74.9%, 68.3%, 74.4%, 77.9%, 80.6%) and 4.5 to 11.8% of the variance was explained in the Prosody: Pitch factor (Waves 1 to 6, respectively: 9.1%, 4.5%, 5.8%, 5.0%, 9.7%, 11.8%).

Discussion

Text reading fluency or oral reading fluency has been widely studied as an important skill in reading development. Although reading prosody has been recognized as an important part of the text reading fluency construct (Kuhn et al., 2010; National Institute of Child Health and Human Development, 2000), its evidence base is substantially more limited compared with that for text reading efficiency (accuracy and speed). Reading prosody has been examined in multiple aspects, measured using either spectrogram or rating scales, which differ in degree of precision and practicality. Spectrographic measurements have strengths such as precise estimation, but wide use of them in the classroom setting is limited because of the specific expertise required to use them and the time-intensive nature in analyzing the data. Rating scales, on the other hand, are more classroom and teacher friendly, but their precision is not comparable with spectrographic measurements. Although multiple aspects of reading prosody have been widely examined using both approaches, no previous studies have investigated the dimensionality of a comprehensive set of reading prosody indicators. To address this gap, we investigated reading prosody in terms of its dimensionality, its relations with word reading and listening comprehension, and the developmental nature of these dimensions and relations, using longitudinal data from Grade 1 to Grade 3.

Dimensionality of Reading Prosody

The findings of the present study advance our understanding of measurement and dimensionality of reading prosody in three important ways. First, we found that reading prosody is multidimensional. When we examined dimensionality by systematically fitting and comparing five alternative models (see Figure 1), neither the unidimensional model (Figure 1a) nor the three-factor model (Figure 1b) was supported, indicating that the eight reading prosody indicators (expression, phrasing, smoothness, pace, pause du-

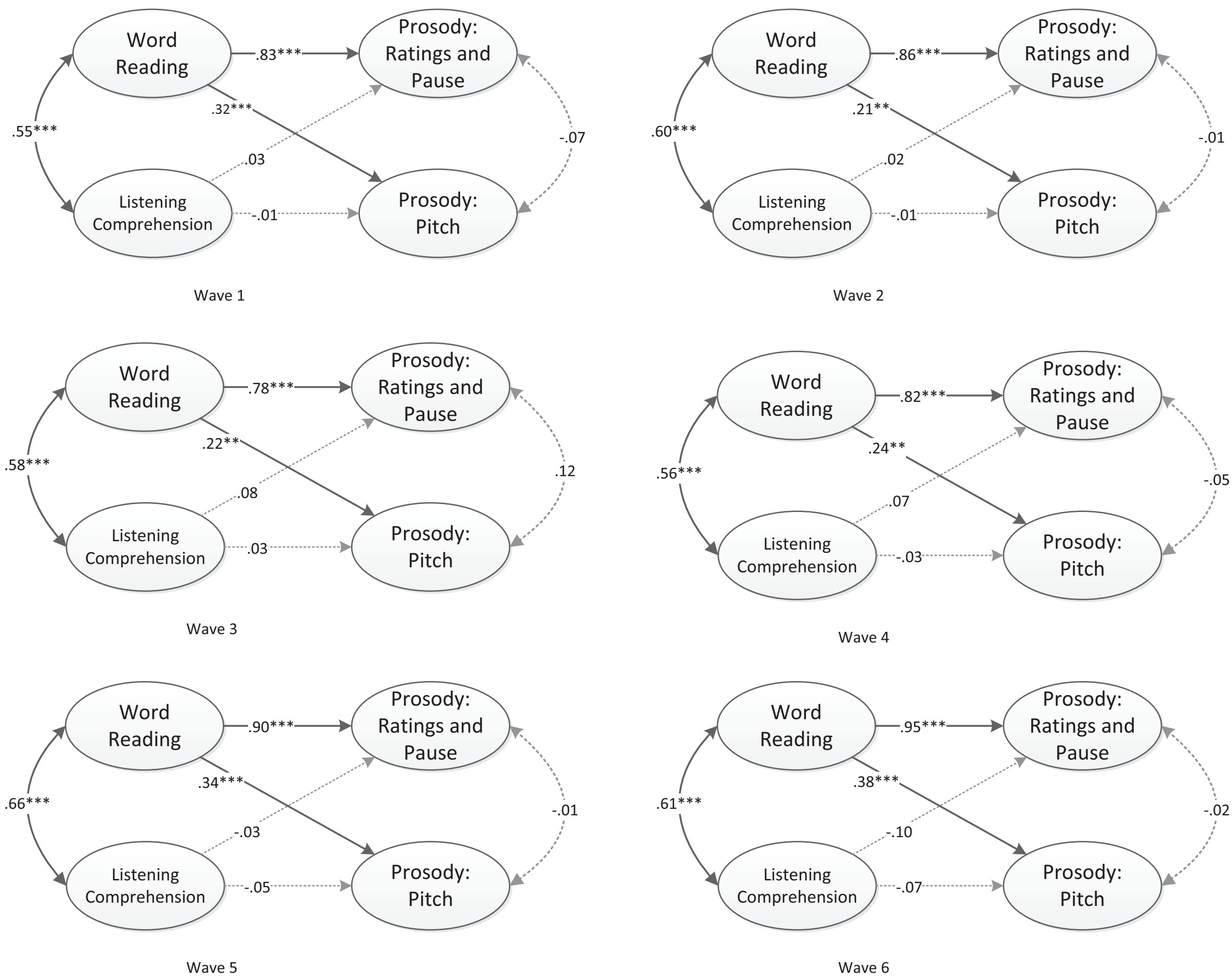

Figure 4. Standardized coefficients of word reading and listening comprehension predicting reading prosody latent variables by wave. Gray, dashed lines are not statistically significant ($p > .05$). Not shown but estimated were measurement models for word reading, listening comprehension, and reading prosody factors and the specific ratings and specific pause factors. $^{**} p < .01$. $^{***} p < .001$.

ration, pause frequency, F_0 change, and intonation contour) do not capture a single underlying dimension or three separate but related dimensions of ratings, pause, and pitch. Three variants of bifactor models (Figures 1c–1e) were also compared, and results revealed that the Figure 1e model, the correlated trait bifactor model, described the data best. That is, the rating scale and pause variables had a bifactor structure composed of a general factor (called Prosody: Ratings and Pause) and a Ratings specific factor and a Pause specific factor. In addition, the pitch-related indicators (intonation contour and F_0 change) did not fit into the bifactor structure, but instead formed a separate but related dimension. The intonation contour and F_0 change variables had very weak or weak relations with pause frequency, pause duration, and the ratings variables (see Appendix A), and consequently did not load on the general factor in the Figures 1c and 1d models. The weak relations are convergent with previous findings (e.g., Binder et al., 2013; Schwanenflugel et al., 2004), but extend previous studies by showing the factor structure of these variables.

Second, our findings showed that the ratings indicators, as measured by MFS, together with pause structure indicators had a bifactor structure. Despite wide use of the MFS, its dimensionality and relations with other prosody indicators remained a black box. Our results revealed that the four indicators of the MFS had moderate relations with pause duration, strong relations with pause frequency, and zero to weak relations with pitch variables[1] (see Appendix A). Moreover, the four indica-

[1] The relations of the Ratings specific factor to the pitch factor varied largely from no statistically significant relations (Wave 1 and Wave 2) to a very strong relation (Wave 5). However, because the Ratings specific factor was mostly unreliable, these results may not be particularly meaningful.

Table 4
Factor Loadings for the Predictive Models, Separated by Wave

Variable	Wave 1			Wave 2			Wave 3			Wave 4			Wave 5			Wave 6		
	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*
Prosody: Ratings and pause																		
Smoothness	.96	.01	<.001	.92	.01	<.001	.91	.02	<.001	.93	.02	<.001	.90	.04	<.001	.92	.13	<.001
Pacing	.94	.01	<.001	.95	.01	<.001	.96	.04	<.001	.97	.02	<.001	.92	.03	<.001	.93	.10	<.001
Phrasing	.96	.02	<.001	.93	.05	<.001	.93	.01	<.001	.93	.03	<.001	.93	.02	<.001	.93	.12	<.001
Expression/volume	.95	.14	<.001	.83	.24	.001	.89	.13	<.001	.83	.20	<.001	.99	.12	<.001	.60	.79	.443
Pause frequency	−.19	.08	.027	−.36	.07	<.001	−.38	.06	<.001	−.42	.06	<.001	−.45	.09	<.001	−.26	.14	.069
Pause duration	−.37	.07	<.001	−.44	.05	<.001	−.39	.13	.002	−.47	.05	<.001	−.49	.06	<.001	−.27	.11	.013
Ratings																		
Smoothness	−.02	.10	.851	.01	.12	.909	−.06	.08	.448	−.01	.11	.955	−.07	.07	.291	−.03	.30	.918
Pacing	−.02	.11	.841	.03	.12	.811	−.15	.09	.086	.08	.11	.483	−.05	.07	.433	−.01	.29	.980
Phrasing	−.08	.11	.472	−.08	.12	.491	−.01	.09	.915	−.03	.11	.758	−.03	.07	.720	−.02	.30	.946
Expression/volume	−.46	.08	<.001	−.51	.12	<.001	.43	.09	<.001	−.48	.12	<.001	.47	.05	<.001	−.63	.42	.132
Pause																		
Pause frequency	.89	.05	<.001	.75	.05	<.001	.61	.11	<.001	.70	.04	<.001	.67	.06	<.001	.80	.11	<.001
Pause duration	.07	.10	.514	.13	.06	.036	.21	.15	.146	.06	.06	.353	.10	.07	.140	.30	.10	.002
Prosody: Pitch																		
F_0 Change	1.00	—	999[a]	1.00	—	999[a]	1.00	—	999[a]	1.00	—	999[a]	1.00	—	999[a]	1.00	—	999[a]
Intonation contour	.36	.06	<.001	.35	.05	<.001	.43	.05	<.001	.41	.05	<.001	.46	.05	<.001	.44	.05	<.001
Word reading																		
SWE	.96	.01	<.001	.95	.01	<.001	.92	.02	<.001	.91	.02	<.001	.87	.02	<.001	.87	.02	<.001
LWID	.91	.01	<.001	.88	.02	<.001	.88	.02	<.001	.86	.02	<.001	.88	.02	<.001	.86	.02	<.001
WIAT	.94	.01	<.001	.92	.01	<.001	.87	.02	<.001	.88	.02	<.001	.86	.02	<.001	.84	.02	<.001
Listening comprehension																		
OWLS	.81	.05	<.001	.72	.05	<.001	.82	.06	<.001	.81	.05	<.001	.80	.04	<.001	.85	.05	<.001
WJOC	.73	.05	<.001	.78	.05	<.001	.78	.05	<.001	.70	.05	<.001	.80	.04	<.001	.73	.05	<.001

Note. F_0 = fundamental frequency; SWE = Sight Word Efficiency; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Test; OWLS = Oral and Written Language Scales; WJOC = Woodcock-Johnson III oral comprehension.
[a] *p* value of 999 indicates loading was fixed to the listed estimate.

tors had very strong loadings to the general factor, Prosody: Ratings and Pause (see Figure 2). Overall, these results indicate that the four aspects of the MFS, although they evaluate somewhat different aspects—expression and volume, phrasing, smoothness, and pace—are largely measuring what is shared with *pause* structures in reading prosody rather than a pitch aspect, at least in early reading development from Grade 1 to Grade 3. These results are also in line with the very strong relation of word reading skill with the Prosody: Ratings and Pause general dimension (see Figure 4).

Third, our speculation of a developmentally changing nature of relations among identified dimensions was supported. Specifically, our data revealed a steady increase in the magnitude of the relation between the Pitch factor and the Ratings and Pause general factor over time—.16, .27, .30, .29, .32, and .37 in the six time points from the beginning of Grade 1 to the end of Grade 3. This is likely attributed to children's development of word reading skill, which is strongly associated with the Prosody: Ratings and Pause general factor, and consequent decrease of the constraining role of word reading, freeing up cognitive resources for semantic processing (LaBerge & Samuels, 1974) and allowing the increasing relation with the Pitch factor.

The Relations of Word Reading and Listening Comprehension to Reading Prosody

Another striking finding is the consistent relation of word reading to reading prosody across the six time points from Grade 1 to Grade 3 (see Figure 4). Beyond this overall finding, there are a couple of important nuanced patterns and magnitudes of relations that are revealing. First, word reading was consistently, strongly, and positively related to the Ratings and Pause general factor (.78–.95). In other words, higher word reading skill was associated with higher performance on the Prosody: Ratings and Pause general factor, indicating that the Prosody: Ratings and Pause general factor is very strongly influenced by word reading development. Second, it is noteworthy that word reading skill was also positively and weakly to moderately related to the Pitch factor (.21–.38), indicating higher word reading skill was associated with higher performance in pitch (greater F_0 change and greater intonation contour). The relation of word reading to the pitch dimension can be interpreted according to the automaticity theory (LaBerge & Samuels, 1974) and the verbal efficiency theory (Perfetti, 1992)—word reading skill frees up cognitive resources to allow one's attention to semantic processing, which, in turn, permits reading with greater pitch variation. The stronger relation of word reading skill with Prosody: Ratings and Pause general factor compared with with the Pitch factor is in line with prior work which showed strong relations of word reading with pause structure variables than with pitch variables (e.g., Binder et al., 2013; Cowie, Douglas-Cowie, & Wichmann, 2002; Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004, 2015). These results underscore the importance of word reading skill in the pause-related dimension of reading prosody, and also add to the literature by showing that the four aspects examined in MFS are also strongly related to word reading skill.

We hypothesized that listening comprehension would be related to reading prosody, particularly after the very beginning phase of reading development when word reading plays a large constraining role because reading prosody theoretically involves semantic processing (e.g., Kuhn et al., 2010). Listening comprehension had a moderate relation with the Prosody: Ratings and Pause general factor and a weak relation with the Prosody: Pitch factor in bivariate correlations (see Appendix D). However, it was not independently related to reading prosody after accounting for word reading in any of the developmental time points from Grade 1 through Grade 3. Furthermore, the magnitude of bivariate correlations between listening comprehension and reading prosody indicators by and large remained similar across the time points (see Appendix D), not supporting the hypothesis about a changing relation between listening comprehension and reading prosody, at least for English-speaking children in Grades 1 to 3. This finding should not be taken as a lack of a relation of listening comprehension to reading prosody. Instead, this result might indicate and underscore the importance of word reading in reading prosody as word reading acts as a bottleneck during the beginning phase of reading development (Kim, 2015; Kim & Wagner, 2015). Future longitudinal studies beyond primary grades are necessary to elucidate whether listening comprehension or language skills make a unique contribution to reading prosody over and above word reading, and if so, when in the developmental phase this occurs. Additionally, studies have shown that text reading efficiency is a mediator of the relations of word reading and listening comprehension to reading comprehension (Kim, 2015; Kim & Wagner, 2015). Given the relation of word reading to reading prosody in the present study and prior work (e.g., Benjamin & Schwanenflugel, 2010; Binder et al., 2013; Schwanenflugel et al., 2004), as well as the relation of reading prosody to reading comprehension (e.g., Arcand et al., 2014; Binder et al., 2013; Calet et al., 2015; Groen et al., 2019; Klauda & Guthrie, 2008; Schwanenflugel et al., 2004; Veenendaal et al., 2014) and the relation of listening comprehension to reading comprehension (Adlof et al., 2006; Florit & Cain, 2011; Hoover & Gough, 1990; Kim, 2017, 2020), a future investigation should shed light on a potential mediating role of reading prosody in the relations of word reading and listening comprehension to reading comprehension.

Limitations, Future Directions, and Implications

As is the case with any study, results should be interpreted with the research design in mind. First, generalizability of the findings is limited to populations that are similar to the sample in the present study—English-speaking children in primary grades. Second, data were not missing completely at random in Waves 1–4 (Grades 1 and 2), and therefore, this should be considered in the generalizability of the findings. Third, although we included a comprehensive set of widely used features or indicators of reading prosody, other prosodic features (e.g., adult-child F_0 match; Comprehensive Oral Reading Fluency Scale, Benjamin et al., 2013) can be included in future studies. Fourth, previous studies indicated that reading prosody is influenced by text features such as syntactic structures (e.g., Miller & Schwanenflugel, 2006). In the present study, we used texts that were normed in the state where the study was conducted and that were not specifically developed for the role of text complexity in reading prosody. Therefore, although these texts are similar to the texts that children are likely to encounter in real life, future studies where sentence structures and types (e.g., exclamation, sarcasm) are intentionally manipulated would be useful for examining questions related to text features.

Another direction for future studies includes longitudinal investigations beyond the primary grades. The dimensionality of reading prosody and the predictive relations of word reading and listening comprehension to reading prosody may change with reading development as the constraining role of word reading decreases. In addition, in the present study we examined listening comprehension as a predictor of reading prosody, given that listening comprehension captures oral comprehension at the discourse level and involves semantic processes and draws on vocabulary and morphosyntactic and syntactic knowledge (e.g., Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Kim, 2015, 2017, 2020; Kim & Phillips, 2014; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012). However, future studies can replicate and extend the present study by examining the relations of vocabulary and syntactic knowledge to reading prosody, controlling for word reading. Finally, previous studies have suggested the relations of reading prosody with lexical-level prosody—prosodic sensitivity (Schwanenflugel & Benjamin, 2017), text reading efficiency (Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004), and reading comprehension (Benjamin & Schwanenflugel, 2010; Calet et al., 2015; Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004, 2015). Future investigations using longitudinal and experimental designs are necessary to further elucidate the nature of their relations.

The dimensionality results suggest that reading prosody instruction may attend to two different aspects: pitch and pause structure. Together with previous suggestions (e.g., Benjamin et al., 2013), these results suggest that evaluation of reading prosody in research and practice should consider both aspects, but keep in mind that prosody is largely a function of word reading skill at least for children in primary grades learning to read English. Also informative was the finding that the different aspects of MFS are largely shared with the pause structure aspect of reading prosody, again for English-speaking students in primary grades; therefore, inferences drawn from MFS can be made with this result in mind.

Importantly, the present study, in conjunction with prior work, indicates that reading prosody instruction should not be isolated from word reading or text reading efficiency. Theoretically, word reading and text reading efficiency are necessary foundations for reading prosody given their constraining roles (Kuhn et al., 2010; LaBerge & Samuels, 1974). Empirically, word reading and text reading efficiency have strong relations to reading prosody in the present study as well as in previous ones (e.g., Calet et al., 2015; Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004, 2015). A study showed that instruction of prosody versus reading rate had somewhat disparate effects on students' reading such that feedback on reading rate, but not reading prosody, had a large effect on reading rate, whereas feedback on prosody had a large effect on students' pause structure (e.g., pause after commas and between sentences; Ardoin, Morena, Binder, & Foster, 2013). However, literature on effective prosody instruction, let alone effective text reading fluency instruction that includes both word reading skill and reading prosody, is thin, and therefore, future studies are required.

References

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19,* 933–958. http://dx.doi.org/10.1007/s11145-006-9024-z

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2015). The role of reading fluency in children's text comprehension. *Frontiers in Psychology, 6,* 1810. http://dx.doi.org/10.3389/fpsyg.2015.01810

Arcand, M.-S., Dion, E., Lemire-Theberge, L., Guay, M.-R., Barrette, A., Gagnon, V., . . . Fuchs, D. (2014). Segmenting texts into meaningful word groups: Beginning readers' prosody and comprehension. *Scientific Studies of Reading, 18,* 208–223. http://dx.doi.org/10.1080/10888438.2013.864658

Ardoin, S. P., Morena, L. S., Binder, K. S., & Foster, T. E. (2013). Examining the impact of feedback and repeated readings on oral reading fluency: Let's not forget prosody. *School Psychology Quarterly, 28,* 391–404. http://dx.doi.org/10.1037/spq0000027

Asparouhov, T., & Muthén, B. (2019). Nesting and equivalence testing for structural equation models. *Structural Equation Modeling, 26,* 302–309. http://dx.doi.org/10.1080/10705511.2018.1513795

Baker, D. L., Park, Y., & Baker, S. K. (2012). The reading performance of English learners in grades 1–3: The role of initial status and growth on reading fluency in Spanish and English. *Reading and Writing, 25,* 251–281. http://dx.doi.org/10.1007/s11145-010-9261-z

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37,* 18–37.

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45,* 388–404. http://dx.doi.org/10.1598/RRQ.45.4.2

Benjamin, R. G., Schwanenflugel, P. J., Meisinger, E. B., Groff, C., Kuhn, M. R., & Steiner, L. (2013). A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly, 48,* 105–133. http://dx.doi.org/10.1002/rrq.43

Binder, K. S., Tighe, E., Jiang, Y., Kaftanski, K., Qi, C., & Ardoin, S. P. (2013). Reading expressively and understanding thoroughly: An examination of prosody in adults with low literacy skills. *Reading and Writing, 26,* 665–680. http://dx.doi.org/10.1007/s11145-012-9382-7

Boersma, P., & Weenink, D. (2015). Praat, doing phonetics by computer (Version 5.4.20) [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Calet, N., Gutierrez-Palma, N., & Defior, S. (2015). A cross-sectional study of fluency and reading comprehension in Spanish primary school children. *Journal of Research in Reading, 38,* 272–285. http://dx.doi.org/10.1111/1467-9817.12019

Calet, N., Gutierrez-Palma, N., Simpson, I. C., Gonzalez-Trujillo, M. C., & Defior, S. (2015). Suprasegmental phonology development and reading acquisition: A longitudinal study. *Scientific Studies of Reading, 19,* 51–71. http://dx.doi.org/10.1080/10888438.2014.976342

Carrow-Woolfolk, E. (2011). *Oral and written language scales* (2nd ed.). Torrance, CA: Western Psychological Services.

Chall, J. (1996). *Stages of reading development*. San Diego, CA: Harcourt Brace College Publishers.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674283947

Cowie, R., Douglas-Cowie, E., & Wichmann, A. (2002). Prosodic characteristics of skilled reading: Fluency and expressiveness in 8–10-year-old readers. *Language and Speech, 45,* 47–82. http://dx.doi.org/10.1177/00238309020450010301

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *The Nation's Report Card: Fourth-grade students reading aloud: NAEP 2002 special study of oral reading*. Washington, DC: National Center for Education Statistics. Retrieved from https://files.eric.ed.gov/fulltext/ED488962.pdf

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice, 30,* 165–175. https://www.jstor.org/stable/1476878. http://dx.doi.org/10.1080/00405849109543497

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review, 23,* 553–576. http://dx.doi.org/10.1007/s10648-011-9175-6

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research, 27,* 285–319. http://dx.doi.org/10.1023/A:1023258301588

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10,* 244–249. http://dx.doi.org/10.1016/j.tics.2006.04.002

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256. http://dx.doi.org/10.1207/S1532799XSSR0503_3

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information bi-factor analysis. *Psychometrika, 57,* 423–436. http://dx.doi.org/10.1007/BF02295430

González-Trujillo, M. C., Calet, N., Defior, S., & Gutiérrez-Palma, N. (2014). Escala de fluidez lectora en español: Midiendo los componentes de la fluidez [Scale of reading fluency in Spanish: Measuring the components of fluency]. *Estudios de Psicología, 35,* 104–136. http://dx.doi.org/10.1080/02109395.2014.893651

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

Groen, M. A., Veenendaal, N. J., & Verhoeven, L. (2019). The role of prosody in reading comprehension: Evidence from poor comprehenders. *Journal of Research in Reading, 42,* 37–57. http://dx.doi.org/10.1111/1467-9817.12133

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127–160. http://dx.doi.org/10.1007/BF00401799

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

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95,* 719–729. http://dx.doi.org/10.1037/0022-0663.95.4.719

Kendeou, P., Bohn-Gettler, C. M., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31,* 259–272. http://dx.doi.org/10.1111/j.1467-9817.2008.00370.x

Kim, Y. G. (2015). Developmental, component-based model of reading fluency: An investigation of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly, 50,* 459–481. http://dx.doi.org/10.1002/rrq.107

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. http://dx.doi.org/10.1016/j.jecp.2015.08.003

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21,* 310–333. http://dx.doi.org/10.1080/10888438.2017.1291643

Kim, Y.-S. G. (2020). Hierarchical and dynamic relations of language and cognitive Skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1016/j.jecp.2020.104813

Kim, Y.-S., Park, C. H., & Wagner, R. K. (2014). Is oral/text reading fluency a "bridge" to reading comprehension? *Reading and Writing, 27,* 79–99. http://dx.doi.org/10.1007/s11145-013-9434-7

Kim, Y.-S. G., & Petscher, Y. (2016). Prosodic sensitivity and reading: An investigation of pathways of relations using a latent variable approach. *Journal of Educational Psychology, 108,* 630–645. http://dx.doi.org/10.1037/edu0000078

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension? *Journal of Educational Psychology, 102,* 652–667. http://dx.doi.org/10.1037/a0019643

Kim, Y.-S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49,* 269–281. http://dx.doi.org/10.1002/rrq.74

Kim, Y. S. G., & Wagner, R. K. (2015). Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading, 19,* 224–242. http://dx.doi.org/10.1080/10888438.2015.1007375

Kim, Y- S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading, 15,* 338–362. http://dx.doi.org/10.1080/10888438.2010.493964

Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology, 113,* 93–111. http://dx.doi.org/10.1016/j.jecp.2012.03.002

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100,* 310–321. http://dx.doi.org/10.1037/0022-0663.100.2.310

Kline, R. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45,* 230–251. http://dx.doi.org/10.1598/RRQ.45.2.4

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3–21. http://dx.doi.org/10.1037/0022-0663.95.1.3

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323. http://dx.doi.org/10.1016/0010-0285(74)90015-2

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- to 6-year old children. *Reading Research Quarterly, 47,* 259–282. http://dx.doi.org/10.1002/rrq.020

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology, 98,* 839–853. http://dx.doi.org/10.1037/0022-0663.98.4.839

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly, 43,* 336–354. http://dx.doi.org/10.1598/RRQ.43.4.2

Muthén, L. K., & Muthén, B. O. (1998–2018). *Mplus user's guide* (8th ed.) Los Angeles, CA: Author.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An* *evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00–4769). Washington, DC: U.S. Government Printing Office.

Paige, D., Rasinski, T., & Magpuri-Lavell, T. (2012). Is fluent expressive reading important for high school readers? *Journal of Adolescent & Adult Literacy, 56,* 67–76. http://dx.doi.org/10.1002/JAAL.00103

Paige, D., Rasinski, T., Magpuri-Lavell, T., & Smith, G. S. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research, 46,* 123–156.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In L. C. Ehri, R. Treiman, & P. B. Gough (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud: Data from NAEP's Integrated reading performance record (IRPR) at grade 4* (NCES 95–726). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. http://dx.doi.org/10.2307/271063

Rasinski, T. (2004). Creating fluent readers. *Educational Leadership, 61,* 46–51. Retrieved from http://educationalleader.com/subtopicintro/read/ASCD/ASCD_364_1.pdf

Rasinski, T., Homan, S., & Biggs, M. (2009). Teaching reading fluency to struggling readers: Method, materials, and evidence. *Reading & Writing Quarterly, 25,* 192–204. http://dx.doi.org/10.1080/10573560802683622

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48,* 350–361. http://dx.doi.org/10.1080/19388070802468715

Ravid, D., & Mashraki, Y. E. (2007). Prosodic reading, reading comprehension and morphological skills in Hebrew-speaking fourth graders. *Journal of Research in Reading, 30,* 140–156. http://dx.doi.org/10.1111/j.1467-9817.2007.00340.x

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42,* 546–567. http://dx.doi.org/10.1598/RRQ.42.4.5

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343–366. http://dx.doi.org/10.1016/j.jsp.2007.06.006

Sabatini, J., Wang, Z., & O'Reilly, T. (2019). Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly, 54,* 253–271. http://dx.doi.org/10.1002/rrq.226

Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice, 30,* 158–164. http://dx.doi.org/10.1080/00405849109543496

Schwanenflugel, P. J., & Benjamin, R. G. (2017). Lexical prosody as an aspect of oral reading fluency. *Reading and Writing, 30,* 143–162. http://dx.doi.org/10.1007/s11145-016-9667-3

Schwanenflugel, P. J., Hamilton, A. M., Wisenbaker, J. M., Kuhn, M. R., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96,* 119–129. http://dx.doi.org/10.1037/0022-0663.96.1.119

Schwanenflugel, P. J., Westmoreland, M. R., & Benjamin, R. G. (2015). Reading fluency skill and the prosodic marking of linguistic focus. *Reading and Writing, 28,* 9–30. http://dx.doi.org/10.1007/s11145-013-9456-1

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading, 17,* 108–133. http://dx.doi.org/10.1080/10888438.2011.618153

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading, 32,* 383–401. http://dx.doi.org/10.1111/j.1467-9817.2009.01401.x

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency – 2nd ed. (TOWRE-2).* Austin, TX: Pro-Ed.

Valle, A., Binder, K. S., Walsh, C. B., Nemier, C., & Bangs, K. E. (2013). Eye movements, prosody, and word frequency among average- and high-skilled second-grader readers. *School Psychology Review, 42,* 171–190.

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2015). What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading, 38,* 213–225. http://dx.doi.org/10.1111/1467-9817.12024

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2014). The role of speech prosody and text reading prosody in children's reading comprehension. *British Journal of Educational Psychology, 84,* 521–536. http://dx.doi.org/10.1111/bjep.12036

Webman-Shafran, R. (2018). Implicit prosody and parsing in silent reading. *Journal of Research in Reading, 41,* 546–563. http://dx.doi.org/10.1111/1467-9817.12124

Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Pearson.

Whalley, K., & Hansen, J. (2006). The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading, 29,* 288–303. http://dx.doi.org/10.1111/j.1467-9817.2006.00309.x

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5,* 211–239. http://dx.doi.org/10.1207/S1532799XSSR0503_2

Wood, C., Wade-Woolley, L., & Holliman, A. J. (2009). Phonological awareness: Beyond phonemes. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 7–23). London, UK: Routledge. http://dx.doi.org/10.4324/9780203877838

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Yildiz, M., Yildirim, K., Ates, S., Rasinski, T., Fitzgerald, S., & Zimmerman, B. (2014). The relationship between reading fluency and reading comprehension in fifth-grade Turkish students. *International Journal of School & Educational Psychology, 2,* 35–44. http://dx.doi.org/10.1080/21683603.2013.854187

Zutell, J., & Rasinski, T. V. (1991). Training teachers to attend to their students' oral reading fluency. *Theory Into Practice, 30,* 211–217. http://dx.doi.org/10.1080/00405849109543502

(*Appendices follow*)

Appendix A

Correlation Matrices at Waves 1, 2, 3, 4, 5, and 6

Table A1
Correlation Matrices at Wave 1 (Below Diagonal) and Wave 2 (Above Diagonal)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. F_0 change	—	−.35	.1	.05	−.16	−.16	−.17	−.21	−.13	−.04	−.21	−.20	−.18
2. Int cont	−.35	—	−.02	.04	.03	.11	.06	.11	.05	.03	.05	.04	−.01
3. Pau freq	.20	.01	—	.46	−.72	−.73	−.75	−.69	−.31	−.29	−.69	−.71	−.77
4. Pau dur	.16	−.04	.32	—	−.44	−.47	−.48	−.44	−.17	−.11	−.42	−.40	−.51
5. Smth	−.23	.10	−.64	−.40	—	.87	.84	.77	.29	.24	.67	.71	.73
6. Pace	−.27	.12	−.61	−.37	.90	—	.88	.80	.31	.29	.71	.73	.75
7. Phra	−.26	.12	−.61	−.37	.92	.90	—	.84	.34	.32	.71	.73	.74
8. Expr	−.34	.15	−.57	−.36	.87	.85	.89	—	.32	.28	.68	.67	.69
9. WJOC	−.07	−.02	−.30	−.24	.34	.35	.31	.30	—	.57	.38	.39	.37
10. OWLS	−.13	.01	−.29	−.18	.35	.29	.31	.28	.58	—	.39	.39	.33
11. LWID	−.27	.08	−.61	−.45	.73	.71	.74	.67	.37	.43	—	.90	.84
12. WIAT	−.26	.09	−.67	−.41	.76	.73	.75	.69	.38	.44	.90	—	.87
13. SWE	−.30	.05	−.69	−.46	.77	.75	.75	.70	.36	.41	.87	.89	—

Note. F_0 Change = fundamental frequency change; Int con = intonation contour; Pau freq = pause frequencies; Pau dur = pause durations; Smth = MFS (Multi-Dimensional Fluency Scale) smoothness; Pace = MFS pacing; Phra = MFS phrasing; Expr = MFS expression and volume; WJOC = Woodcock-Johnson III oral comprehension; OWLS = Oral and Written Language Scales; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Test; SWE = Sight Word Efficiency.

Table A2
Correlation Matrices at Wave 3 (Below Diagonal) and Wave 4 (Above Diagonal)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. F_0 change	—	−.41	.26	.04	−.18	−.21	−.25	−.33	−.07	−.08	−.22	−.22	−.22
2. Int cont	−.42	—	−.10	.01	.15	.13	.09	.18	.05	.06	.05	.03	.04
3. Pau freq	.25	−.10	—	.44	−.74	−.78	−.78	−.71	−.35	−.40	−.70	−.72	−.76
4. Pau dur	.01	.00	.50	—	−.44	−.47	−.45	−.43	−.15	−.15	−.43	−.44	−.55
5. Smth	−.20	.12	−.76	−.49	—	.89	.87	.82	.32	.41	.65	.69	.69
6. Pace	−.24	.18	−.78	−.48	.87	—	.89	.81	.37	.42	.69	.71	.73
7. Phra	−.24	.20	−.76	−.45	.84	.89	—	.84	.33	.39	.69	.71	.74
8. Expr	−.30	.17	−.73	−.52	.79	.79	.83	—	.29	.36	.62	.68	.70
9. WJOC	−.05	.05	−.38	−.23	.39	.41	.35	.31	—	.57	.39	.35	.34
10. OWLS	−.16	.05	−.38	−.17	.40	.43	.39	.37	.63	—	.47	.44	.37
11. LWID	−.22	.06	−.75	−.38	.64	.67	.64	.65	.44	.47	—	.91	.79
12. WIAT	−.25	.05	−.73	−.34	.62	.64	.61	.60	.42	.44	.91	—	.80
13. SWE	−.21	.11	−.79	−.51	.68	.68	.66	.67	.35	.38	.80	.81	—

Note. F_0 Change = fundamental frequency change; Int con = intonation contour; Pau freq = pause frequencies; Pau dur = pause durations; Smth = MFS (Multi-Dimensional Fluency Scale) smoothness; Pace = MFS pacing; Phra = MFS phrasing; Expr = MFS expression and volume; WJOC = Woodcock-Johnson III oral comprehension; OWLS = Oral and Written Language Scales; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Test; SWE = Sight Word Efficiency.

(*Appendices continue*)

Table A3
Correlation Matrices at Wave 5 (Below Diagonal) and Wave 6 (Above Diagonal)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. F_0 change	—	−.44	.27	.10	−.18	−.27	−.28	−.44	−.15	−.13	−.26	−.22	−.32
2. Int cont	−.46	—	−.11	−.02	.10	.15	.14	.27	.09	.10	.03	.04	.11
3. Pau freq	.26	−.04	—	.51	−.72	−.75	−.73	−.68	−.33	−.35	−.70	−.71	−.76
4. Pau dur	.09	.04	.50	—	−.42	−.44	−.41	−.39	−.21	−.24	−.35	−.36	−.48
5. Smth	−.10	−.01	−.73	−.44	—	.84	.86	.74	.26	.30	.65	.64	.70
6. Pace	−.18	.05	−.79	−.49	.84	—	.86	.73	.28	.36	.66	.64	.71
7. Phra	−.20	.03	−.78	−.41	.85	.85	—	.74	.26	.34	.65	.63	.68
8. Expr	−.36	.15	−.70	−.45	.75	.78	.79	—	.23	.33	.59	.58	.65
9. WJOC	−.09	−.01	−.42	−.22	.35	.35	.36	.32	—	.62	.48	.41	.31
10. OWLS	−.21	.03	−.41	−.22	.35	.40	.39	.37	.64	—	.54	.47	.35
11. LWID	−.25	−.02	−.74	−.38	.62	.65	.66	.57	.53	.52	—	.90	.74
12. WIAT	−.26	−.03	−.71	−.41	.58	.61	.61	.55	.51	.49	.89	—	.73
13. SWE	−.29	−.01	−.76	−.50	.64	.71	.67	.62	.41	.40	.76	.76	—

Note. F_0 Change = fundamental frequency change; Int con = intonation contour; Pau freq = pause frequencies; Pau dur = pause durations; Smth = MFS (Multi-Dimensional Fluency Scale) smoothness; Pace = MFS pacing; Phra = MFS phrasing; Expr = MFS expression and volume; WJOC = Woodcock-Johnson III oral comprehension; OWLS = Oral and Written Language Scales; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Test; SWE = Sight Word Efficiency.

Appendix B

Factor loadings for listening comprehension and word reading latent variables by wave

Table B1
Factor Loadings for Listening Comprehension and Word Reading Factors Separated by Wave

	Wave 1			Wave 2			Wave 3			Wave 4			Wave 5			Wave 6		
Variable	λ	*SE*	*p*	λ	*SE*	*p*	λ	*SE*	*p*	λ	*SE*	*p*	λ	*SE*	*p*	λ	*SE*	*p*
Word reading																		
SWE	.93	.01	<.001	.90	.01	<.001	.85	.02	<.001	.83	.02	<.001	.80	.02	<.001	.77	.02	<.001
LWID	.94	.01	<.001	.93	.01	<.001	.95	.01	<.001	.95	.01	<.001	.95	.01	<.001	.98	.01	<.001
WIAT	.96	.01	<.001	.97	.07	<.001	.95	.01	<.001	.96	.01	<.001	.94	.01	<.001	.93	.01	<.001
Listening comprehension																		
OWLS	.82	.05	<.001	.75	.05	<.001	.82	.04	<.001	.83	.05	<.001	.79	.04	<.001	.84	.04	<.001
WJOC	.71	.05	<.001	.76	.05	<.001	.78	.04	<.001	.69	.05	<.001	.81	.04	<.001	.74	.04	<.001

Note. λ = loading; SWE = Sight Word Efficiency; LWID = Letter-Word Identification; WIAT = Wechsler Individual Achievement Test; OWLS = Oral and Written Language Scales; WJOC = Woodcock-Johnson III oral comprehension.

(*Appendices continue*)

Appendix C

Model fit for confirmatory factor models of listening comprehension, word reading, and prosody indicators by wave

Table C1

Model fit for CFA Models of Listening Comprehension, Word Reading, and Prosody Indicators by Wave

Wave	χ^2	*df*	SCF	*p*	CFI	TLI	RMSEA	90% CI	*n*BIC
Wave 1	83.01	48	0.99	.001	0.99	0.98	.05	[.03, .06]	19045.58
Wave 2	80.36	48	0.99	.002	0.99	0.99	.04	[.03, .06]	20912.72
Wave 3	114.90	49	0.99	<.001	0.98	0.97	.06	[.05, .08]	19317.37
Wave 4	117.31	48	0.99	<.001	0.98	0.97	.07	[.05, .08]	18996.82
Wave 5	104.32	49	0.99	<.001	0.98	0.97	.06	[.04, .08]	17784.39
Wave 6	112.23	48	1.01	<.001	0.98	0.96	.07	[.05, .08]	16546.06

Note. CFA = confirmatory factor analysis; SCF = scaling correction factor used for Satorra-Bentler chi-square tests; CFI = confirmatory fit index; TLI = Tucker-Lewis index; RMSEA = root mean squared error of approximation; *n*BIC = sample size–adjusted Bayesian information criterion.

Appendix D

Correlations between reading prosody, word reading, and listening comprehension latent variables

Table D1

Factor Correlations for the CFA Models of Reading Prosody, Word Reading, and Listening Comprehension

	Wave 1			Wave 2			Wave 3			Wave 4			Wave 5			Wave 6		
Factor	*r*	*SE*	*p*	*r*	*SE*	*p*	*r*	*SE*	*p*	*r*	*SE*	*p*	*r*	*SE*	*p*	*r*	*SE*	*p*
Prosody: Pitch factor with																		
Prosody: Ratings and pause	.23	.07	.001	.18	.06	.002	.27	.06	<.001	.23	.06	<.001	.24	.06	<.001	.28	.13	.026
Ratings specific factor	−.26	.08	.001	−.12	.07	.07	−.15	.06	.02	−.27	.07	<.001	.38	.06	<.001	.38	.10	<.001
Pause specific factor	−.05	.07	.52	.05	.06	.44	−.08	.06	.14	−.13	.07	.05	−.12	.06	.03	−.08	.17	.62
Listening comprehension	.15	.08	.07	.12	.07	.07	.15	.07	.03	.11	.07	.09	.19	.07	.007	.17	.06	.005
Word reading	.30	.06	<.001	.19	.06	.001	.26	.05	<.001	.25	.05	<.001	.32	.06	<.001	.34	.06	<.001
Ratings specific factor with																		
Pause specific factor	−.03	.10	.75	.08	.17	.63	n/a	n/a	n/a	−.02	.17	.91	n/a	n/a	n/a	−.19	.40	.64
Listening comprehension	.06	.08	.46	−.07	.09	.46	.11	.07	.11	.05	.09	.61	.01	.07	.86	.06	.15	.67
Word reading	.04	.09	.65	−.02	.11	.85	−.04	.03	.27	−.04	.10	.70	−.02	.04	.63	.10	.25	.71
Word reading with																		
Listening comprehension	.55	.05	<.001	.52	.06	.000	.57	.05	<.001	.56	.06	<.001	.67	.05	<.001	.60	.06	<.001
Prosody: Ratings and pause	.84	.03	<.001	.85	.02	.000	.80	.02	<.001	.85	.02	<.001	.85	.03	<.001	.87	.04	<.001
Pause specific factor	−.22	.04	<.001	−.23	.04	.000	−.35	.04	<.001	−.24	.06	<.001	−.24	.04	<.001	−.26	.09	.003
Prosody: Ratings and pause with																		
Listening comprehension	.49	.06	<.001	.46	.05	.000	.53	.05	<.001	.53	.05	<.001	.55	.05	<.001	.47	.06	<.001
Pause specific factor with																		
Listening comprehension	−.12	.07	.09	−.09	.06	.13	−.08	.06	.19	−.12	.06	.05	−.10	.07	.14	−.08	.08	.31

Note. CFA = confirmatory factor analysis.

(*Appendices continue*)

Appendix E

Standardized coefficients of word reading and listening comprehension predicting all the reading prosody latent variables, including the specific Ratings factor and the specific Pause factor, by wave

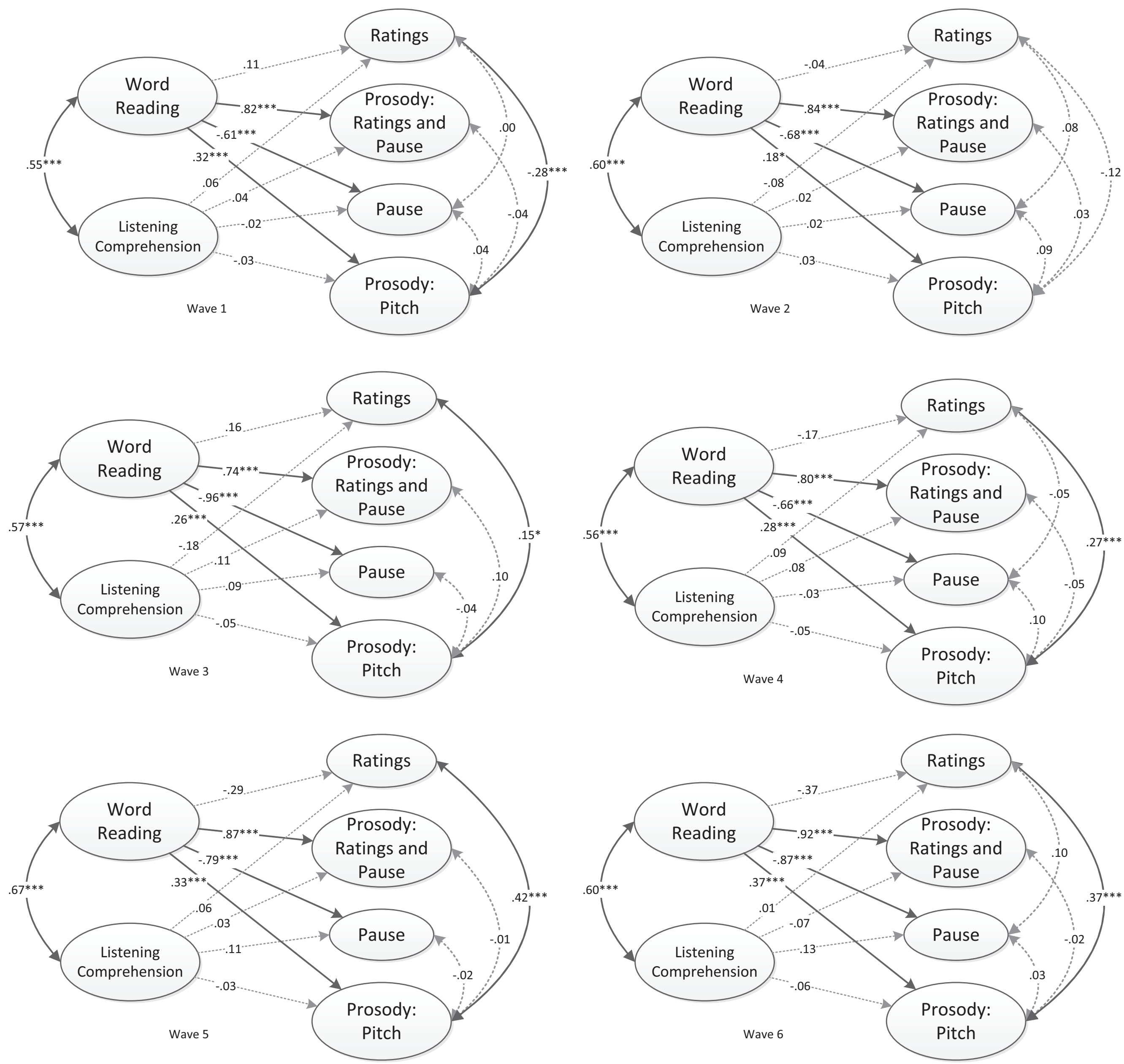

Figure E1. Standardized coefficients of word reading and listening comprehension predicting reading prosody latent variables by wave. Gray, dashed lines are not statistically significant ($p > .05$). Not shown but estimated were measurement models for word reading, listening comprehension, and reading prosody factors. $^{*}p < .05$. $^{***}p < .001$.

(*Appendices continue*)

Results of the original model in Figure 3, which includes the pathways from word reading and listening comprehension to the Ratings specific factor and the Pause specific factor, are presented in Figure E1.

In Wave 1, word reading negatively and independently related to the Pause specific factor ($\gamma = -.61$, $p < .001$), whereas listening comprehension did not ($p = .81$). Neither listening comprehension nor word reading predicted the Ratings specific factor ($p = .52$ and $p = .71$, respectively). The model explained 38.8% of the variance in the Pause specific factor and no significant variance in the Ratings specific factor ($p = .79$). In subsequent waves, word reading was related to the Pause specific factor ($-.96 \leq \gamma\text{s} \leq -.66$, $p\text{s} < .001$), whereas listening comprehension was not ($-.03 \leq \gamma\text{s} \leq .13$, $p\text{s} \geq .05$) after accounting for word reading. The Ratings specific factor was not predicted by word reading or listening comprehension in any subsequent wave. Between 38.8% and 82.9% of the variance was explained in the Pause specific factor across waves (Waves 1 to 6, respectively: 38.8%, 45.0%, 82.9%, 46.1%, 51.5%, 63.1%) and no statistically significant variance was explained in the Ratings specific factor (Waves 1 to 6, respectively: 2.2%, 3.8%, 2.4%, 1.9%, 6.1%, 1.8%).

Received August 26, 2019
Revision received February 17, 2020
Accepted February 21, 2020 ■