Exploring the Dimensionality of Reading Fluency in Grade 4: Prosody and its Role in Supporting Reading Comprehension

Hannah Anderson

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.
EXPLORING THE DIMENSIONALITY OF READING FLUENCY IN GRADE 4: PROSODY
AND ITS ROLE IN SUPPORTING READING COMPREHENSION

by

Hannah Manning, B.A.

A Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

Major: School Psychology

The University of Memphis
December 2023
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Measuring Prosody</td>
<td>2</td>
</tr>
<tr>
<td>Pitch</td>
<td>4</td>
</tr>
<tr>
<td>Pauses</td>
<td>5</td>
</tr>
<tr>
<td>Prosody’s Connection to Reading Comprehension</td>
<td>7</td>
</tr>
<tr>
<td>Dimensionality of Prosody</td>
<td>11</td>
</tr>
<tr>
<td>Purpose</td>
<td>13</td>
</tr>
<tr>
<td>2. Methods</td>
<td>14</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td>Measures</td>
<td>16</td>
</tr>
<tr>
<td>Text Reading Efficiency</td>
<td>16</td>
</tr>
<tr>
<td>Prosody</td>
<td>17</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>19</td>
</tr>
<tr>
<td>Procedure</td>
<td>20</td>
</tr>
<tr>
<td>Analytic Plan</td>
<td>21</td>
</tr>
<tr>
<td>3. Results</td>
<td>22</td>
</tr>
<tr>
<td>Data Processing and Screening</td>
<td>22</td>
</tr>
<tr>
<td>Measurement Model</td>
<td>23</td>
</tr>
<tr>
<td>Structural Models</td>
<td>25</td>
</tr>
<tr>
<td>4. Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Limitations and Future Directions</td>
<td>32</td>
</tr>
<tr>
<td>References</td>
<td>35</td>
</tr>
<tr>
<td>Appendix</td>
<td>43</td>
</tr>
</tbody>
</table>
Abstract

The purpose of the current study was to examine the dimensionality of prosody and its role in supporting comprehension in Grade 4 students (N = 198). Participants were administered a grade-level CBM probe and i-Ready Diagnostic, including the Vocabulary, Literary Text, and Informational Text subtests. Spectrographic analysis was used to extract the prosodic data from audio recordings of the passage reading. A comparison group of 30 adults were recruited. A four-factor measurement model was proposed, with a reading comprehension factor comprised of i-Ready subtest, two prosodic factors comprised of pitch and pausing features gathered from the grade-level passage, and a text reading efficiency (TRE) latent factor comprised of oral reading rate, silent reading rate, and oral reading accuracy. However, results supported a unidimensional fluency model comprised of the prosodic and text reading efficiency indicators, along with a separate reading comprehension factor. An additional ad hoc model with prosody indicator variables was also analyzed. Results indicated that prosody did not constitute a multidimensional construct in this study, and that prosody did not contribute directly or indirectly to reading comprehension. These findings are not consistent with the extant research. Given the sparse literature on this topic, more research is needed to elucidate the relation between prosody and reading comprehension skills in elementary students with varied text difficulty.

Keywords: Prosody, fluency, reading comprehension, text reading efficiency, dimensionality, difficulty, pitch, pause, pausing
Exploring the Dimensionality of Reading Fluency in Grade 4: Prosody and its Role in Supporting Reading Comprehension

Many researchers hold a simple view of fluency in which reading rate and accuracy comprise the construct. Recognition of prosody as a third feature within the fluency umbrella is increasing in academia (Kuhn et al., 2010; National Institute of Child Health and Human Development, 2000), though it is still an understudied area. The terms “text reading efficiency” (to denote rate and accuracy) and “text reading fluency” (to denote rate, accuracy, and prosody) are used to make this distinction in the literature (Kim et al., 2021). Reading prosody is the melodic, expressive aspect of text reading fluency that goes beyond reading with automaticity (Godde et al., 2020). Prosody can take the form of the timing and duration of pauses, the stress or emphasis placed on certain words or syllables, or changes in pitch, among other features.

While text reading efficiency is predictive of reading comprehension time and again across studies (e.g., Jenkins et al., 2003; Kim et al., 2011; Price et al., 2016; Sabatini et al., 2019), prosody also accounts for unique variance beyond the components of rate and accuracy (Benjamin & Schwanenflugel, 2010; Klauda & Guthrie, 2008; Miller & Schwanenflugel, 2006, 2008; Rasinski et al., 2009). Because the goal of fluent reading is comprehension, reading with good prosody is an important aspect of text reading fluency. Two features of prosody, pitch and pauses, have been found to explain unique variance in reading comprehension across several studies both together (Benjamin & Schwanenflugel, 2010; Benjamin et al., 2013; Veenendal et al., 2015) and separately (pitch: Miller & Schwanenflugel, 2006; pauses: Binder et al., 2013).

As such, this literature review will begin by addressing how prosody is measured. Pitch and pausing specifically will be reviewed as the dominant features of prosody measurement. The past theory and research concerning the contributions that prosody makes to reading
comprehension beyond rate and accuracy will then be described. Finally, the role of prosody as either a predictor of reading comprehension or a mediator between text reading efficiency and reading comprehension will be discussed.

**Measuring Prosody**

There are two established methods for measuring prosody: subjective rating scales and spectrographic analyses. Subjective rating scales, such as the Multi-Dimensional Fluency Scale (MDFS; Rasinski, 2004) or the National Assessment of Educational Progress (NAEP) 6-point fluency scale (White et al., 2021), utilize one or more raters who listen to an oral reading by ear and assign numerical ratings. Numbers along the scale are given a label and/or description on what constitutes each skill level, and judgements are made in accordance with subjective views of what constitutes poor or skilled prosody (Pinnell et al., 1995).

Interrater training can be used to consolidate the raters’ concepts of a given point on the scale; however, subjective rating scales do not tend to be sufficiently reliable (Benjamin & Schwanenflugel, 2010; Cowie et al., 2002). Often, the highest and lowest points on these scales are underutilized in favor of more central ratings, which results in a restricted range of prosody measurement (Daane et al., 2005). Further, raters often conflate poor prosody with poor decoding skills (Kim et al., 2021) or good prosody with a faster reading rate (Pinnell et al., 1995) because the raters are trained to listen for global reading skills rather than specific features of prosody. Indeed, Kim and colleagues (2020) found that rating scales and pausing were interrelated enough to form a single latent variable, whereas pitch was a related but separate construct—a finding supported in the meta-analysis by Wolters and colleagues (2020). This finding demonstrates the bias of rating scales toward certain prosodic features and the exclusion of others which can influence the connection of prosody to reading comprehension. Finally, these rating scales were
primarily developed for ease of use by teachers for such purposes as progress monitoring (Godde et al., 2020). Miller and Schwanenflugel (2006) recommend restricting their use to such purposes unless the rating scale has been validated against the second means of prosodic measurement—spectrographic analyses.

Spectrographs utilize computerized analysis of acoustic markers rather than human raters to measure prosodic features and is considered the most direct means of measuring prosody (Binder et al., 2013). Spectrographic analyses allow for specific features of prosody to be targeted and extrapolated using objective measures, such as Hertz values for pitch. Further, this method allows for finer-grained analyses, and it avoids the wholistic approach that rating scales take which can be easily conflated with other variables or lead to range restriction. Due to spectrographic analyses’ greater reliability and analytic precision (Benjamin & Schwanenflugel, 2010; Cowie et al., 2002), this method was chosen for the current study. Further, the lack of studies examining prosody spectrographically in relation to reading comprehension within existing literature presents a need for more spectrographic studies (only 5 of the 35 studies used spectrographs according to the meta-analysis by Wolters et al., 2020). Notably, four of those studies were conducted with students in early elementary (Grades 1 to 3) and one study in upper elementary (Grade 5). Wolters and colleagues (2020) call for increased use of spectrographic analysis of prosody, especially among the understudied upper elementary school years.

However, spectrographic analysis does present its own challenges. This approach is time-intensive to conduct, which can limit sample size. Appropriate prosody can be generated in various ways dependent upon individual factors such as speaking style or dialect (Frazier et al., 2006; Godde et al., 2020), and it can be difficult to account for these individual differences objectively. Finally, variable selection can depend greatly upon the reading stimuli. As such,
different methods of measuring pitch and pausing as well as ways to address individual differences in reading will be discussed below.

**Pitch**

There are many ways in which pitch may be analyzed in a spectrographic analysis. One common approach is to pinpoint the lowest or highest pitch expressed, which can be represented in Hertz. Phrase-final or sentence-final pitch movements are measured as the change in pitch as a phrase or statement is concluded. At the end of a declarative statement or certain phrases (dependent upon the context), pitch should decrease. At the end of some interrogative statements or certain phrases (dependent upon the context), pitch should increase. The Hertz value for measuring sentence-final pitch movements derives from the absolute difference between the pitch at the end of the sentence (the sentence-final pitch) and the nearest preceding pitch peak or valley (Benjamin et al., 2013; Schwanenflugel et al., 2004). Dowhower (1987) specified a change of 15 Hertz as the smallest change indicative of a pitch movement.

The pitch of every syllable throughout the sentence or passage may be analyzed as well. The highest pitch per word may be recorded as a list of values called a pitch contour (F0; Schwanenflugel et al., 2004). Finally, pitch variation derives from the standard deviation of pitch contour or simply as the difference of the highest and lowest pitch values (Benjamin et al., 2013; Cowie et al., 2002). Each of these measurement strategies produces a very different representation of pitch. The selection of the measurement strategies should be guided by the characteristics of the reading stimuli and research questions. Further, these are only the more commonly used strategies that have found significant or mixed results from past research. Among the measures that produced a significant link to reading comprehension in past studies, the most reliable pitch measures were phrase-final or sentence-final pitch movements and pitch
contours (Binder et al., 2013; Miller & Schwanenflugel, 2006; Schwanenflugel et al., 2004), while pitch variation has not been studied alongside reading comprehension.

Pitch is particularly subject to intra-individual differences. Godde and colleagues (2020) noted that pitch contours can differ between children and yet be equally expressive. As such, objective criteria for pitch can be difficult to establish without a balanced group of comparison readers to generate an averaged expectation for pitch values. Although one previous study compared individual children within the sample to the whole child sample itself using within-subjects ANOVA (Schwanenflugel & Benjamin, 2017), most studies opt for an adult sample (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel et al., 2004). Because pitch control during reading develops across the lifespan until finally reaching an approximation of natural speech around adulthood (Godde et al., 2020), an adult sample should serve as a highly consistent baseline which can easily generalize. Alternatively, comparing within or across child samples leaves greater room for variation in reading development and prosodic skill.

**Pauses**

Pauses can be equally complex to measure. The total number of pauses can be tallied across a sentence or passage, or they can be measured in terms of milliseconds. Measuring pauses of 100 milliseconds or greater in duration has become standard due to its reliability for detection (Benjamin et al., 2013; Kim et al., 2021; Miller & Schwanenflugel, 2006, 2008). Following detection, the type of pause must be determined.

Grammatical (or syntactic) pauses reflect punctuation or context properly (Lalain et al., 2016). Ungrammatical (or hesitation) pauses may take the form of intersentential (between-sentence) pauses, intrasentential (within-sentence) pauses, or pausal intrusions. Pausal intrusions
are a form of intrasentential pauses, but they must be ungrammatical. Additionally, they may occur within a word or between connected syntax (e.g., between a noun and its adjective, within a prepositional phrase; Godde et al., 2021). Finally, breath pauses may or may not correspond to punctuation or context depending on the prosodic mastery of the reader (Godde et al., 2021; Grosjean & Collins, 1979).

Once a category is determined, the method of measuring pauses must be chosen. Milliseconds spent pausing may be totaled for a measure of pause duration across the sentence or passage, or they may be averaged together to measure average pause duration. They may be compared on a pause-by-pause basis between readers as well (e.g., against an experienced adult reader; Miller & Schwanenflugel, 2006). The choice of pause measurement should reflect the types of pausing possible from the reading stimuli, the relevant research questions, and the findings of past research.

According to past research, poor readers demonstrate longer pauses, more frequent pauses, and more pauses at all punctuation marks regardless of function (e.g., pausing for commas within a list of adjectives; Binder et al., 2013). According to Binder and colleagues (2013), long and frequent pauses may be a strategy that poor readers employ to slow information intake to comprehend text. Even so, pause frequency and duration are linked to problems with comprehension and automaticity (Benjamin & Schwanenflugel, 2010; Clay & Imlach, 1971; Dowhower, 1987). Additionally, readers with better comprehension tend to demonstrate grammatical pause patterns, whereas poor comprehenders tend to demonstrate poor grammatical pausing (Benjamin et al., 2013).

Pauses, like pitch, are also subject to intra-individual differences. Pause placement tends to vary between individuals as a function of reading style, meaning two people may display
appropriate pauses and yet display very different pause patterns (Godde et al., 2020). Because pausing is influenced by text reading efficiency skill (especially rate), age cannot necessarily predict appropriate development of pause skills (Godde et al., 2020). As such, establishing a baseline with an adult reference group for common pause placements and for the lowest number of pauses while appropriately following context and syntax is important for establishing objective pausal expectations.

**Prosody’s Connection to Reading Comprehension**

Prosody is highly influential in the acquisition and everyday use of spoken language, as it aids in the separation of words and phrases and creates syntactic structure (Whalley & Hansen, 2006). These uses of prosody are thought to carry over into the reading of written language and thus support reading comprehension in similar ways. Proper use of prosody is thought to aid the reader in creating a mental structure as they read, thus scaffolding reading comprehension across the sentence or passage (Frazier et al., 2006; Miller & Schwanenflugel, 2006). However, the mechanism by which this is accomplished is accompanied by several theories.

Early in reading development, cognitive resources are consumed by simply decoding the graphemes into their corresponding sounds, and growth occurs as the accuracy and speed of this decoding increases (Kuhn et al., 2010). Once the cognitive load of decoding becomes negligible, a developing reader can be considered an automatic reader (LaBerge & Samuels, 1974), and cognitive resources may be redirected to prosody. Prosody, in turn, appears to reduce the cognitive load that readers must bear during comprehension (Godde et al., 2020). Indeed, in one study involving Dutch Grade 5 children, reading prosody (measured by the MDFS) was more relevant to reading comprehension than decoding once automaticity had been achieved (Groen et al., 2019).
Other researchers suggest that, because prosodic reading follows syntactic boundaries in a sentence, prosody decreases the cognitive load by creating easy-to-process idea units for working memory (Miller & Schwanenflugel, 2008). Working memory can retain information for longer while reading aloud due to the utilization of the articulatory loop, which is the mechanism responsible for retaining auditory information (Baddeley, 1992). When reading aloud, prosody provides a framework from which working memory may build (Frazier et al., 2006; Swets et al., 2007). In essence, prosody may “chunk” the auditory feedback from reading continuous text aloud into pieces and retain these pieces until other reading processes (such as vocabulary or prior knowledge) can convert the pieces into an overall meaning.

In support of this working memory theory, one study found that less skilled readers could comprehend passages significantly better if allowed to listen to someone else read the passage aloud while reading along silently (Snow et al., 1982). Both adults and children have also demonstrated better recall when a passage is read to them with prosody as compared to a passage being read without prosody (Mira & Schwanenflugel, 2013; Goldman et al., 2006). In conjunction, these studies suggest that prosody may aid in reading comprehension beyond listening comprehension. Additionally, because the findings of increased reading comprehension are present regardless of who generates the prosodic pattern, the role prosody plays in shaping auditory information implicates working memory as a possible mechanism through which comprehension is supported.

Another explanation is found in the implicit prosody hypothesis (Fodor, 1998). It states that a reader may render several potential prosodic patterns for a given text. Then, the reader selects the pattern that most effectively breaks up ideas (using syntactic and contextual boundaries) and helps identify parts of speech (a process called “disambiguation of syntax”) to
reduce confusion during the comprehension stage (Fodor, 2002; Kuhn et al., 2010). For example, pitch and pauses are used to convey discourse information such as switching topics (Noordman et al., 1999; Smith, 2004), and this added context may determine the purpose of the content and influence its meaning. Groen and colleagues (2019) postulated that poor comprehenders may not generate an implicit prosody pattern by default when reading connected text.

Unfortunately, much of the work in the area has been conducted with adult samples only. It is assumed that its conscious use may begin around adolescence, and that children may not be expected to convey disambiguation or discourse information in their prosody yet (Chen, 1998; Snedeker & Trueswell, 2003; Wells & Peppé, 2003). Regardless, studies utilizing children prior to adolescence demonstrate that prosody contributes unique variance to reading comprehension (e.g., Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008), likely without the utilization of disambiguation. This suggests that prosody may aid in comprehension through disambiguation in adolescents and adults, but prosody offers other means of support during childhood.

The unique variance prosody contributes to reading comprehension seems to vary as a function of prosody’s development alongside the other two aspects of text reading fluency. Schwanenflugel and colleagues (2004) found that prosody (using spectrographic indicators of pitch and pausing) is not yet strong enough to bear unique variance in reading comprehension in Grades 1 and 2. The authors postulated, though, that aspects of the passage used may have made some measures unreliable. Conversely, Paige and colleagues (2017) found that prosody (measured wholistically with the MDFS) mediates the relation between text reading efficiency and reading comprehension in Grades 1 through 3, and that prosody bears unique variance in reading comprehension. Further, Sabatini and colleagues (2019) found reading comprehension to
be correlated most with rate and then with prosody (measured by the NAEP 4-Point Fluency Scale) in Grade 4. Both studies used rating scales to measure prosody, meaning replication studies using spectrographic analyses would generate stronger support for these findings. Still, these findings support the hypothesis that prosody contributes to reading comprehension exponentially as children develop their reading skills. However, these somewhat mixed findings and the slight broadening of the grade levels between these two studies raise the question of how early the prosodic contributions to reading comprehension appear and with what degree of variability.

Godde and colleagues (2020) synthesized the nascent body of research conducted on prosody into a proposed developmental timeline, which suggests that the prosodic skills of pause and pitch control begin developing as early as Grade 1. As young readers increase their text reading efficiency, fewer pausal intrusions occur as a reflection of overall reading skill improvement (Clay & Imlach, 1971). Godde and colleagues (2021) suggest pausal intrusions are nearly eliminated by Grade 3, while other forms of pausing are reduced across adulthood. Pitch contour, which is generally quite weak until about Grade 5, becomes akin to natural speech in adulthood. However, readers as young as Grade 3 may begin to utilize unidirectional pitch movements (Godde et al., 2020). Specifically, skilled young readers tend to decrease pitch when concluding declarative statements and increase pitch when concluding interrogative statements that can be answered with a simple yes or no (Binder et al., 2013; Kim et al., 2021; Schwanenflugel et al., 2004). Further, these unidirectional pitch movements are related to greater reading comprehension beyond word or text reading rate (Miller & Schwanenflugel, 2006).

Godde and colleagues (2020) speculate that decoding and rate are the most important skills for comprehension around Grades 1 and 2, but prosody relates to reading comprehension
as reading skills increase around Grade 3 and beyond. One perspective is that, through the reduction of ungrammatical pauses, young readers can efficiently generate idea units and increase reading rate to the benefit of the working memory, thereby increasing reading comprehension (Kuhn et al., 2010). Pitch seems to reflect decoding skill and indirectly contribute to reading comprehension in Grades 1 and 2 (Schwanenflugel et al., 2004). In Grade 3, children have demonstrated an increase in pitch variation and improved grammatical pausing when reading particularly difficult text (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006). Children with high text reading efficiency were significantly better at demonstrating proper prosody with difficult text, but children with low text reading efficiency still demonstrated improved prosody, which reflects prosody’s role in supporting comprehension.

Beginning readers initially develop automaticity at a faster rate than prosodic skills (Paige et al., 2017). Text reading efficiency—particularly accuracy—must begin to develop before prosody, as one cannot read with appropriate expressivity if one cannot read. Findings from spectrographic studies suggest that pause skills in early childhood contribute to pitch skills in supporting text reading efficiency rather than making direct contributions, but that both pitch and pausing make direct contributions to text reading efficiency by Grade 3 (Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2017). These findings demonstrate the scaffolding pitch, pauses, and text reading efficiency provide one another, thus supporting the finding by Cowie and colleagues (2002) that prosody accounts for unique variance in reading comprehension beyond text reading efficiency while being highly interconnected to rate and accuracy.

**Dimensionality of Prosody**

Kim and colleagues (2021) examined the dimensionality of text reading fluency using
longitudinal data from Grades 1 through 3. Text reading efficiency and reading prosody (measured by the MDFS and spectrographic analyses for pitch and pause indicators) were examined. It was found that prosody as measured by the rating scale was strongly related to text reading efficiency, likely due to the conflation of the MDFS with decoding skills. With the spectrographic prosody measures, pitch was weakly correlated with text reading efficiency, whereas pausing was highly correlated. A confirmatory factor analysis provided support for the multidimensional nature of text reading fluency and revealed a trifactor structure of text reading rate, accuracy, and prosody. It was also found, however, that a general text reading fluency factor (reflecting a common skill across rate, accuracy, and prosody) was the most reliable means of predicting reading comprehension. It is worth noting that this study examined Grades 1 through 3, and prosody is thought to make greater contributions to text reading fluency and reading comprehension as development progresses. As such, the specific dimensions of text reading fluency may become more distinctly related to reading comprehension as development continues.

Kim and colleagues (2021) also investigated reading prosody’s relation to reading comprehension in the context of reading variables beyond text reading efficiency (i.e., word reading and listening comprehension). The general text reading fluency factor served to mediate the relation between word reading and reading comprehension as well as between listening comprehension and reading comprehension. The authors speculate that this mediating relation is due to text reading fluency encompassing lower-level reading comprehension skills rather than higher-order skills such as making inferences. It is possible that, as reading development continues and prosody plays a greater role in such tasks as disambiguation of syntax and providing discourse information, the role of text reading fluency in higher-order reading
comprehension may increase. Indeed, studies using prosodic rating scales in Grades 4 and 5 have found that prosody uniquely predicts reading comprehension after controlling for other reading variables (Groen et al., 2019; Veenendaal et al., 2014, 2015). These findings suggest that, sometime around Grade 4, typically developing children may experience a shift toward greater utilization of prosody to support reading comprehension beyond many other aspects of text reading skills. Thus, it is important to explore the dimensions of text reading fluency around this critical point.

**Purpose**

The current study seeks to separate the constructs of text reading fluency to investigate the unique contributions prosody makes to reading comprehension beyond text reading efficiency. Researchers have called for increased use of spectrographic analysis of prosody, especially among the understudied upper elementary school years (Wolters et al., 2020). Further, the ability to parse the features examined within prosody (i.e., pitch and pausing) can influence findings in relation to reading comprehension, especially as a function of development. For instance, ungrammatical pausing tends to reflect the poor decoding skills of young or inexperienced readers (Godde et al., 2020), whereas pitch is more closely related to text reading efficiency and, by extension, reading comprehension in young readers (Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2015). Additionally, as prosody develops, pitch and pausing may assume a greater role in the reading comprehension process (Godde et al., 2020). Although pitch and pausing have been identified as separate but related constructs (e.g., Kim et al., 2021), the literature remains mixed regarding whether prosody should be considered singular or multidimensional within the text reading fluency construct.

The current study fills important gaps in the literature by examining the dimensionality of
prosody and its role in supporting comprehension in Grade 4. Additionally, prosody is separated into the features of pitch and pausing in the current study to determine whether each feature is independent from text reading efficiency (i.e., rate and accuracy) at this grade level. The following research questions were used to guide the current study:

1. What is the dimensionality of prosody for students in Grade 4 (i.e., does prosody present a single dimension, or are pitch and pauses distinct dimensions)?
2. Does prosody mediate the relation between text reading efficiency and reading comprehension in Grade 4?

Although some aspects of the current study are exploratory in nature, it is anticipated that a two-factor solution of prosody will best fit the data, represented by the related but separate constructs pausing, and pitch. Once the dimensionality of prosody is determined using factor analysis (see Figure 1), the relations of the text reading fluency factors and reading comprehension will be examined using structural equation modeling. Two models will be examined, a baseline model where prosody and text reading efficiency are depicted as predictors of comprehension (see Figure 2) and a model where prosody mediates the relation between text reading efficiency and comprehension (see Figure 3).

Methods

Participants

Participants were drawn from a larger study examining benchmark reading measures. Data from 376 Grade 4 students were available, but given the time-consuming nature of spectrographic analyses, it was necessary to select a subsample of participants for the current project ($N = 200$). The following criteria were used to select for study participation: (a) attendance in Grade 4, (b) i-Ready and Lexplore data available within the fall time point
collected no more than 70 days apart, and (c) were not classified as English language learners.

After screening and following the selection criteria ($N = 331$), 17 students were dropped due to missing i-Ready data ($N = 314$). Another seven students were removed due to missing Lexplore data either from administrative error or disambiguating background noise during audio collection ($N = 307$). Finally, two students were dropped due to exceeding the 70-day limit between test administrations ($N = 305$), and 12 participants were not administered the target passage at the correct time point ($N = 293$). A randomizer was used to further select participants from this pool. The resulting subsample included 200 Grade 4 students attending four elementary schools in a public school district in the western part of the United States. After screening the sample of 200, two additional participants were dropped due to missing fluency data. The remaining 198 participants included 4% American Indian/Alaska Native students, 9% Black/African American students, 7% Other identifying students, and 81% White students. Additionally, 50% were boys, and 41% identified as Hispanic/Latino.

To serve as a comparison group for pitch contour and grammatical pause placement, an adult baseline sample was collected using the same passage read by children in this study. According to past studies, approximately 30 adults are sufficient to establish prosodic norms of adult readers (Benjamin & Schwanenflugel, 2010 used 20 adults; Miller & Schwanenflugel, 2006 used 29 adults; Miller & Schwanenflugel, 2008 used 34 adults). As such, 30 adult participants were recruited using the University of Memphis psychology research participant pool. Participants received credit towards completion of a course upon submission of audio recordings. A screener for demographic information (e.g., gender, racial and ethnic group membership) was used to recruit adults with as similar characteristics as possible to the child sample (77% women, 7% Asian, 7% Black/African American, 7% Hispanic, 80% White).
Measures

Data used in this study were collected as part of the computer-based benchmark assessment battery used by the participating school district in the fall, winter, and spring of the school year. However, data available from the fall benchmark was used in this study. Lexplore was used as a measure of text reading efficiency and i-Ready Diagnostic was used to measure foundational reading skills. Audio recordings of students’ oral reading of the Lexplore passages were available for secondary analysis of prosody. Adults will read the same Grade 4 passage read by students to generate the baseline data needed for the prosodic analysis.

Text Reading Efficiency

Data regarding students’ text reading fluency was provided by Lexplore, a new screening measure that uses eye-tracking technology and machine learning models to identify students at-risk for reading difficulties (Seimyr, 2021). Lexplore was administered by school personnel using a laptop computer and an external screen (15.6-inch screen with full HD 1920x1080 resolution and 100% display scale settings) equipped with a Tobii Eye Tracker 4C. After completing a brief eye calibration exercise, students read two brief grade-level passages presented on the computer screen, the first was read aloud and the second silently, and answered brief comprehension questions for each passage to promote reading for understanding. The program is fully automated and takes about 5 minutes per student to complete. Prosodic variables (i.e., pitch and pauses) as well as text reading efficiency variables (oral reading accuracy, oral reading rate, and silent reading rate) were used in the current study. The oral passage at the Grade 4 level has 9 sentences and 62 words in passage. The passage is narrative, describing a child who goes to the movies with her friend, and contains only declarative sentences. The silent passage at the Grade 4 level has 10 sentences and 57 words in passage. The passage is narrative, describing a child who goes
to school while it is snowing, and contains only declarative sentences.

Reading rate was measured as the average number of words read per minute across the passage, calculated as the number of words read in the passage divided by the time in minutes spent reading the passage. Time spent reading each passage is recorded automatically by the Lexplore software, along with children’s oral reading of the passage. Audio recordings were used to score reading errors to calculate the accuracy, determined by the number of words that were not read aloud correctly, as well as the reading rate for each passage. Reading errors included mispronunciations or substitutions of words, skipping words, and pauses of at least three seconds before reading the word. Inter-rater reliability for this scoring was strong ($r = .993$). Data from the technical manual reported an overall test-retest reliability for Grades 1-8 of .85 for passages administered one week apart and strong concurrent relation between Lexplore and i-Ready ($r = .75$) (Seimyr, 2021). When fall Lexplore scores for Grades 3 to 8 were used to identify those at-risk for not passing an end-of-year state test (i.e., the California Assessment of Student Performance and Progress), an overall AUC of .81 was found.

**Prosody**

The prosody features of pitch and pausing were measured using spectrographic analyses within Praat v6.1.16 as applied to the audio of students and adults reading the Lexplore passage. Past research has found that pitch contour (Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004), pitch declinations concluding declarative statements, and pitch rises concluding yes-no interrogative statements differentiate between skilled and poor readers (Binder et al., 2013; Miller & Schwanenflugel, 2006). However, the passage used in the current study does not include interrogative statements. Therefore, pitch contour and pitch declinations will be used to measure the construct of pitch in the current study. Additionally, pitch variation has not been
studied in the context of reading comprehension previously, so it will be included in an exploratory fashion.

First, pitch contour was measured by taking the highest Hertz value within each spoken word throughout the passage to create a list of pitch values. These values were compared to the adult sample by correlating each child participant’s pitch value per word with the adult sample’s averaged pitch value per word. Second, the pitch declinations concluding declarative statements were measured by subtracting the lowest Hertz value from the final spoken word of each sentence from the pitch peak of the preceding word and correlated to the adult sample’s average pitch declinations. Finally, pitch variation was calculated as the absolute difference between the highest and lowest pitch recorded per sentence correlated to the adult sample’s average pitch variation.

In the current study, only pauses indicated by the adult sample were considered grammatical. Pauses exceeding 100 milliseconds (Miller & Schwanenflugel, 2006) were recorded for duration and frequency across the passage. According to the averaged adult sample, no within-sentence pauses were grammatical, though all between-sentence pauses were grammatical. As such, calculating ungrammatical pausing and grammatical pausing separately would be redundant, and therefore intersentential pause duration and intersentential pause frequency were calculated instead for more rich pausal information. Intersentential pause duration was determined by the total number of milliseconds spent pausing above the 100-millisecond cut point between all sentences. Intersentential pause frequency was determined by totaling the number of times at least one >100 millisecond pause was demonstrated between each sentence across the passage. Notably, some child readers repeated or inserted sentences during their reading of the passage, meaning that both a higher or lower frequency of pauses than the
eight intersentential pauses demonstrated in the adult sample constitutes pausal errors.

Reading Comprehension

The i-Ready Diagnostic is adaptive, computerized measure of reading that is used for progress monitoring and benchmark screening within classroom settings (Curriculum Associates, 2018). Grade 4 students are first administered 18 items in the Vocabulary domain, followed by 18 items in the Reading Comprehension: Literature domain, followed by 18 items in the Reading Comprehension: Informational Text domain. After completing these subtests, the test flow is adjusted according to the overall estimated ability score. Additional domains related to basic reading (i.e., phonics, high frequency words) are administered only if a student falls below grade-level expectations on these three domains, and the additional domains contribute to the overall Reading Domain score in those instances. As such, the domain scores rather than the Reading Domain score will be used as indicators of reading comprehension in this study to ensure only the three core domains are included across all participants.

The Vocabulary domain requires students to read a sample sentence containing a target word and then select the meaning of the target word from a list of options. The Reading Comprehension domains require students to click within the passage or drag sentences into a response box to answer presented questions regarding the passage. The number of correct responses influences the number of Reading Comprehension items presented to the student, so the ratio of correct to incorrect responses forms the score for these two domains.

The Reading Domain yields scaled scores for each domain ranging from 100 to 800 and an overall score following interactive Item Response Theory (IRT) modeling. According to the test manual, the overall score produced a test-retest reliability across Grades 2 through 6 between .85 to .86 following a 12-to-18-week time interval (Curriculum Associates, 2018). The mean
standard error of measurement for the overall score fell between 10.0 and 10.5, and marginal reliability fell between .96 and .97. When comparing the overall score from the Reading Domain to the Lexile Linking Test, concurrent validity fell between .88 and .89. When comparing the i-Ready Diagnostic for Grades 3 through 8 to end of year state tests (i.e., the Smarter Balanced Assessment Consortium test, the Partnership for Assessment of Readiness for College and Careers test, the New York State Testing Program test, and the Florida Standards Assessments test), the median correlations fell between .77 to .81 in the fall, .77 to .83 in the winter, and .78 to .84 in the spring. The Vocabulary domain demonstrates a marginal reliability of .88 at Grade 4.

The Reading Comprehension: Literature domain and the Reading Comprehension: Informational Text domain demonstrate a marginal reliability of .90 and .91 at Grade 4, respectively (Curriculum Associates, 2018).

**Procedure**

All data were collected by school personnel (primarily classroom teachers) as part of the school’s routine assessment battery during the 2018-2019 school year. i-Ready Diagnostic reading measures were administered in a group setting, whereas Lexplore measures were administered individually within the classroom or another quiet location in the school. Lexplore reading assessment requires approximately 5 minutes to administer. The i-Ready Diagnostic typically requires over an hour, including both the reading and math subtests. Fall data were collected by the district between September and November, and students’ test dates for i-Ready and Lexplore fell within two months of one another. Due to the lack of identifying information collected during audio recording of students’ readings, these data were deemed not human subjects and submission for IRB approval was not required. The collection of the adult sample was approved by the University of Memphis Institutional Review Board (#PRO-FY2022-196).
Lexplore audio recordings were imported into Praat v6.1.16 and processed by three graduate students and one undergraduate student. These raters delineated boundaries between spoken words for later computerized processing of all pitch and pause data. The raters were given 10 training audio files prior to processing an initial 20 audio files from the current study for intrarater reliability. Intraclass correlation coefficient (ICC) estimates were calculated using SPSS v28.0.1.1 based on a mean-rating (k = 4), absolute-agreement, two-way random-effects model. Values less than 0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.9 indicate excellent reliability. The raters’ ICC was 0.98 with a 95% confidence interval of 0.97 to 0.98, indicating excellent agreement.

**Analytic Plan**

Structural equation modeling was used to examine the dimensionality of prosody and its relation to text reading efficiency and reading comprehension in Grade 4. A confirmatory factor analysis was conducted first to evaluate the proposed measurement models (see Figure 1a-1c). Then, the structural relations among these factors would be analyzed across the proposed baseline and mediation models (see Figures 2a & 2b). All analyses were conducted in MPlus v8.6 (Muthén & Muthén, 2017). Maximum likelihood estimations with robust standard errors (MLR) was used to handle missing data and to account for any multivariate outliers. A total of 13 indicators and four factors were examined, with the sample size of 198 being adequate for analyses in accordance with Tabachnick & Fidell (2019).

Following Kline (2016), model fit was evaluated by several indices including the chi-square and degrees of freedom, Steiger-Lind root mean square error of approximation (RMSEA; Steiger, 1990) with its 90% confidence interval, Bentler comparative fit index (CFI; Bentler,
1990), and weighted root mean square residual (WRMR; Bentler, 1990). A significant chi square value indicated a lack of model fit, as it evaluates the null hypothesis. Regarding RMSEA using a 90% confidence interval, a model would be considered a poor fit to the population if the lower bound of the confidence interval was ≤.05, whereas the null hypothesis of poor population fit would be rejected if the upper bound of the confidence interval equaled or exceeded .10 (Steiger & Fouladi, 1997). Regarding the Comparative Fit Index (CFI), values greater than .95 indicate good model fit as compared to a model in which all variables are assumed to correlate. Regarding SRMR, values below 0.10 indicate adequate fit (Hu & Bentler, 1999).

Standardized and unstandardized regression coefficients were used to examine the direct and indirect relations among the variables within each model (i.e., pitch, pauses, text reading efficiency and reading comprehension). Using Kline’s recommendations (2016), standardized coefficient effect sizes above .05 were considered small, values above .15 considered moderate, and those above .25 considered large. Unstandardized coefficients were used for determining statistical significance while standardized values were used for interpretation.

Results

Data Processing and Screening

Available data were screened for outliers, out of range values, univariate and multivariate normality, multicollinearity, linearity, automated flagging within the i-Ready systems, and extreme noise in the audio recordings via computerized screening of unreadable Hertz values within Praat and manually by human raters. Invalidating errors such as interruption during testing, poor recording quality, or violating instructions, generate “yellow” and “red” flags for responses provided in less than 15 seconds and 11 seconds, respectively. None of the i-Ready data were flagged. Univariate outliers were rare, constituting less than 0.01% of all data (z >
3.29; Tabachnick & Fidell, 2019). All outliers were substituted with the nearest acceptable extreme value in accordance with Tabachnick & Fidell (2019). No data were missing. However, Mahalanobis Distance was used to screen for multivariate outliers, and six multivariate outliers ($p < .001$) were identified and removed. As such, the final data set includes 0.002% missing data. Univariate normality statistics (see Table 1, which displays raw scores prior to adjusting variances or converting pitch variables to z-scores using Fisher’s r to z) fell within acceptable limits for all variables for SEM analyses ($<|3.0|$ for skewness, $<|10.0|$ for kurtosis; Kline, 2016). The correlation matrix was screened for multicollinearity ($r > .80$). Only the correlation between the i-Ready literary and informational subtests exceeded this threshold, both of which were expected to load on the latent reading comprehension factor.

Variables used in this study included response times (intra- and intersentential pause durations), reading rate, reading accuracy, frequency counts (intra- and intersentential pause frequencies), correlations (for pitch variables), and scaled scores (for i-Ready variables). The pitch variables were converted from correlations to z-scores using Fisher’s r-to-z transformation. The pitch variables, reading comprehension and rate variables, and intrasentential pause durations were multiplied or divided by 10 to maintain the recommended 1:10 ratio for variance (Kline, 2016). The pitch variables were reversed coded so that indicators for latent variables would be in the same direction.

**Measurement Model**

Following the nascent literature, a two-factor solution was expected for prosody (see Figure 1b), with three indicators loading onto the pitch factor (i.e., pitch contour, pitch declinations, and pitch variability) and four indicators loading onto the pauses factor (i.e., intrasentential pause duration, intrasentential pause frequency, intersentential pause duration, and
intersentential pause frequency). Three indicators were expected to load onto the reading comprehension factor (i.e., i-Ready vocabulary, i-Ready reading comprehension: Literature, and i-Ready reading comprehension: Informational text), and three indicators were expected to load onto the text reading efficiency factor (i.e., oral reading accuracy, oral rate, and silent rate).

First, the reading comprehension and text reading efficiency latent factors were examined in a confirmatory factor analysis to serve as a basic model (see Figure 1a). The factor loadings of the indicators were analyzed using an asymptotic factor covariance matrix of the correlations and error covariance matrix with reported standardized coefficient estimates. As expected, fit indices suggested that the basic model fit well (see Table 4). The factor loadings, all statistically significant \( p < .001 \), were consistently strong across the two latent variables except for a moderate factor loading of oral reading accuracy on text reading efficiency (see Table 3).

Second, the pitch and pause latent factors were added to the model (see Figure 1b). However, an examination of the correlations among the indicators for the pitch factor suggested that these indicators did not relate well to one another and would likely not constitute a latent variable (see Table 2). Two of the proposed indicators were not related to one another (pitch contour and pitch declination; pitch contour and pitch variation). Further, pitch declination was weakly related pitch contour, but it was not significantly correlated to any other variables in the model across the prosodic (pause), reading efficiency, and reading comprehension indicators. This model yielded an inadmissible solution. Mplus provided a warning statement that the latent variable covariance matrix is a not positive definite, which was caused by the correlation between the pause and pitch latent factors exceeding plausible values (i.e., 1.0). These results indicated model misspecification when separating prosody into two distinct dimensions and instead suggested that prosody may represent a single latent construct.
Next, a single-factor prosody model which combined the pitch and pause factors was conducted (see Figure 1c). As would be expected given the correlational results, pitch declinations did not load significantly on the prosody latent factor and was dropped from the model. Modification indices pointed to the need for correlated residuals among the pause variables, which follows theoretically given that the pause variables were originally conceptualized as a latent factor. As presented in Table 4, the fit indices suggested the single-factor prosody model adequately fit the observed data. All factor loadings for the prosody latent factor demonstrated statistical significance (see Table 3). However, the prosody and text reading efficiency latent factors approached a perfect correlation (\(r = -0.99\)), indicating that these two factors likely represented a single latent construct. Moderate to strong correlations were present between all four pause variables and the two reading rate variables (i.e., oral and silent). It is likely that the nature of the passages used in this study allowed for little difference in pause duration, pause frequency, and reading rate, due to the lack of grammatically indicated pauses. In other words, the pause and rate variables in this study all measured the same latent factor.

Given that the results pointed to strong overlap between the prosody and text reading efficiency latent variables, it was necessary to conduct an ad-hoc model combining these two factors into a single unified fluency factor (see Figure 1d). The fit indices for the unified fluency factor model adequately fit the observed data (see Table 4). All factor loadings demonstrated significance (see Table 3). As anticipated, the reading comprehension factor and the unified fluency factor were significantly correlated with one another (\(r = 0.69\)).

**Structural Models**

The purpose of the current study was to explore the dimensionality of prosody alongside text reading efficiency in relation to reading comprehension. However, the factor analyses did
not support the presence of distinct prosodic latent factors and the only viable model, the unified fluency model, does not align well with the nascent literature. Therefore, it appeared worthwhile to explore contributions of prosody without a latent structure for its contributions to reading comprehension. In previous studies (see Godde et al., 2020; Benjamin & Schwanenflugel, 2010 with Grade 2 children; Binder et al., 2013 with adults; Schwanenflugel et al., 2004 with Grade 2 and 3 children; Valle et al., 2013 with Grade 2 children), pitch contour and intrasentential pause durations are among the variables most related to reading comprehension. As such, using these two commonly implicated prosody variables, a prosody indicator model (see Figure 3) was analyzed. In this ad hoc model, pitch contour and intrasentential pause duration were retained as observed variables alongside the text reading efficiency and reading comprehension latent factors in a final attempt to separate the constructs and link findings to prior literature.

As seen in Table 4, the fit indices suggested that the prosody indicator model demonstrated close approximate fit to the observed data. In analyzing the AIC and BIC values, this model demonstrated a difference greater than 10 compared to the single factor prosody and unified prosody models, which indicated its superior fit. In this model, text reading efficiency was found to contribute significantly to reading comprehension (.50). Notably, neither of the observed prosody variables (pitch contour and intrasentential pause duration) were found to make a significant contribution, directly or indirectly, to reading comprehension (see Table 5). Variables in the model were found to explain 49% of the variation in reading comprehension.

**Discussion**

The purpose of the current study was to explore the dimensionality of reading prosody in relation to reading comprehension in a Grade 4 student population. Many studies researching English reading prosody utilize subjective rating scales (e.g., Bolanos et al., 2013; Lai et al.,
2014; Paige et al., 2017; Klauda & Guthrie, 2008), which have proven to be less reliable than objective measures such as computerized extraction and spectrographic analysis of prosodic features (Benjamin & Schwanenflugel, 2010; Cowie et al., 2002). Of the studies using computerized assessment, most have found that prosody contributes uniquely to reading comprehension separate from text reading efficiency (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008), and that prosody contributes directly to text reading efficiency as well (Schwanenflugel & Benjamin, 2017). Notably, most English language studies have examined children in Grades 1-3 (Benjamin & Schwanenflugel, 2010; Kim et al., 2020; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel et al., 2004, 2015; Schwanenflugel & Benjamin, 2017; Valle et al., 2013), a time when prosody is still early in development due to reliance on a still-budding mastery of text reading efficiency skills (Godde et al., 2020; Miller & Schwanenflugel, 2008, Paige et al., 2017; Sabatini et al., 2019). As such, the current study fills an important gap in the current literature in studying a Grade 4 population using computerized assessment of prosody. Further, the current study sought to address the directionality of prosody, with some prior research and theory (Kuhn & Stahl, 2003; Paige et al., 2017; Sabatini et al., 2019) pointing toward a mediating relation between text reading efficiency and reading comprehension.

This study examined pitch and pause features alongside the text reading efficiency features of oral reading rate, oral reading accuracy, and silent reading rate as well as the reading comprehension features of vocabulary, informational text comprehension, and literary text comprehension. Results from this study were generally inconsistent with most extant literature concerning the dimensionality of prosody and its role relative to text reading efficiency. Instead, support was identified for a unified construct of fluency which combines aspects of prosody and
text reading efficiency. Counter to Kuhn and colleagues’ (2010) conceptualization of unified fluency, support was not found in the current study for an equal contribution of prosody alongside text reading efficiency when analyzing variance in reading comprehension. However, these unexpected results may be explained by characteristics of the sample or the text stimuli.

The first research question regarding the dimensionality of prosody found no support for the existence of prosody as a separate construct from text reading efficiency, much less a multidimensional prosody construct. As such, the second research question regarding mediation of the relation between text reading efficiency and reading comprehension also found no support. Indeed, no direct or indirect effects were found between pitch contour, intrasentential pause duration, and reading comprehension. In pursuit of answering these questions, a series of measurement and structured models were explored.

The first model simply confirmed the text reading efficiency factor and reading comprehension factor prior to introducing prosodic features, and this model demonstrated strong fit. In the second model, a pitch latent factor and a pause latent factor were added, but pitch contour did not correlate with the other pitch variables. Additionally, the pause factor overlapped entirely with the pitch factor, creating a nonpositive definite. As such, a two-factor structure demonstrating multidimensionality within prosody was not supported in this study. When combined into one greater prosody latent factor, pitch declinations was dropped due to low correlations among prosodic variables. Additionally, modification indices suggested adding correlated residuals between the pause variables. Although this model provided adequate fit to the data, the prosody and text reading efficiency latent factors demonstrated near-perfect construct overlap, meaning that prosody was not supported as a separate construct from text reading efficiency in this study. A fourth model was added in which all text reading efficiency
variables, all pause variables, pitch contour, and pitch variability were combined into one unified fluency factor. While this model did adequately fit the data, these results do not explain prosody’s role (or lack thereof) in the data. As such, a fifth model was constructed for exploratory purposes.

The fifth model followed prior literature to select the aspects of prosody most likely to contribute directly to reading comprehension or indirectly through text reading efficiency. Pausal intrusions, particularly intrasentential intrusions, as well as pitch contour appeared most implicated in the literature among pitch and pause variables (Benjamin & Schwanenflugel, 2010 with Grade 2 children; Binder et al., 2013 with adults; Schwanenflugel et al., 2004 with Grade 2 and 3 children; Valle et al., 2013 with Grade 2 children) and were thus selected for structural modeling. Specifically, given that three of the pause variables were allowed to correlate in the previous model, intrasentential pause durations were selected over frequency for this ad hoc model. Notably, Miller & Schwanenflugel (2008) found the largest effect sizes for pitch contour and intrasentential pause frequency among various prosody variables but ultimately only found support for pitch contour in predicting fluency outcomes. Benjamin & Schwanenflugel (2010) utilized pitch contour, sentence-final pitch change, and intrasentential pause frequency (as a ratio of grammatical to ungrammatical pauses) and found that all were significantly correlated to text reading efficiency and reading comprehension, particularly pitch contour. Kim and colleagues (2021) found that pause frequencies as well as text reading efficiency loaded negatively and strongly onto the general prosody factor. Pause durations loaded negatively and moderately onto the general prosody factor, and pitch variation loaded negatively and weakly. Pitch contour did not load at all, indicating that it may be a separate construct from text reading efficiency. However, Miller and Schwanenflugel (2006) found no support for pitch contour or pausing, but
instead, for sentence-final pitch rises and declinations.

This study yielded no support for contributions, direct or indirect, of pitch contour or intrasentential pause durations toward reading comprehension. In explaining the current results, one explanatory factor was readily apparent in reviewing the raw data—the passage used appeared to be relatively easy for a Grade 4 population. This is expected given that the passage used in this study is a CBM probe designed to be relatively simple, which ensures that it was accessible to readers with a wide skill range. Although scores range as low as 44 WCPM, the average Grade 4 student in this study read 138 WCPM, which is well above the common fall benchmark standards (e.g., 94 WCPM corresponds to the 50th percentile; Hasbrouck & Tindall, 2017). However, past studies have indicated that prosodic features are often underutilized in passages that are easily comprehended by the reader, as there is no cognitive demand for additional supports (Benjamin & Schwanenflugel, 2010).

These potential explanatory factors align with prior literature. Benjamin and Schwanenflugel (2010) demonstrated in their sample of Grade 2 children that prosody independently explained reading comprehension of difficult text after controlling for text reading efficiency, whereas it did not explain reading comprehension of easy text. Binder and colleagues (2013) found that adults with low literacy skills demonstrated a significant link between decoding, automaticity, and pausing, indicating that rate and accuracy are likely to account for pauses in lower fluency groups. Pausing also accounted for little variance in reading comprehension. Regarding pitch, pitch contour and sentence-final pitch rises differed greatly between low literacy adults and their skilled counterparts, suggesting that low-skilled readers’ decoding skills better explains their reading and comprehending skills than does prosody. In combination, these two previous studies suggest that the skilled readers in the current study may
underutilize pitch and avoid appropriate pausing because they found the text easy. On the other hand, the text reading efficiency skills of the poorer readers in the current study may completely explain their reading comprehension with little or no room left for prosody (or its absence) to contribute.

Also worth noting is the lack of intrasentential grammatical pause opportunities in the passage, which restricted fluent readers to displaying appropriate prosody only by the lack of intrasentential pause features and the few intersentential pause opportunities. Some researchers investigate grammatical and ungrammatical pausing (Benjamin et al., 2013; Benjamin & Schwanenflugel, 2010). An ungrammatical pause frequency count may be used, or a pause ratio is generated that compares the number of ungrammatical pauses to the number of all pauses (both grammatical and ungrammatical). Although the current study did not provide opportunities for grammatically implicated intrasentential pausing, this study did capture ungrammatical pausing both in duration and in frequency. This study may also point to an argument for increasing the threshold of milliseconds constituting a pause, as the mean intrasentential pause frequency in this study was high compared to the intrasentential pause duration mean. In other words, the intrasentential pauses, on average, were frequent but extremely brief and may not necessarily indicate ungrammatical pausing.

This study seemingly contributes little knowledge to research regarding the dimensionality of prosody, as prosody was—evidently—poorly represented in the current study. However, this study may offer greater insight into the relation between prosody and reading comprehension in combination with text difficulty. Regarding automaticity theory (LaBerge & Samuels, 1974), this study provides support for the idea that prosody is utilized only after text reading efficiency has been mastered, as there are no spare cognitive resources to put toward
prosodic reading. Regarding the working memory and implicit prosody theories (Baddeley, 1992; Miller & Schwanenflugel, 2008; Fodor, 1998; Kuhn et al., 2010), prosody may only be utilized when the cognitive cost to implement prosody for improved text structure is less than that of reading an entire text. In other words, only when the text is particularly challenging and requires prosodic scaffolding is it worthwhile for skilled readers to implement prosody for improved working memory or syntactic disambiguation and, ultimately, reading comprehension. This raises the questions of when average-skill readers may utilize prosody and how to prosodically define an average reader.

**Limitations and Future Directions**

Several limitations of the current study should be considered. As previously discussed, the use of a CBM passage designed to elicit fluent reading served as a limiting factor toward the goals originally presented in this study, not only in content difficulty but also in presenting sufficient opportunities for key aspects of measuring prosody (e.g., pitch rises, ungrammatical pauses). The use of a CBM probe may have provided greater ecological validity in that prosodically designed texts are not representative of text typically used in fluency assessments (whereas CBM probes are often required), but it also introduced several unanticipated limitations.

The demographic fit between the child and adult samples was not ideal. Demographic and regional differences such as dialect are understudied but likely impact the unique profile of prosodic speech and reading (Frazier et al., 2006; Godde et al., 2020). As such, an adult sample with greater similarity to the child sample would likely yield more valid results. Every attempt was made to match the child demographics given the pool of adults who were available for participation. Further, this study features a lack of diversity of English speakers, as English
language learners were excluded from the child sample. According to the U.S. Department of Education, an estimated 13.1% of students in United States public schools are classified as English language learners (Hussar et al., 2020) and future studies should include these readers for enhanced generalizability.

Finally, the quality of recordings, both physically with microphone quality and environmentally with background noise and speech, may have interfered with the raw data’s accuracy in some cases. Severe cases were excluded through human and computerized screening. Manual screening involved rater impressions of clarity, as the inability to interpret a spoken word prevented accurate transcription for spectrographic analysis. Additionally, computerized screening occurred by excluding cases with excessive Hertz values that were returned as “undefined” (i.e., unreadable) by the Praat software. Still, all recordings presented with some degree of noise, which is realistically unavoidable in school-based data collection. The benefit of this method of data collection is, in part, the ecological validity of Grade 4 students reading in a classroom setting with distractions and also, in part, the volume of data collected compared to the smaller, more refined samples collected in most prosodic studies.

Regarding future directions, this study clearly points toward the need for more research examining prosodic dimensionality in late elementary schoolers using varied text complexity with intentional design toward eliciting prosodic reading. Given the converging evidence for the relation between dialect density (i.e., the degree of dialect use) and reading growth (Washington et al., 2018), future studies should also make greater efforts to clearly identify dialect and understand its relation with reading prosody. Further, regarding English language learners, students in years prior to Grade 4 have increasingly higher representation in the school-aged population (up to 15.9% representation in Kindergarten within United States public schools;
Hussar, et al., 2020). Given that most prosodic studies investigate early readers, this population deserves investigation. This point extends to children with learning disabilities as well, as noted by Kim and colleagues (2021). Children with cognitive impairments, reading disabilities, or executive functioning difficulties would, arguably, benefit most from early prosodic interventions that may spawn from specialized investigation.

The current study also points toward the need for combining elements of text difficulty and reader skill, as few studies have examined either, particularly together (Benjamin & Schwanenflugel, 2010 for text difficulty by reader skill; Binder et al., 2013 for reader skill). It may be valuable to separate participants by text reading efficiency skill, reading comprehension, or prosodic skill to examine their influence in the context of various levels of text difficulty. Regarding this study, it may be worthwhile to split this child sample into low and high prosodic groups to examine how prosody may moderate the relation between text reading efficiency by group and subsequent comprehension. Finally, interventions for reading comprehension must consider when and how to prioritize text reading efficiency and prosodic scaffolding, considering both the reader’s text reading efficiency skill and the relative text difficulty.
References


https://doi.org/10.1126/science.1736359


https://doi.org/10.1598/RRQ.45.4.2


https://doi.org/10.1016/S0022-5371(71)80004-X


Godde, E., Bailly, G., & Bosse, M. L. (2021). Pausing and breathing while reading aloud:
Development from 2nd to 7th grade in French speaking children. *Reading and Writing.*

https://doi.org/10.1007/s11145-021-10168-z


http://dx.doi.org/10.1080/02702710600846894


http://dx.doi.org/10.1080/10705519909540118


http://dx.doi.org/10.1080/10888438.2010.493964


physiological data to study silent pauses). *Actes de la conférence conjointe JEP-TALN-RECITAL, 2016*(1), 563–570. https://hal.archives-ouvertes.fr/hal-01462180


Research Online, 9(2), 245-275.


Figure 1a

*Basic Model*

![Diagram of Basic Model]

Figure 1b

*Two Factor Prosody Model*

![Diagram of Two Factor Prosody Model]
Figure 1c

*Single Factor Prosody Model*

**Figure 1d**

*Unified Fluency Model*
**Figure 2a**

*Proposed Baseline Model*

**Figure 2b**

*Proposed Mediation Model*
Figure 3

*Prosody Indicator Model*
Table 1

*Descriptive Statistics for Prosody, Reading Comprehension, & Text Reading Efficiency Variables (N = 198)*

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch Contour</td>
<td>0.27</td>
<td>0.20</td>
<td>0.92</td>
<td>-0.49</td>
<td>0.03</td>
</tr>
<tr>
<td>Pitch Declinations</td>
<td>0.15</td>
<td>0.33</td>
<td>1.57</td>
<td>-0.11</td>
<td>-0.67</td>
</tr>
<tr>
<td>Pitch Variation</td>
<td>0.36</td>
<td>0.16</td>
<td>0.86</td>
<td>-0.22</td>
<td>-0.25</td>
</tr>
<tr>
<td>Intrasentential Pause Duration</td>
<td>3.93</td>
<td>3.81</td>
<td>16.16</td>
<td>1.69</td>
<td>2.38</td>
</tr>
<tr>
<td>Intrasentential Pause Frequency</td>
<td>32.10</td>
<td>19.26</td>
<td>64.00</td>
<td>0.01</td>
<td>-1.47</td>
</tr>
<tr>
<td>Intersentential Pause Duration</td>
<td>3.46</td>
<td>1.89</td>
<td>9.11</td>
<td>1.14</td>
<td>1.27</td>
</tr>
<tr>
<td>Intersentential Pause Frequency</td>
<td>7.56</td>
<td>1.32</td>
<td>7.00</td>
<td>-0.89</td>
<td>1.16</td>
</tr>
<tr>
<td>iReady Vocabulary</td>
<td>530.46</td>
<td>58.58</td>
<td>334.00</td>
<td>-0.31</td>
<td>0.43</td>
</tr>
<tr>
<td>iReady RC Literary Text</td>
<td>527.49</td>
<td>56.40</td>
<td>269.00</td>
<td>-0.50</td>
<td>-0.26</td>
</tr>
<tr>
<td>iReady RC Informational Text</td>
<td>523.28</td>
<td>59.04</td>
<td>275.00</td>
<td>-0.15</td>
<td>-0.63</td>
</tr>
<tr>
<td>Oral Reading Rate</td>
<td>127.30</td>
<td>33.31</td>
<td>177.00</td>
<td>0.04</td>
<td>-0.12</td>
</tr>
<tr>
<td>Silent Reading Rate</td>
<td>140.61</td>
<td>43.34</td>
<td>231.00</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>Oral Reading Accuracy</td>
<td>1.35</td>
<td>1.71</td>
<td>7.00</td>
<td>1.68</td>
<td>2.58</td>
</tr>
</tbody>
</table>

*Note. M = mean, SD = standard deviation.*
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Pitch Contour</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Pitch Declinations</td>
<td>-.06</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Pitch Variation</td>
<td>.11</td>
<td>.25***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Intrasentential Pause Duration</td>
<td>.25***</td>
<td>-.01</td>
<td>.42***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Intrasentential Pause Frequency</td>
<td>.10</td>
<td>-.02</td>
<td>.22**</td>
<td>.46***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Intersentential Pause Duration</td>
<td>.09</td>
<td>-.07</td>
<td>.07</td>
<td>.36***</td>
<td>.17*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Intersentential Pause Frequency</td>
<td>.08</td>
<td>-.01</td>
<td>-.00</td>
<td>.15*</td>
<td>.52***</td>
<td>.45***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>i-Ready Vocabulary</td>
<td>-.22**</td>
<td>-.03</td>
<td>-.29***</td>
<td>-.58***</td>
<td>-.20**</td>
<td>-.27***</td>
<td>.01</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>i-Ready RC Literary Text</td>
<td>-.24***</td>
<td>-.07</td>
<td>-.28***</td>
<td>-.58***</td>
<td>-.21**</td>
<td>-.34***</td>
<td>-.04</td>
<td>.79***</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>i-Ready RC Informational Text</td>
<td>-.22**</td>
<td>-.07</td>
<td>-.27***</td>
<td>-.55***</td>
<td>-.18*</td>
<td>-.30***</td>
<td>-.03</td>
<td>.74***</td>
<td>.84***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Oral Rate</td>
<td>-.23***</td>
<td>.05</td>
<td>-.33***</td>
<td>-.80***</td>
<td>-.38***</td>
<td>-.60***</td>
<td>-.24***</td>
<td>.56***</td>
<td>.60***</td>
<td>.56***</td>
<td>-</td>
</tr>
<tr>
<td>12.</td>
<td>Silent Rate</td>
<td>-.18*</td>
<td>.02</td>
<td>-.32***</td>
<td>-.66***</td>
<td>-.29***</td>
<td>-.48***</td>
<td>-.21**</td>
<td>.57***</td>
<td>.56***</td>
<td>.51***</td>
<td>.80***</td>
</tr>
<tr>
<td>13.</td>
<td>Accuracy</td>
<td>.11</td>
<td>.06</td>
<td>.18*</td>
<td>.36***</td>
<td>.09</td>
<td>.24***</td>
<td>.01</td>
<td>-.33***</td>
<td>-.33***</td>
<td>-.28***</td>
<td>-.38***</td>
</tr>
</tbody>
</table>

*Note.* Correlations presented after converting pitch variables into z-scores using Fisher’s r to z and after adjusting variances. 
*p < .05, **p < .01, ***p < .001.*

48
Table 3
Factor Loadings for the Single Prosody Factor Model

<table>
<thead>
<tr>
<th>Factor Loadings</th>
<th>Factor Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading Comprehension Factor</strong></td>
<td><strong>Reading Comprehension Factor</strong></td>
</tr>
<tr>
<td>i-Ready Vocabulary</td>
<td>.84*</td>
</tr>
<tr>
<td>i-Ready RC Literary Text</td>
<td>.94*</td>
</tr>
<tr>
<td>i-Ready RC Informational Text</td>
<td>.89*</td>
</tr>
<tr>
<td><strong>Text Reading Efficiency Factor</strong></td>
<td><strong>Text Reading Efficiency Factor</strong></td>
</tr>
<tr>
<td>Oral Reading Rate</td>
<td>.96*</td>
</tr>
<tr>
<td>Silent Reading Rate</td>
<td>.82*</td>
</tr>
<tr>
<td>Oral Reading Accuracy</td>
<td>.42*</td>
</tr>
<tr>
<td><strong>Prosody Factor</strong></td>
<td><strong>Prosody Factor</strong></td>
</tr>
<tr>
<td>Pitch Contour</td>
<td>.26*</td>
</tr>
<tr>
<td>Pitch Variation</td>
<td>.38*</td>
</tr>
<tr>
<td>Intersentential Pause Duration</td>
<td>.87*</td>
</tr>
<tr>
<td>Intersentential Pause Frequency</td>
<td>.41*</td>
</tr>
<tr>
<td>Intrasentential Pause Duration</td>
<td>.60*</td>
</tr>
<tr>
<td>Intrasentential Pause Frequency</td>
<td>.24*</td>
</tr>
<tr>
<td><strong>Text Reading Fluency Factor</strong></td>
<td><strong>Text Reading Fluency Factor</strong></td>
</tr>
<tr>
<td>Oral Reading Rate</td>
<td>.96*</td>
</tr>
<tr>
<td>Silent Reading Rate</td>
<td>.82*</td>
</tr>
<tr>
<td>Oral Reading Accuracy</td>
<td>.42*</td>
</tr>
<tr>
<td>Pitch Contour</td>
<td>-.26*</td>
</tr>
<tr>
<td>Pitch Variation</td>
<td>-.37*</td>
</tr>
<tr>
<td>Intersentential Pause Duration</td>
<td>-.86*</td>
</tr>
<tr>
<td>Intersentential Pause Frequency</td>
<td>-.38*</td>
</tr>
<tr>
<td>Intrasentential Pause Duration</td>
<td>-.61*</td>
</tr>
<tr>
<td>Intrasentential Pause Frequency</td>
<td>-.26*</td>
</tr>
</tbody>
</table>

*Note. *p ≤ .001
Table 4
Model Fit Indices

<table>
<thead>
<tr>
<th></th>
<th>$X^2$ (df, p-value)</th>
<th>CFI</th>
<th>RMSEA (90% CI)</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic TRE RC Model</td>
<td>9.83 (8, .278)</td>
<td>.99</td>
<td>.03 (.00, .09)</td>
<td>.03</td>
<td>6012.82</td>
<td>6075.30</td>
</tr>
<tr>
<td>Prosody Indicator Model</td>
<td>23.20 (16, .108)</td>
<td>.99</td>
<td>.05 (.00, .09)</td>
<td>.03</td>
<td>7727.65</td>
<td>7819.72</td>
</tr>
</tbody>
</table>
### Table 5

**Standardized Path Coefficients of Structural Equation Models**

<table>
<thead>
<tr>
<th></th>
<th>Direct effects on RC</th>
<th>Indirect effects on RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text Reading Efficiency</td>
<td>.493**</td>
<td>-</td>
</tr>
<tr>
<td>Pitch Contour</td>
<td>-.186</td>
<td>-.082</td>
</tr>
<tr>
<td>Intrasentential Pause Duration</td>
<td>-.253</td>
<td>-.205</td>
</tr>
</tbody>
</table>

*Note. *p < .05, ** p < .01, ***p < .001*