Scientific Studies of Reading

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hssr20

Is Reading Prosody Related to Reading Comprehension? A Meta-analysis

Alissa P. Wolters , Young-Suk Grace Kim & John William Szura

To cite this article: Alissa P. Wolters , Young-Suk Grace Kim & John William Szura (2020): Is Reading Prosody Related to Reading Comprehension? A Meta-analysis, Scientific Studies of Reading

To link to this article: https://doi.org/10.1080/10888438.2020.1850733

Published online: 08 Dec 2020.

SCIENTIFIC STUDIES OF READING
https://doi.org/10.1080/10888438.2020.1850733

Is Reading Prosody Related to Reading Comprehension? A Meta-analysis

Alissa P. Wolters, Young-Suk Grace Kim, and John William Szura

School of Education, University of California, Irvine, California, USA

ABSTRACT

We examined the relation between reading prosody and reading comprehension, using a systematic review and meta-analysis to estimate the strength of the relation and to understand whether the strength of the relation varies by prosody feature (adult-like contour, F0 sentence-final declination, grammatical pauses, ungrammatical pauses, prosody scale), students' developmental phase of reading skill as examined by grade level, and orthographic depth. A total of 35 studies (N = 9,349; Grades 1–9, 8 languages) met inclusion criteria. Overall a moderate relation (.51) was found between reading prosody and reading comprehension. Furthermore, the strength varied by prosody feature such that the relation was stronger for prosody rating scale than for pitch indicators such as adult-like contour and F0 sentence-final declination. However, grade and orthographic depth were not significant moderators. These results suggest that the relation between reading prosody and reading comprehension is not unitary and should consider specific aspects of reading prosody.

The ability to read connected text with speed, accuracy, and expression (reading fluency) is an important skill for reading comprehension (Kim, 2015, 2020a, 2020b; Kuhn, Schwanenflugel, & Meisinger, 2010; National Institute of Child Health and Human Development [NICHD], 2000; Pikulski & Chard, 2005; Wolf & Katzir-Cohen, 2001). Students who struggle with fluent reading are often found to have difficulty with reading comprehension (Sabatini, Wang, & O'Reilly, 2018). Although the definition of reading fluency includes three aspects, accuracy, speed, *and* expression (i.e., prosody; Hudson, Lane, & Pullen, 2005; Kuhn et al., 2010), the majority of reading fluency research and tools used for classroom assessment, such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (Good & Kaminski, 2002), do not include expression or reading prosody (Dowhower, 1991; Kuhn et al., 2010), and only measure speed and accuracy (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, Petscher, Schatschneider, & Foorman, 2010; Kim & Wagner, 2015; Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Silverman, Speece, Harring, & Ritchey, 2013; Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009). Although extant work indicates a relation between reading prosody and reading comprehension, we do not have a solid understanding of this relation. The current work explores this relation using a systematic review and meta-analysis to estimate the magnitude and to examine potential moderators such as prosody feature (e.g., Benjamin, 2012; Kim, Quinn, & Petscher, 2020; Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004), developmental phase of reading (Calet, Gutiérrez-Palma, & Defior, 2015; Fernandes, Querido, Verhaeghe, & Araújo, 2018; Miller & Schwanenflugel, 2008; Rasinski, Rikli, & Johnston, 2009; Veenendaal, Groen, & Verhoeven, 2016), and orthographic depth (Hussien, 2014; Veenendaal et al., 2016).

CONTACT Young-Suk Grace Kim youngsk7@uci.edu Young-Suk Grace Kim School of Education, University of California, Irvine, CA, USA.

Reading prosody and reading comprehension

Reading prosody is prosodic rendering when reading connected text (i.e., not lexical prosody or prosodic sensitivity, e.g., Kim & Petscher, 2016; Schwanenflugel & Benjamin, 2017; Wood, 2006). There are two alternative views on the relation between reading prosody and reading comprehension. One view is that reading prosody, as a part of the reading fluency construct, plays a role in reading comprehension by facilitating syntactic (i.e., chunking constituents) and semantic (i.e., deriving meaning from read words and phrases) processing (Dowhower, 1991; Kuhn et al., 2010; Schreiber, 1980, 1987). That is, reading prosody allows one "to hold an auditory sequence in working memory" and "assists in maintaining an utterance in working memory until a more complete semantic analysis can be carried out" (Kuhn et al., 2010, p. 237). According to an alternative perspective, reading prosody is an indicator or an outcome of reading comprehension. That is, prosodic reading – reading with appropriate intonation, grouping of words into meaningful units, and pausing in appropriate places – relies on at least some level of text comprehension, such as knowledge of syntactic structures, especially since written text typically does not contain graphic cues to mark constituents (e.g., punctuation to denote Schreiber, 1987, 1991), and, therefore, is a product of comprehension (Davies, 1994; Dowhower, 1991; Ravid & Mashraki, 2007).

Regardless of different views, however, reading prosody is expected to be related to reading comprehension. Indeed, extant studies have found a positive relation between prosodic reading and reading comprehension, and the relation varies in magnitude. For example, for monolingual English speakers, reading prosody and reading comprehension had a weak relation in second and third grade when reading prosody was measured directly by spectrographic analysis (r =.11-.30; Schwanenflugel et al., 2004), a moderate relation in fourth grade using a holistic rating scale where a single score is assigned after evaluating multiple aspects (e.g., NAEP fluency scale; r = .59; Sabatini et al., 2018), and a strong relation in ninth grade using an analytic rating scale where a score is assigned to each of different aspects (e.g., Multidimensional Fluency Scale; r = .71; Paige, Rasinski, Magpuri-Lavell, & Smith, 2014). Somewhat similar magnitudes have been found with monolingual fourth graders of Dutch (r = .41; Veenendaal et al., 2016) and Turkish (r = .44; Yildiz & Çetinkaya, 2017) with analytic rating scales. Variation in the magnitudes of the relation between reading prosody and reading comprehension may be explained by the feature of prosody measured, reading development, and orthographic depth.

Reading prosody features

Prosody is characterized by intonation, stress, duration, and pausing (Couper-Kuhlen, 1986; Kuhn et al., 2010; Schreiber, 1980, 1987). Prosody conveys paralinguistic information that supports a listener's comprehension of a speaker's intended meaning and emotion (Wilson & Wharton, 2006). For example, in English, a Wh-question declines in pitch and volume at the end of the sentence, whereas a yes/no question increases in pitch and volume at the end of a sentence. Reading prosody has been primarily measured in two ways, by rating scale or spectrographic analysis. Rating scales capture the listener's perceptions of reading prosody such as expressiveness, phrasing, smoothness, pace, and deviations from text on a scale (e.g., 1 to 4). On the other hand, spectrographic analysis is done through analysis of a spectrogram, a visualization of sound waves (Denes & Pinson, 1993), by precise measurements of sound waves to directly measure pause structure in milliseconds such as pause duration and frequency (e.g., grammatical and ungrammatical pauses), and pitch changes such as intonation contour (i.e., vocalic nuclei) and F0 changes (e.g., F0 sentence-final declination) in Hertz (number of sound wave cycles per second; Crystal, 2011). Although both approaches measure reading prosody, they measure different features or aspects in different ways, and, thus, it is unclear whether the features and measures are comparable or whether the relation of reading prosody and reading comprehension varies as a function of the feature of reading prosody.

Numerous reading prosody rating scales have been introduced since the 1980s (e.g., Allington, 1983; NAEP oral reading fluency scale by Daane et al., 2005; Six Dimensions Fluency Rubric by Pinnell & Fountas, 2010; the Multidimensional Fluency Scale by Zutell & Rasinski, 1991; prosodic map by Ravid & Mashraki, 2007), but two are most commonly used. One is the NAEP oral reading fluency scale (Daane et al., 2005) which uses a four-point scale to measure phrasing, deviations from text, syntax, and expression holistically together. The other is the Multidimensional Fluency Scale (Rasinski, 2004, adapted from; Zutell & Rasinski, 1991) which uses a 16-point analytic scale where each of the four categories (expression and volume, phrasing, smoothness, and pace) has its own 4-point rating scale that is then added up to get the total score out of 16. Although this is an analytic scale, typically studies only report a total prosody score. These scales have been adapted for speakers of languages other than English such as Spanish (González-Trujillo, Calet, Defior, & Gutiérrez-Palma, 2014) and Turkish (Yildiz et al., 2014), and were found to be reliable and valid (Daane et al., 2005; Moser, Sudweeks, Morrison, & Wilcox, 2014; Rasinski, 2004; Smith & Paige, 2019). A recent study with English-speaking students in Grades 1 to 3 showed that the four aspects of the Multidimensional Fluency Scoring Guide loaded onto a single latent variable together with pause structures measured by spectrographic analysis (i.e., ungrammatical pause duration and frequency; Kim, Quinn, et al., 2020). In contrast, pitch or intonation features such as F0 sentence-final declination and intonation contour formed a separate latent variable (Kim, Quinn, et al., 2020). These results indicate that the Multidimensional Fluency Scale likely captures the decoding-related prosody aspect similar to previous studies that reported a strong relation of pause structure reading prosody to decoding skills (Benjamin & Schwanenflugel, 2010; Binder et al., 2013; Kim, Quinn, et al., 2020; Miller & Schwanenflugel, 2008; Schwanenflugel et al., 2004).

Reading prosody has also been examined using spectrographic analysis. Because spectrographic analysis allows for the exact duration of a pause to be measured in milliseconds and for change in fundamental frequency (F0) to be measured in Hertz, studies using spectrographic analysis measured reading prosody features such as pause structure and pitch variation. Schwanenflugel and colleagues (Benjamin & Schwanenflugel, 2010; Miller & Schwanenflugel, 2006, 2008; Schwanenflugel et al., 2004) captured duration and frequency of pauses, adult-like F0 (pitch) contour (through vocalic nucleic matching), and F0 sentence-final declination (difference in Hertz from one wave peak to the next). Using data from English-monolingual second graders, they found weak to moderate magnitudes for the relations of various aspects of reading prosody to reading comprehension, ranging from r = .03 for ungrammatical pauses to r = .31 for F0 sentence-final declination (Schwanenflugel et al., 2004). However, another study with English-monolingual second graders found stronger magnitudes overall, ranging from rs = .21-.36 for adult-like contour to r = .59 for grammatical pauses (Benjamin & Schwanenflugel, 2010). Clearly, not only are there inconsistent findings across the features of reading prosody, but also across studies examining the same features.

Reading development

Reading fluency is influenced by development because decoding skill – the ability to sound out a real or nonsense words based on grapheme-phoneme correspondence knowledge – constrains the ability to read with speed, accuracy, *and* prosody during connected text reading (Kim, Quinn, et al., 2020; Kuhn et al., 2010; Schwanenflugel et al., 2004). According to the automaticity theory (LaBerge & Samuels, 1974), a reader needs to have proficient lower order skills (i.e., decoding) so that working memory resources are available to chunk text together (i.e., morphosyntax; Schreiber, 1987) and construct meaning (semantic processing) to support prosodic reading. A recent study showed that word reading strongly predicted reading prosody for English-speaking children in Grades 1 to 3 (Kim, Quinn, et al., 2020). If word reading is the primary driver of reading prosody at least in the beginning phase of reading development, then reading prosody captures word reading skills to a large extent in the beginning phase. In the later phase, however, reading prosody is expected to facilitate semantic processing (Kuhn et al., 2010) or is a function of semantic processing (Davies, 1994; Ravid & Mashraki, 2007) as the constraining role of decoding is lifted with reading development (Adlof, Catts, & Little, 2006; Florit & Cain, 2011; Kim, 2015).

If reading prosody as a construct captures both decoding and semantic processes, then the relation between reading prosody and reading comprehension might not change largely as a function of reading development. However, if specific measures of reading prosody capture different *aspects/features* of reading prosody – whether different reading prosody aspects primarily capture decoding or semantic processing – then the relation between reading prosody and reading comprehension would differ as a function of reading development for different features. For reading prosody features that primarily capture decoding skills (e.g., pause structure such as inappropriate or ungrammatical pauses; Arcand et al., 2014; Binder et al., 2013), the relation between reading prosody and reading comprehension will be stronger in the beginning phase of development and will become weaker at a more advanced phase. In contrast, for reading prosody features that primarily capture semantic processing (e.g., pitch indicators such as child-adult pitch [F0] match, or F0 sentence-final declination; Binder et al., 2013; Schwanenflugel et al., 2004), the relation will be weaker in the beginning phase of reading development and will become stronger with reading development.

A small number of studies have explored whether the relation between reading prosody and reading comprehension varies over developmental phases. The limited longitudinal studies yielded inconsistent findings even when reading prosody was measured by the same approach, a rating scale. A study with English-monolingual second graders showed similar strengths in magnitudes in the fall and winter, r = .77 and r = .76, respectively, and a weaker magnitude in the spring, r = .59 (Lai, Benjamin, Schwanenflugel, & Kuhn, 2014). A study in European Portuguese found that the relation was moderate for students in Grade 2 (r = .38) and Grade 3 (r = .31), whereas the relation was weaker for students in upper elementary grades (r = .18 in Grade 4 and r = .06 in Grade 5; Fernandes et al., 2018). A longitudinal study with Dutch-monolinguals reported r = .39 in fourth grade, r = .39 in fifth grade, and r = .60 in sixth grade (Veenendaal et al., 2016).

Orthographic depth

At the center of the theoretical accounts of the relation between reading prosody and reading comprehension is semantic processing (see above; Chafe, 1994; Davies, 1994; Koriat, Greenberg, & Kreiner, 2002; Kuhn et al., 2010), but semantic processing, including reading prosody, is constrained by word reading skill (Kim, Quinn, et al., 2020; Kuhn et al., 2010). Therefore, reading prosody is a function of and captures both word reading and semantic processing. Consequently, two hypotheses with regard to orthographic depth are reasonable. First, the relation between reading prosody and reading comprehension is expected to vary by orthographic depth as a function of reading prosody feature and reading development. The rate at which a reader develops decoding varies as a function of the orthographic depth of the language – word reading acquisition occurs at a faster rate in shallow orthographies (Aro & Wimmer, 2003; Ellis et al., 2004; Katz & Frost, 1992; Seymour, Aro, & Erskine, 2003). Therefore, word reading would play a constraining role for a longer time in deep orthographies (e.g., kindergarten to Grade 2 in English) than in shallow orthographies (e.g., Grade 1). This was supported in a meta-analysis such that the relation between word reading fluency and reading comprehension was .79 for children (Grades 3–5) learning to read in English, a deep orthography, whereas it was .48 (Grades 3–5) for children learning to read in more shallow orthographies (Florit & Cain, 2011). Then, for languages with transparent orthographies, reading prosody features that primarily capture decoding skills (e.g., ungrammatical pausing) would show a short-lived strong relation between decoding-related reading prosody and reading comprehension in comparison to deep orthographies (i.e., relation becomes weaker at an earlier grade). Consequently, reading prosody features that capture intonation modulation (adult-like contour, F0 sentence-final declination) – drawing from semantic processing skills – would have a stronger relation with reading comprehension at an earlier grade in shallow orthographies than in deep orthographies.

The second possibility is that the *overall* relation between reading prosody and reading comprehension is stronger in deep orthographies. In deep orthographies, accurate word reading requires knowledge of morphological and morphosyntactic skills to a greater extent than in shallow orthographies because spellings of words represent morphemes as well as phonemes (e.g., Joshi, Treiman, Carreker, & Moats, 2008/2009; McBride-Chang et al., 2005). If accurate word reading requires morphological and morphosyntactic processing to a greater extent in deep orthographies, and word reading constrains reading prosody, then reading prosody in deep orthographies would reflect morphosyntactic processing to a greater extent. In this case, the relation between reading prosody and reading comprehension would be stronger in deep orthographies than in shallow orthographies, given the role of morphological processing in reading comprehension (e.g., Carlisle, 2000; Frost, 2005; Kieffer, Biancarosa, & Mancilla-Martinez, 2013; Kieffer & Box, 2013; Kim, Guo, Liu, Peng, & Yang, 2020). This certainly does not deny that reading prosody captures sentence-level morphosyntactic and higher order semantic processing (e.g., communicative intent) – this is not expected to differ across orthographic depth. However, if morphological and morphosyntactic processing is captured in reading prosody to a greater extent in deep orthographies by way of word reading, then it seems plausible that the overall relation between reading prosody and reading comprehension may be stronger in deep orthographies.

Extant, although limited, research, however, does not suggest a clear picture about a differential relation by orthographic depth even for studies with students at similar grades. For example, two studies, Fernandes et al. (2018) and Calet et al. (2015), worked with fourth graders in shallow orthographies (European Portuguese and Spanish, respectively), using adapted versions of the Multidimensional Fluency Scoring Guide (Rasinski, 2004). In European Portuguese, a less shallow language than Spanish, reading prosody and reading comprehension had a weak relation, $r = .18$, whereas in Spanish, a more shallow and highly consistent orthography, they had a moderate relation, $r = .47$. Studies with fifth graders learning to read in English that also used the Multidimensional Fluency Scoring Guide found moderate to strong relations ($.49 \le r\text{s} \le .73$; Klauda & Guthrie, 2008; Mokhtari & Thompson, 2006; Rasinski et al., 2009; Sargent, 2002).

The current study

Reading prosody has long been considered as an important aspect of the text reading fluency construct (Kuhn et al., 2010; NICHD, 2000). Studies examined various aspects of reading prosody using different measurement approaches and showed varying relations between reading prosody and reading comprehension. To expand our understanding of the relation between reading prosody and reading comprehension, we addressed the following research questions. First, what is the average magnitude of the relation between reading prosody and reading comprehension? Second, does the strength of the relation differ by prosody feature with and without controlling for grade and orthographic depth? Third, does the relation vary as a function of reading development (using grade as a proxy) and orthographic depth?

We hypothesized that reading prosody and reading comprehension would be related, but the magnitude would differ by prosody feature. We posited that the relation may also vary as a function of reading development, but it would depend on aspects or features of reading prosody. We expected that orthographic depth would also moderate the relation, but it would depend on overall relation versus specific prosody features. For the overall relation, a stronger relation was expected in deep orthographies than in shallow orthographies, whereas for the specific prosody features, the moderation was expected to differ by developmental phase (see above for specifics). It should be noted, however, that we could not examine the latter (moderation by specific prosody features) due to a limited number of effect sizes available (see below).

Method

Literature search

The following databases were searched through ProQuest: Educational Resources Information Center (ERIC), APA PsycInfo, Sociological Abstracts, Linguistics and Language Behavior Abstracts (LLBA), Dissertations & Theses Global, and ProQuest Dissertations & Theses A&I. No study was excluded based on peer review or publication status. In addition to keywords, names of authors who created prosody scales and the titles of the prosody scales (e.g., Rasinski, 2004) were also included. Although this is not conventional, this approach was taken because of the inconsistent use of the term "prosody" (e.g., some texts used "expression" or "reading fluency" instead). The following search terms were paired with *reading comprehension: prosody, Multidimensional AND Rasinski, NAEP scale AND oral reading fluency, spectrograph, Fountas AND Pinnell AND oral reading fluency*, Tindal and Marston (1996), *and* Allington (1983).

The following were inclusion criteria in the current systematic review and meta-analysis: (1) both reading prosody (of a connected text of more than one sentence) and reading comprehension were measured; (2) reading prosody was measured through either spectrographic analysis or a rating scale; (3) the primary participants (over 50% of the sample) were measured in their first language; (4) the primary participants had no severe disabilities such as intellectual disabilities and severe behavioral disabilities (this criterion affected very few studies, and studies with developmental language disorder or learning disabilities were included); (5) the study was published between 2000 and 2019; (6) the study was published in English; and (7) data were not from an experimental group after a reading prosody or reading comprehension intervention – for studies where an intervention was conducted, only the pretest and control group data were included.

Search results were uploaded to an online meta-analysis review tool that allows for two researchers to conduct a double-blind review (i.e., Covidence). The first author and a research assistant reviewed abstracts of results. There was full agreement on 90% of the articles. For the remaining 10%, a mutual decision was made after discussion. In addition, the following journals were digitally hand searched with the terms prosody and reading comprehension: *Reading Research Quarterly, Journal of Research in Reading, Reading and Writing Quarterly*, and *Reading and Writing*. Regarding the studies from the journals, there was 97% agreement between the primary and secondary authors – only two studies had discrepancies in the inclusion decision, and they were resolved in discussion. Finally, the reference sections of all related articles were examined and citation chained.

As can be seen in the PRISMA flow diagram (Figure 1), the systematic review and meta-analysis was conducted over several stages with two double-blind review screenings (Borenstein, Hedges, Higgins, & Rothstein, 2011). The first screening examined titles, keywords, and abstracts. Abstracts were screened for terms that might suggest that reading prosody and reading comprehension were measured (e.g., *reading comprehension, prosody, oral reading fluency, reading outcomes, expression*). Studies that passed the first screening were then reviewed for a second screening as full articles to determine whether they fit the inclusion criteria (i.e., qualitative synthesis, studies that collected the data on both reading prosody and reading comprehension). Studies that met inclusion criteria but did not include correlations between reading prosody and reading comprehension were handled in two ways so that they could be included in the quantitative synthesis. If the study included raw data such as students' reading prosody and reading comprehension scores, the relation was hand calculated. If no such information was given, then the author of the study was contacted via e-mail if their e-mail address could be found online through online search engines – three studies were collected this way (Hussien, 2014; Jefferson, Grant, & Sander, 2017; Schwanenflugel et al., 2004). The final set of studies that met our inclusion criteria with available data included 35 of the studies (52 unique samples, 98 effect sizes; N = 9,349) with 28 journal articles, 6 dissertations (master's and doctoral), and 1 paper presentation. There were no duplicated samples between the two (articles and dissertations) unless they were from different time points (e.g., longitudinal). The final included studies that were used for the quantitative synthesis are the final found in Table 1.

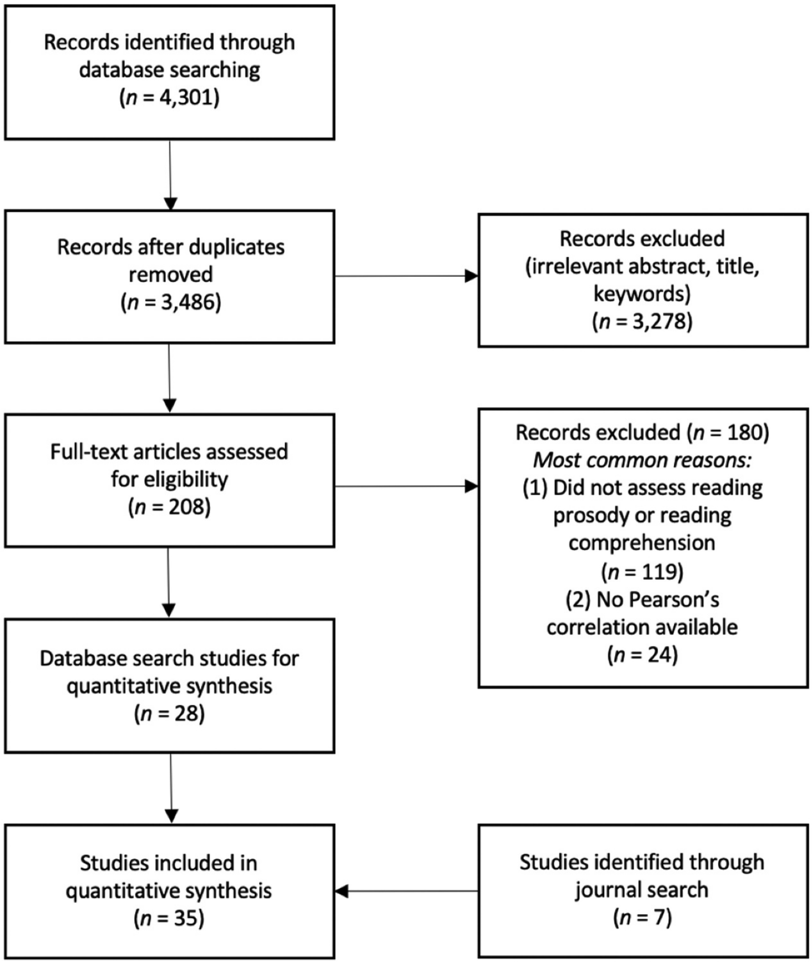

Figure 1. PRISMA flow diagram depicting the literature search process.

Coding procedures

All studies that met the inclusion criteria were coded for the following aspects: sample size, effect size (Pearson's r), participant grade (proxy for reading development), language, measurement method (i.e., rating scale or spectrograph), and feature of prosody measured (e.g., adult-like contour). All studies were coded by both the first author and the third author; there was 100% agreement. Language was also coded as a dichotomous variable (0 = shallow, 1 = deep). Arabic, Hebrew, English, and French were considered as opaque (deep) orthographies; Dutch, Turkish, Spanish, and European Portuguese were considered as shallow (Ellis et al., 2004; Seymour et al., 2003). Grade was examined as a continuous variable. Using the following equations, Pearson's r was converted to Fischer's z and variance was calculated from the sample size (Borenstein et al., 2011).

$$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right). \qquad V_z = \frac{1}{n-3}$$

Table 1. Studies included in meta-analysis.

Study	*N*	Pearson's *r*	Grade	Language	Prosody method
Arcand et al. (2014)	261	.03-.61	2	French (CAN)	GP, UP
Basaran (2013)	90	.10-.85	4	Turkish (TUR)	Scale
Benjamin (2012)[D] – *Study 1*	90	.43-.53	2	English (USA)	Scale
Benjamin and Schwanenflugel (2010)	90	.21-.70	2	English (USA)	ALC, F0, GP, UP
Benjamin et al. (2013) – *Study 2*	60	.46	3	English (USA)	Scale
Brown, Mohr, Wilcox, and Barrett (2018)	25	.51,.58	3	English (USA)	Scale
Calet et al. (2015)	50	.61	2	Spanish (ESP)	Scale
	48	.47	4	Spanish (ESP)	Scale
Dawson (2015)[D]	113	.43	6, 7, 8	African American English (USA)	Scale
Evanchan (2015)[D]	22	.29	2	English (USA)	Scale
Fernandes et al. (2018)*	81	.31,.38	2–3	European Portuguese (POR)	Scale
	76	.06-.18	4–5	European Portuguese (POR)	Scale
González-Trujillo et al. (2014)	74	.61	2	Spanish (ESP)	Scale
	48	.49	4	Spanish (ESP)	Scale
Groen, Veenendaal, and Verhoeven (2018)	63	.52	3	Dutch (NLD)	Scale
Hammer (2003)[D]	13	.06,.44	3	English (USA)	Scale
Hussien (2014)	448	.33	6	Arabic (EGY)	Scale
Jefferson et al. (2017)	83	.33-.44	3	English (USA)	Scale
	30	.19-.27	3	English (USA)	Scale
Kariuki and Baxter (2011)[P]	10	.88	2	English (USA)	Scale
Klauda and Guthrie (2008)	145	.67,.68	5	English (USA)	Scale
Lai et al. (2014)*	154	.59-.77	2	English (USA)	Scale
Marrone (2014)[D]	6	0	1	English (USA)	Scale
May (2014)[D]	68	.04-.54	5	English (USA)	ALC, F0, GP, UP
Miller and Schwanenflugel (2008)*	92	.24-.56	1–2	English (USA)	ALC, UP
Mokhtari and Thompson (2006)	32	.72,.73	5	English (USA)	Scale
Paige et al. (2014)	108	.71	9	English (USA)	Scale
Rasinski et al. (2017)	37	.12,.28	3	English (USA)	Scale
Rasinski et al. (2009)	391	.63	3	English (USA)	Scale
	421	.66	5	English (USA)	Scale
	392	.57	7	English (USA)	Scale
Ravid and Mashraki (2007)	51	.51	4	Hebrew (ISR)	Scale (map)
Sabatini et al. (2018)	1714	.59	4	English (USA)	Scale
Sargent (2002)	52	.22-.57	5	English (USA)	Scale
Schwanenflugel et al. (2004)	120	.11-.29	3	English (USA)	ALC, F0, GP, UP
Taylor, Meisinger, and Floyd (2013)	72	.48	3	English (USA)	Scale
Tortorelli (2018)	2,191	.09	2	English (USA)	Scale
Veenendaal et al. (2016)*	99	.39,.39,.60	4–6	Dutch (NLD)	Scale
Yildirim, Rasinski, and Kaya (2019)	100	.11	4	Turkish (TUR)	Scale
	100	.48	5	Turkish (TUR)	Scale
	100	.49	6	Turkish (TUR)	Scale
	100	.46	7	Turkish (TUR)	Scale
	100	.50	8	Turkish (TUR)	Scale
Yildiz and Çetinkaya (2017)	132	.44	4	Turkish (TUR)	Scale
Yildiz et al. (2014)	119	.45	5	Turkish (TUR)	Scale
Overall (35 studies, 52 samples)	*9,349*	*.51*	*1–9*	*8 languages*	*5 methods/features*

Studies marked by asterisk (*) were longitudinal; if underwent attrition, lowest sample size was reported above. Total meta-analysis sample (*N* = 9,349) counts longitudinal samples measured at different grade levels as unique. If more than two effect sizes were found, a range is reported. African American English is included with American English. Studies marked by superscript (D) are dissertations (master's or doctoral). Studies marked by superscript (P) are paper presentations. Hammer (2003) uses the terms "pre-test" and "post-test" but there was no instruction between assessment; one prosody assessment was a cold read and the other gave students time to practice first. Benjamin (2012) and Benjamin et al. (2013) are the same analysis, two different studies within each publication; due to the way data were presented, the data from Study 1 were coded from Benjamin (2012) and the data for Study 2 were coded from Benjamin et al. (2013). Benjamin (2012) analyzed data from Benjamin and Schwanenflugel (2010) with a different tool, thus, they were treated as the same sample. ALC = Adult-like contour, F0 = F0 sentence-final declination, GP = Grammatical pausing, UP = Ungrammatical pausing. Country abbreviations are as follows: CAN = Canada, POR = Portugal, EGY = Egypt, ESP = Spain, ISR = Israel, NLD = Netherlands, TUR = Turkey, USA = United States of America.

Data analysis

Data were uploaded into R (R Core Team, 2013) and the overall effect size, confidence intervals, and meta-regression were conducted with the *robumeta* package (RVE; Hedges, Tipton, & Johnson, 2010). *Robumeta* accounts for studies with small sample size in R using the robust variation estimator to apply appropriate weighting (Tipton, 2015). *Robumeta* calculates effect sizes using Fischer's z and variance to weight effect size statistics by sample size and then outputs an estimated effect size for interpretable results. Due to the sample variation, to account for samples who are members of a greater population and to accurately weight effect sizes, random effects were used instead of fixed effects (Kreft & de Leeuw, 1998; Viechtbauer, 2005). The statistics I^2 and Q show heterogeneity and whether analysis for a moderation effect is appropriate (Higgins & Thompson, 2002). The I^2 statistic revealed that approximately 93.12% of the total observed variance was due to differences between the studies rather than within-study sampling error. Given the significant heterogeneity, we conducted moderator analysis to identify the source of the between-study variation. The included moderators were prosody features (e.g., scale, adult-like contour), grade (as a proxy for reading development), and orthographic depth (0 = shallow, 1 = deep). *Robumeta* tests whether there is a statistical difference between the overall effect sizes of the groups by moderator and gives an interpretable p-value ($p < .05$ was considered as significant).

The following analytic procedures were carried out to address each research question. For the first research question, an average effect size ($k = 98$) was estimated. If multiple correlations were available from the same sample and time period due to multiple measures of reading prosody and reading comprehension, one average effect size was calculated in *robumeta* for each group. For the second research question, meta-regression was used with prosody features as predictors (a series of dichotomous variables) with and without controlling for grade and orthographic depth. For the third research question, meta-regression was fitted including grade and orthographic depth. Note that the analysis for the third question could not be conducted by reading prosody features (e.g., rating scale, F0 sentence-final declination, ungrammatical pauses) due to insufficient number of effect sizes per prosody feature (particularly for those that examined intonation modulation using spectrographic analysis such as F0 sentence-final declination).

Results

Question 1: what is the average magnitude of the relation between reading prosody and reading comprehension?

A final sample of 35 studies (52 unique samples, $N = 9{,}349$) with 98 effect sizes[1] was used in the analysis (see Table 1). The overall average correlation between reading prosody and reading comprehension was .51 (95% CI = [0.44, 0.57]; Table 2), and there was large variation in correlations, ranging from $r = 0$ to $r = .88$ (Figure 2).

Question 2: does the strength of the relation differ by prosody feature with and without controlling for grade and orthographic depth?

The strength of the relation between reading prosody and reading comprehension differed by reading prosody aspects, ranging from .31 (grammatical pauses) to .53 (rating scale; see Table 2). However, the

Table 2. The relations of various prosody measures with reading comprehension.

Feature	b	SE	CI.LB	CI.UB	k	Levels	I^2	Grades
Overall	0.51	0.03	0.44	0.57	98	52	93.12	1–9
Adult-like contour	0.32	0.03	0.24	0.40	8	6	0	2–5
F0 declination	0.34	0.08	0.09	0.60	6	4	35.00	2–5
Grammatical pauses	0.31	0.14	−0.07	0.69	8	5	89.08	2–5
Ungrammatical pauses	0.38	0.09	0.16	0.60	10	7	82.76	1–5
Prosody rating scale	0.53	0.04	0.46	0.60	66	46	93.58	1–9

The features were tested in separate meta-regression models. The I^2 statistic shows large heterogeneity across all estimated effect sizes except adult-like contour. Levels = number of unique samples. CI.LB = lower bound of 95% confidence interval; CI.UB = upper bound of 95% confidence interval.

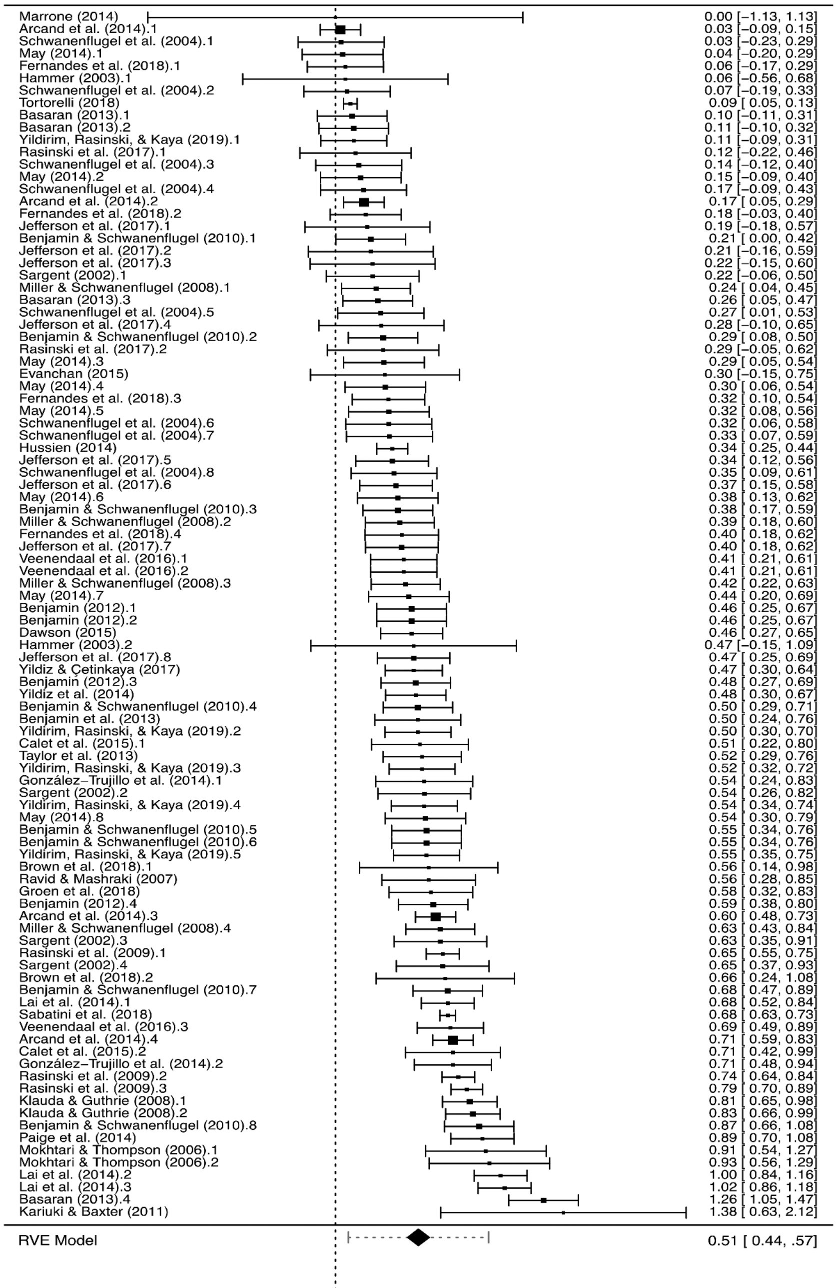

Figure 2. Forest plot of all effect sizes exhibiting the relation between reading prosody and reading comprehension.

only statistical difference in magnitudes was between adult-like contour (r = .33) and prosody rating scale (r = .53, p= .02) such that the relation between reading prosody and reading comprehension is stronger when reading prosody is measured by rating scale than by adult-like contour.

The relations remained essentially the same when grade and orthographic depth were added as covariates in the model (Table 3). The reference group in this model is rating scale (r = .41). Adult-like contour (r = .16) remained statistically significant (p= .03), indicating that its relation with reading comprehension is weaker than that between rating scale (the reference group) and reading

Table 3. The relation between reading prosody features and reading comprehension with prosody scale as the reference controlling for grade and orthographic depth.

Features	b	SE	p	CI.LB	CI.UB
Intercept	0.41	0.10	<.001	0.19	0.62
Adult-like contour	−0.25	0.09	0.03	−0.47	−0.03
F0 declination	−0.27	0.09	0.06	−0.56	0.01
Grammatical pauses	−0.34	0.13	0.06	−0.70	0.02
Ungrammatical pauses	−0.14	0.12	0.28	−0.43	0.15
Grade	0.01	0.02	0.58	−0.03	0.05
Orthographic depth	0.15	0.07	0.04	0.01	0.28

CI.LB = lower bound of 95% confidence interval; CI.UB = upper bound of 95% confidence interval.

comprehension after controlling for grade and orthographic depth. It should be noted that F0 sentence-final declination (*r*= .14) and grammatical pauses (*r* = .07) were just shy of reaching the conventional statistical significance (*p*s = .06).

Question 3: does this relation vary as a function of reading development (grade) and orthographic depth?

As shown in Table 4, the magnitude of relations did not differ by grade (*p* = .40) or orthographic depth (*p* = .18). When grade was controlled for, there was still no significant difference between reading prosody and reading comprehension (*p* = .11) in deep orthographies (*r* = .46) versus in shallow orthographies (*r* = .36).

Sensitivity analysis

Robust variation estimation

Sensitivity analysis was conducted with the *metafor* package (Viechtbauer, 2010). *Metafor* yielded highly similar results as *robumeta* for the relation between reading prosody and reading comprehension with *r* = .50.

Extreme sample size

To control for two studies that had a relatively large sample sizes compared to the other studies (Sabatini et al., 2018: *n* = 1,714, *r* = .59; Tortorelli, 2018: *n* = 2,191), analysis was conducted without these studies. Removing these studies did not change the overall magnitude, *r* = .51, suggesting that the large sample size did not affect the results.

Orthographic depth

European Portuguese has been generally considered as a shallow orthography (Defior, Martos, & Cary, 2002); however, some research has identified it as having intermediate depth (Seymour et al., 2003; Sucena, Castro, & Seymour, 2009). Therefore, analysis was conducted without the study on European

Table 4. The relation of reading prosody and reading comprehension with grade and orthographic depth as moderators.

Features	b	SE	p	CI.LB	CI.UB
Intercept	0.44	0.09	<.0001	0.26	0.62
Grade	0.02	0.02	0.40	−0.02	0.06
Intercept	0.45	0.04	<.0001	0.37	0.54
Deep orthographies	0.09	0.06	0.18	−0.04	0.21
Intercept	0.36	0.10	0.00	0.16	0.56
Deep orthographies	0.10	0.06	0.11	−0.02	0.23
Grade	0.02	0.02	0.25	−0.02	0.06

CI.LB = lower bound of 95% confidence interval; CI.UB = upper bound of 95% confidence interval.

Portuguese (Fernandes et al., 2018). This led to no evidence of a moderation effect by orthographic depth with or without grade as a control (*p*s = .60, .40, respectively).

Publication bias

Studies with statistically significant findings are often favored by journals for publication, which is known as "publication bias" (Sterne, Egger, & Smith, 2001). Effect sizes are expected to be evenly distributed around the overall estimated effect size when there is no publication bias. Figure 3 shows a funnel plot of the distribution of effect sizes. As can be seen from Figure 3, the studies are somewhat symmetric, but the studies do not fall in the white shaded area of the triangle, suggesting the heterogeneity of the studies. A random-mixed effects meta-regression model (weighted regression with multiplicative dispersion with standard error as the predictor; Egger, Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005) was run to statistically test whether correlations were asymmetrical around the mean. However, it did not

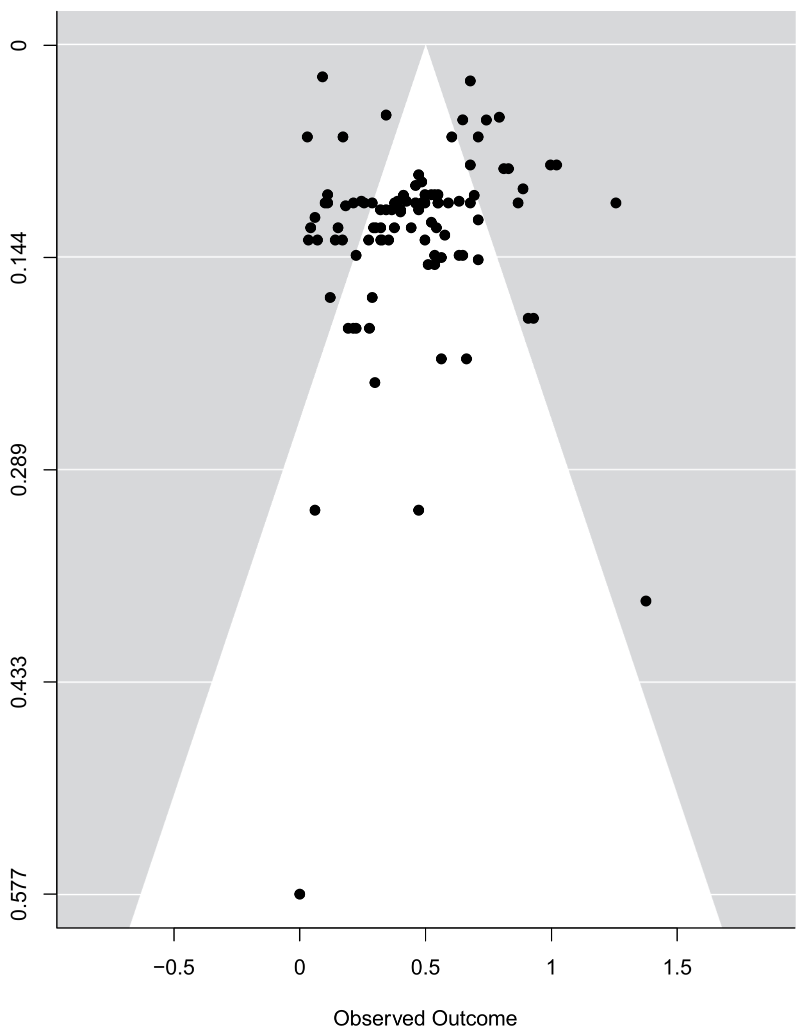

Figure 3. Funnel plot of effect sizes from all included studies.

reach the conventional significance level for publication bias, $z = -1.06$, $p = .29$, suggesting no evidence of publication bias.

Discussion

In this study, we investigated the relation between reading prosody and reading comprehension using a systematic review and meta-analysis. Our final sample consisted of 35 studies (52 unique samples, 98 effect sizes, $N = 9{,}349$), which included five different reading prosody features (rating scale, adult-like contour, F0 sentence-final declination, grammatical pauses, ungrammatical pauses) and readers from Grades 1 to 9 in eight languages with varying levels of orthographic depth (shallow: Turkish, Spanish, European Portuguese, Dutch; deep: French, Arabic, Hebrew, English).

Overall, reading prosody and reading comprehension were moderately related, $r = .51$. Beyond the average magnitude though, there was large variation in the strength of relations. Our hypothesis that this variation would be explained by reading prosody features was partially supported. Once grade and orthographic depth were accounted for, the relations of reading comprehension with adult-like contour were weaker than the relation between reading comprehension and rating scale. In contrast, rating scale did not have a stronger relation with reading comprehension than did ungrammatical pauses, grammatical pauses, and F0 sentence-final declination although for grammatical pauses and F0 sentence-final declination, there was a trend of their weaker relations than rating scale. These results appear to be in line with a recent study which showed that rating scale and pause structure prosody features (e.g., ungrammatical pauses) loaded onto a single latent variable, whereas the pitch aspect of reading prosody (intonation contour and F0 sentence-final declination) was related but a dissociable variable (Kim, Quinn, et al., 2020). Taken together, these results indicate that reading prosody is a multi-dimensional construct and that various measures of reading prosody tap into different aspects or dimensions of reading prosody (Kim, Quinn, et al., 2020), and, therefore, their relations to reading comprehension differ depending on the aspects.

It is unclear why the relation of reading prosody and reading comprehension is stronger when reading prosody is measured by a rating scale than by pitch or intonation measured by spectrographic analysis. One explanation is that rating scales capture multiple aspects. For example, the Multidimensional Fluency Scoring Guide (Rasinski, 2004) examines four categories, expression and volume, phrasing, smoothness, and pace. Although recent studies showed that all of these four aspects essentially capture a single construct with (Kim, Quinn, et al., 2020) and without pause structure indicators (Benjamin & Schwanenflugel, 2010), evaluating multiple aspects somehow might provide a richer picture of reading prosody, which in turn leads to a stronger relation with reading comprehension. Alternatively, the results likely reflect a limitation of the extant literature. As shown in Table 1, the number of studies using spectrographic analysis ($n = 5$) was extremely limited compared to those that employed rating scales ($n = 30$). Moreover, even the majority of studies using spectrographic analysis was conducted with students in primary grades (Grades 1 to 3, with one study in Grade 5), whereas the grade levels of studies that employed rating scales ranged from Grade 1 to Grade 9. In other words, the literature base is too skewed to represent a full picture about the relation of various reading prosody features. Future studies, particularly those that examine reading prosody with spectrographic analysis across the reading development phases (particularly with students in upper elementary and secondary schools), are warranted.

We explored whether orthographic depth moderates the relation between reading prosody and reading comprehension. We hypothesized that the overall relation between reading prosody and reading comprehension might be stronger in deep orthographies because reading prosody in deep orthographies likely reflects morphological and morphosyntactic knowledge to a greater extent than it does in shallow orthographies. However, our results showed no difference in magnitudes of the relations between reading prosody and reading comprehension as a function of orthographic depth ($\Delta r = .10$; $p = .11$). It is important to note that we were not able to address this question by prosody features due to no effect sizes from spectrographic analysis on speakers of shallow languages, and thus, our analysis reflects the relation of reading prosody as a whole to reading comprehension. Future crosslinguistic endeavors are needed, especially on shallow languages (Kuhn et al., 2010) that use spectrographic analysis.

We also explored whether the magnitude of the relation between reading prosody and reading comprehension differs by grade (a proxy for reading development phase), and we found that the relation did not differ by grade. This is in line with our speculation that reading prosody relates to reading comprehension across developmental phases as reading captures both decoding and semantic processes. Similar to the moderation question about orthographic depth, however, differential relations were not addressed by prosody features due to the limited number of studies as noted above.

Limitations and Future Research

There were several limitations of this study. First, it should be noted that our goal in this study was to estimate the magnitude of the relation between reading prosody and reading comprehension, not the directionality of the relation. Directionality inquiry can be best addressed with longitudinal studies and experimental studies. For example, Cypert and Petro (2019) found that university students who received a prosodic reading intervention had significantly better reading comprehension than the control group on the posttest. Note though that research on the causal role of reading prosody on reading comprehension is extremely limited (Ardoin, Morena, Binder, & Foster, 2013) and the present meta-analysis does not allow us to draw inferences on practical implications.

As well, very few studies used spectrographic analysis and all the studies with spectrographic analysis were with students between Grades 1–5 in languages with deep orthographies (English, French). The lack of studies examining reading prosody using spectrographic analysis in shallow orthographies greatly limited the analysis in the present study. Specifically, it would have been ideal to address the third research question by reading prosody features. This would have allowed us to evaluate whether varying magnitudes as a function of orthographic depth differ by reading prosody features. For example, pause structure (e.g., ungrammatical pauses) may have a strong relation with reading comprehension for a longer time in deep orthographies than in shallow orthographies because of the slower rate of word reading development in deep orthographies (Aro & Wimmer, 2003; Ellis et al., 2004; Katz & Frost, 1992; Seymour et al., 2003). On the other hand, the pitch aspect of reading prosody that captures semantic processing may relate to reading comprehension at an earlier grade in shallow orthographies than in deep orthographies, again due to the differences in how long decoding constrains reading processes.

In the present study, we used grade levels as a proxy for developmental phase of reading. However, grade is a rough proxy, and there is variation across education systems in terms of the grade in which reading instruction starts. Furthermore, studies were not be grouped by developmental phases as others have done (e.g., by grades in similar developmental stages, e.g., Florit & Cain, 2011; or by age, e.g., García & Cain, 2014) though this may lead to more robust results (Petscher, 2010) because of the limited number of studies in different developmental phases (e.g., only one older grade used spectrographic analysis, May, 2014) and because developmental phases may vary by orthographic depth.

We included a set of moderators based on theory. However, future studies should investigate other potential moderators. One example is text difficulty. Although the impact of text difficulty on the relation between reading prosody and reading comprehension has been examined (e.g., Benjamin & Schwanenflugel, 2010), there was an insufficient number of studies to conduct reliable moderator analysis (minimum of 4 effect sizes; Borenstein et al., 2011). Another example is moderation by individual skills such as decoding, oral language, and higher order cognitive skills. As noted above, both decoding and semantic skills are needed for reading prosody. Therefore, individuals' decoding and meaning-making ability (e.g., morphological, syntactic, and inferencing ability) might moderate the relation between reading prosody and reading comprehension (e.g., see Ravid & Mashraki, 2007). Future research should more deeply explore these potential moderators. Additionally, future studies should explore whether the relation between reading prosody and reading comprehension varies by features of reading comprehension assessments. Reading comprehension included in the present study was measured in a variety of ways, including open-ended, multiple-choice, oral retell, multiple choice mixed with open-ended. Some of the reading comprehension assessments (e.g., QRI) assessed the same text that was read aloud for the reading prosody measure while others used a different assessment. Given the multiple assessment formats and mixed format, we could not examine whether the relation between reading prosody and reading comprehension differed by assessment format. On a related note, majority of studies also did not indicate oral or silent mode of reading comprehension assessment, and therefore, we could not examine whether the relation between reading prosody and reading comprehension varied by reading mode.

Another direction in future research is an examination of the relation for students with learning disabilities or second language learners as there were very few studies for these populations. For example, people with certain disabilities, such as those on the Autism Disorder Spectrum, have been found to have atypical prosody (McCann & Peppé, 2003); thus, future studies are warranted on the relation of reading prosody and reading comprehension for these populations.

In conclusion, in the present systematic review and meta-analysis, we found a moderate relation between reading prosody and reading comprehension, and differential relations as a function of reading prosody features. Another important finding is a critical gap in the literature: There are insufficient studies measuring different features of reading prosody (e.g., spectrographic analysis), varying orthographic characteristics (e.g., limited shallow orthographies), and reader skills (e.g., beyond primary grades). Thus, our understanding of the exact nature of the relation between reading prosody and reading comprehension is lacking and future studies are needed.

Note

1. Data are available upon request from the first author.

Acknowledgments

The authors declare there is no conflict of interest.

Funding

This research was partially supported by Grants R305A120147 and R305A180055 by the Institute of Education Sciences, US Department of Education, and by the Eunice Kennedy Schriver National Institute of Child Health and HumanDevelopment (P50HD052120) to the second author. The content is solely the responsibility of the authors and does notnecessarily represent the official views of the funding agencies.

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*(9), 933–958. doi:10.1007/s11145-006-9024-z

Allington, R. L. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, *36*(6), 556–561. Retrieved from http://www.jstor.com/stable/20198272

*Arcand, M. S., Dion, E., Lemire-Théberge, L., Guay, M. H., Barrette, A., Gagnon, V., . . . Fuchs, D. (2014). Segmenting texts into meaningful word groups: Beginning readers' prosody and comprehension. *Scientific Studies of Reading*, *18*(3), 208–223. doi: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*(4), 391. doi:10.1037/spq0000027

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635. doi:10.1017/S0142716403000316

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. doi:10.1007/s11145-010-9261-z

*Benjamin, R. G. (2012). *Development and validation of the comprehensive oral reading fluency scale* (Doctoral dissertation). University of Georgia, GA. Retrieved from https://getd.libs.uga.edu/pdfs/benjamin_rebekah_a_201208_phd.pdf

*Basaran, M. (2013). Reading fluency as an indicator of reading comprehension. *Educational Sciences: Theory and Practice*, *13*(4), 2287–2290. doi:10.12738/estp.2013.4.1922

*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*(2), 105–133. doi: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. doi:10.1007/s11145-012-9382-7

Boersma, P., & Weenink, D. (2020, October 13). Praat: Doing phonetics by computer [Computer program]. Version 6.1.27. Retrieved from http://www.praat.org/

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. UK: John Wiley & Sons.

*Calet, N., Gutiérrez-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*(3), 272–285. doi:10.1111/1467-9817.12019

*Brown, L. T., Mohr, K. A., Wilcox, B. R., & Barrett, T. S. (2018). The effects of dyad reading and text difficulty on third-graders' reading achievement. *The Journal of Educational Research*, *111*(5), 541–553. doi:10.1080/00220671.2017.1310711

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, *12*, 169–190. doi:10.1023/A:1008131926604

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago, IL: The University of Chicago Press.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. London, UK: Edward Arnold.

Crystal, D. (2011). *A dictionary of linguistics and phonetics* (Vol. 30). Hoboken, NJ: John Wiley & Sons.

Cypert, R., & Petro, M. (2019). Prosody instruction intervention as a means to improved reading comprehension. *Applied Cognitive Psychology*, *33*(6), 1305–1311. doi:10.1002/acp.3593

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). Fourth-grade students reading aloud: NAEP 2002 special study of oral reading. The Nation's Report Card. NCES 2006–469. *National Center for Education Statistics*. Retrieved from https://files.eric.ed.gov/fulltext/ED488962.pdf

Davies, M. (1994). 'I'm sorry, I'll read that again': Information structure in writing. In S. Cmejrkova & F. Ticha (Eds.), *The syntax of sentence and text: A Festschrift for Frantiek Dane* (pp. 75–89). Amsterdam, The Netherlands: John Benjamins.

*Dawson, A. R. (2015). *Influences of African American English on cognitive processing and reading comprehension for African American adolescent readers* (Doctoral dissertation). University of Arkansas at Little Rock, AR.

Defior, S., Martos, F., & Cary, L. (2002). Differences in reading acquisition development in two shallow orthographies: Portuguese and Spanish. *Applied Psycholinguistics*, *23*(1), 135–148. doi:10.1017/S0142716402000073

Denes, P. B., & Pinson, E. (1993). *The speech chain: The physics and biology of spoken language*. New York, NY: WH Freeman and Co.

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory into Practice*, *30*(3), 165–175. doi:10.1080/00405849109543497

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H., Polyzoe, N., … Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*(4), 438–468. doi:10.1598/RRQ.39.4.5

*Evanchan, G. E. (2015). *The development of fluency and comprehension literacy skills of second grade students by providing regular use of the fluency development lesson* (Doctoral dissertation). University of Akron, OH.

*Fernandes, S., Querido, L., Verhaeghe, A., & Araújo, L. (2018). What is the relationship between reading prosody and reading comprehension in European Portuguese? Evidence from grades 2 to 5. *Journal of Research in Reading*, *41*, S102–S129. doi:10.1111/1467-9817.12248

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, *23*(4), 553–576. doi:10.1007/s10648-011-9175-6

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). Oxford, England: Blackwell.

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. doi:10.1207/S1532799XSSR0503_3

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74–111. doi:10.3102/0034654313499616

*González-Trujillo, M. C., Calet, N., Defior, S., & Gutiérrez-Palma, N. (2014). Scale of reading fluency in Spanish: Measuring the components of fluency/Escala de fluidez lectora en español: Midiendo los componentes de la fluidez. *Estudios de Psicología*, *35*(1), 104–136. doi:10.1080/02109395.2014.893651

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene: Institute for the Development of Educational Achievement.

*Hammer, K. B. (2003). *The effect of oral reading fluency on analysis skills* (Doctoral dissertation). Dominican University of California, CA. Retrieved from https://files.eric.ed.gov/fulltext/ED479195.pdf

*Groen, M. A., Veenendaal, N. J., & Verhoeven, L. (2018). The role of prosody in reading comprehension: Evidence from poor comprehenders. *Journal of Research in Reading*. doi:10.1111/1467-9817.12133

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. doi:10.1002/jrsm.5

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. doi:10.1002/sim.1186

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, *58*(8), 702–714. doi:10.1598/RT.58.8.1

*Hussien, A. M. (2014). The indicating factors of oral reading fluency of monolingual and bilingual children in Egypt. *International Education Studies*, *7*(2), 75. doi:10.5539/ies.v7n2p75

*Jefferson, R. E., Grant, C. E., & Sander, J. B. (2017). Effects of tier I differentiation and reading intervention on reading fluency, comprehension, and high stakes measures. *Reading Psychology*, *38*(1), 97–124. doi:10.1080/02702711.2016.1235648

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. doi:10.1037/0022-0663.95.4.719

Joshi, R. M., Treiman, R., Carreker, S., & Moats, L. C. (2008/2009). How words cast their spell: Spelling is an integral part of learning the language, not a matter of memorization. *American Educator, Winter*, 6–43. Retrieved from https://www.aft.org/sites/default/files/periodicals/joshi.pdf

*Kariuki, P., & Baxter, A. (2011, November). The relationship between prosodic oral reading assessments and standards-based reading assessment in a 2nd grade classroom. Paper presented at the Annual Conference of the Mid-South Educational Research Association, Oxford, Mississippi.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 45–66). Amsterdam, NL: Elsevier. Retrieved from https://files.eric.ed.gov/fulltext/ED359575.pdf#page=157

Kieffer, M. J., Biancarosa, G., & Mancilla-Martinez, J. (2013). Roles of morphological awareness in English reading comprehension for Spanish-speaking language minority learners: Exploring partial mediation by vocabulary and reading fluency. *Applied Psycholinguistics*, *34*(4), 697–725. doi:10.1017/S0142716411000920

Kieffer, M. J., & Box, C. (2013). Derivational morphological awareness, academic vocabulary, and reading comprehension in linguistically diverse sixth graders. *Learning and Individual Differences*, *24*, 168–175. doi:10.1016/j.lindif.2012.12.017

Kim, Y.-S. G., Guo, Q., Liu, Y., Peng, Y., & Yang, L. (2020). Pathways in which compounding morphological awareness is related to reading comprehension: Evidence from Chinese children. *Reading Research Quarterly*, *55*(2), 193–212. doi:10.1002/rrq.262

Kim, Y. S. G., Quinn, J. M., & Petscher, Y. (2020). Reading prosody unpacked: A longitudinal investigation of its dimensionality and relation with word reading and listening comprehension for children in primary grades. *Journal of Educational Psychology*. doi:10.1037/edu0000480

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. doi:10.1037/a0019643

Kim, Y.-S. 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. doi:10.1002/rrq.107

Kim, Y.-S. G. (2020a). 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*, *112*(4), 667–684. doi:10.1037/edu0000407

Kim, Y.-S. G. (2020b). Toward integrative reading science: The direct and indirect effects model of reading (DIER). *Journal of Learning Disabilities*, *53*, 469–491. doi:10.1177/0022219420908239

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. doi:10.1080/10888438.2015.1007375

*Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310. doi:10.1037/0022-0663.100.2.310

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. doi:10.1037/edu0000078

Koriat, A., Greenberg, S. N., & Kreiner, H. (2002). The extraction of structure during reading: Evidence from reading prosody. *Memory and Cognition*, *30*, 270–280. doi:10.3758/BF03195288

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*(2), 230–251. doi:10.1598/RRQ.45.2.4

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

*Marrone, A. (2014). *The effects of enhanced e-books vs. traditional print books on reader motivation, comprehension, and fluency in an elementary classroom* (Doctoral dissertation). The William Paterson University of New Jersey, NJ.

*May, M. C. (2014). *The effects of text complexity on the oral reading prosody of fifth-grade students* (Doctoral dissertation). Fordham University, NY.

*Lai, S. A., Benjamin, R. G., Schwanenflugel, P. J., & Kuhn, M. R. (2014). The longitudinal relationship between reading fluency and reading comprehension skills in second-grade children. *Reading & Writing Quarterly*, *30*(2), 116–138. doi:10.1080/10573569.2013.789785

McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., … Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology*, *92*, 140–160. doi:10.1016/j.jecp.2005.03.009

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. doi:10.1080/1368282031000154204

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, *98*(4), 839. doi:10.1037/0022-0663.98.4.839

*Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Literacy Research and Instruction*, *46*(1), 73–94. doi:10.1080/19388070609558461

*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*(4), 336–354. doi:10.1598/RRQ.43.4.2

Moser, G. P., Sudweeks, R. R., Morrison, T. G., & Wilcox, B. (2014). Reliability of ratings of children's expressive reading. *Reading Psychology*, *35*(1), 58–79. doi:10.1080/02702711.2012.675417

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). U.S. Government Printing Office.

*Paige, D. 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*(2), 123–156. doi:10.1177/1086296X14535170

Petscher, Y. (2010). A meta-analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of Research in Reading*, *33*(4), 335–355. doi:10.1111/j.1467-9817.2009.01418.x

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*(6), 510–519. doi:10.1598/RT.58.6.2

Pinnell, G. S., & Fountas, I. C. (2010). Research base for guided reading as an instructional approach. *Scholastic: Guided Reading Research*. Retrieved from http://www.dps.k12.oh.us/content/documents/Fluency-rubric.pdf.

R Core Team. (2013). *R: A language and environment for statistical computing*. R foundation for statistical computing. http://www.R-project.org

Rasinski, T. (2004). Creating fluent readers. *Educational Leadership*, *61*(6), 46–51. Retrieved from http://educationalleader.com/subtopicintro/read/ASCD/ASCD_364_1.pdf

*Ravid, D., & Mashraki, Y. E. (2007). Prosodic reading, reading comprehension and morphological skills in Hebrew-speaking fourth graders. *Journal of Research in Reading*, *30*(2), 140–156. doi:10.1111/j.1467-9817.2007.00340.x

*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*(4), 350–361. doi:10.1080/19388070802468715

*Rasinski, T., Paige, D., Rains, C., Stewart, F., Julovich, B., Prenkert, D., … Nichols, W. D. (2017). Effects of intensive fluency instruction on the reading proficiency of third-grade struggling readers. *Reading & Writing Quarterly*, *33*(6), 519–532. doi:10.1080/10573569.2016.1250144

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, *42*, 546–567. doi: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. doi:10.1016/j.jsp.2007.06.006

*Sargent, S. E. (2002). *Oral reading fluency: A predictor of reading proficiency in fifth grade students?* (Order No. 3066191). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (276202964). Retrieved from https://search.proquest.com/docview/276202964?accountid=14509

*Sabatini, J., Wang, Z., & O'Reilly, T. (2018). Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly*. doi:10.1002/rrq.226

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior*, *12*, 177–186. doi:10.1080/10862968009547369

Schreiber, P. A. (1987). Prosody and structure in children's syntactic processing. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 243–270). New York, NY: Academic Press.

Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice*, *30*(3), 158–164. doi:10.1080/00405849109543496

Schwanenflugel, P. J., & Benjamin, R. G. (2017). Lexical prosody as an aspect of oral reading fluency. *Reading and Writing*, *30*(1), 143–162. doi:10.1007/s11145-016-9667-3

*Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & 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*(1), 119. doi:10.1037/0022-0663.96.1.119

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England: 1953)*, *94*(Pt 2), 143–174. doi:10.1348/000712603321661859

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. doi:10.1080/10888438.2011.618153

Smith, G. S., & Paige, D. D. (2019). A study of reliability across multiple raters when using the NAEP and MDFS rubrics to measure oral reading fluency. *Reading Psychology*, *40*(1), 34–69.

Sterne, J. A., Egger, M., & Smith, G. D. (2001). Investigating and dealing with publication and other biases in meta-analysis. *BMJ*, *323*(7304), 101–105. Retrieved from https://www.bmj.com/content/bmj/323/7304/101.full.pdf

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. Rothstein, A. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Chichester, UK: John Wiley & Sons, Ltd.

Sucena, A., Castro, S. L., & Seymour, P. (2009). Developmental dyslexia in an orthography of intermediate depth: The case of European Portuguese. *Reading and Writing*, *22*(7), 791–810. doi:10.1007/s11145-008-9156-4

*Taylor, C. D., Meisinger, E. B., & Floyd, R. G. (2013). Variations in directions and overt timing on oral reading accuracy, fluency, and prosody. *School Psychology Review*, *42*(4), 437–448. doi:10.1080/02796015.2013.12087464

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. doi:10.1111/j.1467-9817.2009.01401.x

Tindal, G., & Marston, D. (1996). Technical adequacy of alternative reading measures as performance assessments. *Exceptionality*, *6*(4), 201–230. doi:10.1207/s15327035ex0604_1

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375. doi:10.1037/met0000011

*Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2016). Bidirectional relations between text reading prosody and reading comprehension in the upper primary school grades: A longitudinal perspective. *Scientific Studies of Reading*, *20*(3), 189–202. doi:10.1080/10888438.2015.1128939

*Tortorelli, L. S. (2018). Off to a slow start: Four profiles of slow readers in second grade. *Reading Psychology*, *39*(7), 647–689. doi:10.1080/02702711.2018.1515134

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. doi:10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. doi:10.18637/jss.v036.i03

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*(10), 1559–1579. Retrieved from https://www.researchgate.net/profile/Deirdre_Wilson/publication/222826780_Relevance_and_prosody/links/5b1a4f8baca272021cf2c5ed/Relevance-and-prosody.pdf

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*(3), 211–239. doi:10.1207/S1532799XSSR0503_2

Wood, C. (2006). Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *Journal of Research in Reading*, *29*, 270–287. doi:10.1111/j.1467-9817.2006.00308.x

*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*(1), 35–44. doi:10.1080/21683603.2013.854187

*Yildiz, M., & Çetinkaya, E. (2017). The relationship between good readers' attention, reading fluency and reading comprehension. *Universal Journal of Educational Research*, *5*(3), 366–371. doi:10.13189/ujer.2017.050309

*Yildirim, K., Rasinski, T., & Kaya, D. (2019). Fluency and comprehension of narrative texts in Turkish students in grades 4 through 8. *Education 3-13*, *47*(3), 348–357. doi:10.1080/03004279.2018.1449880

Zutell, J., & Rasinski, T. V. (1991). Training teachers to attend to their students' oral reading fluency. *Theory into Practice*, *30*(3), 211–217. doi:10.1080/00405849109543502