Journal of Research in Reading

UKLA The United Kingdom Literacy Association

Journal of Research in Reading, ISSN 0141-0423
Volume 42, Issue 1, 2019, pp 37–57
DOI:10.1111/1467-9817.12133

The role of prosody in reading comprehension: evidence from poor comprehenders

Margriet A. Groen,
Nathalie J. Veenendaal and Ludo Verhoeven

Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

Text reading prosody and reading comprehension are related, but both rely on decoding. The aim of the current study was, therefore, to disentangle the contribution of decoding from that of prosody skills. We examined the performance on text reading prosody and speech prosody in fifth-grade children with age-appropriate decoding but weak comprehension. We compared their performance with that of chronological-age controls and younger, comprehension-level controls. We found that poor comprehenders scored significantly below the chronological-age controls on all prosody tasks. Importantly, poor comprehenders scored below the younger, comprehension-level controls on a speech rhythm task. Furthermore, speech prosody explained unique variance in predicting reading comprehension status (poor comprehender vs comprehension-level control). This suggests that poor comprehenders have a delay in prosodic development, with an additional indication of a deficiency in perception and production of speech prosody. The results show that the relation between text reading prosody and reading comprehension does not exclusively rely on decoding.

Highlights

What is already known about this topic

- Text reading prosody has consistently been shown to be related to reading comprehension.
- Text reading prosody and reading comprehension both rely on decoding efficiency.
- The role of speech prosody in reading comprehension is less widely investigated.

What this paper adds

- The role of decoding in reading comprehension has been separated from the role of prosody skills in reading comprehension.

- Children with good decoding but poor comprehension (poor comprehenders) performed a text reading prosody task and speech prosody tasks.
- Poor comprehenders scored significantly lower than chronological-age controls on both types of prosody tasks and lower than younger, comprehension-level controls on a speech rhythm task. Producing appropriate speech prosody while telling a story explained unique variance in distinguishing between poor comprehenders and comprehension-level controls.

Implications for practice and/or policy

- Perception and production of speech prosody may be less developed in poor comprehenders and may partly contribute to reading comprehension problems.
- Appropriate use of prosody may be needed for reading comprehension in order to form a correct internal representation of a written text.

Reading comprehension is a complex process that requires children to quickly and accurately recognise the words in a text, while simultaneously constructing meaning. As such, reading fluency is assumed to be a prerequisite for reading comprehension and traditionally has been defined as reading (text) at speed and without errors ('the automaticity aspect'). More recently, however, it has been shown that the degree to which children use appropriate prosody (e.g., intonation, stress placement, word boundaries, pausing and rhythm) when reading aloud is also associated with their reading comprehension ability, and 'text reading prosody' has been added to the concept of reading fluency (adding to reading speed and reading accuracy; e.g., Miller & Schwanenflugel, 2006, 2008; Rasinski, Rikli, & Johnston, 2009; Veenendaal, Groen, & Verhoeven, 2015). However, both text reading prosody and reading comprehension strongly rely on the ability to read well. Indeed, it has been found that efficient and automatised decoding is an essential skill for 'adult-like' text reading prosody to develop in primary school children (Miller & Schwanenflugel, 2006; Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). Therefore, an outstanding question is to what extent the 'automaticity aspect' and the 'prosody aspect' of reading fluency can be separated in their relation to reading comprehension. Is the 'prosody aspect' merely an epiphenomenon of efficient decoding, or does it merit a more independent status? In order to further determine the relations between decoding, prosody and reading comprehension, we study prosodic skills in a group of 'poor comprehenders' – children with age-appropriate decoding but with persisting difficulties in reading comprehension (e.g., Nation, Cocksey, Taylor, & Bishop, 2010).

Decoding efficiency, text reading prosody and reading comprehension

To be able to read fluently, traditionally has meant reading at speed and without errors, but more recently, reading with appropriate intonation has also been suggested to be important (e.g., Kuhn, Schwanenflugel, & Meisinger, 2010; Kuhn & Stahl, 2003). A relation between the 'automaticity aspect' of text reading fluency – defined as speed and accuracy – and reading comprehension has been reported repeatedly (e.g., Berninger et al., 2010; Kim & Wagner, 2015; Kim, Wagner, & Lopez, 2012), but there is a growing literature that emphasises the significance of text reading prosody in this relation (e.g., Klauda & Guthrie, 2008; Miller & Schwanenflugel, 2006, 2008; T. Rasinski et al., 2009; Schwanenflugel et al., 2004). Text reading prosody can be assessed by means of rating scales, to obtain a holistic measure of prosody, or by spectrographic analyses, to measure individual features of prosody. Rating scales assess prosodic aspects such as enthusiasm, phrasing, general smoothness and pace, when children read a text aloud. Studies that used such rating scales showed that text reading prosody was significantly correlated to reading comprehension in children from third, fifth, and seventh grade (T. Rasinski et al., 2009) and that text reading prosody accounted for substantial variance in reading comprehension scores in children from fourth grade (Calet, Defior, & Gutiérrez-Palma, 2015; Veenendaal et al., 2015) and fifth grade (Klauda & Guthrie, 2008). Other studies used spectrographic analyses in order to assess prosodic features such as pauses, intonation contours and end-of-sentence pitch. Miller and Schwanenflugel (2008) showed that children with a decreasing number of inappropriate pauses in their oral reading from first to second grade and an early adult-like intonation contour performed better on a reading comprehension test in third grade. Further, it was shown that skilled readers read with fewer inappropriate pauses and with more intonation than beginning readers (Schwanenflugel et al., 2004).

Because inappropriate pausing is associated with decoding problems, it has been proposed that automaticity in reading, that is, decoding efficiency, is necessary for text reading prosody to develop (Schwanenflugel et al., 2004). However, even though decoding may be a necessary prerequisite for an appropriate text reading prosody performance; no studies have yet investigated whether it is also sufficient. Poor comprehenders are of interest in this context, as their decoding skills dissociate from their reading comprehension skills. If decoding efficiency is sufficient for the relation between text reading prosody and comprehension to develop, then it is expected that poor comprehenders will show an age-appropriate level of text reading prosody. If, however, decoding is necessary, but not sufficient for this relation to develop, and text reading prosody is more tied to reading comprehension, we expect their text reading prosody skills to be weak.

Speech prosody and reading comprehension

Both decoding efficiency and text reading prosody strongly rely on the ability to read well. To further disentangle the contribution of decoding efficiency to the association between prosody and reading comprehension, we investigate speech prosody skills (i.e., the perception and production of speech rhythm, stress placement and word boundaries in oral language) in the same poor comprehenders. Although it was shown that early prosodic speech perception contributed to reading-related skills, such as phonological and morphological awareness (Zhang & McBride-Chang, 2010), little is known about the contribution of early prosody skills to later reading comprehension.

Only a few studies examined the relation between speech prosody and reading comprehension, as a part of general reading skills (Holliman et al., 2014; Kent, 2013; Lochrin, Arciuli, & Sharma, 2015; Whalley & Hansen, 2006). Holliman et al. (2014) showed that the performance on a prosody test that assessed perception of stress placement, intonation and timing was correlated with reading comprehension in children from first and second grade. A comprehensive assessment of speech prosody is the Profiling Elements of Prosodic Systems - Children (PEPS-C) computer test (Peppé & McCann, 2003) that measures both perception and production of speech prosody. Kent (2013) found that the performance on a number of tasks of the PEPS-C test – perception of speech rhythm, stress placement and word boundaries – was associated with children's reading comprehension. Furthermore, the perception of word boundaries and stress placement explained unique variance in reading comprehension scores in fourth grade children (Kent, 2013). Lochrin et al. (2015) also used the PEPS-C test but examined perception and production of speech prosody and related this to word reading and reading comprehension outcomes in children from 7 to 12 years of age. They found that the perception and production of word boundaries and the perception of speech rhythm and stress placement were associated with reading comprehension, but only the production of word boundaries explained unique variance in reading comprehension (Lochrin et al., 2015). Conversely, Whalley and Hansen (2006) used the PEPS-C word boundary task and another, similar word boundary test, and found that these two receptive tasks were not correlated with reading comprehension in fourth grade children. The earlier results show that different aspects of speech prosody might differently relate to reading comprehension, although they also show that results are not consistent between studies. The inconsistent results mainly seem to relate to task-dependent factors with most studies assessing perception of speech prosody (Kent, 2013; Whalley & Hansen, 2006) and some studies assessing production and perception (Lochrin et al., 2015). Furthermore, the age of the children that participated and the control variables included differed (phonological awareness and rhythmic sensitivity, Whalley & Hansen, 2006; age and phonological awareness, Lochrin et al., 2015; decoding and listening comprehension, Kent, 2013). In the current study, therefore, we choose to assess speech prosody using a wide range of perception and production tasks, as well as a more ecologically valid story-telling context.

The present study

The aim of the present study was to disentangle the contribution of decoding efficiency to the association between prosody and reading comprehension, by using two different approaches. First, the current literature proposes that efficient decoding is a necessary skill for the relation between text reading prosody and reading comprehension to emerge, but it is as yet unclear whether decoding is also sufficient. We investigated the performance of children with strong decoding but weak reading comprehension skills ('poor comprehenders') to provide more insight into this issue. Second, we examined whether speech prosody skills – more specifically, storytelling and the perception and production of speech rhythm, stress placement and word boundaries – are related to reading comprehension as strongly as reading-related prosody tasks are.

This lead to the following research question: To what extent do poor comprehenders differ in their prosodic abilities across written and spoken modalities from a chronological-age control group and a younger, comprehension-level control group? The inclusion of a younger control group with similar reading comprehension allowed us to determine the extent to which the prosody skills of poor comprehenders are in line with their level of reading comprehension. If the poor comprehenders would perform at a lower level than the younger, comprehension-level control group, this could provide an indication of a restricting factor in their reading comprehension performance.

Method

Participants

Fifth-grade poor comprehenders were identified in a two-step process. First, we asked teachers from six medium-sized, primary schools in the Eastern part of the Netherlands to refer children to us based on performance on annual national assessments[1] on word recognition (Krom, Jongen, Verhelst, Kamphuis, & Kleintjes, 2010) and reading comprehension (Staphorsius & Krom, 2011). Each poor comprehender was individually matched on chronological-age and word recognition score on the national assessment with a child from the same school and class and matched on reading comprehension level with a younger child, also from the same school. Children with a diagnosis of a language or reading impairment or with speech problems were a priori excluded from participation in the study. Then, in step two, we confirmed the children's reading skills based on our own assessments.

Specifically, the criteria for initial group membership (poor comprehenders vs chronological-age matched controls vs comprehension-level matched controls) as used by the teachers were based on grade-based norm-scores on the national assessments, which provided broad percentile score categories: children received an A (above 75th percentile), B (50th–75th percentile), C (25th–50th percentile, D (10th–25th percentile) or E (below 10th percentile) score. Following selection criteria from previous studies (e.g., Catts, Adlof, & Weismer, 2006; Fletcher et al., 1994; Stanovich & Siegel, 1994), all children (poor comprehenders, chronological-age controls and comprehension-level controls) had average to above-average (i.e., above the 50th percentile) word recognition skills – scoring an A or B – and within the triads, children were matched on that score. Furthermore, the poor comprehenders had lower score than average reading comprehension skills (below the 25th percentile; scoring D or E), whereas both groups of controls scored above the 50th percentile on reading comprehension (again, scoring A or B). To identify the younger controls, who performed at the same reading comprehension level as the poor comprehenders, but above the 50th percentile for their grade, reading-age scores (as derived from the national assessments) were compared. Children in the younger control group were predominantly from third grade, with a few children from second or fourth grade. Triads of 28 'poor comprehenders' along with 28 children in each of the control groups, making a total of 84 children participating, were initially referred by teachers.

Second, to confirm the status of 'poor comprehenders' as indicated by the schools, we assessed reading comprehension, word recognition and pseudoword decoding ourselves for all children. See for a full description the section. Data from poor comprehenders with reading comprehension scores above the sample-based 50th percentile on the reading comprehension measure we administered ourselves were rejected. This led to the exclusion of seven children in the poor comprehender group and their chronological-age and comprehension-level controls, resulting in 63 participants remaining (21 per group); 12 girls and 9 boys in the poor comprehender group, 14 girls and 7 boys in the chronological-age control group and 9 girls and 12 boys in the comprehension-level control group. All participating children were native speakers of Dutch. Parental informed consent was obtained for all participating children.

Materials

Selection variables

The level of reading comprehension, word recognition and pseudoword decoding was assessed in order to confirm the status of the poor comprehenders as referred by teachers.

Reading comprehension

The reading comprehension test was constructed of two standardised reading comprehension tests (Aarnoutse & Kapinga, 2006). One of these tests was a reading comprehension test for children from first, second and third grade, and the other one was a test for children from fourth, fifth and sixth grade. Questions from both tests were included to prevent floor and ceiling effects. The reading comprehension test presented the children with seven short texts; each followed by three multiple choice questions and two to four 'true or false' questions about each text. Four texts were informative, and three texts were narratives. The total number of items for this test was 44. Cronbach's alpha reliability coefficient has been calculated for this sample and was .83.

Word recognition

Word recognition efficiency (rate) was assessed with a standardised test by Brus and Voeten (1973). Three columns with a total of 116 words were presented to the children, who were given 1 minute to read as many words as possible, as quickly and clearly as they could. The wordlist included one-syllable words and two-syllable and multi-syllable words. Raw scores were converted to standard scores (M = 10, standard deviation [SD] = 3). Cronbach's alpha reliability was reported to be between .73 and .92 (Brus & Voeten, 1973).

Pseudoword decoding

Efficiency of pseudoword decoding was measured with a standardised test by Van den Bos, Lutje Spelberg, Scheepstra, and de Vries (1994). The pseudowords have been created based on the existing words from the word recognition task described earlier and were therefore similar in number of syllables and phonotactic complexity. Children had 2 minutes to read as many items, as quickly and clearly as they could. Raw scores were converted to standard scores (M = 10, SD = 3). Cronbach's alpha was reported to be between .63 and .80 for the pseudoword decoding test (Van den Bos et al., 1994).

Prosody assessment

The use of prosody during reading (text reading prosody) and during storytelling (storytelling prosody) was assessed by means of a rating scale that distinguishes enthusiasm, phrasing, smoothness and pace. Additionally, three subtasks with a receptive and a productive part of each of the Dutch version of the PEPS-C computer task (Peppé & McCann, 2003) were used to assess speech prosody: two speech rhythm tasks (PEPS-C: long-item discrimination and imitation), two word boundary tasks (PEPS-C: chunking) and two stress placement tasks (PEPS-C: contrastive stress). In the receptive parts of each subtask (perception of prosody), children listened to sound samples, presented via the speakers of a computer, whereas in the expressive parts (production of prosody), children had to produce prosodic utterances themselves. There were 16 items per task, plus two practice items to start each task with. The practice items were not included in the scores. Because of the problems with the expressive word boundary task, this task was not included in the analyses. Because of the translation into Dutch, half of the items became too lengthy and therefore too complex for the children. Therefore, we report the results on the five remaining PEPS-C subtasks. Each of these prosody tasks is described in more detail in the succeeding text.

Text reading prosody

To assess text reading prosody, we used two short narratives (approximately 100 words each). Word frequency was based on a selection of wordlists naming the most frequent words for Dutch schoolchildren (Vermeer, 2000) to make sure that the texts were not too difficult for the younger children participating. Children were first asked to read the two short stories silently and then to read these aloud. They were asked to read the way they would normally read aloud in class. The reading was recorded on a digital voice recorder and scored at a later time by means of the Multidimensional Fluency Scale (Rasinski, 2004). This scale assesses four aspects of prosody: expression (making the text sound like natural language, adequate expression and enthusiasm), phrasing (marking clause and sentence units), smoothness (resolves word and structure difficulties easily) and pace (pleasant conversational pace, not too fast and not too slow). On each of these sections, children could receive between 1 and 4 points, resulting in a total score ranging from 4 to 16. An average score over the two stories has been used for analysis. The reliability of this task has been calculated, and the sample-based Cronbach's alpha was .85. Twenty percent of the data were scored by an independent rater, and inter-rater reliability was calculated using intraclass correlation coefficients (ICC), using the analysis of variance (ANOVA) two-way mixed model and an absolute agreement definition (rather than consistency). The ICC on the average score for the two texts (as used for analysis) was excellent: ICC = .940, $F(11,11) = 29.625$, $p < .001$.

Storytelling prosody

Speech prosody was firstly assessed by using two storytelling cards (Verhoeven & Vermeer, 2001). Each card showed a sequence of six pictures. Children were asked to look at these pictures and to tell a story about what happened. The child was asked to make the story sound interesting for a younger child that would not see the pictures. The stories were recorded on a digital recorder, and prosody was scored at a later time. The Multidimensional Fluency Scale (Rasinski, 2004) was adapted to make it more suitable for assessing storytelling prosody. In the adapted version, the four sections refer to expression (making it sound like a natural story, adequate expression and enthusiasm), phrasing (adequate indication of sentence, phrase and passage boundaries), smoothness (generally smooth speech, structure difficulties resolved quickly) and pace (consistently conversational pace, not too fast and not too slow). Performance on each section was marked with 1–4 points, so total scores per story ranged from 4 to 16. An average score over the two stories has been used for analysis. Reliability of this task has been calculated, and Cronbach's alpha was .86. Twenty percent of the data were scored by an independent rater, and inter-rater reliability was calculated. The ICC on the average score for the two stories (as used for analysis) was excellent: ICC = .873, $F(11,11) = 16.777$, $p < .001$.

Speech rhythm

The receptive task was the long-item discrimination task of the PEPS-C computer test, which assessed the ability to hear differences in rhythmic patterns of filtered speech. On each trial, children heard pairs of short phrases (six to seven syllables), taken from other PEPS-C tasks (word boundaries and stress placement). These phrases were low-passed filtered and, therefore, lacked any phonemic content, sounding as if someone was talking in a room next door. Pitch, loudness and length variation was preserved, though. The child was asked to indicate whether the two phrases sounded the same (which was the case on half of the trials) or different from each other. The child received one point per correct answer. The internal reliability of this task has been calculated. After removing four items, Cronbach's alpha was reasonable, $\alpha = .59$. Twelve items were therefore included in further analyses.

The expressive speech rhythm task was the long-item imitation task. Children heard short phrases and had to repeat not only the words but also the speech pattern of the phrase as precisely as possible. The sentences had six to seven syllables and were based, in structure, on the sentences in the word boundary and stress placement tasks, without being identical to those. An example is 'I wanted YELLOW shoes'. The tester decided whether the imitation was correct, and children received either one or a half point for their performance. Twenty percent of the data were scored by an independent rater, and inter-rater reliability was calculated. The ICC on the imitation task was excellent: ICC = .851, $F(11,11) = 11.726$, $p < .001$. Reliability was calculated, and after removing two items, Cronbach's alpha was .69. The remaining 14 items have been included in further analyses.

Stress placement

The two contrastive stress tasks of the PEPS-C computer test assessed receptive and expressive use of stress placement. The first task was a receptive task. The child heard a short story about someone who went to a store to buy socks but later realised that she had forgotten to buy one specific colour of socks. The child heard sentences such as 'I wanted BLUE and black socks'. Children had to decide which colour of socks the speaker had forgotten to buy. Half of the time, stress was placed on the first word and the other half, on the second word. Reliability of this task has been calculated, and Cronbach's alpha was .80.

Secondly, children performed an expressive contrastive stress task where they had to place stress on certain words themselves. Children saw a picture and heard an incorrect commentary. An example is a picture of a white cow with a ball, and the speaker saying 'The red cow has got the ball'. This was said in a neutral tone of voice, without any pitch or stress changes. The child had to correct the speaker by saying 'No, the WHITE cow has got the ball!' The tester decided whether the stress placement was appropriate. Twenty percent of the data were scored by an independent rater. Inter-rater reliability analysis was performed, and the ICC on the productive focus task was excellent: ICC = .914, $F(11,11) = 32.133$, $p < .001$. Reliability analysis of this task showed a Cronbach's alpha of .76.

Word boundaries

The receptive chunking task of the PEPS-C computer test assessed the perception of word boundaries. Children saw two pictures on a computer screen and heard either a compound noun and a noun or a string of nouns (e.g., 'Chocolate-cake and jam' vs 'Chocolate, cake and jam'). Children had to select the corresponding picture on the screen. Every correct answer resulted in a point for the child. The reliability of this task was calculated, and after removing four items, Cronbach's reliability coefficient was fair, $\alpha = .59$. Twelve items have therefore been included in further analyses.

Control variables

A productive vocabulary task and a nonverbal reasoning test (Raven) were added to the test battery as general control measures.

Vocabulary

Productive vocabulary was assessed with a subtest of the Wechsler Intelligence Scale for Children III; Dutch edition (Kort et al., 2005). Children were aurally presented with a word and were asked for a spoken definition of this word. Raw scores were converted to standard scores ($M = 10$, $SD = 3$). Cronbach's alpha reliability coefficient has been reported to be .79 for this test (Kort et al., 2005). Twenty percent of the data were scored by an independent rater, and inter-rater reliability was calculated. The ICC on the vocabulary task was excellent: ICC = .972, $F(11,11) = 75.602$, $p < .001$.

Nonverbal reasoning

Nonverbal reasoning was assessed using the Raven (1976) Progressive Matrices Test. Children received a booklet with 60 incomplete patterns and were asked to identify the missing element that completed the pattern. Raw scores were converted to percentile scores. Cronbach's alpha has been reported to be .90 for this test (Raven, 1976).

Procedure

All assessments were carried out during school hours. The tests to assess reading comprehension and nonverbal reasoning were administered groupwise in two sessions of 40 minutes each. All participating children of one school sat together in one room to make these tests silently. The other assessments were performed on an individual basis and were administrated in two separate sessions by the first author and two trained master students. Individual testing was carried out in a separate room, provided by the schools. In the first individual session, performance on text reading prosody and speech prosody was assessed and three tasks not discussed in the current paper. The order of the two narratives and the two story cards was counter-balanced. During the second session, performance on word recognition, pseudoword decoding, vocabulary and the subtasks of the PEPS-C computer task was assessed.

Data analyses

Firstly, data were visually inspected, and skewness and kurtosis values (Table 1) and the Shapiro–Wilk test were examined to determine whether the data were normally distributed.

Table 1. Descriptive statistics of selection, control and prosody measures.

	Chronological-age controls (n = 21)				Poor comprehenders (n = 21)				Comprehension-level controls (n = 21)			
	M (*SD*)	Median (min–max)	Skewness	Kurtosis	*M* (*SD*)	Median (min–max)	Skewness	Kurtosis	*M* (*SD*)	Median (min–max)	Skewness	Kurtosis
Age (years)	10.87 (0.37)	10.77 (10.37–11.75)	0.20	−1.31	10.82 (0.41)	10.76 (10.30–11.46)	0.96	−0.14	8.72 (0.81)	8.64 (7.30–10.44)	0.27	−0.40
Selection												
Reading comprehension[a]	38.24 (2.30)	38.00 (34.00–42.00)	−0.07	−1.12	28.71 (3.94)	30.00 (19.00–34.00)	−0.84	−0.08	29.95 (5.58)	28.00 (19.00–38.00)	0.04	−1.33
Word recognition[b]	13.05 (2.18)	13.00 (10.00–19.00)	0.91	0.69	11.95 (1.99)	12.00 (8.00–16.00)	0.10	−0.52	13.43 (1.91)	14.00 (10.00–16.00)	−0.22	−1.41
Pseudoword decoding[b]	14.33 (2.74)	14.00 (10.00–19.00)	−0.08	−1.38	12.95 (2.48)	13.00 (9.00–18.00)	0.67	−0.48	13.61 (2.42)	13.00 (8.00–19.00)	0.01	0.02
Control												
Vocabulary[b]	10.95 (2.20)	11.00 (7.00–15.00)	−0.02	−1.03	9.00 (2.37)	9.00 (6.00–15.00)	0.54	−0.26	10.33 (2.56)	11.00 (4.00–13.00)	−0.77	−0.43
Nonverbal reasoning[c]	76.81 (20.35)	90.00 (25.00–99.00)	−0.96	−0.19	49.76 (26.57)	50.00 (5.00–95.00)	−0.18	1.32	76.38 (22.20)	75.00 (5.00–99.00)	−1.62**	2.53**
Text-related prosody												
Text reading prosody[a]	13.14 (1.33)	13.00 (10.50–16.50)	0.52	−0.01	11.40 (1.17)	11.50 (9.50–13.50)	−0.16	−1.21	10.73 (1.67)	11.00 (6.50–13.00)	−0.81	−0.08
Speech prosody												
Storytelling prosody[a]	12.57 (2.05)	12.50 (8.00–16.00)	−0.17	−0.57	10.43 (2.14)	10.50 (5.50–13.50)	−0.51	−0.59	9.64 (1.68)	9.50 (5.50–12.50)	−0.42	−0.25
Speech rhythm												
Receptive[a]	10.86 (1.59)	11.00 (6.00–12.00)	−1.55**	1.87	9.29 (1.93)	9.00 (5.00–12.00)	−0.51	−0.70	10.86 (1.28)	11.00 (7.00–12.00)	−1.26*	1.48

(*Continues*)

Table 1. (Continued)

	Chronological-age controls (n = 21)				Poor comprehenders (n = 21)				Comprehension-level controls (n = 21)			
	M (*SD*)	Median (min–max)	Skewness	Kurtosis	*M* (*SD*)	Median (min–max)	Skewness	Kurtosis	*M* (*SD*)	Median (min–max)	Skewness	Kurtosis
Expressive[a]	10.71 (2.17)	11.00 (6.00–14.00)	−0.50	−0.78	8.83 (2.60)	9.50 (3.00–12.50)	−0.76	−0.33	10.05 (2.24)	10.50 (2.00–13.00)	−2.00***	5.28***
Stress placement												
Receptive[a]	15.24 (1.18)	16.00 (12.00–16.00)	−1.31**	0.61	13.38 (2.94)	14.00 (7.00–16.00)	−0.83	−0.60	13.48 (2.98)	15.00 (7.00–16.00)	−0.84	−0.82
Expressive[a]	13.62 (3.07)	15.00 (4.00–16.00)	−1.70***	2.33	11.67 (3.15)	12.00 (5.00–16.00)	−0.60	−0.95	13.81 (1.78)	14.00 (10.00–16.00)	−0.44	−0.78
Word boundaries												
Receptive[a]	10.71 (1.38)	11.00 (8.00–12.00)	−0.59	−1.03	9.14 (2.06)	10.00 (5.00–12.00)	−0.54	−1.15	9.10 (1.61)	9.00 (5.00–12.00)	−0.42	0.01

Note: *SD*, standard deviation.
[a]Raw scores.
[b]Standard scores (M = 10, SD = 3, range = 1–19).
[c]Percentiles.
*$p < .05$.
**$p < .01$.
***$p < .001$.

The speech rhythm (receptive: chronological-age controls $W = 0.75$, $p < .001$, comprehension-level controls $W = 0.81$, $p = .001$; expressive: comprehension-level controls $W = 0.78$, $p < .001$), stress placement (receptive: chronological-age controls $W = 0.71$, $p < .001$, poor comprehenders $W = 0.83$, $p < .001$, comprehension-level controls $W = 0.81$, $p < .001$; expressive: chronological-age controls $W = 0.75$, $p < .001$, poor comprehenders $W = 0.91$, $p = .049$) and word boundaries (receptive: chronological-age controls $W = 0.83$, $p = .002$, poor comprehenders $W = 0.89$, $p = .025$) variables, as well as nonverbal reasoning (chronological-age controls $W = 0.83$, $p = .002$, poor comprehenders $W = 0.90$, $p = .04$, comprehension-level controls $W = 0.78$, $p < .001$), were not normally distributed, characterised mostly by negative skew, and transformation did not solve this. We therefore used a non-parametric ANOVA (Kruskal–Wallis) to determine group differences for these variables. In a second step, we examined correlations between variables (non-parametric where appropriate) to inform a multinomial logistic regression to determine whether prosody measures accounted for unique variance in predicting group membership (poor comprehenders vs chronological-age controls and poor comprehenders vs comprehension-level controls), when variation in nonverbal reasoning, vocabulary and word recognition[2] were taken into account. To constrain the number of variables in the model in light of our modest sample size, only prosody measures that showed significant correlations with reading comprehension were added to the model including the control variables (nonverbal reasoning and vocabulary) as well as word recognition.

Results

Descriptive statistics for all measures are presented in Table 1.

Group differences in reading, vocabulary and nonverbal reasoning

There were group differences on the scores of the reading comprehension task, as expected, $F(2,60) = 32.52$, $p < .001$, $\omega = .58$. The reading comprehension scores were similar for the poor comprehenders and for the younger, comprehension-level control group, $t(20) = 0.96$, $p = .339$, $r = .21$, but the poor comprehenders scored significantly lower than the chronological-age control group, $t(20) = 7.42$, $p < .001$, $r = .86$. Differences between the three groups on word recognition were marginally significant, $F(2,60) = 3.00$, $p = .058$, $\omega = .24$). Chronological-age controls showed marginally higher scores than poor comprehenders, $t(20) = 1.75$, $p = .085$, $r = .36$, and comprehension-level controls showed significantly higher scores than poor comprehenders, $t(20) = 2.36$, $p = .02$, $r = .47$. The groups did not differ on pseudoword decoding, $F(2,60) = 0.67$, $p = .515$, $\omega = .13$. Further, there were significant group differences on the productive vocabulary test, $F(2,60) = 3.69$, $p = .031$, $\omega = .28$. Poor comprehenders had significantly lower scores on the vocabulary test than the chronological-age control group, $t(20) = 2.66$, $p = .010$, $r = .51$, and marginally lower scores than the comprehension-level control group, $t(20) = 1.82$, $p = .074$, $r = .38$. Nevertheless, all children performed within the average range for their age, according to norms. There were also significant group differences on the percentile scores on the Raven; the nonverbal reasoning test, $H(2,63) = 14.81$, $p = .001$. Poor comprehenders had lower scores on this test than the chronological-age controls ($p = .003$) and also lower scores than the younger, comprehension-level controls ($p = .003$).

Group differences in prosody measures

Group differences were found on the reading-related prosody task, that is, text reading prosody, $F(2,60) = 16.36$, $p < .001$, $\omega = .57$. Poor comprehenders achieved significantly lower scores on text reading prosody than the chronological-age control group, $t(20) = 4.00$, $p < .001$, $r = .67$, but similar scores to the comprehension-level control group, $t(20) = -1.54$, $p = .130$, $r = .33$. The results of the performance on the speech prosody tasks, firstly storytelling prosody, also showed group differences, $F(2,60) = 12.49$, $p < .001$, $\omega = .52$. Poor comprehenders had a weaker performance on storytelling prosody than the chronological-age controls, $t(20) = 3.53$, $p < .001$, $r = .62$, but a similar performance to the younger comprehension-level controls, $t(20) = -1.30$, $p = .200$, $r = .28$. The results on the other speech prosody tasks, that is, the PEPS-C tasks, showed that there were group differences on all tasks, except one; on the receptive stress placement task, only marginally significant group differences were found, $H(2,63) = 5.94$, $p = .051$. The chronological-age control group scored at ceiling level on this task, with a high median score and a small range (*Mdn* = 16/16, *Range* = 12–16), whereas the scores from the poor comprehenders and the comprehension-level control group had a wider range. Group differences were found on the expressive stress placement task, $H(2,63) = 7.71$, $p = .021$, the receptive word boundary task, $H(2,63) = 11.15$, $p = .004$, and the expressive speech rhythm task, $H(2,63) = 6.48$, $p = .039$. Pairwise comparisons showed that poor comprehenders scored significantly lower than the chronological-age control group ($p = .027$, $p = .006$, $p = .036$, respectively) but similar to the comprehension-level control group ($p = .105$, $p = 1.00$, $p = .325$, respectively) on these tasks. Finally, groups also scored significantly different on the receptive speech rhythm task, $H(2,63) = 11.68$, $p = .003$. On this task, poor comprehenders scored lower than the chronological-age control group ($p = .006$) but also lower than the younger, comprehension-level control group ($p = .015$).

Logistic regression

Parametric and non-parametric correlations (uncorrected for multiple comparisons) are presented in Table 2.

Reading comprehension, vocabulary and nonverbal reasoning, but not word recognition or pseudoword decoding, were significantly associated with prosody measures, which were also partly associated with each other.

To examine whether prosody measures accounted for unique variance in predicting group membership (poor comprehenders vs chronological-age controls vs comprehension-level controls), when variation in nonverbal reasoning, vocabulary and word recognition were taken into account, multinomial logistic regression was carried out. First, we created a model with the control variables (nonverbal reasoning and vocabulary) and word recognition as predictors (Table 3). This model was significant, $X^2(2) = 20.55$, $p < .01$, but the explained variance was low (McFadden $R^2 = .15$). Nonverbal reasoning was the only significant predictor in both the comparison of poor comprehenders versus chronological-age controls, and the comparison of poor comprehenders versus comprehension-level controls. Based on the correlations, several prosody measures were subsequently included. When performance on the text reading prosody task, storytelling prosody task, expressive stress placement task and the receptive word boundaries task were added (Table 4), this resulted in a significant model fit, $X^2(2) = 70.77$, $p < .01$, and a higher amount of explained variance (McFadden

Table 2. Pearson (above the diagonal) and non-parametric (Kendall's tau; below the diagonal) correlations for all participants (n = 63) between selection, control and prosody measures.

		1	2	3	4	5	6	7	8	9	10	11	12
Selection	1. Reading comprehension	–	.05	.14	.26*	.32*	.52***	.41***	.11	.03	.26*	.19	.37**
	2. Word recognition	.01	–	.70***	.01	.22	.11	−.09	.14	.20	.08	.04	.02
	3. Pseudoword decoding	.09	.56***	–	−.01	.11	.20	−.04	.16	.24	.01	.05	−.03
Control	4. Vocabulary	.20*	−.01	.02	–	.45***	.41***	.35***	.21	.47***	.45***	.56***	.07
	5. Nonverbal reasoning	.22*	.17	.07	.32**	–	.26*	.43***	.36**	.28*	.25	.32*	.18
Text-related prosody	6. Text reading prosody	.41***	.08	.14	.29**	.18	–	.59***	.06	.22	.37**	.21	.36**
Speech prosody	7. Storytelling prosody	.30***	−.07	−.03	.24**	.31**	.46***	–	.13	.22	.33**	.32**	.31*
	8. Speech rhythm (rec.)	.10	.10	.13	.15	.20*	.05	.07	–	.48***	.18	.28*	.03
	9. Speech rhythm (exp.)	.01	.10	.12	.34***	.15	.17	.13	.33***	–	.27*	.40**	.04
	10. Stress placement (rec.)	.17	.05	.06	.22*	.21*	.29**	.30**	.12	.22*	–	.15	.17
	11. Stress placement (exp.)	.19*	.01	.01	.48***	.29**	.22*	.27**	.14	.33***	.10	–	.12
	12. Word boundaries (rec.)	.29**	−.01	−.01	.06	.14	.29**	.26**	−.04	.08	.18	.12	–

Note: rec., receptive; exp., expressive.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 3. Results for the multinomial logistic regression with nonverbal reasoning, vocabulary and word recognition as predictors.

		95% CI for odds ratio		
	B (*SE*)	Lower	Odds ratio	Upper
Poor comprehenders vs chronological-age controls				
Intercept	−7.55 (2.97)*			
Nonverbal reasoning	0.04 (0.02)*	1.01	1.04	1.07
Vocabulary	0.20 (0.16)	0.88	1.22	1.67
Word recognition	0.26 (0.19)	0.90	1.30	1.89
Poor comprehenders vs comprehension-level controls				
Intercept	−7.40 (2.92)*			
Nonverbal reasoning	0.04 (0.02)*	1.01	1.04	1.07
Vocabulary	0.09 (0.15)	0.80	1.09	1.49
Word recognition	0.33 (0.19)	0.96	1.40	2.02

Note: CI, confidence interval; *SE*, standard error.
*$p < .05$.

R^2 = .51). In the comparison of poor comprehenders versus chronological-age controls, performance on the text reading prosody task and the receptive word boundaries task was significant predictors. In the comparison of poor comprehenders versus comprehension-level controls, performance on the nonverbal reasoning task and on the storytelling prosody task was significant predictors, whereas performance on the expressive stress placement task was marginally significant.

Discussion

In the current study, we examined to what extent poor comprehenders differed in their prosodic abilities across written and spoken modalities from a chronological-age control group and a younger, comprehension-level control group, in order to disentangle the contribution of decoding skills to the association between prosody and reading comprehension. The first main result was that, despite having age-appropriate decoding efficiency, poor comprehenders were outperformed on the text reading prosody task by the chronological-age control group. Indeed, text reading prosody performance predicted group membership in the logistic regression, even when variance in nonverbal reasoning, vocabulary and word recognition was taken into account. In fact, their performance on text reading prosody was in line with the performance of the younger, comprehension-level control group. This suggests that although decoding may be necessary for text reading prosody performance, it is, in itself, not sufficient for text reading prosody to develop. This conclusion is reinforced by the second main result that poor comprehenders performed weaker than the chronological-age controls, but similar to the comprehension-level controls, on most speech prosody tasks. While accounting for

Table 4. Results for the multinomial logistic regression with prosody measures as predictors, in addition to nonverbal reasoning, vocabulary and word recognition.

		95% CI for odds ratio		
	B (*SE*)	Lower	Odds ratio	Upper
Poor comprehenders vs chronological-age controls				
Intercept	−33.05 (10.43)**			
Nonverbal reasoning	0.02 (0.02)	0.98	1.02	1.07
Vocabulary	−0.01 (0.24)	0.62	0.99	1.59
Word recognition	0.23 (0.24)	0.79	1.26	2.01
Text reading prosody	1.55 (0.68)*	1.23	4.72	18.05
Storytelling prosody	0.21 (0.32)	0.66	1.24	2.32
Stress placement (expressive)	0.02 (0.18)	0.73	1.03	1.45
Word boundaries (receptive)	0.67 (0.33)*	1.03	1.96	3.74
Poor comprehenders vs comprehension-level controls				
Intercept	−4.10 (4.80)			
Nonverbal reasoning	0.08 (0.03)*	1.02	1.08	1.15
Vocabulary	−0.12 (0.28)	0.51	0.89	1.52
Word recognition	0.29 (0.24)	0.84	1.34	2.15
Text reading prosody	−0.13 (0.40)	0.40	0.87	1.93
Storytelling prosody	−0.98 (0.42)*	0.17	0.38	0.85
Stress placement (expressive)	0.51 (0.27)^	0.98	1.66	2.81
Word boundaries (receptive)	0.13 (0.31)	0.62	1.14	2.08

Note: CI, confidence interval; *SE*, standard error.
^$p < .06$.
*$p < .05$.
**$p < .01$.

variance in nonverbal reasoning, vocabulary and word recognition, only performance on the storytelling prosody task significantly predicted group membership when comparing poor comprehenders with comprehension-level controls, indicating that an impairment in the ability to produce appropriate prosody might restrict reading comprehension development.

Reading comprehension is a complex process that requires children to quickly and accurately recognise the words in a text (‘the automaticity aspect’) while simultaneously constructing meaning. It has been proposed that reading fluency – as a combination of accuracy, automaticity and text reading prosody – facilitates the reader’s construction of meaning (Kuhn et al., 2010). The results from the current study, however, suggest that the ‘automaticity aspect’ of reading is a distinct process from the construction of meaning. The construction of meaning seems more closely tied to text reading prosody than to decoding efficiency, at least, when children have mastered automaticity in reading. The theoretical rationale behind this is that text reading prosody may facilitate the unification between the retrieved words on a phonological, syntactic and semantic level. This is in line with neurocognitive models of language processes that propose that memory retrieval and unification processes comprise two distinct brain areas that operate in parallel (Hagoort, 2007). This proposed facilitation in unification processes could have implications for the construct of text reading fluency in relation to reading comprehension. We suggest that in mature readers that have automatised decoding, assessment of text reading fluency should always include the component of text reading prosody. Text reading prosody performance could provide an insight in how well a child manages to unify the phonological, syntactic and semantic levels, and therefore, how well he or she constructs meaning from the text.

The second main result concerns the performance on the speech prosody tasks, partly distinguishing poor comprehenders from both chronological-age and comprehension-level controls, and provides further evidence for a relation between prosody and reading comprehension. Deficiencies in perceiving and producing speech prosody in poor comprehenders – as is evident on some of the speech prosody tasks – could obstruct the use of implicit prosody (an inner representation of what a text should sound like) while reading a text silently. It has been suggested that implicit prosody may facilitate text comprehension (Kentner, 2012; Kuhn et al., 2010; T. Rasinski et al., 2009). Fodor (1998, 2002) proposed the implicit prosody hypothesis stating that a default prosodic contour is projected onto a text, in order to help solve syntactic ambiguity when reading silently. The results from the current study raise the question whether poor comprehenders may not have access to a default prosodic contour. Given that the results of the groups' comparisons and the logistic regression gave somewhat inconsistent results, it is not entirely clear which aspects of prosody are affected in poor comprehenders. Although group comparisons suggested that poor comprehenders performed more poorly than chronological-age controls on most speech prosody measures, only performance on the receptive word boundaries subtask (indicating the ability to successfully discriminate between 'Chocolate-cake and jam' vs 'Chocolate, cake and jam') predicted group membership in the logistic regression. When comparing poor comprehenders with comprehension-level controls, only storytelling prosody predicted group membership. It is likely that shared variance across prosody tasks and with non-verbal reasoning and vocabulary skills partly explains these inconsistencies. Further research, using tasks and designs that carefully control differential reliance on wider language and memory skills across different aspects of prosody, is required to clarify which aspects of prosody may be consistently delayed and/or impaired in poor comprehenders.

Although the prosodic abilities of poor comprehenders have, as far as we know, not been examined before, a considerable body of research has investigated speech-language problems in poor comprehenders. It has been shown that poor comprehenders have weaker grammatical, syntactic and semantic skills (Nation, Clarke, Marshall, & Durand, 2004; Nation & Snowling, 2000). The current study adds an impairment in prosody skills in this group. It would be interesting to investigate relations between prosody and grammatical, syntactic and semantic skills in poor comprehenders in future research, as it has been proposed that one of the functions of prosody is the attribution of syntactic roles to words within sentences (Chafe, 1988; Koriat, Greenberg, & Kreiner, 2002). Furthermore, an appropriate use and understanding of prosody may assist in segmenting a sentence into syntactically and semantically correct chunks (Kintsch, 1998; Snedeker & Trueswell, 2003; Snedeker & Yuan, 2008).

Importantly, longitudinal research showed that the weaker grammatical, syntactic and semantic skills in poor comprehenders are persisting and can, in retrospect, be related to speech-language impairments in earlier school years (Catts et al., 2006; Nation et al., 2010). Because speech prosody develops well before learning to read, it seems plausible to assume that delays in prosodic development may be observed well before poor comprehenders start to fail at reading comprehension. This could potentially provide valuable information for early language screening, as delayed development in speech prosody skills may hinder later reading comprehension. The relation between prosody and reading comprehension would be in line with the developmental trajectory of speech prosody and its influence on later literacy development, such as phonological and morphological awareness, as outlined by Zhang and McBride-Chang (2010). Future studies, however, are needed to further examine the relation between early speech prosody and later reading comprehension and the potential of this for early screening possibilities.

The fact that some subtasks of the PEPS-C had a low internal reliability is a limitation of the study, and this means that the findings of the current study need to be interpreted with some caution. Possibly, reduced range in scores on some of them contributed to the low internal reliability. Nevertheless, it should be mentioned that there are currently very few highly reliable measures of prosodic sensitivity (Holliman et al., 2014). A second limitation of the study concerns the significant group differences in nonverbal reasoning, vocabulary and word recognition that were not controlled for in the non-parametric ANOVAs. As a consequence, those group comparison should be interpreted with caution. Nevertheless, both text reading prosody and some speech prosody measures were found to predict group membership in subsequent logistic regression models in which these variables were taken into account. In future work, the use of non-parametric analyses of covariance (Akritas, Arnold, & Du, 2000) in group comparisons, or the use of a regression model (see for instance, Tong, Deacon, Kirby, Cain, & Parrila, 2011) to identify those children who struggle with comprehension, rather than a case-control design, would more easily allow for control of nonverbal reasoning and language skills. Finally, analysing the acoustic profile of children's prosody in future work would be a useful addition to the use of the Multidimensional Fluency Scale. The current study is the first to examine prosodic abilities in poor comprehenders, and the results should therefore be taken as a first step. Further research is needed to confirm these results and to examine differences in prosody development in more depth.

Conclusion

The current study provides evidence for a delay in both text reading prosody and speech prosody in poor comprehenders, compared with typical readers of the same age. Because poor comprehenders have age-appropriate decoding skills but weak reading comprehension skills, we were able to show that decoding efficiency in itself is not sufficient to establish the relation between text reading prosody and reading comprehension. It is therefore proposed that text reading prosody may be more strongly related to the level of reading comprehension. This was also shown by our finding that poor comprehenders had impairments in speech prosody. Poor perception and production of speech prosody may hinder an internal representation of what a text should sound like, which is suggested to obstruct comprehension of written text.

Acknowledgements

We thank Imke Schepers and Jolijn Wolbers for assistance with data collection and the children, parents and staff from the participating schools.

Notes

1. In these national assessments, the word recognition test reflected timed single-word reading, whereas in the reading comprehension assessment, children were asked to answer multiple-choice comprehension questions about several texts. As such, these measures were highly similar in terms of the procedures as the word recognition and reading comprehension measures administered by us and described in the succeeding text. The items used in the assessments were completely different though.
2. We chose to include word recognition rather than pseudoword decoding because the groups differed on the former, but not the latter. The same variables predicted group membership though, when pseudoword decoding was included instead of word recognition.

References

Aarnoutse, C. & Kapinga, T. (2006). *Begrijpend lezen 345678*. Ridderkerk: Onderwijs Advisering.

Akritas, M.G., Arnold, S.F. & Du, Y. (2000). Nonparametric models and methods for nonlinear analysis of covariance. *Biometrika*, 87(3), 507–526. https://doi.org/10.1093/biomet/87.3.507.

Berninger, V.W., Abbott, R.D., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S. et al. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment*, 28(1), 3–18. https://doi.org/10.1177/0734282909336083.

Brus, B.T. & Voeten, M.J.M. (1973). Een-minuut test [EMT]. Vorm A en B. *Verantwoording en handleiding*. Nijmegen: Berkhout.

Calet, N., Defior, S. & Gutiérrez-Palma, N. (2015). A cross-sectional study of fluency and reading comprehension in Spanish primary school children. *Journal of Research in Reading*, 38(3), 272–285. https://doi.org/10.1111/1467-9817.12019.

Catts, H.W., Adlof, S.M. & Weismer, S.E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, 49(2), 278–293. https://doi.org/10.1044/1092-4388(2006/023.

Chafe, W. (1988). Linking intonation units in spoken English. In *Clause combining in grammar and discourse*, (pp. 1–27). Amsterdam: John Benjamins B.V.

Fletcher, J.M.H., Saywitz, S.E., Shankweiler, D.P., Katz, L., Liberman, I.Y., Stuebing, K.K. et al. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*, 86(1), 1–23. https://doi.org/10.1037/0022-0663.86.1.6.

Fodor, J.D. (1998). Learning to parse. *Journal of Psycholinguistic Research*, 27(2), 285–319. https://doi.org/10.1023/A:1023258301588.

Fodor, J.D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the 32nd Annual Meeting of the North East Linguistic Society*, (pp. 113–132). Amherst, MA: GLSA, University of Massachusetts.

Hagoort, P. (2007). The memory, unification, and control (MUC) model of language. In T. Sakamoto (Ed.), *Communicating skills of intention*, (pp. 259–291). Tokyo: Hituzi Syobo.

Holliman, A.J., Williams, G.J., Mundy, I.R., Wood, C., Hart, L. & Waldron, S. (2014). Beginning to disentangle the prosody-literacy relationship: A multi-component measure of prosodic sensitivity. *Reading and Writing*, 27(2), 255–266. https://doi.org/10.1007/s11145-013-9443-6.

Kent, W.M. (2013). *Intonation And Reading Skills In Fourth-Grade Students*. Wayne State University Dissertations. Paper 773.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, 123(1), 1–20. https://doi.org/10.1016/j.cognition.2011.11.012.

Kim, Y.S.G. & Wagner, R.K. (2015). Text (speech) 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(3), 224–242. https://doi.org/10.1080/10888438.2015.1007375.

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

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge UK: Cambridge University Press.

Klauda, S.L. & Guthrie, J.T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, 100(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310.

Koriat, A., Greenberg, S. & Kreiner, H. (2002). The extraction of structure during reading: Evidence from reading prosody. *Memory & Cognition*, 30(2), 270–280. https://doi.org/10.3758/bf03195288.

Kort, W., Schittekatte, M., Bosmans, M., Compaan, E.L., Dekker, P.H., Vermeir, G. et al. (2005). *Dutch version of Wechsler Intelligence cale for Children-WISC-III*. Amsterdam: Pearson Assessment and Information B.V.

Krom, R., Jongen, I., Verhelst, N., Kamphuis, F. & Kleintjes, F. (2010). *DMT en AVI. Groep 3 tot en met 8*. Arnhem: Cito.

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. https://doi.org/10.1598/rrq.45.2.4.

Kuhn, M.R. & Stahl, S.A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3–21. https://doi.org/10.1037/0022-0663.95.1.3.

Lochrin, M., Arciuli, J. & Sharma, M. (2015). Assessing the relationship between prosody and reading outcomes in children using the PEPS-C. *Scientific Studies of Reading*, 19(1), 72–85. https://doi.org/10.1080/10888438.2014.976341.

Miller, J. & Schwanenflugel, P.J. (2006). Prosody of syntactically complex sentences in the speech reading of young children. *Journal of Educational Psychology*, 98(4), 839–853. https://doi.org/10.1037/0022-0663.98.4.839.

Miller, J. & Schwanenflugel, P.J. (2008). A longitudinal study of the development of reading prosody as a dimension of speech reading fluency in early primary school children. *Reading Research Quarterly*, 43(4), 336–354. https://doi.org/10.1598/rrq.43.4.2.

Nation, K., Clarke, P., Marshall, C.M. & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research*, 47(1), 199–211. https://doi.org/10.1044/1092-4388(2004/017).

Nation, K., Cocksey, J., Taylor, J.S. & Bishop, D.V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, 51(9), 1031–1039. https://doi.org/10.1111/j.1469-7610.2010.02254.x.

Nation, K. & Snowling, M.J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied PsychoLinguistics*, 21(02), 229–241. https://doi.org/10.1017/s0142716400002046.

Peppé, S. & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, 17(4–5), 345–354. https://doi.org/10.1080/0269920031000079994.

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. https://doi.org/10.1080/19388070802468715.

Rasinski, T.V. (2004). Creating fluent readers. *Educational Leadership*, 61(6), 46–51.

Raven, J.C. (1976). *Raven's standard progressive matrices*. Oxford: Oxford Psychologists Press Ltd.

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 speech reading of young readers. *Journal of Educational Psychology*, 96(1), 119–129. https://doi.org/10.1037/0022-0663.96.1.119.

Snedeker, J. & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48(1), 103–130. https://doi.org/10.1016/S0749-596X(02)00519-3.

Snedeker, J. & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58(2), 574–608. https://doi.org/10.1016/j.jml.2007.08.001.

Stanovich, K.E. & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, 86(1), 24–53. https://doi.org/10.1037/0022-0663.86.1.24.

Staphorsius, G. & Krom, R. (2011). *CITO toetsen begrijpend lezen*. Arnhem: Cito.

Tong, X., Deacon, S.H., Kirby, J.R., Cain, K. & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology*, 103(3), 523–534.

Van den Bos, K.P., Lutje Spelberg, H.C., Scheepstra, A.J.M. & de Vries, J.R. (1994). De Klepel. Vorm A en B. Verantwoording, Handleiding, Diagnostiek en behandeling. Nijmegen: Berkhout.

Veenendaal, N.J., Groen, M.A. & Verhoeven, L. (2015). What speech text reading fluency can reveal about reading comprehension. *Journal of Research in Reading*, 38(3), 213–225. https://doi.org/10.1111/1467-9817.12024.

Verhoeven, L. & Vermeer, A. (2001). *Taaltoets Alle Kinderen [TAK]*. Arnhem: Cito groep.

Vermeer, A. (2000). Lexicale rijkdom, tekstmoeilijkheid en woordenschatgrootte: Beschrijving van de MLR, een woordeschat-analyseprogramma. *Toegepaste*

Whalley, K. & Hansen, J. (2006). The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading*, 29(3), 288–303. https://doi.org/10.1111/j.1467-9817.2006.00309.x.

Zhang, J. & McBride-Chang, C. (2010). Auditory sensitivity, speech perception, and reading development and impairment. *Educational Psychology Review*, 22(3), 323–338. https://doi.org/10.1007/s10648-010-9137-4.

Margriet A. Groen is Assistant Professor at the School of Pedagogical and Educational Science at the Radboud University in Nijmegen, the Netherlands. She studies language and literacy development in typically and atypically developing children. The latter includes children with dyslexia, Specific Language Impairment, hearing impairments or Down syndrome. She is particularly interested in how speech perception and production contribute to the development of phonological representations and processes.

Nathalie J. Veenendaal currently works as educational trainer and consultant at Utrecht University. Before that, she worked at the Behavioural Science Institute of the Radboud University in Nijmegen. She studied the role of text reading fluency in literacy processes, with a special interest in the relation between prosody and reading comprehension.

Ludo Verhoeven is full professor in Psychology and Education at the Radboud University in Nijmegen, the Netherlands. Within the Behavioural Science Institute, he coordinates a research group on Learning and Plasticity. His research interests focus on the acquisition of language and literacy by first-language and second-language learners, including children with learning difficulties. He is also director of the National Language Education Centre that aims at improving the teaching and learning of Dutch language arts.

Received 24 January 2016; revised version received 22 September 2017.

Address for correspondence: Margriet A. Groen, Behavioural Science Institute, Radboud University, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: *m.groen@pwo.ru.nl*