Rhythm discrimination and metronome tapping in 4-year-old children at risk for developmental dyslexia

Marina Kalashnikova[1, 2, 3], Denis Burnham[3] and Usha Goswami[4]

1. BCBL. Basque Center on Cognition, Brain and Language, Spain
2. IKERBASQUE, Basque Foundation of Science, Spain
3. MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
4. Centre for Neuroscience in Education, University of Cambridge, UK

Corresponding author: Marina Kalashnikova, BCBL. Basque Center on Cognition, Brain and Language, Mikeletegi Pasealekua, 69, Donostia, Gipuzkoa, 20009, Spain; t: +34 943 309 300, m.kalashnikova@bcbl.eu

Acknowledgements: This research was supported by the Australian Research Council grant DP110105123, “Seeds of Literacy”, to the 2nd and 3rd authors. We thank Maria Cristou-Ergos, Scott O’Loughlin, and Hana Zjakic for their assistance with participant recruitment, data collection, and data analyses. We also thank all the children and their parents for their valuable time and interest in this research.

Highlights

- Rhythmic abilities were assessed in children at- and not at-risk for dyslexia.
- A temporal rate of 2Hz was used, hypothesised as important for reading development.
- At-risk children showed poorer performance in the rhythm discrimination task.
- Rhythm production related to non-word repetition, vocabulary, and letter knowledge.
- Rhythmic difficulties in at-risk children relate to early reading-readiness skills.

Abstract

Temporally accurate perception and production of rhythmic patterns are key factors related to language development and reading acquisition. Here we investigate rhythm discrimination and rhythm production in children who are at family risk or not at family risk for dyslexia, to compare group performance in these tasks prior to the start of reading instruction and to investigate the relation between individual children's rhythmic abilities and pre-reading skills. Four-year-old children completed a rhythm discrimination task and a rhythm production task, both utilizing a temporal rate of 2Hz, and also received pre-reading measures of non-word repetition, vocabulary, and letter knowledge. Controls outperformed at-risk children in the rhythm discrimination task and in the pre-reading measures. No group differences were observed in the rhythm production task, but individual differences in this task were related to scores of non-word repetition, vocabulary size, and letter knowledge. The data are discussed in terms of Temporal Sampling theory (Goswami, 2011).

Keywords: Dyslexia; Rhythm; Temporal sampling; Reading; Phonological awareness; Letter knowledge

Introduction

The accurate perception and production of rhythmical patterns is associated with language and reading development, in both typically-developing children and children with dyslexia or developmental language disorder (Cumming, Wilson, Leong, Colling, & Goswami, 2015; Dellatolas, Watier, Le Normand, Lubart, & Chevrie-Muller, 2009; Douglas & Willatts, 1994; Forgeard et al., 2008; Goswami, Huss, Mead, Fosker, & Verney, 2013a; Goswami et al., 2013b; Huss, Verney, Fosker, Mead, & Goswami, 2011; Lundetrae & Thomson, 2018; Overy, 2000, 2003; Overy, Nicolson, Fawcett, & Clarke, 2003; Waber et al., 2000; Wolff, 2002; Wolff, Michel, Ovrut, & Drake, 1990). Regarding early reading, individual differences in rhythm production by pre-readers are related to measures of reading readiness (Rios-Lopez, Molinaro, & Lallier, 2019; Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014). Further, training young children's rhythm perception and production skills can improve their early reading performance (e.g., Bhide, Power, & Goswami, 2013; Cancer et al., 2020; Degé & Schwarzer, 2011; Flaugnacco et al., 2015; Moritz, Yampolsky, Papadelis, Thomson, & Wolf, 2013). However, the developmental mechanisms underlying the relationships between individual differences in rhythm perception, rhythm production, reading and language are currently unclear. In order to study these causal mechanisms, longitudinal studies are required, which would allow for systematic assessments before children acquire reading skills, ideally beginning in infancy. Here, we present a cross-sectional study with children participating in such a longitudinal study, designed to test a theoretical framework linking rhythm, language, and reading (Temporal Sampling theory, Goswami, 2011; described further below). The current data were collected during participants' fourth year of life, the final year before schooling and the onset of reading instruction.

Rhythm and Reading Development

Both rhythm perception and rhythm production have been shown to be related to reading development in typically-developing children. Rhythmic non-speech beat perception measured prior to schooling predicts reading development, letter-sound knowledge, and phonological awareness from 3 to 6 years (Corriveau, Goswami & Thomson, 2010; Ozernov-Palchik, Wolf, & Patel, 2018), and the perception of rhythm in music is also related to later reading and phonological awareness. For example, Anvari et al. (2002) created musical tasks based on piano tones requiring either rhythm pattern copying or same/different rhythm discrimination and administered them to a pre-reading sample of 4-year-olds. All the musical tasks loaded on a single factor in a principal component factor analysis, and this factor was significantly associated with later phonological awareness and single word reading.

Rhythm production in young children is most typically measured by asking them to drum, clap or tap in time to a rhythmic beat, a measure often termed rhythmic synchronisation. Individual differences in clapping, drumming and tapping tasks are reliably related to reading development across languages (e.g., Bonacina, Krizman, White-Schwoch & Kraus, 2018 [English]; Dellatolas et al., 2009 [French]; Lundetrae & Thomson, 2018 [Norwegian]). Rhythmic synchronisation is also related to pre-reading skills across languages, for example letter-name knowledge (Rios-Lopez et al., 2019) and phonological awareness (Woodruff-Carr et al., 2014). If the link between rhythmic skills and reading is a causal one, then training pre-school children in rhythm should improve their reading and phonological awareness, and this is the case. For example, Degé and Schwarzer (2011) trained German preschoolers with a musical programme that included joint singing, joint drumming, metrical training, dancing, and other rhythmic exercises. The children subsequently showed significant gains in phonological awareness in comparison to a sports training control group. Similar training effects have been reported in English (Moritz et al., 2013). Five-year-old children who were following a music curriculum based on the Kodaly rhythm method showed better phonological awareness at the end of their pre-school year than a matched group following a musical curriculum that did not focus on rhythm, and also performed better in experimental tasks such as rhythm copying on a bongo drum. Two years later, significant time-lagged correlations between rhythm copying and later phonological awareness, non-word repetition, reading and spelling were found (Moritz et al., 2013). Accordingly, individual differences in both rhythm perception and production are related to reading and phonological development in typically-developing children (see also Rautenberg, 2015).

Rhythm and Dyslexia

Developmental dyslexia is a disorder of learning that primarily affects the skills involved in accurate and fluent word reading and spelling (Rose, 2009). The characteristic cognitive features across languages are difficulties in phonological awareness, verbal memory and verbal processing speed (Snowling, 2000). Critically, beat perception and rhythmic synchronisation are reliably impaired in children who have developmental dyslexia (Flaugnacco et al., 2014; Forgeard et al., 2008; Goswami et al., 2013a; Huss et al., 2011; Overy, 2000; 2003). Individual differences analyses show that accuracy in rhythm discrimination tasks accounts for over 40% of the variance in both single-word reading and reading comprehension, and over 30% of the variance in phonological awareness, even after controlling for age and nonverbal IQ (Huss et al., 2011; Goswami et al., 2013a). Children with dyslexia even show poorer accuracy in rhythm discrimination tasks than do younger children matched by reading level (Goswami et al., 2013a). Children with developmental dyslexia also show impairments in rhythmic synchronisation tasks, and individual differences are again related to phonological awareness and reading (Overy et al., 2003; Thomson & Goswami, 2008; Flaugnacco et al., 2014). Moreover, rhythm-based training for children with developmental dyslexia or with poor reading skills has been shown to improve both phonological awareness and reading (Bhide et al., 2013; Flaugnacco et al., 2015; Cancer et al., 2020).

The developmental associations between rhythm perception and production in children with dyslexia and their poor reading outcomes are thought to relate in part to the role of rhythm in developing children's language skills. Rhythm perception is considered a universal precursor of language acquisition, as infants across numerous languages begin parsing the speech signal by using rhythm and stress patterns (Mehler et al., 1988). The hierarchically-nested levels of rhythmic organisation in speech are intimately related to the identification of phonological units at different linguistic levels such as stressed syllables, syllables and rhymes (Liberman & Prince, 1977; Leong & Goswami, 2015). The child's ability to consciously isolate and manipulate phonological units in speech such as syllables, phonemes and rhymes is called phonological awareness, and this cognitive construct is causally related to reading development across languages (Ziegler & Goswami, 2005). Children with developmental dyslexia show reduced awareness of phonology at all levels in the linguistic hierarchy (Goswami, 2018). In the next section, we introduce Temporal Sampling theory (Goswami, 2011) as a neural, sensory and cognitive framework for understanding the well-documented links between rhythm, phonology, reading and dyslexia in children.

Temporal Sampling (TS) Theory of Dyslexia

TS theory links rhythm, phonology and reading via novel psychoacoustic and neural data, and was originally proposed to explain phonological difficulties in dyslexia (Goswami, 2019, for most recent update). TS theory proposes that the automatic alignment of endogenous brain rhythms with rhythm patterns in speech is critical for linguistic and phonological development, and that this unconscious alignment (or sampling) process is atypical in children with developmental dyslexia from birth (Goswami, 2011, 2015, 2018). This occurs in part because the sensory cues that trigger automatic neural alignment are perceived poorly. Consequently, the development of phonology is also atypical, which leads to impaired reading. TS theory builds on data from neural (brain rhythms), sensory (perception of speech amplitude envelope cues), and cognitive (behavioural) studies. Briefly, it is known that in the adult brain, accurate tracking of the speech amplitude envelope (the critical low-frequency acoustic information that assists linguistic decoding at multiple time scales) is triggered by the automatic phase-resetting of brain rhythms by acoustic "edges" in the speech signal (Doelling, Arnal, Chitza, & Poeppel, 2014; Giraud & Poeppel, 2012; Gross et al., 2013). These perceptual edges are the rise times in amplitude associated with syllables, and amplitude rise times are important perceptual cues to both speech and non-speech rhythm (Greenberg, 2006). Rise time is typically called 'attack time' in the musical literature (the rise time of a musical note; Gordon, 1987). Children with dyslexia are known to show impaired discrimination of amplitude rise times across languages (Goswami, 2015, for review), and the degree of their impairment is associated with individual differences in language-relevant phonological awareness tasks. In addition, rise time impairments are associated with their poorer rhythm discrimination and synchronisation skills (Thomson & Goswami, 2008; Huss et al., 2011). TS theory thus proposes that impaired sensory discrimination of rise times in dyslexia is related to atypical automatic speech-brain alignment, and that this atypical sensory/neural processing affects the perception of rhythm and the development of phonological awareness (Goswami, 2018).

As noted above, longitudinal studies are required to study causal mechanisms in development. The sample of children in the current study are participating in an ongoing longitudinal study of TS theory, and have received a series of auditory, language and phonological tasks at different ages, beginning at 5 months. In prior publications, we have reported that the participants at family risk for dyslexia (AR group) showed impaired rise time discrimination at the age of 10 months, and that individual differences in rise time discrimination in infancy predicted vocabulary size at 3 years (Authors, b, c). We have also reported impaired novel word learning in the AR group (Authors, e), and a delay in the ability to encode phonetic variability in familiar lexical items such as words produced in a foreign accent (Authors, d). The current study investigated the perception and production of rhythmical patterns assessed when these children were 4 years of age, the year prior to school entry.

The Current Study

This study pursued two objectives. First, it investigated the effects of family history of dyslexia (children at-risk for dyslexia vs. children not at-risk for dyslexia) on children's rhythm perception and production skills prior to the onset of reading instruction. Tapping and other rhythmic synchronisation behaviours typically show large variability in child populations (Drake, Jones, & Baruch, 2000; McAuley, Jones, Holub, Johnston, & Miller, 2006; Provasi & Bobin-Begue, 2003), and so 4 years was judged the earliest at which rhythmic skills could be measured in our sample. Second, this study investigated the relations between individual children's rhythm perception and production (children at-risk for dyslexia *and* not at-risk for dyslexia) and their developing reading-related skills. In particular, we assessed the extent to which early rhythm abilities predicted individual reading readiness outcome measures over and above the variance explained by early language and phonological skills.

To assess rhythm perception, children completed a rhythm discrimination task that measured their ability to detect rhythmic disruptions caused by differences in the duration of accented notes in rhythmical musical sequences. The sequences comprised different arrangements of a series of musical notes with an underlying pulse rate of 500 msec (120 bpm). Each series was delivered twice within one trial, and children's task consisted in detecting whether the series were the same or different. For half of the trials (Different trials), there was a change in rhythmic structure in the second sequence, either caused by adding 100 msec to the accented notes (Short Duration change) or by adding 166 msec to the accented notes (Long Duration change). This variation was always used in the prior literature, and it allowed us to additionally test children's discrimination of differences that were perceptually easier versus more difficult to detect - the Long Duration changes (166 msec) were expected to be perceptually more salient and thus easier to differentiate. In prior work, older children with dyslexia have performed poorly in this rhythm discrimination task irrespective of whether Short or Long Duration changes disrupted the rhythm (Huss et al., 2011; Goswami et al., 2013a). Whether this would be the case with younger participants is unknown.

To assess rhythm production, children completed a metronome tapping task that measured their ability to produce finger taps synchronised to a metronome beat. Following our prior work in dyslexia, both the rhythm discrimination and metronome tapping tasks were organised around a frequency of 2 Hz (a beat rate of 500 msec; Goswami, 2019, for review). Selection of a beat rate of 2 Hz was thought most likely to reveal group differences between AR and NAR pre-readers, given prior neural data suggesting that the dyslexic brain may not set up a reliable internal representation of rhythmic inputs organised around a 2 Hz rate (Colling, Noble

& Goswami, 2017; Molinaro, Lizarazu, Lallier, Bourguignon, & Carreiras, 2016; Power, Mead, Barnes & Goswami, 2013; Power, Colling, Mead, Barnes, & Goswami, 2016).

Children had also completed non-word repetition, vocabulary size, and letter knowledge tasks. Non-word repetition has proved a reliable measure of phonological sensitivity with younger children. Letter knowledge is a task that assesses children's reading readiness since letter knowledge scores are known to be a strong predictor of later reading development (Byrne, 1998; de Jong & van der Leij, 1999; Edwards, Beckman & Munson, 2004; Hulme et al., 2007; Lervåg et al., 2009; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Snowling, Chiat & Hulme, 1991). Finally, vocabulary was assessed as a measure of general individual linguistic proficiency, which is related to reading outcomes but not a characteristic cognitive deficit associated with dyslexia.

Our predictions were based on TS theory and prior data with older children. First, with regards to the at-risk vs. not at-risk group comparisons, we predicted impaired performance for the at-risk children in both measures of rhythm perception and production. This prediction was made as prior studies using the same rhythm discrimination and metronome tapping tasks with older children have shown impaired performance by those with dyslexia in both measures (Thomson & Goswami, 2008; Huss et al., 2011). Second, for the correlational and regression analyses comprising our entire sample, we predicted significant relationships between individual differences in the two rhythm measures and the phonological sensitivity and reading readiness measures. This prediction was made as correlations between rhythm and phonology and reading measures have been found with similar samples of older children (Thomson & Goswami, 2008; Huss et al., 2011; Goswami et al., 2013).

Method

Participants

All children were originally recruited to participate in the longitudinal project [blind for review] at the age of 5 months and were included in this particular cross-sectional study based on their availability to visit the lab at the age of 4 years. This study was approved by the [blind for review] Human Research Ethics Committee (approval: H9142, title: [blind for review]). Fifty-six children were included (26 female, *M* age = 48 months, *SD* = .43 months). Out of this total sample, 55 children contributed data for the language related tasks and the metronome tapping task, and 45 children contributed data for the rhythm discrimination task. The differences in sample sizes were due to individual children's failure to contribute analysable data for both experimental tasks (see Procedure for individual tasks below for further detail). The samples included in the analyses for each task and across tasks are reported in Table 1.

Table 1. *Sample sizes of the NAR and AR groups included in the group comparisons for the rhythm discrimination, metronome tapping, and language related tasks; and in the correlational and regression analyses across tasks.*

Task and Planned Statistical Analyses	NAR	AR	Total
Rhythm discrimination (group comparisons)	26	19	45
Metronome tapping (group comparisons)	29	26	55
Non-word repetition, vocabulary, and letter knowledge (group comparisons)	29	26	55
Rhythm discrimination, non-word repetition, vocabulary, and letter knowledge (correlation analyses)	25	18	43

Metronome tapping, non-word repetition, vocabulary, and letter knowledge (correlation analyses)	29	26	55
Metronome tapping, non-word repetition, vocabulary, and letter knowledge (regression/mediation analyses)	29	26	55

All children were growing up in English-language dominant families. Children's allocation to the AR and NAR groups was based on their parents' existing dyslexia diagnoses and/or performance on a comprehensive screening battery that included language, reading, and cognitive tasks, which they completed at the start of the [blind for review] project when their children were 5 months of age. Using this screening battery, a child was allocated to the AR group if one of their parents (1) obtained a score of 1.5 *SD* below the average in a measure of word or non-word reading and in at least two of the following tests – oral reading (accuracy, fluency, and rate), spelling, Rapid Picture Naming (RAN), and digit span, (2) indicated history of experiencing reading difficulties in childhood, and (3) obtained an average score (within .5 *SD* from the standardised mean) on a measure of non-verbal IQ. Of the 26 AR children, 11 had a parent with a previous formal diagnosis of dyslexia, and 17 had a parent who screened as dyslexic on our battery (in the case of three parents, this was in addition to the dyslexia diagnosis). Only one child was included for whom neither parent had a diagnosis nor screened as dyslexic, but in this case the child had two biologically-related older siblings with a formal diagnosis of dyslexia. A child was allocated to the NAR group if both their parents obtained scores within .5 *SD* from the average on all screening tests. Detailed information about parents' screening performance is presented in Authors (b) and summarised here in Appendix A. Maternal education was used as a proxy for the families' socio-economic-status. The median education level was a university degree, and it did not differ between mothers of children in the at-risk and not at-risk groups (Kolmogorov-Smirnov $Z = .443$, $p = .990$).

Children's non-verbal IQ was measured as part of the longitudinal study's test battery at the age of 3 years using the Fluid Reasoning Object Series/Matrices sub-test in the Routing Non-Verbal Domain of the Stanford-Binet Intelligence Scales-5th Edition (Roid, 2003), and it was confirmed that children's non-verbal IQ was within the normal range. Independent-samples *t*-tests comparing IQ scores between the two groups were not significant, $t(53) = 1.905$, $p = .060$, $d = .523$ (NAR $M = 11.31$, $SD = 2.48$; AR $M = 10.19$, $SD = 1.77$).

Rhythm Discrimination Task

Stimuli and apparatus. The stimuli and procedure for this task are identical to Goswami et al. (2013a) who used an adaptation of the Perception of Musical Meter task originally developed by Huss et al. (2011).

The auditory stimuli consisted of 48 rhythmic musical sequences that mixed accented and non-accented sounds. The sequences comprised sounds that were generated using a Sibelius Version 4 from a sound set produced by Native Instruments (Kontakt Gold) (Huss et al., 2011). The individual sounds were sampled from a vibraphone, sounded like naturally produced musical notes, and had an underlying pulse rate of 500 msec and an identical pitch (G, 392 Hz). Half of the sequences were constructed in 3/4 time and half in 4/4 time. This variation was introduced to maintain children's attention to the subtle changes in the stimuli throughout the task. The duration of the sequences ranged from 4.8 to 6.4 seconds ($M = 5.6$ seconds, $SD = .74$). Each sequence had one repeating accentuated sound every 2, 3, 4 or 5 notes per bar, created by increasing its intensity by 5dB. The change in the rhythm structure of each 'Different' sequence was created by elongating this accentuated sound by 100 msec or by 166 msec.

The 48 sound sequences were used to create 24 experimental trials. In each trial, a child heard two sequences and was required to indicate whether they were the same or different. Twelve trials contained two identical sequences (Same trials). Twelve trials contained two different sequences (Different trials). In the Different trials, the sequences were identical except for their rhythm structure determined by the elongation of the accentuated sound (see above). In 6 Different trials, the sound was elongated by 100 msec (Short Duration change) and in 6 trials by 166 msec (Long Duration change). The Different trials using 166 msec elongation were expected to be easier, as the longer accented sound should be perceptually more salient. Figure 1 depicts the structure of a Different trial used in the task. The sound stimuli used in each trial are available for download at [blind for review].

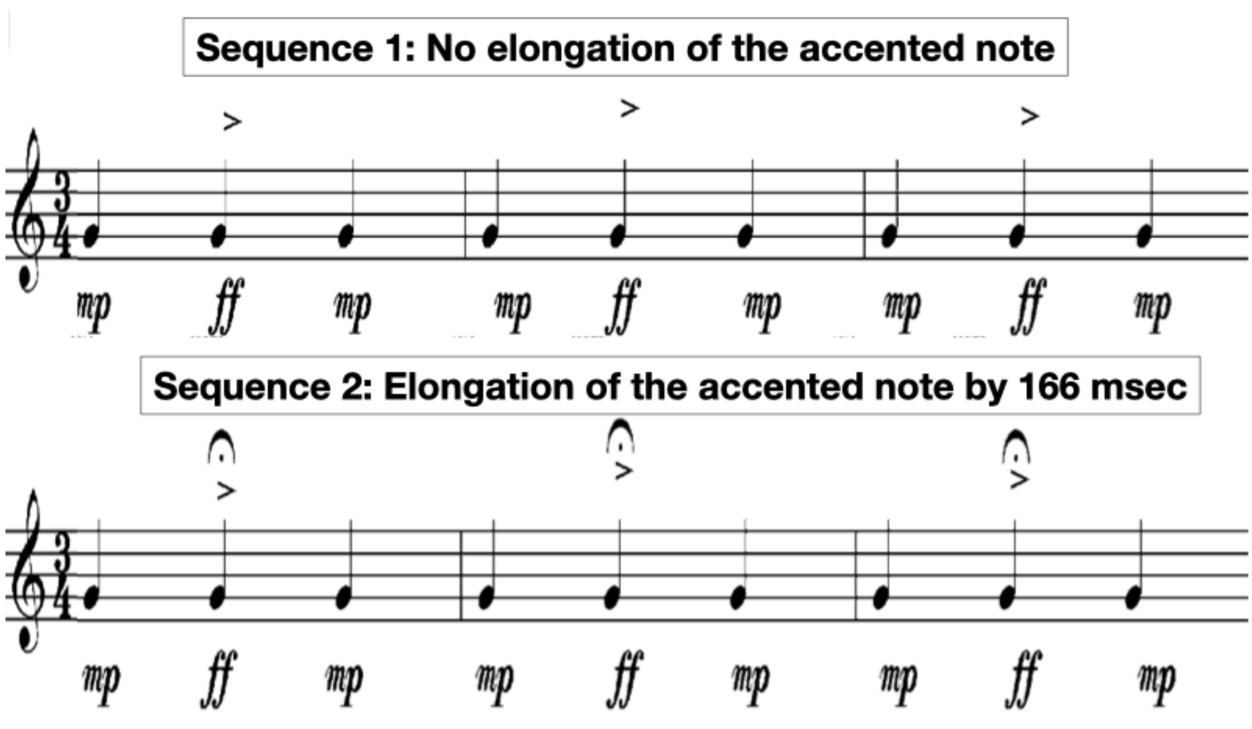

Figure 1. Sample auditory sequences of a Different trial of the rhythm discrimination task comprising 3 notes per bar. The two sequences are constructed in 3/4 time. The second sound in

both sequences is accentuated (>), and in Sequence 2, it is repeatedly elongated by 166 msec (Long Duration change), creating the difference in rhythm structure between Sequences 1 and 2.

The stimuli were embedded into a child-friendly interface designed in Matlab and presented on a Windows laptop computer. The auditory stimuli were presented over loudspeakers located to the right and left sides of the computer screen. A volume level that was clearly audible and comfortable for the participants was determined during piloting, and it was kept identical for all participants. The visual interface consisted of two pairs of circles presented side by side (each circle approximately 100 x 100 pixels in size, pairs displayed on the right and left sides along the horizontal plane on the screen). In one pair, the circles were identical in size and colour, and in the other pair, the circles were identical in size, but differed in colour.

Procedure. During the task, the child sat at a desk facing the laptop computer. An experimenter sat next to the child and controlled the stimuli presentation and recorded the child's responses using a computer mouse. Children were instructed to point to the two circles that looked the same when they heard two sounds that sounded the same, and to point to the two circles that looked different when they heard two sounds that sounded different. Children were told that sometimes it would be difficult to tell the different sounds in a pair apart, but that they had to pay attention and spot as many different pairs as possible. When children correctly identified a difference (i.e., provided a 'different' response on Different trials), they received a visual reward in the shape of a colourful sticker on the screen and a cheering sound. When they failed to identify a difference correctly (i.e., provided a 'same' response on Different trials or a 'different' response on Same trials), they heard a sighing sound. Correct responses to the Same trials did not trigger any feedback. The experimenter recorded the child's responses by clicking on the circles using a computer mouse. The experimenter encouraged the child to attend to the task but did not offer any feedback on their performance.

Given that children received feedback for correct and incorrect ‘different’ responses, several children refused to continue the task after repeatedly hearing the sighing sound and failing to collect any stickers for correct responses. Children with partial data were included in the analyses (a minimum of 5 completed trials was required for inclusion, see Results for full detail on the number of completed trials by AR and NAR children). Fourteen children failed to contribute any analysable data as they refused to try the task or failed to complete at least 5 trials, so they were excluded from the analyses (8 AR and 6 NAR).

Data analyses. Raw scores for this task consisted of the total correct responses that each child obtained on each type of trial (Same trials, Short duration change Different trials, and Long duration change Different trials). These raw scores were used to calculate two types of scores for analyses. First, raw scores were converted to proportions of correct trials out of the total completed trials of each type in order to account for individual differences in the total number of trials completed by each child (see Procedure). Second, raw scores were used to calculate *d* prime (*d*’) scores in order to capture precisely children’s ability to detect accurately the differences between two sequences while controlling for potential individual biases to ‘same’ or ‘different’ responses. These *d*’ scores were computed as the difference between children’s correct responses on Different trials (hits) and incorrect responses on Same trials (false alarms). The hit and false alarm raw scores were transformed to *z*-scores to compute the *d*’ ($d' = z(\text{hits}) - z(\text{false alarms})$).

Metronome Tapping Task

Stimuli and apparatus. This task is a simplified version of the Expressive Rhythm: Metronome task used by Thomson and Goswami (2008). A 20-second string of pure tone 10 msec events produced at the rate of 2 Hz was used as the auditory stimulus. The string was presented twice resulting in two experimental trials. Presentation software (Version 0.92, www.neuro-bs.com) was used to present the stimuli over loudspeakers (same volume settings as described above), and children produced their taps by pressing the Space bar on a computer keyboard.

Procedure. Children sat a desk next to an experimenter, with a computer keyboard on the desk in front of them. First, children completed a familiarisation phase in which they were asked to tap their index finger on the desk together with the experimenter. Then, they were asked to repeat the tapping, this time using the computer keyboard. The experimenter specifically focused on reminding children that they had to tap at the same time as she did, and that they should press the button every time instead of holding it down. When the child was comfortable with the procedure, they proceeded to the task in which the keyboard taps were recorded as input by the software. Four children were excluded from analyses for this task for failure to contribute analysable data as they refused to complete the task or could not understand the instructions during the familiarisation phase (2 AR and 1 NAR).

Data analyses. In order to include only taps that were preceded and followed by metronome events, the first and last taps were removed from the sequence, and as a result only taps that could be synchronised to the metronome events were analysed. Two measures were extracted for this task in order to enable comparison to earlier tapping studies (Colling et al., 2017; Corriveau & Goswami, 2009; Cumming et al., 2015). First, a synchronisation score was calculated for each child's performance on each trial. These scores capture the absolute distance in msec of each tap produced by the child from the expected tap rate (Inter-Tap-Interval, ITI; every 500 msec following the metronome rate), and they were computed as the |(median ITI-500)|. Given that the resulting scores could not be negative, and this resulted in a skewed distribution, the scores were log transformed. In addition, +1 was added to each score before the transformation to avoid infinite values, and +6 was added after the transformation, so that all the final scores would be positive. Hence, the synchronisation scores ranged from 0 to 6, and they were averaged across the two trials completed by each child. Therefore, a score of 6 denoted that a child's median ITIs were identical to the rate of the metronome (every 500 msec), and lower scores denoted how far the child's medians were from the metronome rate, with 0 being the furthest. The second measure consisted in calculating the standard deviation (*SD*) of the ITIs for each trial completed by each child as an index of inter-subject variability in performance. This score captures the consistency of the rates of the taps produced by each child independently of whether these rates were close to the target rate of the metronome (i.e. a high synchronisation score) or far from it (i.e. a low synchronisation score).

Non-word repetition, vocabulary, and letter knowledge tasks

Children also completed tasks assessing non-word repetition, vocabulary size, and letter knowledge. The non-word repetition task was completed during a visit to the lab prior to the rest of the tasks, children were 3 years 6 months when they completed this task.

Non-word repetition. The experimental measure of non-word repetition used here is designed for use with children between 2 and 3 years of age (Authors, c). In this task, children were introduced to a hand puppet, and told that the puppet was an alien from a different planet who spoke an alien language. Children were then asked to learn some of the words of the alien language by repeating them after the puppet. While the experimenter manipulated the puppet, the stimuli were played over loudspeakers. The task included four practice items, and 16 test items. The test items were split into three categories: 1-syllable (6 items), 2-syllables (6 items), and 3-syllables (4 items). The full list of items is presented in Appendix B. Each item was presented twice before the child was asked to respond. If the child failed to respond, they were given the opportunity to attempt the trial one more time. In this case, the experimenter said, “That was a hard one to say, have another go”, and the item was presented again (presented twice as on the first attempt). If the child produced an incorrect repetition on the first attempt, they proceeded to the next trial. The experimenter coded whether the child’s production of the word matched or mismatched its target form and wrote down the exact production by the child. Only identical repetitions were coded as correct. The numbers of first attempt correct, second attempt correct, total correct, total incorrect, and missing responses were computed. The proportion of total correct responses out of the total attempted items was used for analyses.

Vocabulary. The Knowledge Vocabulary sub-test in the Routing-Verbal Domain of the Stanford-Binet Intelligence Scales-5th Edition (Roid, 2003) was administered. This test assesses children’s receptive and expressive vocabulary skills. In the receptive vocabulary items, children hear a word and are asked to point to its referent on a picture (or on themselves in case of body parts), and in the expressive vocabulary items, children hear a word and see its image and are asked to provide the word’s definition. A single scaled score (M = 10; SD = 3) computed based on the combination of receptive and expressive vocabulary test items was calculated for each child and used for analyses.

Letter knowledge. The letter knowledge section from the Early Reading Skills subtest of the Wechsler Individual Achievement Test, WIAT III (Wechsler, 2009) was used. In this test, children were shown a booklet depicting letters from the alphabet and were asked to name the letters and find the letters named by the experimenter. This section consists of 13 items, and each child received a score of 1 for a correct response on each item, so the total scores could range from 0 to 13.

Results

Planned Statistical Analyses

The data from the rhythm discrimination, metronome tapping, non-word repetition, vocabulary, and letter knowledge tasks were subject to two types of analysis to address our hypotheses regarding group differences and inter-task correlations. First, NAR and AR performance was compared for each task using *t*-test analyses and Analyses of Variance (ANOVA) in order to assess the effects of children's risk status on rhythm discrimination, metronome tapping, non-word repetition, vocabulary, and letter knowledge. Given that two scores were calculated for each rhythm task capturing different aspects of children's performance (rhythm discrimination: proportion scores and *d*' scores; metronome tapping: synchronisation score and Median ITI), the group comparisons were conducted separately for each score. Second, correlational followed by regression analyses were used to assess task associations for the entire sample. First, correlations between each rhythm task and the non-word repetition, vocabulary, and letter knowledge tasks were assessed. Second, the results of these correlational analyses were used to identify the variables of interest for a subsequent hierarchical linear regression analysis that aimed to test directly the extent to which children's rhythm skills were uniquely associated with performance in the early measure of reading readiness (letter knowledge).

Rhythm Discrimination Task

This task was used to assess children's rhythm perception abilities. TS theory would predict an overall group difference, with better rhythm perception for the NAR group. The analyses in this section include data from 43 children (Table 1). Given that children could contribute partial data to this task, the number of completed trials was first compared across the NAR and AR groups. Children in the NAR group (M = 23.46, SD = 2.74, range 10 to 24) completed significantly more trials than children in the AR group (M = 17.84, SD = 8.5, range 5 to 24), $t(43) = 3.156$, $p = .003$, $d = .963$. Critically, all subsequent analyses of children's performance in this task employed proportion and *d*' scores. As described in the Data Analysis section for this task, the computation of proportion and *d*' scores took into consideration the number of trials completed by each child, so it is not the case that NAR children had higher scores because they completed more trials (as would be the case if raw scores were analysed)[1]. For completeness, the Online Supplementary Materials present additional analyses of the relation between the number of trials completed by children and their performance in this task.

First, group performance for proportion correct across the three types of trials was assessed using a repeated-measures ANOVA with Trial Type (Same, Short duration change different, Long duration change different) as the within-subjects factor and Group (NAR, AR) as the between-subjects factor. There was no main effect of Trial Type, $F(2, 86) = 2.549$, $p = .084$, $\eta^2p = .056$, but a main effect of Group, $F(1, 43) = 5.402$, $p = .025$, $\eta^2p = .112$, and no Trial Type by Group interaction, $F(1, 43) = 1.195$, $p = .308$, $\eta^2p = .027$. Across trial types, AR children provided lower proportions of correct responses than NAR children (Figure 2).

[1] In order to further test this possibility, we compared the proportion of correct responses produced by NAR and AR children in the first 5 trials, which was the minimum number of trials completed by a child on this task. NAR children (M = .577, SD = .207) obtained higher scores compared to AR children (M = .453, SD = .239). This difference did not reach statistical significance, $t(43) = 1.866$, $p = .069$, likely due to the inclusion of such a reduced number of trials for all participants, but it yielded a medium effect size, Cohen's $d = .563$.

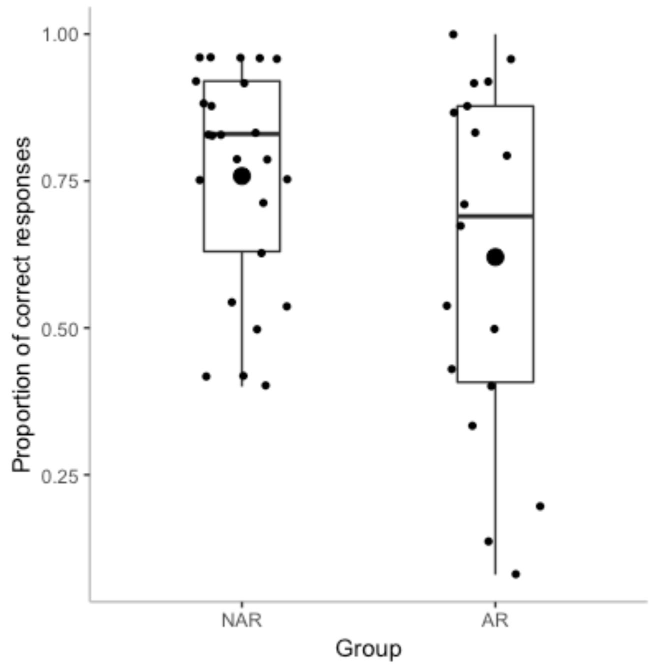

Figure 2. Proportion of correct responses on trials completed by children in the AR and NAR groups in the rhythm discrimination task (the internal large circles represent the Mean, the internal line represents the Median, and the hinges extend to the first and third quartiles).

Second, group performance was assessed using *d*' scores, which in addition to capturing children's overall success in the task also controlled for potential individual biases to 'same' or 'different' responses. Given that the analysis above did not yield a significant effect of elongation for different trials (Short duration change vs. Long duration change), these data were collapsed here. An independent-samples *t*-test confirmed the significant group difference in performance; NAR children (M = .719, SD = 1.01) obtained significantly higher *d*' scores compared to AR children (M = -.137, SD = 1.72), $t(43) = 2.035$, $p = .048$, $d = .621$.

Metronome Tapping Task

This task was used to assess children's rhythm production abilities. TS theory would predict an overall group difference, with better rhythm production for the NAR group. The analyses in this section include data from 55 children (Table 1). The synchronisation scores and standard deviations (*SD*) for the inter-trial-intervals (ITIs) for children in the NAR and AR groups are presented in Figure 3. The average synchronisation score in the NAR group was 2.71 (*SD* = 1.04) and 2.52 (*SD* = 1.12) in the AR group, and an independent-samples *t*-test comparing these scores was not statistically significant, $t(53) = .675$, $p = .503$, $d = .185$. As can be seen in Figure 3, children in the AR group produced taps that varied to a greater extent than children in the NAR group, but the *t*-test comparing the ITI *SD*s between groups was not statistically significant, $t(52) = -.82$, $p = .418$, $d = .227$.

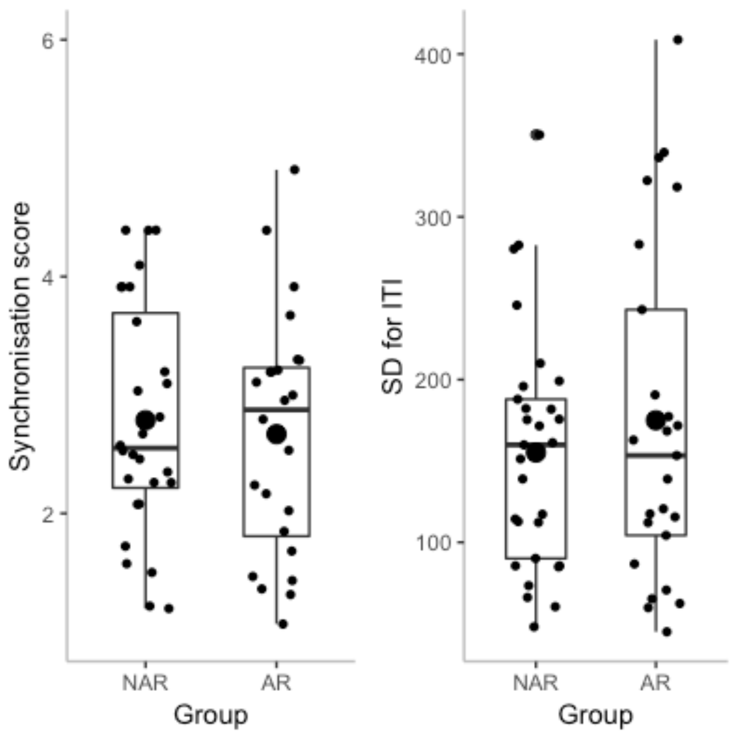

Figure 3. Metronome tapping performance for children in the AR and NAR groups. The synchronisation scores are displayed on the left panel and standard deviation (*SD*) for the inter-trial-intervals (ITI) on the right panel (the internal large circles represent the Mean, the internal line represents the Median, and the hinges extend to the first and third quartiles).

Non-word Repetition, Vocabulary, and Letter Knowledge Tasks

The non-word repetition, vocabulary, and letter knowledge tasks were used to assess children's phonological sensitivity, linguistic proficiency, and reading readiness, respectively. The analyses in this section include data from 55 children (Table 1). Table 2 presents NAR and AR children's scores for the non-word repetition, vocabulary, and letter knowledge tasks and the results of independent-samples *t*-tests comparing group performance. Children in the NAR group significantly outperformed children in the AR group in all these tasks.

Table 2. *Mean (SD; range) scores for the non-word repetition, vocabulary, and letter knowledge tasks, and independent-samples t-test (df = 53) analyses comparing performance in the NAR and AR groups.*

	NAR	AR	*t*	*p*	*d*
Non-word repetition	.61[1] (.21; .19-.94)	.47 (.23; 0-.81)	2.442	.017	.671
Vocabulary	11.21[2] (2.21; 8-19)	9.65 (2.59; 3-15)	2.398	.020	.659
Letter knowledge	6.21[3] (3.76; 1-13)	3.96 (3.99; 0-13)	2.148	.036	.590

[1]Proportion correct score; [2]standardised test score (M = 10; SD = 3); [3]raw score out of 13.

Relations between Rhythm Tasks, Non-word Repetition, Vocabulary, and Letter Knowledge

TS theory would predict significant relations between individual differences in the rhythm measures (rhythm discrimination and metronome tapping), phonological sensitivity (non-word repetition) and reading readiness (letter knowledge). First, correlational analyses were conducted to assess the relationship between rhythm discrimination performance (*d*' scores), metronome tapping performance (synchronisation scores) and non-word repetition, vocabulary, and letter knowledge scores for the two groups of children combined (see Table 3). Given that not all scores were normally distributed, Spearman correlations are reported. The correlations involving metronome tapping scores were conducted on data from 43 children, and the remaining correlations were conducted on data from 55 children. Additional correlational analyses were conducted separately for the AR and NAR sub-groups, and these results are reported in the Online Supplementary Materials.

Table 3. *Spearman correlation analyses for children's rhythm discrimination, metronome tapping, non-word repetition, vocabulary, and letter knowledge scores. (*p<.05, **p<.01)*

	Rhythm discrimination	Non-word repetition	Vocabulary	Letter knowledge
Metronome tapping	-.221	.318*	.266*	.304*
Rhythm discrimination		-.056	.054	-.018
Non-word repetition			.337*	.448**
Vocabulary				.388**

Inspection of Table 3 shows that, contrary to prediction, individual differences in children's rhythm discrimination performance did not show significant relations with their non-word repetition and letter knowledge scores, nor with metronome tapping. A relation between rhythm discrimination and metronome tapping was also absent. However, children's metronome tapping scores were significantly related to non-word repetition, letter knowledge and vocabulary scores (all p-values < .05). These significant relations are plotted in Figure 4.

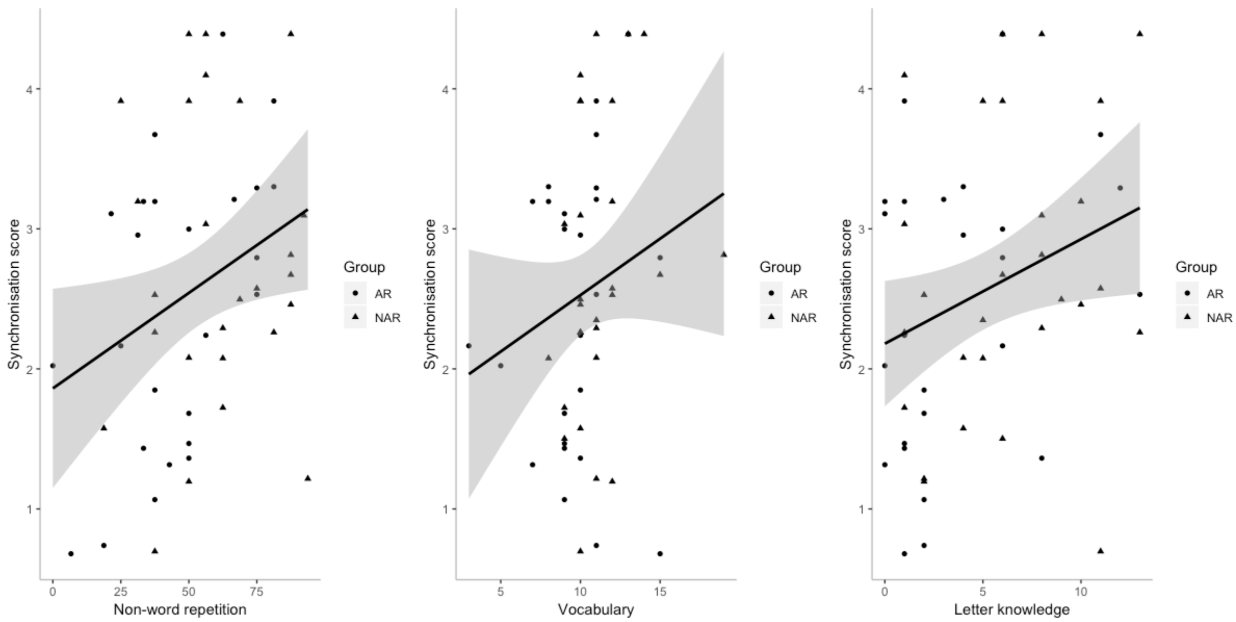

Figure 4. Significant relations between NAR and AR children's performance in the metronome tapping task and non-word repetition (left), vocabulary (center), and letter knowledge (right) (regression lines represent linear model fits, and Standard Error is shown by the shaded area).

Relations between Children's Rhythm Production and Reading Readiness

Next, regression analyses were conducted in order to assess whether children's rhythm production skills predict their emerging literacy skills over and above the variance explained by their phonological sensitivity. A series of regression models were conducted to assess the unique variance explained by metronome tapping scores in children's non-word repetition, letter knowledge, and vocabulary scores. Rather than report each model here, these results can be found in the Online Supplementary Materials. As can be seen in Table SI 2, even though metronome tapping was significantly correlated with vocabulary, when entered into a linear regression model controlling for non-word repetition, it no longer accounted for significant variance in individual vocabulary scores. Given this result and our prediction that children's rhythm abilities would relate to phonological sensitivity and reading readiness only, vocabulary was not included in the regression model presented here.

This linear regression model was constructed with letter knowledge as the dependent variable, and independent variables non-verbal IQ (Step 1), non-word repetition (Step 2), and metronome tapping score (Step 3). The resulting model accounted for a significant 20.1% of variance (see Table 4). Children's non-word repetition was a significant predictor of their letter knowledge, but their metronome tapping score was not. The standardised coefficients and model fit were identical when metronome tapping was entered at Step 2, indicating that these results are not influenced by the order of the predictors.

Table 4. *Results of the fixed-order linear regression model with non-verbal IQ, non-word repetition, and metronome tapping score as the predictor variables and letter knowledge scores as the dependent variable*

Model fit: R^2 = .201, $F(3, 50)$ = 5.434, p = .003					
	R^2 *change*	β	*SE*	*t*	*p*
Step 1: IQ	-	.128	.232	1.012	.316
Step 2: Non-word repetition	.164	.362	.023	2.753	.008
Step 3: Metronome tapping	.026	.172	.496	1.312	.195

Taken together, these regression analyses indicate that non-word repetition scores act as a mediator of the observed relation between metronome tapping and letter knowledge performance observed in the full sample of 55 children. Theoretically, this may reflect a developmental pathway in which rhythmic processing is one determinant of a child's phonological sensitivity as measured by non-word repetition, with individual differences in phonological sensitivity then determining early reading readiness skills, such as letter knowledge. To test this possibility, a mediation analysis was conducted using the PROCESS macro in SPSS (Hayes, 2019). The mediation model was specified with letter knowledge score as the outcome variable, metronome tapping score as the predictor variable, and non-word repetition score as the mediator variable. The model is summarised in Figure 4. As shown, the standardised regression coefficients between metronome tapping scores and non-word repetition, and non-word repetition and letter knowledge were significant, but this was not the case for the effect of metronome tapping scores on letter knowledge (.722, shown in parenthesis in Figure 4), which reflected the findings of our regression model (Table 4). The mediation analysis yielded an indirect effect of .472, and its statistical significance was confirmed using bootstrapping procedures showing that its confidence interval ranged from .068 to 1.074. The indirect effect is the product of the coefficients of the two independent effects, and it captures the full mediation pathway (metronome tapping to non-word repetition to letter knowledge). These results indicate that while the temporal accuracy of children's tapping synchronisation as measured by the metronome tapping task was significantly related to their emerging literacy skills, this relationship was mediated by children's phonological abilities as measured by the non-word repetition task.

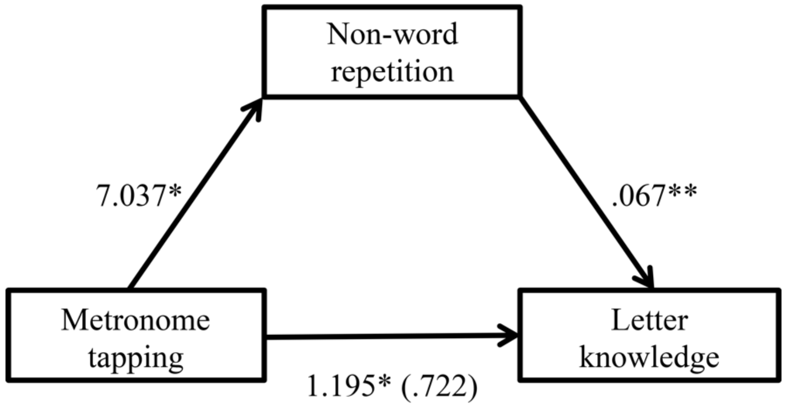

Figure 4. Regression coefficients for the mediation analysis (the coefficient for the relation of metronome tapping and letter identification controlling for non-word repetition is shown in parenthesis); $p < .02^*$, $p < .01^{**}$.

Discussion

The present study set out to investigate rhythm perception and rhythm production abilities in children aged 4 years who were either at family risk (AR) for developmental dyslexia or not at-risk (NAR), by using rhythm discrimination and metronome tapping tasks that were previously used with older children with developmental disorders of language (Corriveau & Goswami, 2009; Cumming et al., 2015; Flaugnacco et al., 2014; Goswami & Thomson, 2008; Goswami et al., 2013a; Huss et al., 2011). On the basis of TS theory and these prior studies, we predicted that the AR children would show significantly poorer performance in both these tasks assessing rhythm perception and production abilities, and that individual differences in these tasks in the whole sample of participants would be significantly associated with the measures of phonological sensitivity and reading readiness.

For the rhythm discrimination task based on musical events, which has not previously been administered to 4-year-olds, as predicted, the AR group showed significant impairments.

Firstly, AR children completed significantly fewer trials in this task compared to the NAR children. In the group analysis of *d*' sensitivity, which accounted for the differences in the number of trials completed by individual children, the AR group showed significantly less accurate rhythm perception than did NAR children, which is consistent with data from older dyslexic children (Huss et al., 2011; Goswami et al., 2013a). Secondly, accuracy in the AR group was not affected by whether the Different rhythm trials depended on a Short Duration change or a Long Duration change, again consistent with data from older dyslexic children for age-matched comparisons (Huss et al., 2011; Goswami et al., 2013a). Contrary to expectation, however, individual differences in rhythm discrimination performance within our sample were not related to performance in the non-word repetition or letter knowledge tasks. This is inconsistent with findings from samples of older children speaking English or Italian (Cumming et al., 2015; Flaugnacco et al., 2014; Goswami et al., 2013; Huss et al., 2011). All prior studies with English-speaking children have found significant relations between individual differences in rhythm perception as measured by the current task, reading and phonological awareness, although not with vocabulary. However, the children in those studies were considerably older than the 4-year-olds in this sample, namely 9 years (Huss et al., 2011; Cumming et al., 2015), 11 years (Goswami et al., 2013a), and 8 - 11 years (Flaugnacco et al., 2014). Accordingly, it is possible that this particular rhythm perception measure lacks sensitivity for younger participants.

To investigate this possibility post-hoc, we examined individual responding (proportion correct) for children in each group. In the NAR group, 22 out of 26 children responded correctly on more than 50% of trials they completed, binomial test $p = .001$. This suggests that the task is not too difficult for NAR 4-year-olds. For the AR group, however, only 8 out of 19 children responded correctly on more than 50% of trials they completed, binomial test $p = .648$. Hence as a group, the AR children were at chance in the rhythm discrimination task. Taken together with the finding that it was significantly more challenging for AR children to proceed through all the trials of the task, the chance level performance characterising this group makes it less surprising that the hypothesised relations between rhythm perception, phonological sensitivity and reading readiness were not significant. It is possible that, if the task were to be re-administered when the children are older and are performing above chance, significant relations with reading and phonology may emerge.

In the metronome tapping task, where similar rhythmic synchronisation tasks involving tapping or drumming have been used previously with typically-developing children of this age group (Rios-Lopez et al., 2017, 4-year-olds; Woodruff Carr et al., 2014, 3-4 year-olds), the AR children did not show a group-level impairment compared to the NAR children. This was contrary to our hypothesis, but of course not all the AR children may later turn out to qualify for a diagnosis of developmental dyslexia. Indeed, visual inspection of Figure 3 shows that individual metronome tapping scores were widely dispersed within each group, with higher variability in the AR group. Individual differences in tapping to the beat were significantly related to all the outcome measures administered at around 4 years for the full sample, namely vocabulary development, non-word repetition, and letter knowledge. Figure 4 depicts the scatterplots in each case. However, the mediation analysis suggested that the overall developmental picture was that rhythm production related to phonological sensitivity and phonological sensitivity related to the reading readiness measure (letter knowledge). It can be noted that significant relations between beat synchronisation (tapping or drumming in time with a beat), letter name knowledge and phonological awareness have been reported in similar studies relying only on typically-developing samples (Rios-Lopez et al., 2019; Woodruff-Carr et al., 2014).

The rhythm discrimination and the metronome tapping tasks used here employed a 2 Hz beat rate. The frequency of 2 Hz has been hypothesised to be an important periodic structure across human languages (based on the timing of stressed syllables, see Goswami et al., 2013a, b) to which the infant brain may synchronise during linguistic processing. Indeed, there is evidence that a beat rate of 2 Hz may be biologically privileged for human behaviour. For example, in speech, stressed syllables occur at approximately 500 msec intervals across languages (Dauer, 1983), and when mothers sing “playsongs” such as lullabies to their infants, the average tempo is 498 msec (~2 Hz; Trainor, Clark, Huntley, & Adams, 1997). Spontaneous applause that is rhythmically synchronised also converges on a similar (493 msec) average (~2 Hz; Neda, Ravasz, Brechet, Vicsek, & Barabasi, 2000). Developmentally, McAuley et al. (2006) demonstrated that children aged 8 years and above show spontaneous tapping rates centered around 500 msec. Biologically, these convergent findings may suggest that an underlying beat rate of 500 msec emerges because of physiological factors, possibly related to neural synchronisation to acoustic inputs such as rhythmic speech or music. It is notable that speech directed to infants emphasises this temporal rate, showing a modulation peak at 2 Hz, whereas speech directed to adults peaks at 5 Hz (Leong, Kalashnikova, Burnham, & Goswami, 2017). Temporal modulations (amplitude modulations) at ~2 Hz also sit at the apex of the prosodic hierarchy of stress feet, syllables, onset-rime units and phonemes, as demonstrated via computer modelling of nursery rhymes (Leong & Goswami, 2015). It is thus plausible that neural synchronisation at this pulse rate may be impaired in developmental disorders of language such as dyslexia (Goswami, 2019). Accordingly, remedial rhythmic programs for young children with developmental disorders of language are likely to be of benefit developmentally.

Indeed, a number of remedial programs that included tapping synchronisation practice have produced positive results regarding improvements in phonological awareness (Degé & Schwarzer, 2011; Bhide et al., 2013, Flaugnacco et al., 2015; Ozernov-Palchik et al., 2018). The most recent of these studies showed an advantage for metrically organised rhythmic patterns (temporal patterns with an underlying isochronous grid) over non-metrical sequences, even though the latter were hypothesised *a priori* to be more characteristic of speech rhythmic structure, which is only quasi-rhythmic (Ozernov-Palchik et al., 2018). On the other hand, according to TS theory, it is precisely the experience of linguistic input with relatively isochronous metrical patterning, such as infant-directed speech (Leong et al., 2017) and the rhythmic routines of the nursery (hand clap games and nursery rhymes, see Leong & Goswami, 2015), that support early language acquisition. Isochronous routines provide a metrically-organised scaffold of acoustic landmarks to which the infant brain can entrain (see discussions in Goswami et al., 2013b; Cumming et al., 2015; Richards & Goswami, 2019). Once this scaffold is in place, typically during the first year of life, then the neural temporal prediction system can begin to learn the non-metrical rhythmic patterning that is more characteristic of natural language and its quasi-rhythmic nature. One way to explore these temporal rate and isochrony issues in more depth would be to contrast the effects on children's phonological awareness of general musical remediation (e.g., Slater et al., 2014) with remediation focused specifically on 2 Hz rhythms versus other rhythms, along with remediation focused specifically on metrically-structured language (see Bhide et al., 2013).

The current study had a number of limitations. In particular, it is unfortunate that our testing schedule in the children's fourth year in the ongoing longitudinal project did not enable a test of phonological awareness. However, non-word repetition has also been used as a phonological sensitivity measure in other rhythm-based studies with children at this age (Moritz et al., 2013). We do note that even though non-word repetition scores were significantly correlated with children's rhythm production as predicted, children overall showed low performance on this task (around 60% for NAR children). This confirms that assessing phonological sensitivity and phonological awareness is challenging in young children, so it will be critical to assess these skills again when this sample becomes older. A second limitation is that children were not encouraged to continue until the end of the rhythm discrimination task. The AR children found this task difficult, and some were discouraged by the lack of positive feedback, refusing to complete the task. Given that the typically-developing children were able to complete the task, in future work the task could be adapted to include motivators every few trials. Finally, the sample size of 55 children could be increased in future studies. Longitudinal studies by their nature cannot change the sampling once the study has commenced, but it is notable that some recent studies of rhythm synchronisation in slightly older children have had larger samples than the current study (Bonacina et al., 2018, 5 – 7 year-olds, $N = 64$), while studies with 4-year-olds have had similar samples (Anvari et al., 2002, 4-year-olds, $N = 50$; Rios-Lopez et al., 2017, 4-year-olds, $N = 43$; Woodruff Carr et al., 2014, 3-4 year-olds, $N = 35$).

Conclusion

This study investigated rhythm perception and production abilities in children at-risk and not at-risk for dyslexia at the age of 4 years, one year before they began formal education and reading instruction. In line with previous studies involving older children with dyslexia, risk status affected rhythmic abilities, with at-risk children showing significantly poorer rhythm perception compared to the not at-risk controls. Analyses that included the entire sample of children demonstrated that individual differences in rhythm production significantly predicted individual reading readiness abilities over and above the variance explained by early phonological sensitivity. Specifically, phonological sensitivity mediated the relation between rhythm and reading readiness. The developmental pathway demonstrated in this study is consistent with Temporal Sampling theory, which provides a mechanistic framework that accounts for these relations between rhythm, phonology, and reading-related skills. As our participants are now entering school, in the future we will be able to assess whether the rhythm discrimination and rhythm synchronisation tasks used here are predictors of their later reading acquisition.

References

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002) Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology,* 83(2): 111-130, 2002

Authors (a) Reference removed for blind review.

Authors (b) Reference removed for blind review.

Authors (c) Reference removed for blind review.

Authors (d) Reference removed for blind review.

Authors (e) Reference removed for blind review.

Bhide, A., Power, A.J., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain & Education,* 7 (2), 113-123. doi.org/10.1111/mbe.12016

Bonacina, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Clapping in time parallels literacy and calls upon overlapping neural mechanisms in pre-readers. *Annals of the New York Academy of Sciences, 1423*, 338-48. doi.org/10.1111/nyas.13704

Byrne, B. J. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle*. Psychology Press.

Cancer, A., Bonacina, S., Antonietti, A., Salandi, A., Molteni, J., & Lorusso, M.L. (2020). The effectiveness of interventions for developmental dyslexia: Rhythmic reading training compared with hemisphere-specific stimulation and action video games. *Frontiers in Psychology, 11*, 1158. doi: 10.3389/fpsyg.2020.01158

Colling, L.J., Noble, H.L. & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience, 11,* 360. doi.org/10.3389/fnins.2017.00360

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex,* 45*,* 119-130. doi.org/10.1016/j.cortex.2007.09.008

Corriveau, K. H., Goswami, U., & Thomson, J. M. (2010). Auditory processing and early literacy skills in a preschool and kindergarten population. *Journal of learning disabilities*, *43*(4), 369-382.

Cumming, R., Wilson, A., Leong, V., Colling, L.J., & Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments: A new look at a perceptual hypothesis. *Frontiers in Human Neuroscience,* 9*,* 672. doi.org/10.3389/fnhum.2015.00672

Dauer, R. (1983). Stress-timing and syllable timing revisited. *Journal of Phonetics, 11*, 51-62.

De Jong, P. F., & Van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of educational psychology*, *91*(3), 450.

Degé, F & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology,* 2*,* 124.

Dellatolas, G., Watier, L., Le Normand, M-T., Lubart, T., & Chevrie-Muller, C. (2009). Rhythm reproduction in kindergarten, reading performance at second grade, and developmental dyslexia theories. *Archives Clinical Neuropsychology, 24*, 555-563. doi.org/10.3389/fpsyg.2011.00124

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85*, 761–768. doi.org/10.1016/j.neuroimage.2013.06.035

Douglas, S. & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading*, *17 (2),* 99-107. doi.org/10.1111/j.1467-9817.1994.tb00057.x

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*(3), 251-288. doi.org/10.1016/S0010-0277(00)00106-2

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421-436.

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S. et al. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience, 8,* 392. doi.org/10.3389/fnhum.2014.00392

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., & Schon, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: An RCT. *PLoS One, 10*(9), e0138715. doi.org/10.1371/journal.pone.0138715

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4)*,* 383-390. doi.org/10.1525/mp.2008.25.4.383

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, *82*(1), 88-105.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Science, 15*, 3-10. doi.org/10.1016/j.tics.2010.10.001

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43-54.

Goswami, U. (2018). A neural basis for phonological awareness? An oscillatory temporal-sampling perspective. *Current directions in psychological science*, *27*(1), 56-63.

Goswami, U. (2019). Speech rhythm and language acquisition: an amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, *1453*(1), 67-78.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading & Writing*, *23*, 995-1019. doi.org/10.1007/s11145-009-9186-6

Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. (2013a). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex, 49* (5), 1363-1376. doi.org/10.1016/j.cortex.2012.05.005

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013b). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language, 69 (1),* 1-17. doi.org/10.1016/j.jml.2013.03.001

Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In S. Greenberg & W. Ainsworth (Eds.), Understanding speech: An auditory perspective (pp. 411–434). Mahweh, NJ: LEA.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11(12*), e1001752. doi.org/10.1371/journal.pbio.1001752

Hayes, A. F. (2019). Process for SPSS and SAS (Version 3.3)[Macro].

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of experimental child psychology*, *96*(2), 150-166.

Huss, M., Verney, J.P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, 47, 674-89. doi.org/10.1016/j.cortex.2010.07.010

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Psychology, 102* (3), 299-314. doi.org/10.1016/j.jecp.2008.07.005

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, *8*(2), 249-336.

Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One, 10 (12),* e0144411. doi.org/10.1371/journal.pone.0144411

Leong, V., Kalashnikova, M., Burnham, D. & Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind, 1 (2), 78-90*. doi.org/10.1162/OPMI_a_00008

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: a Norwegian latent variable longitudinal study. *Developmental psychology*, *45*(3), 764.

Lundetrae, K., & Thomson, J.M. (2018). Rhythm production at school entry as a predictor of poor reading and spelling at the end of first grade. *Reading & Writing, 31*, 215 – 237. doi.org/10.1007/s11145-017-9782-9

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348. doi.org/10.1037/0096-3445.135.3.348

Mehler J., Jusczyk P., Lambertz G., Halsted N., Bertoncini J., & Amiel-Tison C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *37*, 2767-2783. doi.org/10.1002/hbm.23206

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between early rhythm skills, musical training and phonological awareness. *Reading & Writing, 26*, 739 – 769. doi.org/10.1007/s11145-012-9389-0

Neda, Z., Ravasz, E., Brechet, Y., Vicsek, T., & Barabasi, A.L. (2000). The sound of many hands clapping: Tumultuous applause can transform itself into waves of synchronized clapping. *Nature, 403*, 849-850.

Overy K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. *Psychology of Music, 28*, 218-229. doi.org/10.1177/0305735600282010

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences, 999,* 497-505. doi.org/10.1196/annals.1284.060

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. *Dyslexia*, *9*(1), 18-36. doi.org/10.1002/dys.233

Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *Journal of Experimental Child Psychology*, *167*, 354-368. doi.org/10.1016/j.jecp.2017.11.009

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*, 777. doi.org/10.3389/fnhum.2013.00777

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain & Language, 160,* 1-10. doi.org/10.1016/j.bandl.2016.06.006

Provasi, J., & Bobin-Begue, A. (2003). Spontaneous motor tempo and rhythmical synchronization in 2 1/2- and 4-year-old children. *International Journal of Behavioural Development, 27*, 220 – 231. doi.org/10.1080/01650250244000290

Rautenberg, I. (2015). The effects of musical training on the decoding skills of German-speaking primary school children. *Journal of Research in Reading, 38* (1), 1-17. doi.org/10.1111/jrir.12010

Richards, S., & Goswami, U. (2019). Impaired recognition of metrical and syntactic boundaries in children with developmental language disorders. *Brain Sciences, 9,* 33. doi.org/10.3390/brainsci9020033

Rios-Lopez, P., Molinaro, N., & Lallier, M. (2019). Tapping to a beat in synchrony predicts brain print sensitivity in pre-readers. *Brain & Language, 199*, 104693. doi.org/10.1016/j.bandl.2019.104693

Roid, G. H. (2003). Stanford-Binet Intelligence Scales–Fifth Edition. *Itasca, IL: Riverside Publishing*.

Rose, J. (2009) Identifying and Teaching Young People with Dyslexia and Literacy Difficulties; An Independent Report From Sir Jim Rose to the Secretary of State for Children, Schools and Families. Nottingham, DCSF Publications.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728-43.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of educational psychology*, *96*(2), 265.

Slater, J., Strait, D.L., Skoe, E., O'Connell, S., Thompson, E., & Kraus, N. (2014). Longitudinal effects of group music instruction on literacy skills in low-income children. *PLoS One, 9* (11), e113383. doi.org/10.1371/journal.pone.0113383

Snowling, M. (2000). Dyslexia. Oxford: Blackwell Publishing.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(3), 369-373.

Stefanics, G., Fosker, T., Huss, M., Mead, N., Szűcs, D., & Goswami, U. (2011). Auditory sensory deficits in developmental dyslexia, a longitudinal ERP study. *Neuroimage,* 57 (3), 723-32.

Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS One, 10* (6), e0128839. doi.org/10.1371/journal.pone.0128839

Thomson, J., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris,* 102, 120-129. doi.org/10.1016/j.jphysparis.2008.03.007

Trainor, L.J., Clark, E.D., Huntley, A., & Adams, B.A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development, 20,* 383-396.

van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of dyslexia. *Journal of Speech, Language, and Hearing Research*, *60*, 937-949.

Waber, D.P., Weiler, M.D., Bellinger, D.C., Marcus, D.J., Forbes, P.W., Wypij, D., & Wolff, P.H. (2000). Diminished motor timing control in children referred for diagnosis of learning problems. *Developmental Neuropsychology, 17*(2): 181-197. doi.org/10.1207/S15326942DN1702_03

Wechsler, D. (2009) *Wechsler Individual Achievement Test—Third Edition*. *San Antonio*, *TX*: NCS *Pearson*.

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing, 15*, 179-206. doi.org/10.1023/A:1013880723925

Wolff, P. H., Michel, G. F., Ovrut, M., & Drake, C. (1990). Rate and timing precision of motor coordination in developmental dyslexia. *Developmental Psychology, 26*, 349-359.

Woodruff Carr, K., White-Schwoch, T., Tierney, A.T., Strait, D.L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in pre-schoolers. *Proceedings of the National Academy of Sciences, 111*, 14559-14564. doi.org/10.1073/pnas.1406219111

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3. doi.org/10.1037/0033-2909.131.1.3

Appendix A

Parental scores on the screening battery to determine children' AR and NAR group allocation

	AR parent with dyslexia	AR parent without dyslexia	NAR parents
TOWRE[1]- Word reading subtest	87.00 (7.36) [a,b]	108.05 (12.38)	108.37 (12.37)
TOWRE[1]- Non-word reading subtest	81.37 (13.41) [a,b]	102.25 (14.69) [c]	110.94 (8.95)
WIAT III[2] - Spelling subtest	83.58 (17.63) [a,b]	103.65 (8.20) [c]	111.81 (8.88)
WIAT III[2] - Oral reading fluency	85.00 (11.69) [a,b]	104.35 (8.95)	106.50 (8.88)
WIAT III[2] - Oral reading accuracy2	81.78 (19.93) [a,b]	99.05 (14.49) [c]	108.88 (11.29)
WIAT III[2] - Oral reading rate2	91.00 (27.37) [a,b]	104.05 (9.04)	105.37 (8.21)
Woodcock-Johnson[3] - RAN	90.63 (7.12) [a]	99.05 (13.63)	101.75 (13.45)
WAIS[4] - Digit span	8.63 (2.81) [a]	10.00 (2.59)	11.12 (2.56)
WAIS[4] - Non-verbal IQ subtests	11.53 (2.62)	12.05 (3.06)	12.56 (2.19)

[1]Test of Word Reading Efficiency (standardised score $M = 100$, $SD = 15$); [2]Wechsler Individual Achievement Test (standardised score $M = 100$, $SD = 15$); [3]Woodcock-Johnson test (standardised score $M = 100$, $SD = 15$); [4]Wechsler Adult Intelligence Scale (standardised score $M = 10$, $SD = 13$)

One-way ANOVA with post-hoc Tukey tests: [a]AR with dyslexia < NAR; [b]AR with dyslexia < AR without dyslexia, [c]AR without dyslexia < NAR

Appendix B

Items used in the non-word repetition task

	Practice		
Pemmie	Nape	Diff	Metton
	Test		
Nuck	Gick	Dinnick	Fean
Pame	Gattom	Katapet	Pennell
Hom	Baddep	Suppennack	Derappin
Sep	Hammett	Megatess	Sallan