

A Rhythmic Musical Intervention for Poor Readers: A Comparison of Efficacy With a Letter-Based Intervention

Adeetee Bhide¹, Alan Power¹, and Usha Goswami¹

ABSTRACT— There is growing evidence that children with reading difficulties show impaired auditory rhythm perception and impairments in musical beat perception tasks. Rhythmic musical interventions with poorer readers may thus improve rhythmic entrainment and consequently improve reading and phonological skills. Here we compare the effects of a musical intervention for poor readers with a software intervention of known efficacy based on rhyme training and phoneme-grapheme learning. The research question was whether the musical intervention would produce gains of comparable effect sizes to the phoneme-grapheme intervention for children who were falling behind in reading development. Broadly, the two interventions had similar benefits for literacy, with large effect sizes.

Even with explicit instruction, 5%–17.5% of schoolchildren do not become fluent readers (Shaywitz, 1998). Remedial educational packages which target the component skills of grapheme-phoneme and phoneme-grapheme learning combined with training in phonological awareness can be very effective (Bus & van Ijzendoorn, 1999). Even so, some poor learners do not benefit significantly from such interventions (Torgesen, 2000). A deeper understanding of the underlying causes of poor reading could enable the development of alternative remediation packages which benefit reading by training the cognitive and perceptual abilities that may underpin poor phonology and poor reading, rather than training reading per se. Here, we explore the utility of a

rhythmic musical intervention for poorer readers. The musical intervention was developed on the basis of temporal sampling theory, which proposes that an underlying difficulty in neural rhythmic entrainment found across the IQ spectrum is one cause of the poor phonological skills developed by children who go on to become poor readers (Goswami, 2011; Kuppen, Huss, Fosker, Fegan, & Goswami, 2011).

Temporal sampling theory was developed to explain why poor rhythmic entrainment (tapping to a beat), poor perception of acoustic rhythm, and poor perception of amplitude envelope rise time are all associated with developmental dyslexia. The perception of amplitude envelope rise time (hereafter rise time) is impaired in children with developmental dyslexia in many languages (see Goswami, 2011, for a summary), and individual differences in rise time perception are usually associated with individual differences in phonological skills in those same languages. The wholeband amplitude envelope of speech is a power-weighted summary of the amplitude modulations in different frequency bands in the speech stream. As these variations in intensity are made by the different articulators, the amplitude envelope can be viewed as a slowly-varying “modulator” that dynamically controls the energy level of its quickly-varying phonetic content. The envelope is dominated by the low frequency fluctuations that result from the production of syllables, and thus carries speech rhythm. Accurate perception of the amplitude envelope is thought to be critical for speech intelligibility (Zion Golumbic, Poeppel, & Schroeder, 2012) and for phonological development (Goswami, 2011).

Rise time perception is also related to the deliberate *rhythmic timing* of speech. When we speak rhythmically, we are timing the rise times of the vowels in different syllables, rather than timing the onset of the syllable itself. This is because we are timing the “perceptual center” of the syllable, the point in time when we experience a sound to occur, which is governed by

¹University of Cambridge

Address correspondence to Usha Goswami, Centre for Neuroscience in Education, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK; e-mail: ucgl0@cam.ac.uk.

rise time (the P-center, see Scott, 1998). To speak to a deliberate rhythm, we need to begin saying a syllable that starts with a longer onset, such as “street”, earlier than a syllable that has a shorter onset, such as “seat”. Longer onsets before the vowel (e.g., “skate”) move the P-center temporally to the left, whereas long codas (e.g., “banks”) move it to the right (Port, 2003). Musical notes also have different rise times, depending on how they are produced. For example, the note C produced by a slow bow movement on a violin string will have a longer rise time than if it is produced by blowing sharply into a trumpet. In music, rise time is called “attack time”, and musicians in an orchestra have to co-ordinate the attack times of their instruments to play together. In children, rise time perception is a longitudinal predictor of individual differences in the ability to hear metrical structure in music (patterns of beat perception, see Goswami, Huss, Mead, Fosker, & Verney, 2012), as well as a significant concurrent predictor of individual differences in the ability to hear metrical structure in language (strong and weak syllable stress, Goswami, Gerson, & Astruc, 2010).

As individual differences in rise time perception are associated with so many aspects of rhythmic performance, as well as with poor reading and phonology, it seems likely that musical training focused on rhythm may offer benefits to poor readers. In the current small-scale intervention study, we therefore explored whether training children’s rhythmic abilities via musical games and linking musical rhythms to rhythm in language would impact rise time sensitivity, reading, and phonology. The musical intervention was developed on the basis of temporal sampling theory, and aimed to train all the components of rhythm perception found to be impaired in developmental dyslexia in an engaging and fun manner.

There are already some studies using music training (combining rhythm and pitch) in the reading/phonology literature, and some of these studies have shown significant effects. For example, Degé and Schwarzer (2011) found that German preschoolers (4–5-year-olds) who received musical training showed improvement in phonological awareness at the large grain sizes of the syllable and rhyme that was equivalent to improvements made by a group of children who received direct training on phonological awareness. Both groups showed significantly more improvement than a control group who received a sports intervention. Overy (2003) found that boys with dyslexia, 8–9 years of age, showed significant improvement in both phonological skills and spelling after receiving musical training. More generally, music lessons have been shown to be superior to both painting and swimming lessons in terms of their effects on reading skills and phonological awareness, respectively (Moreno, Marques, Santos, & Santos, 2009; Rauscher & Hinton, 2011).

Nevertheless, studies without a training component have varied in whether an association between rhythmic awareness and reading/phonology has been found. For example, Anvari, Trainor, Woodside, and Levy (2002) gave 100

typically developing 4- and 5-year-old American children musical tasks based on piano tones which required either rhythm or pitch judgments (e.g., same/different rhythm discrimination, same/different melody discrimination), and examined associations with rhyme awareness and single word reading. At age 5, the musical tasks loaded on to two separate factors, a pitch perception factor and a rhythm perception factor. Although the rhythm perception factor was significantly associated with phonological awareness, it was not significantly related to single word reading. Anvari et al. (2002) concluded that the relationship between rhythm perception and reading was unclear. In contrast, a study of 1,028 French children aged 5–6 years showed a linear relationship between rhythm perception and attainment in reading (Dellatolas, Watier, Le Normand, Lubart, & Chevrie-Muller, 2009). The French children were asked to produce 21 rhythmic patterns modeled by the experimenter by tapping with a pencil on a table. The rhythm reproduction task showed a normal distribution, and individual differences in rhythmic performance were a significant predictor of reading at age 7–8 years (second grade), even after controlling for attention and linguistic skills. Studies comparing rhythmic and tonal awareness have found that rhythmic skills are better predictors of literacy skills (Douglas & Willatts, 1994; Strait, Hornickel, & Kraus, 2011). Therefore, on balance there is broad support for a link between musical rhythm perception, phonological processing, and progress in written language development.

The musical intervention designed for this study was focused on rhythm, but in addition incorporated the theoretically related factors of syllable stress and rise time discrimination (Goswami, 2011). The novel musical intervention was compared to a computer-assisted reading intervention designed to enhance reading and phonological skills, GraphoGame Rime. Its precursor, GraphoGame, was developed for the transparent Finnish language, where it has been shown to lead to significant gains in reading, with large effect sizes (Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011). GraphoGame Rime is an English version of this game based on rhyme analogy theory (Goswami, 1986). In a small-scale intervention study, GraphoGame Rime led to larger effect sizes for English poor readers aged 6–7 years compared to GraphoGame Phoneme, another English version of the game, which taught phoneme-grapheme correspondences (PGCs) via “synthetic phonics” (Kyle, Kujala, Richardson, Lyytinen, & Goswami, 2013). Kyle et al. also demonstrated that GraphoGame Rime led to significant improvements (with large effect sizes) in reading, spelling, non-word reading, rhyme and phoneme awareness in comparison to an untreated control group. Kyle et al. (2013) trained poor readers from similar schools and catchment areas to those 6–7-year-old poor readers trained here. As GraphoGame Rime is known to be an effective reading intervention for our target age group, we were interested to see whether a theoretically driven musical

Table 1
Baseline Measures by Intervention Group

	Musical group	GraphoGame group	<i>t</i> -Statistic	<i>p</i> -Value
N	10	9	—	
Age (years)	6.89 (0.28)	6.74 (0.36)	1.02	.32
WISC vocabulary subtest	7.30 (4.00)	6.11 (3.10)	0.72	.48
WISC block design subtest	8.20 (4.05)	8.33 (4.03)	0.07	.94
BPVS (standard score)	103.20 (16.70)	99.78 (7.46)	0.56	.58
TOWRE word reading (standard score)	95.50 (7.50)	94.33 (13.05)	0.24	.81
TOWRE non-word reading (standard score)	97.70 (7.93)	99.22 (10.96)	0.34	.73
BAS spelling (standard score)	98.70 (9.97)	98.11 (13.01)	0.11	.91
WAIS digit span (out of 30)	10.60 (4.74)	12.33 (3.20)	0.92	.37
Rhyme oddity (out of 20)	9.50 (4.86)	8.89 (4.37)	0.29	.78
Phoneme deletion (out of 20)	5.30 (3.53)	7.11 (4.65)	0.96	.35
Rapid naming speed (s)	56.80 (15.19)	58.11 (13.27)	0.20	.84
Rise time discrimination threshold (ms)	208.29 (71.47)	180.74 (87.39)	0.75	.46
Duration threshold (ms)	160.26 (23.67)	131.97 (51.05)	1.57	.13
Frequency threshold (semitones)	1.55 (0.52)	1.37 (0.47)	0.76	.46
Intensity threshold (dB)	9.74 (5.91)	5.89 (5.23)	1.49	.15

Note. Data are presented as mean (standard deviation). Both intervention groups were compared in a series of two-tailed, equal variance *t*-tests and the resulting *p*-values are shown. The groups do not significantly differ on any of the pre-test measures (all *ps* > .1).

intervention could achieve similar effect sizes for literacy and phonology outcome measures for the same age group.

Children were referred by their schools to the study if they were showing a failure to progress in reading at the expected level for their peer group, and were aged 6–7 years. As will be seen, the children nominated had relatively low language skills and relatively low cognitive ability. Half of the children were assigned to the musical intervention, and the other half were assigned to play GraphoGame Rime.

METHODS

Participants

Nineteen children aged 6–7 years participated in the study. All of them were identified by their class teachers as struggling readers. Ten children participated in the musical intervention (mean age = 6.9, *SD* = 0.3, 6 male) and nine children in the GraphoGame intervention (mean age = 6.7, *SD* = 0.4, 5 male). The children came from four schools, and participants from each school were divided evenly between the two intervention groups to minimize differences on pre-test measures. There were no group differences at pre-test on the outcome measures of interest, described below (all *ps* > .1) (see Table 1). Although all participants were nominated by their teachers as poor readers relative to their classmates, it should be noted that the British Ability Scales (BAS) used here to assess spelling were standardized before the National Literacy Strategy was introduced into the United Kingdom in 1997, while the Test of Word Reading Efficiency (TOWRE) used to measure word and non-word reading is standardized on a U.S. population. Typically-developing children in the United Kingdom now usually score well above the standard score (SS) of 100 on the

BAS (see also Kuppen et al., 2011; Kyle et al., 2013). The current participants were comparable to the poor readers studied by Kyle et al. (2013) for their BAS performance (average SS in Kyle et al.'s sample, 104.6, average SS here 98.1), and were worse in terms of their TOWRE performance (average TOWRE sight word SS in Kyle et al.'s sample, 103.9; average TOWRE SS score here, 94.3). Hence, in terms of current U.K. reading performance profiles, the nominated students were poor readers.

Intelligence and Vocabulary Tests

Children's IQs were estimated using the Vocabulary and Block Design subtests of the Wechsler Intelligence Scale for Children (WISC-III; Wechsler, 1991; given at pre-test only). Receptive vocabulary was measured using the British Picture Vocabulary Scale-II (BPVS-II; Dunn, Dunn, Whetton, & Burley, 1997; given at pre-test only).

Literacy Tests

Standardized tests of literacy were given at both pre-test and post-test. Reading was assessed using two subsets of the Test of Word Reading Efficiency, Form A: Sight Word Efficiency and Phonemic Decoding Efficiency (TOWRE SWE and PDE; Torgesen, Wagner, & Rashotte, 1999). Spelling was assessed using the spelling test from the British Ability Scale II (BAS II; Elliott, 1997).

Memory Test

The digit span from the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997) was used to assess memory at both pre- and post-test.

Phonology Tests

Phonological awareness was measured using experimental rhyme oddity and phoneme deletion tasks at both pre-test and post-test. For the rhyming task, the children listened to three words or non-words, two of which rhymed. The children had to select the item which did not rhyme (see Goswami et al., 2012). For the phoneme deletion task, the children listened to a non-word and were told to remove one phoneme from it to create a real word (e.g., “splo without the ‘p’ is slow”) (see Corriveau, Pasquini, & Goswami, 2007). For both tasks, scores out of 20 were used for analysis. In addition, a Rapid Automatized Naming (RAN) task was administered at both pre-test and post-test (see Kuppen et al., 2011).

Auditory Thresholds

Auditory processing of sound rise time, duration, frequency, and intensity was also matched across intervention groups. All of the auditory measures employed the “Dinosaur game” threshold estimation program which was originally created by Dorothy Bishop (University of Oxford) and adapted by Martina Huss (University of Cambridge; see Huss, Verney, Fosker, Mead, & Goswami, 2011). The amended program utilized an adaptive staircase procedure (Levitt, 1971). For the intensity, frequency, and duration tasks, the participants had to choose the stimulus which made a softer, higher, or longer sound, respectively. For the rise time task, children were asked to judge which sound began more softly. A lower auditory threshold indicates better performance. The rise time auditory threshold task was also an outcome measure. The task was completed twice at post-test and an average threshold was calculated.

Intervention Procedure

Participants were seen for 19 sessions of approximately 25 min in length delivered over a period of approximately 2 months. One participant in the GraphoGame group only completed 15 sessions, but given the small group sizes we included his data. All of the children in the musical intervention group were seen one-on-one. Children in the GraphoGame group were either seen one-on-one or in pairs. All interventions, as well as pre- and post-testing, were carried out by the same researcher, AB.

Musical Intervention

The musical intervention consisted of numerous tasks, of which children would do 4–5 in each session. The tasks are listed below (fuller detail is given in the appendix):

- 1 Tapping a space bar at the same time as a metronome. Five different rates were presented during training: 60, 80, 100, 120 and 140 bpm, or 1,000, 750, 600, 500, and 428.57 ms. Timings of taps were captured using Presentation software and were analyzed using circular statistics.

- 2 Performing a same-different judgment on two metronome tempos.
- 3 Performing a same-different judgment on two short rhythms.
- 4 Mimicking a short rhythm.
- 5 Rise time discrimination task (see auditory thresholds).
- 6 Clapping and marching to the beat of a song.
- 7 Learning to chant and play hand-clap games.
- 8 Listening to a poem and answering questions about its rhythm.
- 9 Playing the Dee-Dee game (see Goswami et al., 2010). The child sees a picture of a famous character/movie and the computer names the picture twice in “DeeDees”, once correctly and once incorrectly. Every syllable is replaced by the sound “dee”. The child needs to listen to the syllable stress pattern in order to decide which answer choice is correct.

GraphoGame Intervention

GraphoGame is a child-friendly computerized reading intervention originally developed for the orthographically-consistent Finnish language by scientists at the University of Jyväskylä (Saine et al., 2011). The GraphoGame Rime version of the game was developed for the orthographically inconsistent English language (Kyle et al., 2013). GraphoGame Rime teaches children PGCs via the psycholinguistic unit of the rime. The children hear sounds, rimes, and words spoken by the computer and have to match them to spellings or make decisions about whether words rhyme (see Kyle et al., 2013, for more detail about the words in the game and the progression levels). The game teaches PGCs via rhyme families and rhyme analogies, also showing children how rime units can be segmented into individual PGCs. The rhyme family format means that PGC information is always linked to oral rhyming patterns; therefore rhyme awareness is trained at the same time as phoneme awareness. Further detail about the game can be found in Kyle et al. (2013; all GraphoGame materials are available for research purposes from the GraphoWorld Network, <http://grapholearning.info/graphoworld>).

RESULTS

Effect of Intervention

To examine changes in auditory, phonological, and literacy skills as a function of intervention group, a series of paired *t*-tests were performed. Since the sample sizes were small, we first used the Shapiro–Wilk test to make sure that all assumptions for the *t*-test were met (i.e., the pre-test scores, post-test scores, and the difference between the scores were all normally distributed). If they were met, we carried out the paired *t*-test. If they were not met, we used a Wilcoxon signed rank test. Results are displayed in

Table 2
Pre- and Post-Test Measures by Intervention

Measure	Pre-test score	Post-test score	<i>t</i> -Statistic ^a or <i>z</i> -score ^b	<i>p</i> -Value	Effect size (<i>d</i> ^a or <i>r</i> ^b)
Musical intervention					
TOWRE word reading (raw score) (out of 104) ^a	21.00 (7.01)	25.00 (7.83)	2.30	.05*	0.73
TOWRE non-word reading (raw score) (out of 63) ^a	8.00 (3.40)	11.30 (3.71)	3.01	.01*	0.95
BAS spelling (ability score) (out of 200) ^a	57.70 (9.88)	63.20 (8.27)	2.86	.02*	0.90
Digit span (out of 30) ^a	10.60 (4.74)	12.10 (3.48)	2.24	.05*	0.71
Oddity rhyme (out of 20) ^a	9.50 (4.86)	12.80 (4.10)	3.19	.01*	1.01
Phoneme deletion (out of 20) ^a	5.30 (3.53)	8.10 (4.15)	2.47	.04*	0.78
RAN speed (s) ^a	56.80 (15.19)	51.70 (9.86)	1.39	.20	0.44
Rise time discrimination threshold (ms) ^b	208.29 (71.47)	128.69 (90.13)	-1.99	.05*	0.44
Overall reading measure ^a	-2.52 (4.80)	2.35 (4.09)	7.16	.01*	2.27
GraphoGame intervention					
TOWRE word reading (raw score) (out of 104) ^a	16.56 (9.13)	24.44 (11.48)	6.09	<.01*	2.03
TOWRE non-word reading (raw score) (out of 63) ^a	7.22 (3.15)	10.56 (3.84)	3.85	<.01*	1.28
BAS spelling (ability score) (out of 200) ^a	53.89 (12.17)	64.00 (13.16)	4.19	<.01*	1.40
Digit span (out of 30) ^b	12.33 (3.20)	12.89 (3.06)	-1.19	.24	0.28
Oddity rhyme (out of 20) ^a	8.89 (4.37)	11.22 (3.38)	1.66	.14	0.55
Phoneme deletion (out of 20) ^a	7.11 (4.65)	9.22 (4.60)	3.03	.02*	1.01
RAN speed (s) ^b	58.11 (13.27)	48.44 (12.75)	-2.43	.02*	0.57
Rise time discrimination threshold (ms) ^b	180.74 (87.39)	106.72 (91.34)	-1.72	.09	0.41
Overall reading measure ^a	-2.59 (5.79)	2.77 (4.94)	8.46	<.01*	2.82

Note. Pre- and post-test scores are displayed as mean (standard deviation). The overall reading measure is the sum of the *z*-scores of the eight measures (word reading, non-word reading, spelling, rhyme oddity, phoneme deletion, rapid naming speed, rise time discrimination, and digit span) known to be correlated with reading ability.

^aTest scores were normally distributed so the data were analyzed using a paired *t*-test. The *t*-statistic, the *p*-value, and the effect size (*d*) are shown.

^bTest scores were not normally distributed so the data were analyzed using a Wilcoxon signed rank test. The *z*-score, the *p*-value, and the effect size (*r*) are shown.

**p* ≤ .05.

Table 2. Inspection of the table shows that the gains from the musical intervention showed large effect sizes for non-word reading ($d = 0.95$), spelling ($d = 0.90$), rhyme awareness ($d = 1.01$) and phoneme deletion ($d = 0.78$). The gains from the GraphoGame intervention showed a very large effect size for word reading ($d = 2.03$), and large effect sizes for non-word reading ($d = 1.28$), spelling ($d = 1.40$) and phoneme deletion ($d = 1.01$). Given our small sample size, however, we were underpowered in terms of running a separate analysis of variance (ANOVA) for each measure.

Therefore, we decided to aggregate the *z*-scores of the eight measures (word reading, non-word reading, spelling, rhyme oddity, phoneme deletion, rapid naming speed, rise time discrimination, and digit span) into one measure, henceforth known as the overall reading measure, because they are known to be highly correlated. These measures were mostly highly correlated in our sample as well (see Table 3). The *z*-scores for each individual outcome measure were calculated by listing the scores for both intervention groups at both pre- and post-test and then converting them to *z*-scores. Please note that these *z*-scores are different than the *z*-scores calculated when performing the Wilcoxon-signed rank test. These *z*-scores were then resorted by intervention group and time point and the means for each intervention group at each time point were calculated. The mean *z*-scores of the different outcome measures were then added together to form the overall reading

aggregate scores. The overall reading aggregate scores are also shown in Table 2. As can be seen from the table, the aggregate scores were very similar for each group.

We ran an Intervention Group × Time (pre- or post-test) repeated measures ANOVA on the overall reading aggregate measure to see if the participants improved significantly over the course of the intervention and if there were any differences between the two intervention groups. There was a significant effect of Time, $F(1, 17) = 119.56$, $p < .01$, with a large effect size, partial $\eta^2 = 0.88$. Participants improved significantly between pre- and post-test. The effects of Intervention Group and the interaction between Interaction Group and Time were both non-significant, $p = .94$ and $p = .61$, respectively, suggesting that in both groups participants improved by comparable amounts.

Rhythmic Entrainment

We analyzed the tapping data using circular statistics to test whether or not the participants in the musical intervention were improving in their rhythmic accuracy through the course of the intervention. Circular statistics transform each inter-stimulus interval (ISI) into a unit circle, with the stimulus aligned at 0 radians. In this way, every response can be plotted along the circumference of the circle.

From these responses, the mean vector *R* can be calculated. It has two non-parametric components, the mean direction

Table 3
Correlations Between Measures in the Overall Reading Score

	RAN speed	Oddity rhyme	Phoneme deletion	Rise time discrimination threshold	TOWRE word reading (raw score)	TOWRE non-word reading (raw score)	Digit span
Oddity rhyme	.014						
	.953						
Phoneme deletion	-.245	.531*					
	.312	.019					
Rise time discrimination threshold	.104	-.301	-.467*				
	.673	.210	.044				
TOWRE word reading (raw score)	-.535*	.403	.217	-.275			
	.018	.087	.373	.254			
TOWRE non-word reading (raw score)	-.254	.407	-.076	-.192	.324		
	.294	.084	.758	.432	.176		
Digit span	-.269	.617**	.590**	-.461*	.314	.203	
	.266	.005	.008	.047	.191	.404	
BAS spelling (ability score)	-.470*	.494*	.369	-.213	.805**	.459*	.532*
	.042	.031	.120	.382	.000	.048	.019

Note. The Pearson correlation (top value) and significance (bottom value) of the measures aggregated to form the Overall Reading Measure. The pre-test scores were used to calculate the correlations. Every measure is significantly correlated with 1-5 of the other measures.

* $p \leq .05$; ** $p \leq .01$.

θ , which is analogous to the mean asynchrony, and the mean resultant length \bar{R} . \bar{R} always varies between 0 and 1 and is inversely related to variance in asynchronies. An \bar{R} of 1 implies that responses always occur at the same time relative to the stimulus (perfect synchrony). We assumed for our data that an \bar{R} of 0 meant that the responses were evenly distributed around the circle (low synchronization accuracy), though there can be other interpretations (see Kirschner & Tomasello, 2009).

For each participant, we calculated his/her \bar{R} and θ during each rhythmic entrainment session. In our calculation, we included all responses that fell after the fifth stimulus, to allow some time for entrainment, and before where the 36th stimulus would have occurred. As each participant experienced each rate (60, 80, 100, 120, and 140 bpm) twice throughout the intervention, we could measure improvement in rhythmic entrainment throughout the course of the intervention (see Figure 1 for example data from one participant).

An interesting pattern emerged regarding improvement; participants who had the most variable rhythmic entrainment initially were the most likely to show temporal improvement throughout the intervention. We ran a correlation between the mean \bar{R} of the first attempts at each rate (a measure of initial aptitude) and the mean improvement (mean difference in \bar{R} between the second and first attempt at each rate). This correlation was significant, $r = -0.90$, $p < .01$, suggesting that participants who showed the poorest synchronization at the beginning of the intervention showed the greatest temporal improvement (see Figure 2). Similarly, for θ , we ran a circular correlation between the mean absolute value of θ of the first attempts at each rate (a measure of initial aptitude) and the

mean difference in absolute value of θ between the second and first attempt at each rate (a measure of improvement). The circular correlation trended towards significance, $r = -0.68$, $p = .09$, suggesting that participants who had the greatest asynchronies at the beginning of the intervention showed the greatest improvement. This suggests that, educationally, the training is affecting the appropriate target behavior.

An extension of temporal sampling theory is that there should be a correlation between improvement in rhythmic entrainment and improvement in reading. We ran a correlation between the mean difference in \bar{R} between the second and first attempt at each rate (a measure of improvement in rhythmic entrainment) and the difference between the post-test and pre-test overall reading measure (a measure of improvement in reading). The two measures were strongly correlated, $r = 0.57$, and this correlation was significant using a one-tailed test, $p = .04$ (see Figure 3). Hence, we found modest support for the hypothesis that improvement in rhythmic entrainment should be correlated with improvement in reading.

We were also curious to see whether our participants showed the same general trends found in previous tapping studies using populations with different demographics. Previous tapping literature has shown that nonmusicians tapping in synchrony with a metronome tend to tap before the stimulus (there is a negative mean asynchrony) (see Repp, 2005 for review). When we averaged the θ values for each participant across all ten trials, eight of the ten participants had negative mean asynchronies. Hence, our participants did follow the same trends noticed previously. They were entraining to the rhythm and attempting to predict when the stimulus would

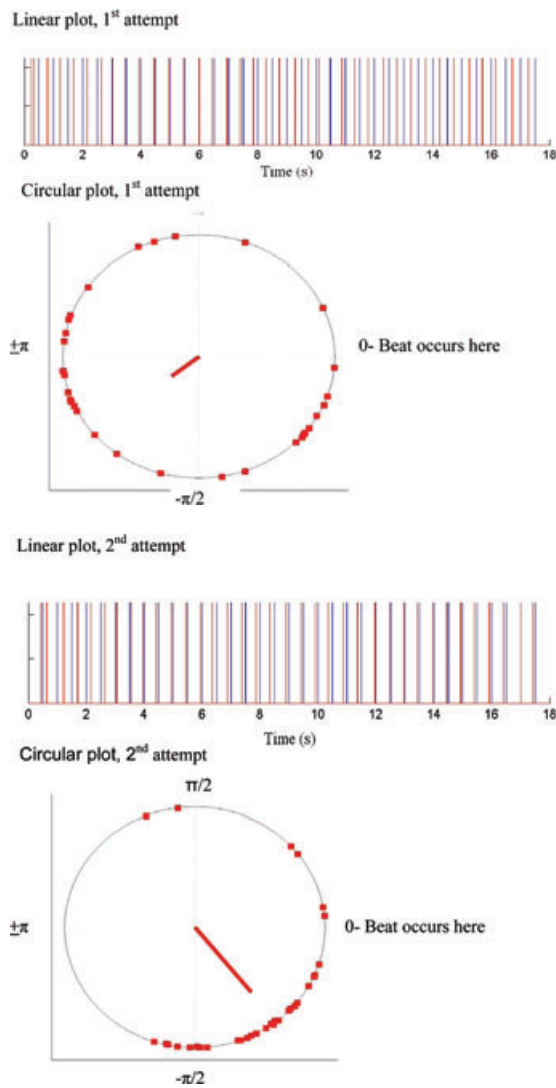


Fig. 1. Example figures from the circular statistics calculations at 120 bpm rate. For the linear plots, the blue lines represent the stimuli and the red lines represent the responses. For the circular plots, 0 is pointing east and the circle reads counter-clockwise. The red squares around the circumference represent each response during the trial. The vector R is depicted with the red line within the circle. Note how in the linear plot of the first attempt, the participant sometimes responded with the stimulus and was sometimes completely out of phase with the stimulus. By the second attempt, the participant has significantly improved and his/her responses were almost always near the stimulus. This pattern is captured by the circular analyses. In the first attempt, the responses are widely distributed around the circle, thus the vector has a small magnitude. Furthermore, the vector is very far from 0 radians, the location of the stimulus, and close to $\pm\pi$, which is completely out of phase with the stimulus. In contrast, the responses during the second attempt are tightly clustered, thus the vector has a larger magnitude. Furthermore, the vector is much closer to 0 radians. The circular plot of the second attempt also shows that the participant tends to anticipate the beat, which matches the findings of other tapping studies.

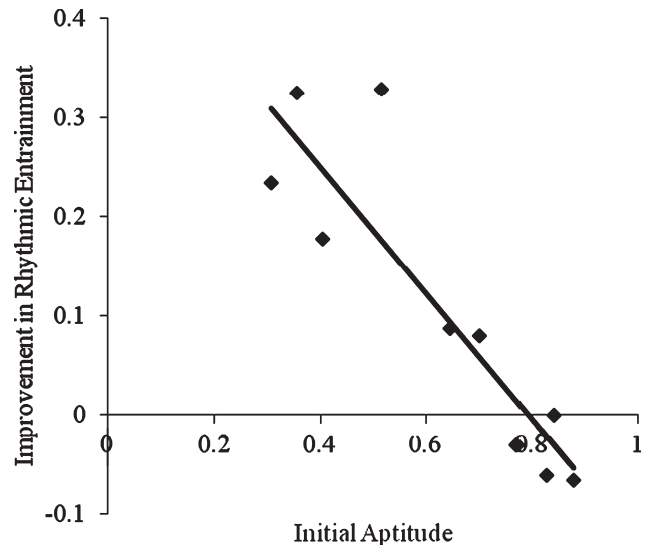


Fig. 2. Correlation between initial aptitude in rhythmic entrainment and mean improvement in rhythmic entrainment. Initial aptitude is defined as the mean \bar{R} during the first attempt at each rate. Improvement in rhythmic entrainment is defined as the mean difference in \bar{R} between the second and first attempt at each rate. The correlation was strong, $r = 0.90$, $p < .01$.

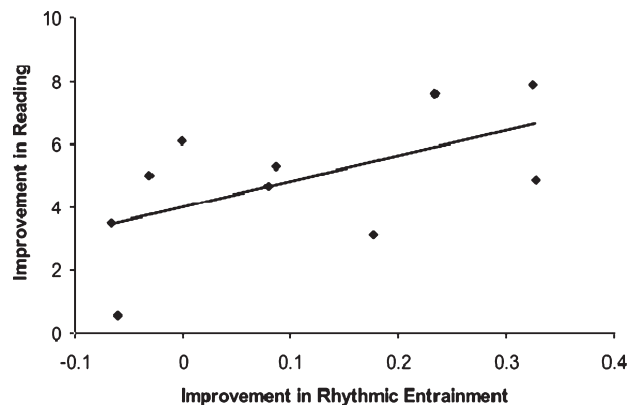


Fig. 3. Correlation between improvement in rhythmic entrainment and improvement in reading. Improvement in rhythmic entrainment is defined as the mean difference in \bar{R} between the second and first attempt at each rate. Improvement in reading is defined as the difference between the post-test and pre-test overall reading measure. The correlation was strong, $r = 0.57$, and significant on a one-tailed test, $p = .04$.

occur, rather than simply responding to the stimulus. This negative mean asynchrony can be clearly seen in Figure 1 (circular plot, second attempt).

DISCUSSION

The research question underpinning this study was whether training the perceptual and cognitive skills that may underpin

poor phonological development and poor reading via a musical intervention would show benefits for reading and phonological development, and whether these benefits would be equivalent to those that result from the *direct* training of reading and phonology. Broadly, the data suggested that the musical training was as effective as the direct intervention (GraphoGame Rime). This is theoretically interesting, as sub-lexical phonological awareness and letter-sound correspondences were not being taught in the musical intervention. The effects found suggest that giving children rhythmic training, and linking non-linguistic rhythms to rhythms in language, has a positive effect on literacy acquisition and on phonological skills.

The main limitations of this study are the small sample size and the fact that we were not able to include a control group which did not receive any training. This was because participating schools requested that all their poor readers receive some kind of intervention. When both intervention groups show comparable benefits, it cannot be determined if this benefit is due to the interventions per se or to other factors. However, previous research with 6–7-year-old children that did use an unseen control group (i.e., research utilizing children who were the same age as the children in the current study, and who were attending similar schools and experiencing similar reading curricula) did find significant benefits from the GraphoGame Rime intervention used here (see Kyle et al., 2013). Kyle et al. reported gains from GraphoGame Rime in comparison to their unseen controls that showed medium to large effect sizes for some of the same outcome measures used here (word reading, spelling, non-word reading, rhyme and phoneme awareness). For the GraphoGame group in this study, these measures also showed gains with medium to large effect sizes. Given that the previous intervention study by Kyle et al. (2013) used an unseen control group of the same age and reading profile as the children trained here, we can conclude that GraphoGame Rime is an effective intervention. Therefore, since the musical intervention used in the current study produced gains comparable to those produced by GraphoGame Rime in this study, and if we assume that GraphoGame is an effective intervention, then we can tentatively conclude that both interventions benefit literacy to an equal degree, and that these benefits are due to the interventions themselves. Nevertheless, we cannot rule out potential Hawthorne effects in the current study. All of our children were receiving personalized interventions that we expected to yield benefits for reading. We were not able to include a control group who also received personalized intervention for art or sport, which would not be expected to yield benefits for reading (Degé & Schwarzer, 2011).

The musical intervention was based on temporal sampling theory. Therefore, the focus of the intervention was to match musical rhythms with language rhythms, and to improve children's rhythmic entrainment. This intervention

study provided modest support for the underlying theory. Firstly, the musical intervention led to literacy gains of comparable effect sizes to direct training in letter-sound correspondences. Secondly, we found that improvement in rhythmic entrainment over the course of the intervention was strongly correlated with an increase in the overall reading score between pre- and post-test.

CONCLUSION

This small-scale intervention study suggests that a theoretically-driven musical intervention based on rhythm and on linking metrical structure in music and language can have benefits for the development of literacy and phonological awareness. Further studies including an unseen control group are needed to explore whether these gains are greater than those that occur with the natural passage of time, and to see whether such benefits would accrue to all children who receive musical training, or only to children with lower language skills and lower cognitive ability (those trained here). Furthermore, larger sample sizes are needed so that each individual outcome measure can be analyzed separately. It would also be interesting to study whether a combination of the two approaches would be most beneficial, and whether (for example) the musical intervention should precede the letter-based training. Since the musical intervention trains the rhythmic perception and rhythmic entrainment skills that, by hypothesis, are important for the development of phonological awareness (Goswami, 2011), it seems reasonable that the musical intervention should be delivered first. Then, once those prerequisite auditory abilities have been strengthened, the GraphoGame intervention could be delivered, so that children can learn PGCs. These research questions are important for maximizing the educational benefits of musical training to progress in literacy.

Acknowledgments—We would like to thank the schools for their participation and the Churchill Scholarship for funding the research. We would also like to thank Victoria Leong for help writing the Presentation scripts, John Verney and Mark Haggard for useful discussion, Natasha Mead and Lisa Barnes for assistance with standardized tests, and Nichola Daily for help with equipment and logistics.

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, 111–30.
- Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.

- Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and Specific Language Impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research, 50*, 647–666.
- Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology, 2*, 1–7.
- 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 of Clinical Neuropsychology, 6*, 1–9.
- Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading, 17*(2), 99–107.
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale* (2nd ed.). Windsor, England: NFER-Nelson.
- Elliott, C. D. (1997). *British abilities scale II*. Windsor, England: NFER-Nelson.
- Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42*, 73–83.
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*, 3–10.
- Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing, 23*, 995–1019.
- Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2012). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex, 1–14*. doi:10.1016/j.cortex.2012.05.005
- 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*(6), 674–689.
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology, 102*, 299–314.
- Kuppen, S., Huss, M., Fosker, T., Fegan, N., & Goswami, U. (2011). Basic auditory processing skills and phonological awareness in low-IQ readers and typically developing controls. *Developmental Psychology, 15*, 211–243.
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Comparing the effectiveness of two theoretically motivated computer-assisted reading interventions in the UK: GG Rime and GG Phoneme. *Reading Research Quarterly, 48*, 61–76.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.
- Moreno, S., Marques, C., Santos, A., & Santos, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19*, 712–723.
- Overy, K. (2003). Dyslexia and music. *Annals of the New York Academy of Sciences, 29*, 1–9.
- Port, R. F. (2003). Meter and speech. *Journal of Phonetics, 31*, 599–611.
- Rauscher, F., & Hinton, S. (2011). Music instruction and its diverse extra-musical benefits. *Music Perception: An Interdisciplinary Journal, 29*, 215–226.
- Repp, B. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review, 12*, 969–992.
- Saine, N. L., Lerkkanen, M.-K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Development, 82*, 1013–1028.
- Scott, S. K. (1998). The point of P-centres. *Developmental Psychology, 61*, 4–11.
- Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine, 338*, 307–312.
- Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions, 7*(44), 1–11.
- Torgesen, J. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*(1), 55–64.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.
- Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed. manual). New York, NY: The Psychological Corporation.
- Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed. manual). San Antonio, TX: The Psychological Corp.
- Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*, 151–161.

APPENDIX

THE MUSICAL INTERVENTION TASKS

Rhythmic Entrainment: Tapping Along to a Metronome

The stimuli were delivered over headphones and the children were instructed to tap the space bar at the same time as the stimulus. The stimuli were programmed using Audacity software. Presentation software was used to capture the timings of children's taps. The tempos used were 60, 80, 100, 120, and 140 bpm. Despite the tempo, the participant heard 35 taps. Therefore, faster tempos had shorter durations. Each tap rate was administered twice, once during the first 10 sessions and again during the subsequent 9 sessions. This allowed us to see if the children were becoming more accurate with their rhythmic entrainment. The tapping data was analyzed using circular statistics. This task was administered once every other session for a total of 10 sessions.

Differentiating Between Two Tempos

Stimuli were delivered using TempoPerfect Metronome Software. The children heard two beats and they had to determine whether or not they were the same speed.

If the two tempos were different, the difference was much greater at the beginning of the intervention than at the end to make the task more difficult as the intervention progressed. If the child answered incorrectly, the researcher played the two beats for him/her again and encouraged him/her to either look

at the metronome (where graphics displayed a ball bouncing on the beat) or clap along to the beat so that he/she could see or feel the correct answer. This task was administered 9 times throughout the intervention.

Differentiating Between Two Rhythmic Sequences

The children heard two short rhythmic sequences, each containing four beats and separated by four beats of rest. There was no tonal information in the rhythmic sequences; all notes were a high G in the treble clef. The children had to determine whether the two rhythmic sequences were the same or different. The rhythmic sequences were composed by the first author and played using ForteFree Software. The rhythmic sequences became more complex, and the differences less salient, as the intervention progressed. If the child was incorrect, the researcher encouraged him/her to look at the sheet music presented on the computer while listening to the rhythmic sequences so that he/she could visually see the rhythms. The software highlights the notes on the screen as it plays through the rhythmic sequences, making the rhythmic differences more visually salient. This task was administered 10 times throughout the intervention.

Singing a Rhythmic Sequence

Participants heard a short rhythmic sequence which ranged from 4 to 8 beats long. The rhythmic sequence had no tonal information; all of the notes were the high G in the treble clef. The participants were encouraged to mimic the rhythm using the syllable “la”. If required, the researcher assisted them in learning and practicing the rhythmic sequence. All of the rhythmic sequences were composed by the first author and played using ForteFree software. The rhythmic sequences were designed to become more difficult as the intervention progressed. This task was administered 9 times throughout the intervention.

Rise Time Discrimination Task

See auditory thresholds for a description of this task. The children did the task without the practice trials. This task was administered 7 times throughout the intervention.

Clapping and Marching to a Song

The children would hear the same song in two consecutive sessions. During the first session, they would be asked to clap along to the beat. During the second, they would be asked to march along to the beat. The researcher would demonstrate if they were struggling. The songs chosen were popular, so as to be enjoyable to the children, but with age-appropriate lyrics. The songs had a variety of tempos and salient rhythms. The clapping and marching tasks were administered 9 and 10 times throughout the intervention, respectively.

Chanting and Playing Hand Clap Games

The children would work on the same hand clap game during two consecutive sessions. During the first session, the children would learn to chant a portion of the song. During the second session, the children would learn the hand motions that accompanied it. The chants were approximately ordered by level of difficulty.

The first author composed pieces using ForteFree Software which mimicked the rhythm of the chant, but had no tonal information (all of the notes were the high G in the treble clef). The researcher would first chant the entire song to the child. Then, the researcher would teach the child the first couple of lines. The researcher read out loud all of the lyrics and encouraged the child to memorize them, so that the child did not have to be able to read to perform this task. The child was especially encouraged to make sure that all of his/her words lined up with the notes in the music. The child was then asked to chant the first couple of lines along with the music. During the second session, the researcher taught the children the hand motions that accompanied the chant that they had learnt. The chanting and clapping tasks were administered 7 and 6 times throughout the intervention, respectively.

Poetry

The researcher read a poem to the child and then asked the child three questions about it. The poems were chosen to be highly rhythmic and child-friendly. The first question regarded syllable stress. A recorded voice of the researcher would say one word, which was in the poem, with both the correct and incorrect syllabic stress patterns. The child had to determine the correct way of saying the word. If the child got the answer wrong, the researcher explained the correct answer to him/her. The second question regarded the number of syllables in a word. The child was asked whether or not replacing a word in the poem with another word of similar meaning would “mess up” the rhythm of the poem. The words were chosen so that if they had the same number of syllables, then they also had the same syllabic stress pattern. Therefore, if the words had the same number of syllables the rhythm would not be messed up, and vice versa. A recorded voice of the researcher would read a portion of the poem, first with the original word and then with its replacement. The child had to determine whether the rhythm in the second reading was “messed up”. If the child answered incorrectly, the researcher encouraged him/her to clap out the number of syllables in the two words, so that he/she could hear whether or not the words had the same number of syllables. The third question regarded the overall rhythm of the poem. A recording of the researcher presented a portion of the poem twice, once with the correct rhythm and once with a distorted rhythm. The child had to choose the reading that “fit the poem better”. If the child got it wrong, the researcher helped him/her understand why the

other option was correct. This task was administered 6 times throughout the intervention.

Dee-Dee Game

The Dee-Dee game was used to enhance the participant's understanding of linguistic prosody. In this game, the child sees a picture of a popular fictional character or movie. The child has to name the picture so that the researcher can ascertain whether or not the child knows the reference. If the child does not know the reference, the researcher will provide the name. The computer then says the name of the picture twice, once correctly and once incorrectly. However, all distinctive phonological information is removed, as every syllable is replaced by the sound "dee". The child needs

to listen to the syllabic stress pattern in order to decide which answer choice is correct (see Goswami et al., 2010). The children played this game four times throughout the intervention. Before the children played for the first time, they looked at all of the pictures and tried to name them. If they did not know any, the researcher provided the name. This was done to minimize the number of times the researcher needed to assist the children during the game. Due to very poor performance seen in all participants during the first game, the researcher helped the children during the second and third games. During the second game, the researcher repeated the name of the picture while emphasizing the prosody. During the third game, the researcher clapped the syllables for the children. The fourth game the children played unaided.