International Journal of Multidisciplinary Comparative Studies
Volume 2 No. 1, 2015, pp. 57—72

Can reading skills which are developed through the reading of music be transferred to benefit the early decoding of text?

MARK BETTENEY
University of Greenwich, United Kingdom

GREG BROOKS
University of Sheffield, United Kingdom

ABSTRACT This study reports on a randomised control trial which examined whether a 6-week intervention of music reading through recorder playing would have an effect on phonic decoding skills in children (n=50) aged 5-6. The study was conducted by recruiting matched randomised intervention and control groups from two Year 1 classes in a Primary school in North West Kent, England. Pre- and post-tests measured the recognition fluency of single-letter graphemes; clusters and digraphs; and nonsense words. The children in the intervention group showed greater gains in their decoding of clusters and digraphs, and of nonsense words, and overall, than did those in the control group. Although the overall result was not statistically significant (t = 1.061; df = 48; one-tailed p = 0.147), there was a modest positive effect size of d = 0.29. Trends in the results suggested a hypothesis (which would, however, need further testing for stronger support) that the synchronous learning of simple formal music notation can have a beneficial effect on the development of phonic decoding skills.

Keywords: Phonics; decoding; fluency; music/literacy connection; RCT; notation

The aims of the research

This article reports on one focus of a piece of independent research (Betteney, 2012), which was led by a hypothesis that there is a possible correlation, or even a causal connection, between children's ability to read music and their ability to decode text during the early stages of reading. This hypothesis was selected because the literature in this field is somewhat sparse, a view confirmed by Dehaene (2009), who, when writing about the effects of various interventions on reading scores, observed, of musical notation training, that "*although well-designed studies on this particular topic remain rare, early musical training does seem to have a positive impact on reading scores*" (p.242).

Tallal and Gaab (2006) also emphasised the lack of published work in this area. Urging caution in making claims about the influence of musical training on language and literacy skills correlation or causality, they noted that:

Additional research aimed at understanding better the potential role that auditory and/or musical training might have in improving language and reading skills is also needed … Music and speech represent the most cognitively complex uses of acoustic information by humans and utilizing one to improve the other seems to be an auspicious and promising approach. (Tallal and Gaab, 2006: 388 & 389).

Context

Within this field of music/literacy connection, the approaches that researchers have taken have tended to fall into two camps. The first is a neurological approach. For example, Bishop-Liebler et al. (2014) demonstrated that "*impairments in basic auditory processing show particular links with phonological impairments*" and suggested that "*musical training might have a beneficial effect on the … basic auditory processing skills which are found to be deficient in individuals with dyslexia*" (abstract). Thomson and Goswami (2008) explored the way rhythm and auditory processes impact upon reading experiences of people with dyslexia. Similarly, Tierney and Kraus (2013) identify five sub-skills in reading, which they link "*… through a unifying biological framework, positing that* [the links] *share a reliance on auditory neural synchrony*" (p209). These studies, and others like them, have taken an explanatory approach, identifying patterns in how the brain functions when processing information, often from a desire to understand a deficit in reading capacity, or to suggest solutions or alleviations to difficulties that learners might experience, either though inherited neurological conditions, or perhaps after a trauma. Welch et al. (2012) observe that "*studies which have attempted to enhance literacy skills through experiences with music have generally worked with children with recognised literacy difficulties, such as dyslexia, rather than with normally developing children*" (Welch et al., 2012: 13).

The second approach (for example, Hansen et al, 2014) identifies parallels between the structures of music and literacy (such as reading and writing strategies, or symbol recognition), and between the pedagogies of teaching the reading and writing of text, and of music. Within this approach the focus is invariably an identification of the effect that a music intervention can have on the development of literacy skills, and never the other way round. Researchers using this second approach tend to measure the impact of an intervention on pupil attainment and progress. For example, the Act, Play, Sing project (Education Endowment Foundation, 2014) sought to evaluate whether music workshops had a greater impact on numeracy and literacy attainment than did drama workshops. This Act, Play, Sing study is a recent example of attainment effect studies in the field by different researchers that date back to 2000.

For example, in 2000, Butzlaff undertook a meta-analysis of a number of studies on attainment effect, and concluded that "*there is indeed a strong and reliable association between the study of music and performance on standardized reading/verbal tests*" (Butzlaff, 2000: 172), and in the last 15 years this focus on attainment-driven synergy between music and literacy has since been built upon by several researchers (Hansen et al., 2014; Welch et al, 2012; Hallam 2010; Gromko, 2005; Hansen and Bernstorf, 2002; Copple et al., 2000). All of these writers refer to auditory aspects of phonemic awareness. For example Gromko (2005) considered that the "*ability to recognize that a spoken word consists of individual sounds or phonemes may be the mechanism that explains the relationship of music instruction to reading skill*" (Gromko, 2005: 203). There have been more recent research studies in this field (Rautenberg, 2015; Tierney and Kraus, 2013; Runfola et al., 2012; Bugaj and Brenner, 2011; Long, 2007; Gromko, 2004) but there remains a great deal of scope for further study.

Rautenberg (2015) reported on research conducted in Germany, looking at, amongst other things, "*the effects of musical training on word-level reading abilities*" (abstract). Her study, like the one reported here, used matched groups with randomised sampling, a music intervention, and controlled testing which produced numeric data. She focused on three areas, the first of which is directly comparable to this study:

• Accuracy—words articulated correctly, both concerning single phonemes and prosody/word stress, were counted.

• Speed—average reading time per word per child was measured.

• Prosody—when words were articulated correctly regarding phoneme—grapheme correspondences but word stress/prosody and thus rhythm was incorrectly realised, these readings were classified as 'sound synthesising strategy readings'(Rautenberg, 2015: 8).

Rautenberg's study concluded that "*musical training had a significant effect on reading accuracy in word reading*" (Rautenberg, 2015: abstract).

Long's (2007) study was similar in length to that reported here (six weeks), but smaller in size ($n = 24$). It also had a different focus, concentrating as it did on children's development of comprehension skills at the age of 8-10, but reported similar effects. She found that "*very brief training (10 minutes each week for 6 weeks) in stamping, clapping and chanting in time to a piece of music while following simple musical notation had a considerable impact on reading comprehension in children experiencing difficulties in reading"* (Long, 2007, in Hallam, 2010: 10). Long (2007) was also able to determine that the positive effects of the intervention were sustained six months after the intervention had finished, "*particularly amongst children with below average reading capability*" (p.113).

Gromko's (2005) study influenced the present one markedly. Her study was larger in terms of participants ($n = 103$) and longer in duration (4 months) but used unmatched groups, comparing the effect of the intervention on children in two separate kindergarten settings. Her aim was to "*determine whether music instruction was related to significant gains in the development of young children's phonemic awareness ... An analysis of the data revealed that kindergarten children who received ... music instruction showed significantly greater gains in development of their phoneme segmentation fluency when compared to children, who did not receive music instruction*" (Gromko, 2005: 203, abstract). Her data showed that in the context of her research study, music instruction had a greater influence on phoneme-segmentation fluency than on letter-naming fluency, or nonsense-word fluency. The data in the present study also revealed a stronger influence on phoneme-segmentation fluency than in other categories.

Method

Choice of age group

The research was conducted with Year 1 children in order to focus upon participants who were in specific formative stages of their reading development. The time of year was therefore important. Had the opportunity for research fallen at the end of the academic year, a Reception class would have been chosen. The Early Years and Key Stage 1 documentation in place at the time of the research was clear about what was expected of children at the start and end of each academic year. The Early Years Foundation Stage (DCSF, 2008) expected children aged between 40- and 60+ months (that is, by the end of the Foundation Stage, and the beginning of Key Stage 1) to be able to engage with the following aspects of "*linking sounds and letters*" (pp.52-54), which are presented in a developmental sequence:

• Continue a rhyming string.

• Hear and say the initial sound in words and know which letters represent some of the sounds.

• Hear and say sounds in words in the order in which they occur.

• Link sounds to letters, naming and sounding the letters of the alphabet.

• Use their phonic knowledge to write simple regular words and make phonetically plausible attempts at more complex words.

(DCSF, 2008: 53. The use of bold font is from the document, and by ‘phonetically’, ‘phonemically’ must be meant, as in DfES, 2007, below.)

These developmental skills tied in closely with the focus of the research. By contrast, the expectation for children by the end of Year 1 (for whom phase six of Letters and Sounds (DfES, 2007) was deemed appropriate) was that they should be able to do the following:

- Children should know most of the common grapheme–phoneme correspondences (GPCs).
- They should be able to read hundreds of words, doing this in three ways:
  - reading the words automatically if they are very familiar;
- decoding them quickly and silently because their sounding and blending routine is now well established;
  - decoding them aloud.
- Children’s spelling should be phonemically accurate, although it may still be a little unconventional at times. (DfES, 2007: 168)

These learning and teaching aspirations were clearly in advance of the focus of the study (formative stages of reading), and so, had the research project been conducted in the summer term, a sample drawn from Year 1 would clearly have been inappropriate.

Design

In a matched-pairs randomised control trial (RCT), a group of 5- to 6-year-old children received a six-week intervention in music reading and playing, while a control group did not, in order to test the hypothesis that it would benefit the development of the intervention group’s phonic decoding skills. This RCT mirrored the focus, but not the design, of Gromko (2005), whose work provided the original impetus for this study.

For ethical reasons (not depriving any of the children of the experience), the control group were given the intervention after the completion of the RCT, but no data from that phase of the study were analysed.

Sample size and randomisation

There was a constraint on the sample size, because the school which agreed to host the study operated only two-form entry, with 30 children in each class. Increasing the statistical power of the experiment would have required not just one, but several, more schools, in order to adopt a cluster RCT design; logistical, financial and time constraints prevented this. It was acknowledged from the outset that, with a maximum sample of 60 participants, the probability of a statistically significant result was low.

The 60 children were arranged in pairs from lists provided by the host school’s teachers who, from information taken from end-of-Foundation Stage assessments, had allocated each of the children to one of four reading groups (lower; lower middle; upper middle; and higher). These groups each helpfully held an even number of children, and so the children were simply paired in the order in which they appeared on the lists. Thus each child was paired with another from the same group. For each pair a die was thrown; if it came down odd, the first-named child was placed in the intervention group, and the second in the control group. If the die came down even, the opposite allocation was made.

The parents/carers of all the children were asked to give permission for their child to participate. Five permission slips were not returned, and so the data of the five relevant children, together with those of their allocated pairs, were not used. The sample size was therefore $n = 50$.

Intervention

Following pre-testing, the intervention group children received 30-minute music lessons on six consecutive Tuesday afternoons. The total 'contact' teaching time for the intervention was therefore 6 x 30 minutes = 3 hours. However, the children were practising, and talking about their experiences to their peers and parents/carers, throughout the six weeks of the programme. The pre- and post-tests and the music lessons were all carried out by a team of five research assistants. They were all student teachers who received course credit, but no remuneration, for this work.

The lessons were divided into three sections. The first engaged the children in the reading of two sets of written stimuli, beginning with combinations of crotchets, quavers and rests, which the research assistants wrote large on a white board, and the rhythms of which the children clapped. Following this activity, the children clapped and then chanted the rhythm of the piece of music to be learnt during that lesson, although it was only once they were familiar with the rhythm that they were told that this was the rhythm of the piece (the title of the first four pieces reflected the rhythm).

In the second section the children played on recorders the piece they had learnt and practised from the week before, and then returned to the new piece, with more clapping and singing to remind them of what they had learnt in section 1.

Thirdly, the children were placed in groups of 3 or 4 (each research assistant now taking one small group), and the children were taught to play the new piece, firstly by ensuring that each child understood the rhythm (more clapping), then by interpreting the notes on the stave and learning the fingering. The instructors were rotated across the groups over the course of the experiment.

Over the six weeks the pieces that were taught developed more in terms of rhythmic and structural complexity than melodic. The children were taught combinations of three note types (crotchet, quaver, rest) and four notes (B, A, G, and, right at the end of the intervention, D). Generally, the note selection was intentionally repetitive within each phrase, whereas the rhythms of the phrases, although often identical one to another, used a variety of note types (i.e. it was common to include all three note types, crotchet, quaver, and rest, in a single phrase, but rare to use all three notes, B, A and G).

Figure 1: Piece 4

Figure 1 gives an example of the pieces that the children were taught, showing the level of musical demand that the programme attained. The example was the fourth piece of six (Appendix 1 contains all six pieces). The pieces were composed on a weekly basis, building on progress made during the previous week, and this progression gave children planned incremental opportunities to interpret and internalise repetitive symbols, and kinaesthetic opportunities to demonstrate their understanding of those symbols, much in the same way as *Letters and Sounds* (DfES, 2007), and *Jolly Phonics* (2011) employ repetition of a narrow selection of graphemes in the early stages of reading acquisition. In this way there was a clear and identified hierarchy of musical reading demands, which became progressively more complex in direct response to the children's engagement both with the scores and with the written clapping exercises in which the assistants engaged the children at the beginning of each lesson.

Assessment instruments

At pre- and post-tests the children in both groups were assessed on recognition of single-letter graphemes, clusters and digraphs, and nonsense words (appendix 2). These categories were selected as they are the foci of the phonics schemes *Letters and Sounds* (DfES, 2007), and *Jolly Phonics* (2011), both of which were used in the school, and both of which develop the recognition of grapheme/phoneme correspondence, the blending and segmenting of clusters and digraphs, and the whole-word recognition of "high-frequency" (Letters and Sounds, p.193) and two- or three-syllable words (p.149). Participating Year 1 children would therefore be experienced and comfortable being asked to give a response to such stimuli. The maximum possible score on each test was 80, comprised of the 26 letters of the alphabet, 9 consonant clusters/ digraphs and 45 nonwords. The letters of the alphabet, clusters and digraphs were, by necessity, identical at the two stages, By contrast, the pre- and post-tests of nonsense words were different; however, the structure of the words in each test was identical (for example, in one section, pre-test words 'hig', 'ost', 'shib', 'fluz', and 'hibe' were replaced at post-test with 'hos', 'ist', 'shog', 'flam', and 'kibe', appendix 2).

The use of nonsense or pseudo-words as a device for measuring decoding skills is not new (for example Oney-Kusefoglu and Durgunoglu's (1997) *Phonological Awareness and Pseudoword Recognition Test*; Peynircioglu et al.'s (2002) examination of the correlation between musical aptitudes, phonological awareness, and pseudo-word recognition abilities). However, the use of nonwords has been a contentious area in recent years (Reedy et al., 2011, Coldwell et al., 2011), given the nationwide introduction in England in 2012 of a mandatory phonics screening check for six-year-olds. This national screening employs nonwords as half of the test items (the other half being real words), but has received a mixed reception from teachers, since the results are wrongly employed as an indication of proficiency in reading rather than (as in this study) exclusively as a test of a range of decoding skills.

Some children in this study refused to read the nonsense words at all, because they were 'not real words', and this refusal was respected (although it is interesting that the children must have meaningfully engaged with each pseudo-word in some manner in order to be able to make the judgement to refuse to read it). Interestingly, this response to nonsense words parallels findings with regard to pseudo-words contained in Caldwell et al.'s (2011) evaluation of the 2011 pilot of the Phonics Screening Check for 6-year-olds. The evaluators reported that:

Most pupils felt that the use of pseudowords … was a 'fun', novel aspect. However the majority (60%) of pilot schools said that pseudowords caused confusion for some pupils, while 12% said they caused confusion for most pupils.
(Coldwell et al., 2011: 2)

In this study the format and content of the tests were carefully explained to the children prior to their experiencing them. However, the research assistants did observe that all the children noticed that the nonsense words were not real words (some children liked this, thinking the words to be funny), but a small proportion (4%) refused to read them as a result. In this circumstance, all non-attempted items were attributed a score of zero. This was an appropriate attribution because there was no pattern to be found in scoring or failing to score on any one item for those students who continued beyond the point at which those who refused were allowed to withdraw, and so an additional score post-withdrawal was unlikely. Data from all 50 children for all sections of the testing were used in the statistical analysis of the RCT.

The testing procedures and conditions were the same for every child on every occasion. All tests were conducted individually, and the testing of each child took approximately ten minutes. This amount of time was deliberately set because the five research assistants had 60 children to test in each assessment session. Therefore, on average, each research assistant had twelve children to test, and they had two hours in which to complete the task.

Procedure

The research assistants were given clear instructions about which responses were acceptable as correct, and the ground rules for the tests were established. These were that:

- No child should be made to take the test if s/he did not want to (although this eventuality did not occur);
- No child would be coerced to go further in the test than s/he wanted to (s/he could withdraw at any point, and this did occur);
- The test should be terminated at the point at which it became clear that a child was unlikely to score any further.

Where single-letter graphemes were concerned, it was agreed before the testing began that a point would be awarded if a child produced either the letter name or the phoneme, but that they should be encouraged to articulate the phoneme. This was done because, although the vast majority of the children articulated the phoneme of a single-letter grapheme as a default approach, not all did, and some articulated the phoneme incorrectly, even though they recognised the grapheme. If such an error arose, it was because the phoneme (/m/ for example) was articulated as /mə/ ('muh'), not /əm/ ('mmm'). We considered it unfair to penalise the obvious recognition of a grapheme because of inconsistency of pronunciation, and so a child received a point for the articulation of /mə/ or /əm/. Without exception, if a child recognised the letter name, s/he could also offer a corresponding phoneme, so there was no need to have a protocol about the awarding of a point if the letter name was the child's only response.

The clusters and digraphs were less problematic with regard to pronunciation, and there was great consistency of articulation from those children who recognised each. However, some children articulated <th> as /f/, in line with prevalent local diction. We had anticipated this, and asked the assistants to ask children who recognised <th> but articulated it as /f/ to name a word which started with <th>. If the children said (for example) /fruː/ ('froo', frequent local pronunciation of 'through'), a mark was awarded. If they said /fɒg/ ('fog'), it was not. When asking for confirmation, the testers were instructed to point to the digraph and say "Can you think of a word which starts with this sound", and not to say "Can you think of a word which starts with /θ/?", or "…which starts with '*tee-aitch*'?"

Results

Table 1 (page 64) shows the pre-test scores for the two groups, for the three tests and overall; all differences were non-significant, thus demonstrating satisfactory matching of the groups.

		Mean scores (s.d's in brackets)			
	n	Overall	Single-letter graphemes	Clusters & digraphs	Nonsense words
Maximum		80	26	9	45
Interven-tion group	25	33.96 (13.47)	23.36 (1.75)	3.16 (2.44)	7.12 (10.07)
Control group	25	33.64 (14.63)	22.48 (3.65)	3.36 (2.99)	8.00 (11.09)
t-value		0.080	-0.114	-0.259	-0.294
p-value		0.936	0.886	0.797	0.770

Table 1: Pre-test scores

Table 2 shows the pre- and post-test and gain scores, and the effect sizes, for the three tests and overall. The intervention group made a greater gain on clusters & digraphs, on nonsense words, and overall, but the control group made a greater gain on single-letter graphemes. An explanation for the last result can be found in the data of two children, both in the control group, who each scored only eleven marks in the pre-test, all of which were gained exclusively from recognition of single-letter graphemes. However, the single-letter grapheme recognition scores of these two children improved greatly during the period of the intervention (post-test scores of 16 and 19 respectively), a level of improvement which was mathematically impossible for the other children, because their knowledge in this area at pre-test was very close to the maximum.

Analyses of variance showed that the effect of time was highly significant for all three tests and overall (overall: ($F(1,48) = 43.144$, $p<0.001$; single-letter graphemes: $F(1,48) = 20.797$, $p < 0.001$; clusters & digraphs: $F(1,48) = 20.794$, $p < 0.001$; nonsense words: $F(1,48) = 25.495$, $p < 0.001$), as would be expected. However, the anovas also showed that the factor of interest, the time x group interaction, was non-significant in all four cases (overall: $F(1,48) = 3.113$, $p = 0.084$; single-letter graphemes: $F(1,48) = 2.465$, $p = 0.123$; clusters & digraphs: $F(1,48) = 3.407$, $p = 0.071$; nonsense words: $F(1,48) = 4.600$, $p = 0.037$. The latter results were confirmed by t-tests of the differences in post-test means, all of which were also non-significant (overall: $t = 1.061$; $df = 48$; one-tailed $p = 0.147$; single-letter graphemes: $t = 0.074$; $df = 48$; one-tailed $p = 0.471$; clusters & digraphs: $t = 1.435$; $df = 48$; one-tailed $p = 0.079$; nonsense words: $t = 1.055$; $df = 48$; one-tailed $p = 0.149$).

By Cohen et al.'s (2007: 521) categorisation, the overall effect size, and those for clusters & digraphs and nonsense words, are "*modest*", though that for clusters & digraphs falls at the very upper end of Cohen's categorisation. The effect size of $d = -0.45$ for single-letter grapheme recognition again demonstrates that the control group made greater progress than the intervention group. Given the oddity of that result, further analysis was carried out on the combined scores for clusters & digraphs and nonsense words (i.e. on the overall scores minus those for single-letter graphemes). The relevant data are shown in Table 3.

Again, by Cohen et al.'s (2007: 521) categorisation, this effect size is "*modest*". An anova showed the effect of time was highly significant ($F(1,48) = 33.159$, $p < 0.001$). On this occasion the interaction of time and group was statistically significant ($F(1,48) = 4.368$, $p = 0.042$), albeit narrowly. The difference between the mean post-test scores was non-significant ($t = 1.089$; $df = 48$; one-tailed $p = 0.181$).

			Pre-test		Post-test		Gain	Effect size (Cohen's d)
Test	Maxi-mum	Group	mean	(s.d.)	mean	(s.d.)	mean	
Overall	80	Interven-tion	33.96	(13.47)	45.20	(16.59)	11.24	0.29
		Control	33.64	(14.63)	40.12	(17.26)	6.48	
Single-letter graphemes	26	Interven-tion	23.36	(1.75)	24.16	(1.46)	0.80	-0.45
		Control	22.48	(3.65)	24.12	(2.26)	1.64	
Clusters & digraphs	9	Interven-tion	3.16	(2.44)	5.52	(2.74)	2.36	0.49
		Control	3.36	(2.99)	4.36	(2.97)	1.00	
Nonsense words	45	Interven-tion	7.12	(10.07)	15.64	(14.43)	8.52	0.37
		Control	8.00	(11.09)	11.44	(13.73)	3.44	

n=25 for both groups for all tests

Table 2: Pre- and post-test and gain scores, and effect sizes

Group	*n*	Pre-test		Post-test		Gain	Effect size
		mean	(s.d.)	mean	(s.d.)		
Intervention	25	10.60	(12.56)	21.04	(16.31)	10.44	*d* = 0.35
Control	25	11.12	(12.99)	16.00	(16.43)	4.88	

Table 3: Pre- and post-test and gain scores, and effect size, for clusters & digraphs and nonsense words combined

	Pre-test	Post-test
Children unable to recognise any nonsense word (Intervention) / Control	16 (9) / 7	9 (3) / 6
Children recognising 1-9 nonsense words	22 (9) / 13	19 (8) / 11
Children recognising 10-19 nonsense words	6 (5) / 1	5 (4) / 1
Children recognising 20-29 nonsense words	3 (1) / 2	8 (4) / 4
Children recognising 30 or more nonsense words	3 (1) / 2	9 (6) / 3
Total	50 (25) / 25	50 (25) / 25

Table 4: Further analysis of nonsense word recognition scores

In these calculations, specific data (single-letter grapheme recognition) have been excluded, and this could lead to an accusation of manipulating the data; however, there is precedent for this approach. Wiliam (2008), keen to maintain as much reliability within quantitative studies as possible, suggested that:

> … items that all students answer correctly, or ones that all students answer incorrectly, are generally omitted, since they do not discriminate between students, and thus do not contribute to reliability. In this circumstance it is quite standard practice to omit such data from the analysis … [In so doing] the reliability of the test is increased.
> (Wiliam, 2008: 255)

Further analysis of the nonsense-word recognition data revealed some interesting additional patterns, presented in Table 4.

The most noticeable difference between the two groups was the pre/post reduction in the number of children unable to recognise a single nonsense word. Of the seven children who moved up from a score of zero in the pre-test, six were from the intervention group. Numbers are not very different in the recognition of 1-9 and 10-19 words, but again there were noticeable differences in the increases between 20-29 and above 30. Of the six children who moved from a score of less than 30 at pre-test to 30 or more at post-test, five were from the intervention group. It was true that the participant numbers were small in some categories, but in all situations where there was a noticeable difference between the groups, the data show that more children in the intervention group were able to progress into a higher category band than in the control group.

It is worthy of comment that, in all categories other than single-letter recognition, positive effect sizes in favour of the intervention group were found, albeit of modest size (by Cohen et al.'s, 2007 classification). Wiliam (2008: 254) would describe all the positive effect sizes in this study as "*substantively useful*", and he would consider that the categories showing an effect size >0.3 suggest that the progress which might be expected in these areas over a year has been condensed into a six-week period.

Conclusions

This study raised as many questions as it answered. Were further research to be pursued in this field (and it is a field with a great deal of scope for further research) one might explore whether the effect size would have been as great, or greater, if children had been taught only rhythmic notation, rather than the melodic as well. The children thoroughly enjoyed the physicality of clapping or beating rhythms, and for the majority of the children the physical understanding of how to play the recorder was evident. The connection had been made between the contours of the notation and the mechanics of the instrument, but the physical ability to do it was not within every child's compass. This research project was an enterprise to try to demonstrate the possibly of a transfer between two codes of symbol recognition, not whether the playing of an instrument enhances children's ability to decode. In that respect, and in hindsight, two interventions were applied, the clapping of rhythms and the playing of an instrument, where one intervention might have been more effective.

Approaches very different from those used here might also have been considered. The potential of connections between music and literacy for children with dyslexia could be explored. Alternatively, studies have been made of reading development which take a neurological approach. For example, Thompson and Goswami (2008) gave an overview of studies involving brain imaging which "*offers a new technology for understanding the acquisition of reading by children*" (p.67). Welch et al. (2012), in their identification of relationships between music and literacy, considered neurobiological ways in which the brain processes music, including the brain's plasticity when accommodating new experiences. Similarly, Besson and Friederici (2005) note that:

From a neuroscience perspective, the most challenging findings from brain imaging studies may be that common networks are activated in tasks that were first thought to involve specialized brain areas and mechanisms. The neural networks for language and music processing do show a large overlap.

(p.57)

It might also be interesting, for example, to compare the eye movements of experienced readers of text and music when engaging with previously unseen texts and pieces of music.

This was a small study, but the findings point in the hypothesised direction. Validation would require a larger study. Meanwhile, the results are at least compatible with Gromko (2005), and resonate with Rautenberg (2015). Allowing children opportunities to interpret a variety of incremental symbol combinations in this way seems to have helped them to decode, and this RCT in music reading serves to illustrate that point. It is not possible to claim to have proved a connection, but it is possible to claim to have contributed to the body of evidence by illustrating the effect such an intervention can have.

Correspondence
Mark Betteney
Faculty of Education & Health
University of Greenwich, UK
Email: m.w.betteney@greenwich.ac.uk

References

Betteney, M. (2012) Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text? A randomised controlled trial. (EdD thesis, University of Sheffield, England).

Bishop-Liebler, P., Welch, G., Huss, M., Thomson, J. M. and Goswami, U. (2014) Auditory Temporal Processing Skill in Musicians with Dyslexia, Dyslexia, 20, 3, pp. 261-279.

Bugaj, K. and Brenner, B. (2011) The Effects of Music Instruction on Cognitive Development and Reading Skills - An Overview, Bulletin of the Council for Research in Music Education, 189, pp. 89-104.

Butzlaff, R. (2000) Can Music Be Used to Teach Reading? Journal of Aesthetic Education, 34, 3/4, pp. 167-178.

Cohen, L., Manion, L., and Morrison, K. (2007). Research Methods in Education (6th Edition), London and New York, RoutledgeFalmer.

Coldwell, M., Shipton, L., Stevens, A., Stiell, B., Willis B., and Wolstenholme C. (2011). Process evaluation of the Year 1 Phonics Screening Check Pilot, Sheffield, Centre for Education and Inclusion Research.

Copple, C., Bredekamp, S., and Neuman, S. (2000) Learning to read and write: Developmentally appropriate practices for young children, Washington DC, National Association for the Education of Young Children.

Davies, H. and Crombie, I. (2009) What are confidence intervals and p-values? (2nd Edition), Dundee, Scotland, Hayward Medical Communications.

DCSF (2008) Practice Guidance for the Early Years Foundation Stage, London, HMSO.

Dehaene, S. (2009) Reading in the Brain, New York, Viking Penguin.

DfE (2010) The Importance of Teaching - The Schools White Paper, London, HMSO.

DfES (2007) Letters and Sounds: Principles and Practice of High Quality Phonics – six-phase teaching programme, London, DfES Publications.

Easton, J.E. and McColl, J.H. (2012) Statistics Glossary vol 1.1, STEPS.

Education Endowment Foundation (2015) Creative Futures: Act, Sing, Play, Evaluation Report and Executive Summary, Retrieved from https://educationendowmentfoundation.org.uk/uploads/pdf/Act__Sing__Play.pdf

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., and Rosen, S. (2002) Amplitude envelope onsets and developmental dyslexia: A new hypothesis, Proceedings of the National Academy of Sciences of the United States of America, 99, 16, pp. 10911-10916.

Gromko, J. (2004) Predictors of Music Sight-Reading Ability in High School Wind Players, Journal of Research in Music Education, 52, 1, pp. 6-15.

Gromko, J. (2005) The Effect of Music Instruction on Phonemic Awareness in Beginning Readers, Journal of Research in Music Education, 53, 3, pp. 199-209.

Hallam, S. (2010) The power of music: its impact on the intellectual, social and personal development of children and young people, International Journal of Music, 28, 3, pp. 269-289.

Hansen, D. and Bernstorf, E. (2002). Linking music learning to reading instruction, Music Educators Journal, 88, 5, pp. 17-25.

Hansen, D., Bernstorf, E, and Stuber, G. M. (2007) The Music and Literacy Connection, Lanham, Maryland, Rowman and Littlefield.

Hansen, D., Bernstorf, E, and Stuber, G. M. (2014) The Music and Literacy Connection (2nd edition), Lanham, Maryland, Rowman and Littlefield.

Jolly Phonics (2011) Parent/Teacher guide. Jolly Phonics Ltd, Retrieved from http://www.jollylearning.co.uk/Parent%3ATeacher%20Guide.pdf

Lamb, S., Gregory, J., and Andrew, H. (1993) The relationship between music and reading in beginning readers, Educational Psychology, 13, 1, pp. 19-27.

Oney-Kusefoglu, B. and Durgunoglu, A.Y. (1997) Beginning to read in Turkish: A phonologically transparent orthography, Applied Psycholinguistics, 18, 1, pp. 1-15.

Peynircioglu, Z.F., Durgunoglu, A.Y. and Oney-Kusefoglu, B. (2002) Phonological awareness and musical aptitude, Journal of Research in Reading, 25, 1, pp. 68-80.

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, pp. 1-17.

Reedy, D., Parkin, P., Coe, J., and Tutt, R. (2011) Open letter to Mr Gove, The Independent, 7th July 2011.

Runfola, M., Etopio, E., Hamlen, K. and Rozendal, M. (2012) Effect of music instruction on preschoolers' music achievement and emergent literacy achievement, Bulletin of the Council for Research in Music Education, 192, pp. 7-27.

Tallal, P. and Gaab, N. (2006) Dynamic auditory processing, musical experience and language development, Trends in Neuroscience, 29, 7, pp. 382-390.

Thomson, J.M. and Goswami, U. (2008) Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling, Journal of Physiology – Paris, 102, 1-3, pp. 120-129.

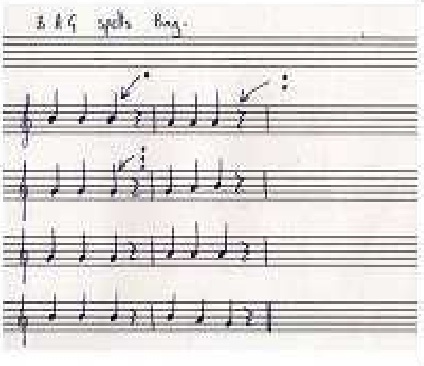
Tierney, A. and Kraus, N, (2013) Music Training for the Development of Reading Skills.
In: M.M. Merzenich, M. Nahum, and T.M. van Vleet (eds.) Progress in Brain Research, 207, pp. 209-241.

Welch, G., Saunders J., Hobsbaum, A., and Himonides, E. (2012) Literacy through Music: A Research Evaluation of the New London Orchestra's Literacy through Music Programme, London, International Music Research Centre.

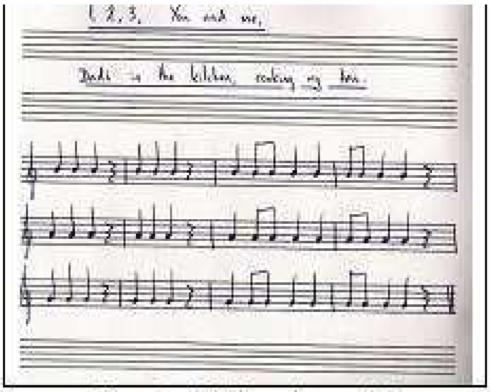
Wiliam, D. (2008) International comparisons and sensitivity to instruction, Assessment in Education: Principles, Policy and Practice, 15, 3, pp. 253–257.

Appendix 1: The taught pieces

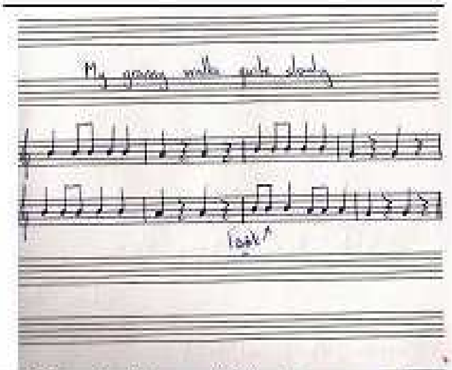
Piece 1: BAG spells bag



Piece 2: 1,2,3, You and me.

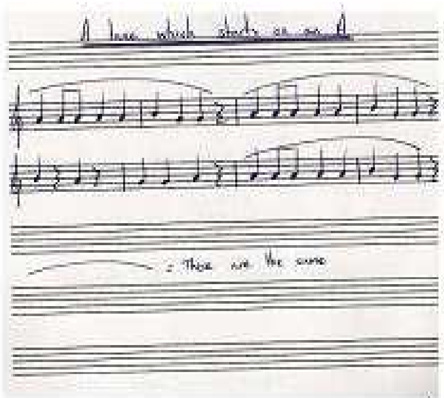


Piece 3: My granny walks quite slowly

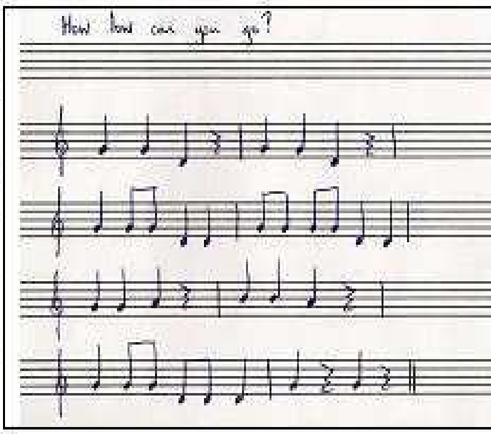


Piece 4: 1,2, How do you do?

Piece 5: A tune which starts on an A



Piece 6: How low can you go?



Appendix 2: Test materials

Single-letter graphemes (presented in this randomly-selected order for all children):
c, g, o, r, t, e, k, n, u, s, d, I, j, y, a, x, z, f, l, q, m, p, b, v, h, w

Clusters and digraphs (items were presented in the order shown):

Clusters	Digraphs
bl, dr, sp, str	ch, sh, th, ph, ing (the latter preceded by the most frequent vowel letter before word-final <ng> to support identification)

Nonsense words

Pre test	description	Post test
hig, reb, tov	c-v-c	hos, riv, taz
ost, arn, ent	v-c-c	ist, orn, ent
shib, thun, quop	Digraph-v-c	shog, thid, quat
fluz, grad, scug	Cluster-v-c	flam, grun, scir
hibe, vome, yake	Split digraph	kibe, wome, vake
nugfim, fetzum, jumdop	c-v-c- x2	nagfot, fitzud, jimdap
fiss, gudd, wobb	Double final consonants	liss, fudd, tobb
leab, waig, joam	Medial vowel digraphs	weab, baig, moam
tremp, drint, sculp	Clusters, beginning and end	slemp, brint, frulp
chish, shoth, phash	Digraphs, beginning and end	thish, photh, chash
remar, disfug, prejox	c-v-c with prefix	demar, risfug, pregox
zining, yogful, hitate	c-v-c with suffix	zinate, yoging, hitful
luttle, gimming, vesser	Double medial consonants	muttle, timming, nesser
ginkly, fornter, duntness	Two syllables, separate medial consonants	minkly, pornter, tuntness
buntingham, slantering, preflopation	Multi-syllable	lartingham, blintering, trendigation