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

A randomised controlled trial.

Mark William Betteney

Registration Number 070220799

Submitted in part fulfillment of the award of Doctor of Education (Ed.D.)

Department of Education

University of Sheffield

March 2012

For Claire, without whom, none of this would have happened.

Also, for Edward and Arthur, who have been constantly encouraging and proud.

# Acknowledgements

I am grateful to the following people, all of whom helped enthusiastically.

Leigh Bailey

John Baker

Dianne Barrett

Jacquie Conetta

Jane Chandler

Marcia Gough

Kelly Green

Prof. Peter Hannon

Emma Horan

Jenny McMann

Nickie Michie

Grace Moore

Prof. Cathy Nutbrown

Dr. Sammy Rashid

Kat Sankey

James Simmonds

Prof. Graham Welch

Patrick Wilson

A huge debt of gratitude is owed to Prof. Greg Brooks, whose tutoring was supportive, generous, and demanding – a perfect combination. Thank you.

# Abstract

*This study reports on an RCT which examined whether a 6-week intervention of music reading through recorder playing had an effect on phonic decoding skills in children (n=50) aged 5-6. The study was conducted during the Autumn term of 2010, by recruiting matched randomised intervention and control groups from two combined parallel 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 & digraphs; and nonsense words, and analysis was made of these categories when combined, and separately.*

*Analysis of the data revealed that the children who received six weeks of music instruction showed significantly greater gains in their development of decoding clusters and digraphs, and of nonsense words, than did those in the control group. Pre-test data showed that the children’s recognition of single-letter graphemes was at an advanced stage, and a negative effect was revealed in that category as a result of the randomised allocation of matched pairs. The data showed a modest overall effect size of d = 0.29. However, the result was not statistically significant (t = 1.061; df = 48; one-tailed p = 0.147). A systematic audit of the RCT was undertaken.*

*Children were also interviewed about their experiences of learning to read music. Analyses of these interviews compared and contrasted the children’s awareness of themselves as readers, and as readers of music, and compared their experiences of learning the two genres. The interviews demonstrated that the majority of children had little difficulty in correctly interpreting crotchets, quavers and rests through the clapping of series and combinations of these three symbols, but the recognition of the names of notes (G, A, B) was more problematic. Interviews with research assistants, together with analyses of the incremental difficulty of the pieces of music that the children were taught, confirmed the presence of developing skills in music reading.*

*The concept of near-transfer was explored, and the results supported an hypothesis that the synchronous learning of simple formal music notation can have a beneficial effect on the development of phonic decoding skills.*

# Contents

[Acknowledgements 2](#_Toc332731429)

[Abstract 3](#_Toc332731430)

[Contents 4](#_Toc332731431)

[List of Tables and Figures 10](#_Toc332731432)

[Chapter 1, Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text? 13](#_Toc332731433)

[1.1 Introduction 13](#_Toc332731434)

[1.2 Current national priorities 14](#_Toc332731435)

[1.3 The Schools White Paper (2010) and the Henley Review (2011) 17](#_Toc332731436)

[1.4 The teaching of the early reading and writing of music 21](#_Toc332731437)

[1.5 The focus of this study 24](#_Toc332731438)

[1.6 The notion of transfer 25](#_Toc332731439)

[1.7 The research question 26](#_Toc332731440)

[Chapter 2 Literature Review 28](#_Toc332731441)

[2.1 Introduction 28](#_Toc332731442)

[2.2 The literature, 1960-1990 30](#_Toc332731443)

[2.3 The literature, post-1990 31](#_Toc332731444)

[2.4 Lamb et al. (1993) 32](#_Toc332731445)

[2.5 Butzlaff (2000) 34](#_Toc332731446)

[2.6 Orthography and transparency 39](#_Toc332731447)

[2.7 The vocabulary of the field of phonics 42](#_Toc332731448)

[2.8 Orthographic depth and syllable structure 44](#_Toc332731449)

[2.9 Attitudes to the isolation of decoding in children’s reading 45](#_Toc332731450)

[2.10 What is it, to be a reader? 49](#_Toc332731451)

[2.11 The notion of transfer 50](#_Toc332731452)

[2.12 Philosophical positions on the concept of transfer 56](#_Toc332731453)

[2.13 Taxonomies of transfer 57](#_Toc332731454)

[2.14 Reading skills that the children brought with them into the study 63](#_Toc332731455)

[2.15 Children’s articulation of their experiences of learning to read 67](#_Toc332731456)

[2.16 The assessment of music and of literacy 67](#_Toc332731457)

[2.16.1 Auditory skills (i): Auditory association 69](#_Toc332731458)

[2.16.2 Auditory skills (ii): Auditory blending 69](#_Toc332731459)

[2.16.3 Auditory skills (iii): Auditory discrimination 70](#_Toc332731460)

[2.16.4 Auditory skills (iv): Auditory memory 71](#_Toc332731461)

[2.16.5 Visual skills (i): Visual memory 71](#_Toc332731462)

[2.16.6 Visual skills (ii): Visual analysis 72](#_Toc332731463)

[2.16.7 Visual skills (iii): Visual blending 74](#_Toc332731464)

[2.16.8 Visual skills (iv): Special functions 76](#_Toc332731465)

[2.16.9 Visual skills (v): Visual associations 76](#_Toc332731466)

[2.17 Physiological links between music and other forms of learning 78](#_Toc332731467)

[2.18 Relevant literature with regard to dyslexia 80](#_Toc332731468)

[2.19 An analysis of three studies: Long (2007), Standley and Hughes (1997) and Gromko (2005) 84](#_Toc332731469)

[2.19.1 Long (2007) 86](#_Toc332731470)

[2.19.2 Standley and Hughes (1997) 88](#_Toc332731471)

[2.19.3 Gromko (2005) 92](#_Toc332731472)

[Chapter 3 Methods and Methodology 98](#_Toc332731473)

[3.1 Introduction 98](#_Toc332731474)

[3.2 Design 100](#_Toc332731475)

[3.3 Ethics 103](#_Toc332731476)

[3.4 The selection of school 105](#_Toc332731477)

[3.5 Selection of age range 107](#_Toc332731478)

[3.6 Enlisting participants 108](#_Toc332731479)

[3.7 Pairing and randomisation methods 109](#_Toc332731480)

[3.8 Absentees at pre-test 111](#_Toc332731481)

[3.9 The phases of the study 114](#_Toc332731482)

[Phases 1 and 2. The Randomised Controlled Trial, and follow-up 118](#_Toc332731483)

[3.10 Power calculation 118](#_Toc332731484)

[3.11 The size of the research programme 121](#_Toc332731485)

[3.12 Research assistants 123](#_Toc332731486)

[3.13 The format and content of the music teaching 126](#_Toc332731487)

[3.14 Analysis of the pieces 129](#_Toc332731488)

[3.14.1 Piece 1: B A G spells Bag 130](#_Toc332731489)

[3.14.2 Piece 2: One, two, three, You and me, (Dad’s in the kitchen cooking my tea) 130](#_Toc332731490)

[3.14.3 Piece 3: My granny walks quite slowly 131](#_Toc332731491)

[3.14.4 Piece 4: One, two, How do you do 132](#_Toc332731492)

[3.14.5 Piece 5: A tune which starts on an A 132](#_Toc332731493)

[3.14.6 Piece 6: How low can you go? 133](#_Toc332731494)

[3.15 The activities of the control group during the intervention 135](#_Toc332731495)

[3.16 The testing materials 135](#_Toc332731496)

[3.17 The testing procedure 137](#_Toc332731497)

[Phase 3. The interpretation of existing data from the participating school’s internal tracking records 141](#_Toc332731498)

[3.18 The design of Phase 3 141](#_Toc332731499)

[Phase 4: Semi-structured interviews with the participant children 143](#_Toc332731500)

[3.19 Use of mixed methods 143](#_Toc332731501)

[Phase 5: Structured interviews with research assistants 148](#_Toc332731502)

[3.20 Rationale 148](#_Toc332731503)

[3.21 Conclusion 150](#_Toc332731504)

[Chapter 4 Data Analysis (Findings) 152](#_Toc332731505)

[4.1 Introduction 152](#_Toc332731506)

[Section 1: Datasets and forms of analysis 154](#_Toc332731507)

[4.2 Categories of test data used for statistical analysis 154](#_Toc332731508)

[4.3 Non-attempted items 156](#_Toc332731509)

[4.4 Forms of statistical analysis used 157](#_Toc332731510)

[4.4.1 *t*-tests of pre-test differences 157](#_Toc332731511)

[4.4.2 Effect sizes 158](#_Toc332731512)

[4.4.3 Analyses of variance 159](#_Toc332731513)

[4.4.4 *t*-tests of post-test differences 159](#_Toc332731514)

[4.4.5 Confidence intervals around effect sizes 159](#_Toc332731515)

[4.5 Effect sizes and statistical significances 160](#_Toc332731516)

[Section 2: Phase 1 – The randomised controlled trial 164](#_Toc332731517)

[4.6 Pre-test equivalence of samples 164](#_Toc332731518)

[4.7 Presentation of SPSS data 164](#_Toc332731519)

[4.8 Overall (total) mean scores 165](#_Toc332731520)

[4.9 Single-letter grapheme recognition mean scores 166](#_Toc332731521)

[4.10 Overall (total) mean scores, excluding single-letter grapheme recognition 168](#_Toc332731522)

[4.11 Cluster and digraph recognition mean scores 169](#_Toc332731523)

[4.13 Presentation of confidence intervals around effect size 173](#_Toc332731524)

[4.14 Comparison of this study with that of Gromko (2005) 174](#_Toc332731525)

[4.15 Overall scores by gender 176](#_Toc332731526)

[4.16 A systematic audit of the quality of the RCT 177](#_Toc332731527)

[4.17 Section 3: Analysis of phase 3 – the school’s internal tracking records 180](#_Toc332731528)

[4.18 Absentees 181](#_Toc332731529)

[4.19 National Curriculum levels 182](#_Toc332731530)

[4.20 Imputation of numeric values 183](#_Toc332731531)

[4.21 Presentation of Phase 3 data 184](#_Toc332731532)

[4.22 Comparison of degrees of progress, by National Curriculum sublevels 185](#_Toc332731533)

[Section 4: Analysis of semi-structured interviews with participating children 186](#_Toc332731534)

[4.23 Rationale 186](#_Toc332731535)

[4.24 Children’s attitude to the testing 187](#_Toc332731536)

[4.25 The children’s attitude to learning to read music, and the respective levels of difficulty in learning to read music and/or text 190](#_Toc332731537)

[4.26 The respective levels of difficulty associated with learning to read music through the clapping of rhythms, or through the playing of an instrument 196](#_Toc332731538)

[4.27 Children’s perceptions of themselves as readers of texts, and as readers of music 200](#_Toc332731539)

[4.28 The lack of correlation between children’s given reading levels, and their ability to articulate their thoughts about reading acquisition 205](#_Toc332731540)

[4.29 Section 5: Interviews with research assistants 211](#_Toc332731541)

[4.30 Question 1, 212](#_Toc332731542)

[4.31 Question 2, 212](#_Toc332731543)

[4.32 Question 3, 213](#_Toc332731544)

[4.33 Question 4 214](#_Toc332731545)

[4.34 Question 5, 214](#_Toc332731546)

[4.35 Summary and conclusions 215](#_Toc332731547)

[Chapter 5: Conclusion 217](#_Toc332731548)

[5.1 The context of the study 217](#_Toc332731549)

[5.2 The skills of the researcher 221](#_Toc332731550)

[5.3 Possible adjustments to the tests 224](#_Toc332731551)

[5.4 The public dimension 225](#_Toc332731552)

[5.5 The limitations of the study 227](#_Toc332731553)

[5.6 Other factors which may have contributed to the effect 228](#_Toc332731554)

[5.7 Further research 229](#_Toc332731555)

[5.8 And finally ... 232](#_Toc332731556)

[References 233](#_Toc332731557)

[Glossary 245](#_Toc332731558)

[Appendix 1: Nonsense words (pre-test) 247](#_Toc332731559)

[Appendix 2: Nonsense words (post-test) 248](#_Toc332731560)

[Appendix 3: Nonsense words (second post-test) 249](#_Toc332731561)

[Appendix 4: Overview of Reading Developmental Continuum (Rees, 1994) 250](#_Toc332731562)

[Appendix 5: Explanatory leaflet 251](#_Toc332731563)

[Appendix 6: Permission slip (parent/carer) 253](#_Toc332731564)

[Appendix 7: Permission slip (child) 254](#_Toc332731565)

[Appendix 8: Calculation of imputed data for children 29 and 30 255](#_Toc332731566)

[Appendix 9: Power calculation (i) 256](#_Toc332731567)

[Appendix 10: Power calculation (ii) 257](#_Toc332731568)

[Appendix 11: Interview with research assistants G and E 258](#_Toc332731569)

[Appendix 12: Interview with research assistant J 263](#_Toc332731570)

[Appendix 13: Pre-test attempts 267](#_Toc332731571)

[Appendix 14: Post-test attempts 268](#_Toc332731572)

[Appendix 15: Summary of pre- and post-test endpoints 269](#_Toc332731573)

[Appendix 16: RCT, pre-test *t*-test, overall scores 270](#_Toc332731574)

[Appendix 17: RCT, pre-test *t*-test, single letter grapheme recognition 271](#_Toc332731575)

[Appendix 18: RCT, pre-test *t*-test, overall scores less single-letter grapheme recognition 272](#_Toc332731576)

[Appendix 19: RCT, pre-test *t*-test, clusters and digraphs 273](#_Toc332731577)

[Appendix 20: RCT, pre-test *t*-test, nonsense words 274](#_Toc332731578)

[Appendix 21: RCT anova and unrelated t-tests, overall scores 275](#_Toc332731579)

[Appendix 22, RCT anova and unrelated t-tests, single-letter graphemes 276](#_Toc332731580)

[Appendix 23: RCT anova and unrelated t-tests, overall scores minus single-letter graphemes 277](#_Toc332731581)

[Appendix 24: RCT anova and unrelated t-tests, Clusters and Digraphs 278](#_Toc332731582)

[Appendix 25: RCT anova and unrelated t-tests, Nonsense words 279](#_Toc332731583)

[Appendix 26. Year 1 SATs results in reading: comparison between the participating cohort with the previous year’s cohort. 280](#_Toc332731584)

[Appendix 27: Phase 3, pre-test *t*-test 281](#_Toc332731585)

[Appendix 28: Phase 3 SPSS outputs 282](#_Toc332731586)

[Appendix 29: Evidence of Ethical Approval 283](#_Toc332731587)

# List of Tables and Figures

|  |  |  |
| --- | --- | --- |
| **Tables** | **Page** | **Figures** |
| **Chapter 1** | | |
|  | 14  15  20 | The Searchlights Model  The Simple View of Reading  The Tale of the Tadpole |
| **Chapter 2** | | |
| 2.1 Hypothetical classification of participating languages relative to the dimensions of syllabic complexity (simple, complex) and orthographic depth (shallow to deep)  2.2 Percentages of error in reading common words, age 7, by country  2.3 Barnett and Ceci’s taxonomy for near & far transfer  2.4 Analysis of the distance of transfer contained in this research design  2.5 Schunk’s taxonomy of diverse aspects of transfer  2.6 Assessments of auditory and visual processes important in music reading and music learning  2.7 Gromko’s pre-test and post-test means for DIBELS subtests by group  2.8 Mean gains (and s.d’s) in Gromko’s study  2.9 Effect sizes for Gromko’s study | 34  40  41  58  61  65  68  79  86  93  94  95 | 2.1 Jianzipu score  2.2 Kodaly tonic sol-fa hand sequence  2.3 Swanwick and Tillman’s (1986) sequence of early musical development  2.4 Gromko’s pre-test and post-test means for DIBELS subtests by group  2.5 Mean gains in Gromko’s study |
| **Tables** | **Page** | **Figures** |
| **Chapter 3** | | |
| 3.1 Original matched pairs of children within classes and recorder groups  3.2 Phases 1 and 2 of the study | 100  110  114  117  130  130  131  132  132  133  146  149 | 3.1 The four selected methods  3.2 The phases of the study  3.3 Piece 1  3.4 Piece 2  3.5 Piece 3  3.6 Piece 4  3.7 Piece 5  3.8 Piece 6  3.9 Interview questions for children  3.10 Interview questions for research assistants |
| **Chapter 4** | | |
| 4.1 Mean pre-test scores, t-test values, and statistical significances  4.2 Overall (total) mean scores  4.3 Single-letter grapheme recognition mean scores  4.4 Overall (total) mean scores excluding single-letter grapheme recognition  4.5 Cluster and digraph recognition mean scores | 155  155  164  165  166  168  169 | 4.1 Distribution of total scores, pre and post  4.2 Distribution of single-letter grapheme recognition scores, pre and post  4.3 Overall (total) mean scores  4.4 Single-letter grapheme recognition mean scores  4.5 Overall (total) mean scores excluding single-letter grapheme recognition  4.6 Cluster and digraph recognition mean scores |
| **Tables** | **Page** | **Figures** |
| 4.6 Nonsense word recognition mean scores  4.7 Further analysis of Nonsense word recognition scores  4.8 Confidence intervals around effect sizes  4.9 Comparison of this study with that of Gromko (2005), by calculated effect sizes  4.10 Overall results, girls  4.11 Overall results, boys  4.12 Mean pre-test scores, phase 3  4.13 National Curriculum Attainment Targets  4.14 Numeric values allocated to National Curriculum sub-levels  4.15 Phase 3, pre- and post-test mean scores, and effect size  4.16 95% confidence interval of effect size, phase 3  4.17 Tally of degrees of progress, by National Curriculum sublevel  4.18 Tally of children whose responses feature in the presentation of data | 170  171  173  174  175  176  177  177  178  181  182  183  184  185  185  205  210 | 4.7 Nonsense word recognition mean scores  4.8 Confidence intervals around effect sizes  4.9 Comparison of this study with that of Gromko (2005), by calculated effect sizes  4.10 Characteristics of a valid RCT  4.11 Wordle; from group interviews |

# Chapter 1, Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text?

## 1.1 Introduction

This thesis was written in interesting times, by a practitioner of Primary education, interested in enhancing the reading experiences of emergent readers. The central tenet of this thesis was to explore the extent to which the teaching of simple music notation can enhance children’s awareness of grapheme/phoneme correlation in the early stages of learning to read; in other words to see whether the reading of music can enhance children’s early decoding of text. As such, I found myself writing before a backdrop of one of the most topical and monitored aspects of current Primary and Early Years education – the learning and teaching of synthetic phonics (glossary, p.246).

The focus became narrow, but the intended study did not begin that way. Initially I was hoping to consider broad aspects of children’s reading: finding possible parallels between textual engagement and musical structure (perhaps comparing children’s engagement with written sentences or clauses to their recognition of musical phrases; comparing story structure with musical form; exploring patterns of preference for particular written or musical genres). These were grand ideals, but they carried insurmountable methodological challenges, and so were rejected.

By contrast, during my many years as a Primary school practitioner, I found the developmental stages of reading acquisition to be just as fascinating. Here are met not only exciting and diverse cognitive and pedagogic processes, but also governmental interest, policy, directives and documentation. Of late, this field has become contested ground, and to contextualise this study, I was obliged to begin by briefly exploring the educational climate and political landscape within which the research was couched, because an understanding of the context informed not only the selection of focus, but also the consideration of method; the significance of any findings; and the identification of a potential audience for publication.

## 1.2 Current national priorities

**Figure 1.1: The Searchlights model**

|  |
| --- |
|  |
| Source: National Literacy Stategy (DfEE, 1998: 4) |

The central place of the teaching of synthetic phonics in English Primary schools has its beginnings in the introduction of the National Literacy Strategy (DfEE, 1998), since when this centrality has been consistently held in the attention of successive governments. That document introduced the Searchlights model (Figure 1.1), which demonstrated recognition that reading is a multi-faceted enterprise, and which identified specific aspects of reading to be taught systematically and separately at word, sentence and text level, each level being afforded equal status.

**Figure 1.2: The simple view of reading**

|  |
| --- |
|  |
| Source: Letters and Sounds: Principles and Practice of High Quality Phonics (DfES (2007a: 9) |

The National Literacy Strategy was supplemented by the Rose Report (DfES, 2006b), a flagship education initiative of the Labour government, which countenanced the prescriptive and systematic teaching of phonics, and foreshadowed the systematic synthetic phonics programme (glossary, p.246) entitled *Letters and Sounds* (DfES, 2007a). In this document, the Searchlights model was replaced by ‘The Simple View of Reading’ (DfES, 2007a: 9, Figure 1.2), a model which presented two axes of reading development (comprehension and word recognition). During the last Labour government, the Rose Report (DfES, 2006b) had a major influence on the way in which early reading was taught, because the published pedagogy which followed it (and which was seen by many as prioritising the status of the teaching of systematic synthetic phonics over all other aspects of early reading acquisition) became additionally an identified interest of OFSTED inspections (OFSTED, 2011). The systematic teaching of synthetic phonics thus focused the attention of Primary head teachers, teachers, Literacy coordinators, and heads of Initial Teacher Training departments alike.

So successful had Jim Rose been with the Rose Report (DfES, 2006b), that he was commissioned to oversee a root and branch review of the Primary curriculum, which again contained specific affirmations about the place of systematic synthetic phonics:

Primary schools should continue to build on the commendable progress many have made in teaching decoding and encoding skills for reading and spelling through high quality, systematic phonic work as advocated by the 2006 reading review as the prime approach for teaching beginner readers.

(DCSF, 2009: 21, Recommendation 10)

Synthetic phonics is here advertised as being ‘*the prime approach for teaching beginner readers*’, an affirmation which was repeated in later documents. The report did not have sufficient time to be fully implemented in Primary schools before the 2010 general election, and failed to survive the incoming coalition government. However, although the review foundered, the place of the systematic teaching of synthetic phonics not only survived in the policies of the incoming coalition, but flourished, to the extent that the systematic teaching of the correlation between letters and sounds (graphemes and phonemes, glossary, p.245) secured a prominent place in the Schools White Paper (DfE, 2010), which set out the coalition’s agenda for education. For the purposes of this thesis, this White Paper had to be explored because it unequivocally expressed central tenets of government policy with regard to the decoding of texts and the pedagogy of early reading which served to establish the prevailing learning environment surrounding the field in which I conducted my research. A second review (the Henley Review, DfE, 2011a) was also potentially influential, and also needed to be explored, because its aims and recommendations were also apposite.

## 1.3 The Schools White Paper (2010) and the Henley Review (2011)

There was no obfuscation in The Schools White Paper (DfE, 2010). Even a cursory discourse analysis disclosed an overt agenda with regard to the teaching of phonics:

Ensure that there is support available to every school for the teaching of systematic synthetic phonics, as the best method for teaching reading. (p.11)

New teachers report that they are not always confident about ... the teaching of systematic synthetic phonics as the proven best way to teach early reading. (pp.22-23)

The evidence is clear that the teaching of systematic synthetic phonics is the most effective way of teaching young children to read. (p.43)

Note how the phrase ‘systematic synthetic phonics’ regularly heralds a correlative and superlative affirmation of status. One might also observe how the paper consistently holds that systematic synthetic phonics is the best way to *teach* reading, although it is less vocal about whether such an approach is the best way to *learn* early reading skills.

Literacy is not the only area in which reviews are commissioned of course. The Henley Review on the teaching of music in English schools (DfE, 2011a) was commissioned on 24th September 2010 by Michael Gove, Secretary of State for Education, expecting “*implementation [of the recommendations] from 2012*” (p.40). The timing of the review’s implementation coincided with the submission of this thesis. The review contained a variety of recommendations, and was of direct relevance and topicality to my study. For example, it acknowledged the contribution of the learning and teaching of music across and beyond the curriculum, and made specific reference to the beneficial relationship music can have with the development of literacy skills:

The benefits of a quality music education are those of increased self-esteem and aspirations; improved behaviour and social skills; and improved academic attainment in areas such as numeracy, literacy and language.

(DfE, 2011a: 42)

The Henley Review called for the status of music within Primary schools and Early Years settings to be enhanced, and made several recommendations to that end:

All children at Key Stage 2 should have the opportunity to learn an instrument through whole class ensemble teaching. Ideally, this would be for a period of one year, but at the barest minimum, one term of weekly tuition should be offered.

(Recommendation 3, p.31)

Currently, musical instruments are purchased on an ad hoc basis by music services and schools. This should be replaced by one centralised national purchasing system.

(Recommendation 16, p.34)

Much primary school classroom teaching of music is provided by non-specialist teachers. The amount of time dedicated to music in most Initial Teacher Training courses is inadequate. ... It is recommended that a new minimum number of hours of ITT for primary music teachers be spent on the delivery of Music Education.

(Recommendation 21, p.35)

All primary schools should have access to a specialist Music teacher. (Recommendation 22, p.36)

There should be a clear pathway through from Early Years Provision for all children. (p.13)

I have given this recent brief history of the status of synthetic phonics pedagogy in the White Paper, and have also heralded these recommended changes in music provision in the Henley Review, because they demonstrated not only that the field in which I was operating was topical, contentious, contested and shifting, but also that they had huge implications for the reading and learning experiences of children. In truth, I, like Plowden (Central Advisory Council for Education, 1967), was less interested in the politics than in the child. I sought not only to make reading a satisfying experience for children, but also to make the learning of reading a diverse, interesting, achievable and welcome process. I sought to “*overcome potential barriers to learning and assessment for individuals and groups of pupils”* (DfEE, 1999: 33) by exploring whether providing specific and timely additional stimuli might help the acquisition of reading skills for children, or perhaps for those children who struggle to make consistent grapheme/phoneme connections. Reading is an immersive act, not a mechanical one, but my research question sought to make a contribution to this one aspect of reading (the recognition and interpretation of symbols), acknowledging that the decoding of graphemes is not the only reading strategy that might be taught.

As such, I am at pains to state overtly that this thesis was not written as an intended addition to the synthetic phonics debate. Although I personally hold strong views on the subject, this thesis neither contributes to, nor subscribes to, nor denies, the argument that systematic synthetic phonics is the ‘best’ or ‘most effective’ method for the teaching of reading. My own view, to make my positionality overt, is that reading is a multifaceted act, and the learning and teaching of reading skills needs to reflect that. The ability and desire to learn to read comes from social as well as cognitive experiences. My experience of teaching children tells me that they learn best when enthusiasm for reading has been developed. There is more to reading than the decoding of graphemes and grapheme clusters. Even if a child struggles to decode, s/he can still engage with a story book or a pictorial non-fiction book in a meaningful way, making connections with the images and symbols of the book, perhaps even assigning sequence to a pictorial narrative. S/he will also make connections with previous books that have been encountered.

|  |
| --- |
| C:\Users\Mark & Claire\Pictures\9781405347655H.jpg |

**Figure 1.3 Tale of a tadpole (Wallace, 2009)**

Take, for example, the book shown in figure 1.3. Without being able to recognise or decode the word ‘Tadpole’ or even without knowing that text carries meaning, a child can engage with this book; can know that this book is about frogs and tadpoles: can share this book with a friend or adult; can reinforce or learn something new about tadpoles; can satisfy an interest in non-fiction books; can navigate through the book, knowing where to start and in which direction to continue; and can take enjoyment from engaging with the book. The decoding of the text is thus demonstrated to be one facet of reading, and not even, chronologically, the first facet that a child develops. One might therefore argue that if a child is to want read this book, an interest in tadpoles must precede an interest in phoneme/grapheme correspondences.

Meek (1994) observed that “*everything that children eat, wear, play with or pass in the streets has a sign or symbol*” (p.2), and so the act of reading is not confined to the interpretation of books, or even to organised texts. Children learn to read through storytelling, reenactment, prediction, character empathy, author and genre preference, logo recognition, mark making, play, role play, phoneme/grapheme correspondences, television, websites, mobile phones – the list could continue extensively. So let us put aside for one moment comparative claims and arguments about which method is ‘best’ (as if a better way will get us to a better destination). This thesis simply acknowledges that the teaching of synthetic phonics is recognised as one way of teaching one aspect of reading. The fact that it is such a contentious field at present was purely coincidental to my choice of focus. Were it not so topical, the focus would still have been selected. The single aim of this thesis was to explore whether the acquisition of those reading skills which are supported by the teaching (and learning) of synthetic phonics would be contributed to, and measurably enhanced by, the concurrent teaching of simple musical notation.

## 1.4 The teaching of the early reading and writing of music

Interestingly, tucked in amongst the noise and thrust for the singular and unrivalled status of the teaching of synthetic phonics, a question quietly hides, seemingly undisturbed, as to when, at what age, children should learn these skills of symbol recognition and blending (glossary, p.245). This question is rarely asked. Examples of current Primary and Early Years documentation (EYFS, DCSF, 2007a; *Letters and Sounds*, DfES, 2007b; The Primary Framework, DfES, 2006a) contain specific instructions about what should be taught and learnt, and when and how, but nowhere do these documents explain why. Why are these skills of decoding given priority over all other reading skills; and why should they be taught to children at this specific age? If an educationalist in 2012 were to suggest that the formal teaching of phonic reading skills in Early Years and Key Stage 1 was inappropriate, or too early, and unnecessary, that practitioner would be shouted down as a misfit and a heretic. Yet the same question when applied to young children’s reading of music, a question which is considered in most published literature on the general teaching of Primary music, receives guarded responses.

For example, Mills (2009) is representative of most writers on the subject when she states that “*pupils should not be introduced to music reading until a musical need has arisen*” (p.69). Hallam (2010) reinforces this view, considering that any form of notation, even if fluent, “*interferes with creativity*” (p.117) and so should be used only if a student chooses to use it. Odam (1995) was of the same view. Accordingly, there is no mention of the reading and writing of music in current Primary or Early Years documentation. The unused Rose Review (DCSF, 2009) was articulate on the need for children to learn to read and write texts through the systematic teaching of synthetic phonics, yet in his proposed restructuring of the curriculum under the heading ‘Understanding the Arts’, Rose had nothing to say on children’s reading and writing of music.

It can be seen from these omissions that reading and writing is deemed to be an optional skill for those children who learn an instrument, but it was not always thus. Prior to the original National Curriculum (1988), HMI (DES, 1985) stated that “*7-year-olds should be able to associate sounds with symbols; to show a readiness to see the relationship between performed music and various forms of notation (pictorial, graphic and conventional*” (p.3).

Glover and Young (1999) considered that since the ability of children in Primary phases to compose music far outstrips their ability to notate it in any form, the need for the formal teaching of notation is redundant. Jones and Robson (2008) agreed:

Music notation can be seen as something that could be learnt when needed. The motivation to read music lies in the desire to learn an instrument. ... Wholesale music notation teaching to primary children would be more of a hindrance to becoming musical because it will take away valuable time that can be used to experience actual music.

(Jones and Robson, 2008: 92)

Are the two codes of reading and writing so dissimilar in both form and function that such polar differences of opinion should be afforded to each? Chapter 2 of this thesis (Literature Review) contains an identification of the many ways in which the two codes are both similar and different, and Chapter 4 (Data Analysis) demonstrates that a specific intervention of the teaching of formal music notation did make a difference to the decoding skills of a cohort of Year 1 children, thus demonstrating that in this particular intervention there was evidence of cognitive links between the two codes. If there are developmental links, it is entirely plausible and justifiable that an argument might be made for pedagogic links as well, although whether the teaching of literacy might change in the light of arguments by Mills (2009), Hallam (2010), Odam (1995), Glover and Young (1999); and Jones and Robson (2008) is politically unlikely.

## 1.5 The focus of this study

To this end, and subject to important methodological changes (identified in Chapter 3), the research method of this thesis was an adaptation of a piece of published research by Gromko (2005), in which she conducted a two-group pre-test/post-test study using unmatched groups, producing data which demonstrated some correlation between children’s reading of text and their reading of music. I adapted her method by ensuring matched groups in order to produce a Randomised Controlled Trial (RCT). Gromko’s was a piece of American research, and it is of interest that there is a body of American literature on the subject of textual/notational transfer (Lamb et al., 1993; Butzlaff, 2000; Copple et al., 2000; Hansen et al., 2007; Hansen and Bernstorf, 2002), but an extensive literature search on my part found nothing published for a non-American audience. It is to address this curious incidence of geographically-confined publication history that I elected to recreate Gromko’s work, employing her research focus (but redesigning aspects of her methodology), and applying it to a semi-urban Primary school in North West Kent, England. I have related my methods and data to the documentation currently in use in England and Wales, and interpreted the findings in the light of the education initiatives prevailing in England in 2012.

The only piece of research I found from outside America was an unpublished English PhD thesis (Long, 2007). Long embarked upon doctoral research fieldwork similar in length to my own (six weeks), but smaller in size (*n* = 24). Although her doctoral thesis had a different focus to mine, concentrating as it did on children’s development of comprehension skills at the age of 8-10, she found that:

Very brief training (10 minutes each week for six weeks) in stamping, clapping and chanting in time to a piece of music while following simple music notation had a considerable impact on reading comprehension in children experiencing difficulties in reading.

(Long, 2007, in Hallam, 2010: 10)

Although her study was smaller in sample size and contained more variables than my own, Long (2007) was also able to determine whether the positive effects of the intervention were sustained six months after the intervention had finished, which they were, “*particularly amongst children with below average reading capability*” (p.113).

## 1.6 The notion of transfer

There is also the notion of transfer to be considered, and this is explored at some length in chapter 2 (section 2.11). ‘Transfer’ is the theory that a particular skill (or a piece of knowledge; or conceptual understanding) which is found or developed in one contextualised learning situation can enhance or influence learning in another, separate and therefore different learning situation, and is not a new one. Cases for and against transfer were argued by (again) American theorists at the beginning of the twentieth century (Thorndike and Woodworth, 1901; Judd, 1908; Dewey, 1916). Yet currently, even here is found a peculiarly American interest, not regularly reproduced in other parts of the world. For example, although the case for transfer has yet to be proved, three taxonomies of transfer produced by five American academics (Bransford and Schwartz, 1999; Barnett and Ceci, 2002; Schunk, 2004) were helpful in categorising ways in which seemingly disparate skills can overlap into other areas of learning, but all were produced by American writers and publication houses.

This thesis is therefore an attempt to do two things. Firstly, it aspires to determine whether such an overlap can be demonstrated to be consistent and measurable in a given learning environment; and secondly, it seeks to broaden the debate geographically (see chapter 5, section 5.2), giving it a wider contextual arena by wresting it away from the exclusive use of our American colleagues.

## 1.7 The research question

The research question has evolved in the writing of this thesis. I began by asking ‘What correlation can be found between children’s ability to read music, and their ability to decode text?’ However, the seeking of correlation was found to be nebulous; not well-served by the mainly quantitative methods I had in mind. I sought a research question which demonstrated impact: one which might make policy-makers sit up and take notice. As such, the question ‘Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text?’ promised to provide a more satisfying outcome, because the premise can be proved or not proved, and was therefore adopted. The intentions are finite, the results quantifiable. This research question was crafted in order to unambiguously inform a potential reader what the thesis is about; to limit the scope of the study, and to identify the approximate age of participant children.

Such a question engenders others, of course. What are these reading skills? How are they to be identified? Are they the same for each genre, or different? If different, how can the idea of transfer be justified? How is transfer, if it exists, to be demonstrated? These questions, and others like them, are considered in the following literature review.

# Chapter 2 Literature Review

## 2.1 Introduction

It is traditional for a doctoral thesis to contain a literature review (indeed some theses contain little else), but before embarking on such an enterprise it would be useful to briefly consider what such a review seeks to achieve, and what the parameters might be. This is somewhat problematic because the word ‘review’ has become a diverse misnomer. A book, play or film review is a disclosure of subjective opinion. An organisational review (for example the Rose Review, DCSF, 2009) seeks to impose change, or to restructure. However, none of these definitions is helpful in a doctoral thesis.

Hart (1998) considers the doctoral literature review to serve two purposes – “*to develop the skills of the researcher, [and to have] a public dimension”* (Hart, 1998: 27). This public dimension is for the doctoral student to demonstrate comprehensive understanding of both the history and breadth of the chosen study, and to contribute to the sum of human knowledge.

To this end, this literature review focuses on four areas. Firstly, it outlines the history of research into a field which seeks to demonstrate correlation between diverse aspects of music (in particular the reading of notation) and reading (in particular, decoding). It demonstrates that the sum of published writings on this area of research does not constitute a rich seam, and that therefore my research, although by no means pioneering, contributes to a previously sparsely-considered field. To categorically claim that there is little existing published research within a given field is a brave thing to do when writing a doctoral thesis, so I was pleased to have this view reinforced 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 reinforced the lack of published work in this specific area. They noted that:

Current results on the potential influence of musical training on language and literacy skills must be interpreted cautiously, pending further research. ... 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)

Secondly, the notion of cognitive transfer is explored, and to this end, this review briefly identifies parallel studies which sought to make correlation between aspects of reading and aspects of music, and which made connections between the mutual development of perceptual, language and literacy skills. Here, I draw extensively on the work of Hallam (2010). Care is needed, however, since section 2.11 will show that the notion of transfer is not universally recognised or proved. Welch et al. (2012) for example, are, at best, guarded on the subject. They write “*There does not seem to be any research showing a direct, causal relationship between music and the improvement of literacy, although there is a growing body of inferential evidence*” (Welch et al., 2012: 13). This thesis seeks to add to that body of evidence.

Thirdly, relevant published literature on auditory aspects of the early reading experiences of children with dyslexia will be explored (section 2.18). Although this thesis is not focused upon children with dyslexia, studies in this area (Thompson and Goswami, 2008; Forgeard et al., 2008; Huss et al., 2010) consider connections between music and skills of phonological processing (glossary, p.245), and such a focus on dyslexia serves to effectively inform this literature review.

Fourthly, this review critiques three primary sources: Long (2007); Standley and Hughes (1997); and Gromko (2005), by whose work my own was influenced in content, and method.

## 2.2 The literature, 1960-1990

The history of the field begins in the mid-1960s, and I have chosen that particular cut-off point for two reasons. Firstly, it was difficult to secure primary sources before that date, and I was loath to use secondary sources unless it was unavoidable. Secondly, the psychology and vocabulary of education prior to 1960, and between 1960 and 1980 disclosed an attitude to developmental education which was not helpful for my purposes. Before 1980, much of the writing about the development of reading through musical experiences was motivated by the desire to understand the reading problems of children who had special needs. This was laudable. However, phrases within the titles of books and journal articles disclosed a commitment to labelling which a modern reader finds both distasteful and (for the purposes of this thesis) counter-productive. For example, Bradley and Bryant (1978: 253) explored “*Reading backwardness*”, and this was echoed by Parker (1970: 244), who wrote in consideration of “*Musical perception and backwardness in reading*”. As late as the mid-1980s this attitude remained prevalent. Atterbury (1985: 114) wrote an article entitled “*Musical differences in learning-disabled and normal-achieving readers, aged seven, eight and nine*”. The history is relevant to my study, because such a study helps to categorise and identify areas of the field which promise to be rich seams, or which threaten to constitute unhelpful cul-de-sacs. However, the findings of data procured before 1980 had different agenda and rationales from my own study, and for that reason I have not dwelt on them beyond the identification of focus.

## 2.3 The literature, post-1990

By contrast, research studies since the early 1990s, particularly those produced by Gromko (2004, 2005); Lamb et al. (1993); Standley and Hughes (1997); Butzlaff (2000); Hansen et al. (2007); and Long (2007) were helpful in terms of both findings and methodology. A literature search, using databases such as ERIC (Educational Resources Information Center), Education Research Complete, Google Scholar, and SWETSWISE, revealed that during the last two decades an interest in the study of the demonstrable relationships between music and literacy has existed and been developed in the United States of America. Gromko’s (2005) work is based in Bowling Green University, Ohio. Standley and Hughes’ (1997) study was conducted at the Center for Music Research, Florida State University. Hansen et al.’s (2007) research, together with that of Hansen and Bernstorf (2002) was undertaken in two Kansas universities, and in the Kansas Department of Education. In contrast, the database searches revealed nothing published outside the United States. This can be partly explained by two observations. Firstly, ERIC (the most fruitful source of citations, numerically) is an American database. Secondly a paper trail of citation references from the publications of those authors cited above revealed a closely knit (American) research community. Nor did the published articles on the subject refer to any non-American sources. If such existed, it would surely have been in the interests of these academics to make comparative analyses, and this supposition on my part strengthened my claim that the location of research in this area is exclusively in the USA.

## 2.4 Lamb et al. (1993)

Since the 1960s, aspects of reading which have been investigated as having a possible connection with the reading of text have included such things as tonal memory, chord analysis and rhythm perception (Bentley, 1966). However, of those studies whose authors considered that their data pointed to a demonstrable correlation between musical skills and the development of reading ability, auditory discrimination has been found to be the most promising area of focus. Lamb et al. (1993) wrote:

It is generally considered that the acquisition of listening skills is a prerequisite of learning to read … However it has not been demonstrated whether the required listening skills need to be specifically phonemic in nature.

(Lamb et al., 1993: 19)

Lamb et al.’s study went on to demonstrate such a correlation. However, they acknowledged that their research design, being correlational, “*precludes the possibility of drawing causal inferences from the data*” (p.25). In that study, Lamb et al. explored two discrete skills in the area of auditory processes – the discrimination both of pitch, and of timbre. However, there are other auditory areas. The National Curriculum (DfEE, 1999: 124) lists seven – pitch (high and low), timbre (tone), tempo (speed), duration (the length of individual sounds), dynamics (loud and soft), structure (organisation), and texture (the identification and separation of individual sounds amongst a plurality of parts). For Early Years settings, Ouvry (2004) categorises these as “*loud/quiet sounds; fast/slow sounds; high/low sounds; hard/soft sounds; thick/thin sounds; and pulse and patterns*” (p.23). It is unrealistic to suggest that a single study of the size of my own (*n* = 55) might explore each and all of these, but the list shows the breadth of possibility that might be considered.

Lamb et al. (1993) did acknowledge that other approaches are equally valid, and they went on to recognise other studies which took different approaches to their own. For example, in terms of texture, McMahon (1979) trained young children to discriminate between pairs of three-note chords, some differing by only a semitone, and this was reported to be linked to improvements in their word recognition, reading and general phonic skills compared with a matched untrained group.

Not all research is designed to show how music can enhance aspects of literacy, of course. Approaching the argument from a different direction, Spychiger et al. (1993), demonstrated that when the amount of time spent teaching music was increased, whilst at the same time the amount of literacy teaching was reduced, no detrimental effect was observed on children’s language and reading skills.

During the 1990s, an area of regularly researched interest with regard to the link between music and reading was the enhancement of reading comprehension through the technique of asking children to sing the text whilst reading it (Ericson and Scjuliebo, 1998; Fetzer, 1994; Fountas and Pinnell, 1999; Hansen and Bernstorf, 2002; Miller and Coen, 1994; Page, 1995; Smith, 2000; Yopp and Yopp, 1997). However, although these studies and publications each suggested a beneficial correlation through these interventions or activities:

[many of these authors] suggested that phonemic awareness, defined as 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).

## 2.5 Butzlaff (2000)

**Figure 2.1: Jianzipu score**

|  |
| --- |
|  |

Butzlaff (2000: 167) summarised a literature analysis by listing four hypothetical reasons why instruction in music may help children acquire reading skills at different levels. Firstly he observed that both written text and written music must be read from left to right, and in both cases the symbols map directly to a specific sound. Of course, not all languages or musical notations are written from left to right. For example, Arabic and Hebrew are written from right to left, and the jianzipu scores (Figure 2.1), written for the qin (a Chinese seven-stringed zither), are written from top to bottom, in keeping with the traditional written script (Yung, 1997: 17). The jianzipu symbols can be transliterated to be read from left to right for a western readership, because the symbols themselves do not rotate.

Butzlaff’s (2000: 167) second suggestion was that “*skill in reading requires a sensitivity to phonological distinctions, and skill in music listening requires a sensitivity to tonal distinctions.*” This echoes Lamb et al.’s (1993) reported correlation of reading and listening skills. Butzlaff (2000: 167) offered a supposition that there may be a “*general kind of auditory sensitivity*” which could be applied to both genres. Evidence for this supposition was taken from Douglas and Willats (1994: 99), whose data pointed to “*[children’s] multi-sensory awareness and response to sounds*”. This notion of there being a link between listening skills and reading skills was given weight by the structure of *Letters and Sounds* (DfES, 2007b), where, under aspect 1 of Phase 1, the very beginnings of organised developmental experiences for children, is found the directive for children to be encouraged to develop “*general sound discrimination* [under the categories of] *environmental sounds; … instrumental sounds; … and body percussion”* (DfES, 2007b: 4).

Butzlaff’s third suggestion related to the repetitive nature of books designed for children’s early reading, which is mirrored when children read the words of songs – the repetition (of, for example, songs with many verses interspersed with a chorus) encourages and allows prediction at word, sentence, and structural levels. In Butzlaff’s view, such predictable text may train reading skills. Young children will happily engage with a song, and if they don’t know a section will hum or ‘la’, or wait until a familiar section returns, at which point they will resume singing without any sense of failure or lack of engagement. This is particularly true of chorus songs. “*Wouldn’t it be interesting if children were allowed to say “la, la, la” when they couldn’t read a word in order to keep moving in the sentence*”, observed Hansen et al. (2007: 47). Perhaps this is a pedagogic strategy that could be explored.

Finally, Butzlaff offered a motivational approach, which my experience of teaching instrumental lessons supports. He suggested that the peer experience when playing in a group or ensemble has a motivational effect, since the success of the whole is dependent on the team effort of the parts. “*Perhaps personal responsibility ... leads to heightened academic responsibility and performance*” (Butzlaff, 2000: 167).

Butzlaff’s (2000) work was pertinent to my own study. He used seven electronic databases, and scrutinised 94 articles or books, which were reduced to 31 by the application of the following three criteria – “*a standardized measure of reading ability was used as the dependent variable; a test of reading followed music ‘instruction’; statistical information was sufficient to allow for a calculation of an effect size*” (Butzlaff, 2000: 168). The design of my study was informed by this categorisation, and my methodology (chapter 3) shows that each of these features was present.

Butzlaff also helpfully divided his analyses between considerations of correlational studies and experimental studies. The correlational studies all differed from my own study because in each case (24 studies) a single reading score was ascertained for each student, and the reading performances of students with some music experience were compared with the reading performances of students without musical experience. Butzlaff’s analysis of the results showed that there was a “*strong and reliable association*” (Butzlaff, 2000: 172) between reading development and musical instruction, but Butzlaff himself suggested that the results could be influenced by a variety of factors:

[There is] the possibility that students who are already strong in reading choose to study music; students who are interested in [reading] music are already interested in reading; or it is possible that a causal relationship exists, such that either music instruction transfers to reading achievement, or the reverse.

(Butzlaff, 2000: 172)

This interest in ‘transfer’ was repeated in Gromko’s (2005: 199) work, and has informed the phrasing of the title of my own study. Indeed, I have set aside an entire section of this literature review (section 2.11) to explore the notion and the evidence for, and against, the existence or possibility of transfer.

This perceived correlation between strength of reading, and participation in the reading of music, did not apply in my study, because the participating cohorts were each populated by children with a matched and broad range of reading ability, and there was no attrition (glossary, p.245) after the paired cohorts were finalised, and so it could not be considered here whether strong or weak readers were more or less attracted by the reading of music. However, as a brief aside, the concept of choice in what skills are being learnt is a stark area of divergence between the experiences of children learning to read and those learning to read music. I have known children to lose interest in learning an instrument, and so give up. They therefore choose not to learn to read music, and the default position of the music teacher is to respect that decision, and the child walks away, no longer encumbered by the need to interpret the symbols. Educational values differ between the two genres of reading. One is compulsory, the other is deemed to be optional.

A criticism that might be lodged against Butzlaff’s work, however, is that he was too simplistic. This notion of “*transfer*” (Butzlaff, 2000:172) is attractive and plausible if there are few variables between that which is being transferred ‘from’ and ‘to’. In music, the written symbols map onto specific notes (ignoring the possibility of tonal transposition, which was an unnecessary complication for the purposes of my research), and these symbols are afforded specific rhythms by specific notational conventions. Moreover, the written conventions of music are, with only a few exceptions, homogenous. By this I mean that a piece of music transcribed in standard notation by an Italian can be read by a Scot, a Russian or a Peruvian without the need for translation. There is not the diversity of language within conventionally-written music that there is within conventionally-written language. Furthermore, the writing of music is an exact science (a feature I have always found ironic, given the range of emotions that can be touched and described by music). In music a written symbol maps directly to a given pitched sound for a given duration, given an identified tempo.

## 2.6 Orthography and transparency

However, in the decoding of text, the same notion that written symbols (graphemes) map directly onto specific sounds (phonemes) is demonstrated to be wrong, at least in English. Three sources are of interest here: Goswami (2005), Seymour et al. (2003) and Dehaene (2009). Goswami (2005) spoke of the ‘transparency’ of diverse European languages, by which she referred to the consistency of alphabetic orthographies (glossary, p.245). In some languages:

one letter or letter cluster can have multiple pronunciations (e.g. English, Danish). In others, it is always pronounced in the same way (e.g. Greek, Italian, Spanish). In some alphabetic orthographies, a single speech sound (phoneme) can have multiple spellings (e.g. English, French, Hebrew). In others, it is almost always spelled the same way (e.g. Italian).

(Goswami, 2005: 274)

In this article, Goswami drew heavily on Seymour et al.’s (2003) comparative study of European orthographies, and she amplified two important aspects of their work: the first being the alphabetic consistency of written European languages’ orthographies, and the second, the syllabic structure of each as a spoken language. Seymour et al. (2003) preferred to speak of “*shallow orthographies*” (p.143) as opposed to Goswami’s “*transparent*” ones. By ‘shallow orthographies’, Seymour et al. referred to those European orthographies “*with consistent grapheme–phoneme correspondences*” (p.144), and such languages were listed to include Finnish, Greek, Italian, Spanish and German. Deep orthographies, which require the reader to engage in “*the implementation of a dual (logographic and alphabetic) foundation*” (p.143), include Portuguese, French, Danish and English. In their analyses and categorisations of the transparency or depth of European languages, Goswami (2005) and Seymour et al. (2003) do not offer starkly different lists.

**Table 2.1: Hypothetical classification of participating languages relative to the dimensions of syllabic complexity (simple, complex) and orthographic depth (shallow to deep)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Orthographic depth  Shallow Deep | | | | |
| Simple syllabic structure | Finnish | Greek  Italian  Spanish | Portuguese | French |  |
| Complex syllabic structure |  | German  Norwegian  Icelandic | Dutch  Swedish | Danish | English |

Source: Seymour et al. (2003: 146)

Seymour et al. (2003; see Table 2.1) categorise many European languages using this dual layer of consideration, which contrasts syllabic complexity with orthographic depth. Syllabic complexity:

refers principally to the distinction between the Romance languages, which have a predominance of open CV [Consonant-Vowel] syllables with few initial or final consonant clusters (e.g. Italian, Spanish), and the Germanic languages, which have numerous closed CVC syllables and complex consonant clusters in both onset and coda position (e.g. German, Danish, English).

(Seymour et al., 2003: 145)

Seymour et al.’s classification of orthographic depth can be compared with that of Dehaene (2009), who refers to English as being “*very opaque*” (p.231) and who reports the following percentage of errors made by children at the end of first grade (or 7 years old) when reading common words in the home language (Table 2.2).

**Table 2.2: Percentages of error in reading common words, age 7, by country**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Denmark | England | Finland | France | Greece | Germany |
| 29% | 67% | 2% | 28% | 3% | 3% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Holland | Italy | Norway | Portugal | Spain | Sweden |
| 7% | 5% | 8% | 23% | 6% | 6% |

Source: (Adapted from Dehaene, 2009: 231)

Unsurprisingly, the three most opaque languages identified by Seymour et al. (English, Danish and French) are shown to be the most problematic when readers of these languages are required to accurately read common words at the age of 7. Finnish, being the most transparent, is shown to be least problematic. Note how Swedish and Dutch are classified closely in both analyses; so too, Greek, Italian and Spanish.

As such, it is clear that the emergent reader of English has not only to grapple with inconsistency of grapheme-phoneme correspondence, but also with a range and variability of structure even within the syllables of the words s/he is learning to read. Take, for example, the monosyllabic homophones ‘poor’, ‘paw’, ‘pore’ and ‘pour’. Within these examples there is to be found no consistent grapheme/phoneme correlation in the vowel sound. It is true that the onset (glossary, p.245) of each is consistent, but the vowel grapheme is different in each (<oor, aw, ore, our>), while the vowel phoneme is the same, namely /ɔː/ (in many accents in England, that is). These words have no spoken coda (glossary, p.245). Similarly the words ‘scam’ and ‘skim’ share an ‘initial consonant cluster – short medial vowel – single consonant coda’ structure, yet the onset clusters, though audibly identical, are different graphically, because the phoneme /k/ is spelt differently (though according to rule: /k/ is mostly spelt <c> before <e, i, y>, <k> elsewhere).

## 2.7 The vocabulary of the field of phonics

At this point it is necessary to define terms, particularly the word ‘blend’. In current documentation (EYFS guidance, DCSF, 2008; Primary Framework, DfES, 2006a; *Letters and Sounds*, DfES, 2007b) this word is used only as a verb, which describes the process, within synthetic phonics for reading, once the constituent parts (graphemes) of a word have been recognised and sounded out, of pushing them together to read the word as a whole. This use of the word is perceived as unproblematic:

[Practitioners should] look listen and note children’s confidence in blending and segmenting and in using grapheme-phoneme knowledge to read and spell regular consonant-vowel-consonant (CVC) words, including consonant digraphs and long vowels ... [and] how children read simple words by sounding out and blending the phonemes all through the word from left to right.

(DCSF, 2008: 54)

Key features are to teach beginner readers to apply the highly important skill of blending (synthesising) phonemes in order, all through a word in order to read it.

(DfES, 2006a: 7)

The main objective should be segmenting words into their component sounds, and especially blending the component sounds all through a word.

(DfES, 2007b: 4)

However, in many published books and articles, the word ‘blend’ has commonly been used as a noun to depict instances within words where two or more consonant graphemes stand sequentially, but are not separated by a vowel grapheme, and represent separate consonant phonemes. Coincidentally the word ‘blend’ begins with a blend by this definition:

An additional test of phonic skills was provided by ...`Standard Reading Tests', involving consonantal blends respectively at the beginnings and ends of phonically simple words.

(Lamb et al., 1993: 23)

See how one of the most influential commercially produced phonics programmes, *Jolly Phonics*, uses the word as a noun, muddling the terminology:

You will find it helpful to be able to distinguish between a blend (such as st) and a digraph (such as sh). In a blend the two sounds, ‘s’ and ‘t’ can each be heard. In a digraph this is not so... When sounding out a blend, encourage children to say the two sounds as one unit, so fl-a-g not f-l-a-g.

(*Jolly Phonics*, 2011: 5)

However, such use of the word is ambiguous and ill-defined, since, by definition, if two or more objects are blended, they merge to become a single new item, losing their previous separate identities. That is not the case with the use to which Lamb et al. (1993) and *Jolly Phonics* (2011) put the word. This notion may lead a nascent reader to consider that there may somehow be a way in which the two separate sounds in a consonant blend should be treated phonically in the same way in which a consonant digraph (such as ‘sh’) might be approached, which in developmental terms is unhelpful. A more accurate and helpful term for the erroneously-named ‘blend’ is ‘consonant cluster’, which accurately describes both the constitution of, and the relationship between, two independent but consecutive consonant phonemes or graphemes. In the remainder of this thesis, instead of using the word ‘blend’, I shall refer to this combination of independent consecutive consonants as a ‘cluster’.

## 2.8 Orthographic depth and syllable structure

Returning to the consistency or otherwise of grapheme-phoneme correlation, my research design took Seymour et al.’s (2003) distinction between orthographic depth and syllable structure into consideration. The nonsense words that the children were asked to decode (appendices 1-3) have zero orthographic depth (since reading them aloud by standard grapheme-phoneme rules would always produce the same pronunciation), but a hierarchic variety of syllable structures – the format of the tests moving from simple syllable structures (CVC, VCC) to more complex ones (cluster-vowel-cluster; vowel-consonant digraph; words with repeated syllable structures – see methodology, section 3.16). This allowed children to use and demonstrate their understanding of ‘analogies’, a phrase used by Goswami (1995:139) “[to describe using] *the spelling-sound system of one word, such as ‘beak’, as a basis for working out the spelling-sound of a corresponding word, such as ‘peak’.*”

Seymour et al.’s work highlighted the additional consideration of the relationship between the phonological complexity of a given spoken language, and the spelling consistency of the written language. In both of these respects, the child who is learning to decode English is disadvantaged. English is both non-transparent and phonologically complex by comparison to all the other European languages Seymour et al. studied.

These inconsistent factors have combined and evolved to produce spellings of words in English which are (rather quaintly) referred to in the Primary National Strategy (DfES, 2007b: 194) as “*tricky words*”, i.e. words which are not fully synthetically decodable. These include such high-frequency words as ‘one’, ‘said’, ‘could’, and ‘come’. This observation had a methodological implication for my work, and it is for this reason that I confined my research to exploring children’s recognition of three categories of stimulus: single-letter graphemes, clusters and digraphs, and non-ambiguous nonsense words (sometimes referred to as pseudo-words, e.g.. Dehaene, 2009: 231). To expand the study to include either words or phrases which may be contextualised semantically or syntactically (glossary, p.246), or to include words which need to be learnt as ‘tricky’, would have increased the number of variables to a level which would have threatened the validity of the research. I was careful to disambiguate the chosen nonsense words, to allow children to generate correct or plausible pronunciations based upon children’s use of analogies.

## 2.9 Attitudes to the isolation of decoding in children’s reading

In this study I was seeking to explore whether the teaching of the reading of music can have a positive influence upon the development of decoding skills. For correlation to be demonstrated, it was necessary for levels of ‘transparency’ (Goswami, 2005) to be high. As such, there were no ‘tricky words’ in this research: indeed, there were no recognisable words at all. This was because the study focused on the act of decoding. All the words used in the tests were ‘transparent’ and ‘shallow’, orthographically. However, informed by the work of Seymour et al. (2003), the syllabic complexity of the words in the tests developed incrementally as the children progressed through them. There are many principles to learn and understand in the process of learning to read, and in this piece of research I was isolating just one of them. Because of the influence of Goswami (2005), my research methods were specifically chosen to exclude the possibility of data being muddied by children bringing with them differing abilities with regard to established or emerging reading attributes such as whole-word recognition, sentence construction, or comprehension. I needed the children to decode, and nothing else.

However, this study was undertaken within a period of direct government intervention in the field, during which the diverse processes involved in learning to read had been officially and hierarchically categorised. Although the selection of test focus was informed and adapted from the research of Gromko (2005), it is of interest that the mandatory phonics screening for Year 1 children, piloted in a number of schools in England during the week beginning 13th June 2011, was based upon the same premise of requiring children to decode some nonsense words (in addition to some real words). The benefit to children of these nationally administered tests is contested ground, and again I do not want to get involved with the politics of the subject, because the question of whether or not the teaching of systematic synthetic phonics is the best way to teach children to read was irrelevant to my study. I was trying to show whether children’s phonic development can be enhanced by the transfer of skills, knowledge and understanding from another form of reading. I was not making a case for the enhanced place of the teaching of phonics in current Early Years and Key Stage 1 education practice. Nonetheless, for the purposes of this doctoral thesis, I think it worthy of reaffirmation that my study was topical, and was focused on an area which was of major significance to current drivers of education policy.

For this reason, this study was neither diminished nor enhanced by a recent announcement from the United Kingdom Literacy Association (UKLA), amongst whose membership I am pleased to be numbered, which publicly described the mandatory phonics screening as “*bonkers*” (Reedy, 2011). This is the combative vocabulary of political lobbying, not of research analysis. My tests were not designed to provide diagnostic data on individual children, but to provide evidence of developmental possibilities within the field of phonic decoding. The mandatory screening had parallels with my own, but the rationales were completely different.

However, because the place of the teaching of systematic synthetic phonics was a current national imperative, the need to clarify and justify the differences in rationale was essential. This need for justification was reinforced to me over breakfast during the UKLA conference in 2010, where I found myself surrounded by indignant fellow members who had (initially) been kind enough to ask about my research. To my astonishment I found myself obliged to forcefully refute their (aggressively voiced) linear syllogism that since I was a UKLA member, and since the tests in my study held similarities to the proposed mandatory phonics screening check for six-year-olds, and since right-minded UKLA members tended to consider the mandatory check to be a ‘bad thing’, my research was, by definition, methodologically inappropriate. Fixation on the enhanced hierarchical place of synthetic phonics amongst the canon of other developmental reading processes, demonstrated by successive governments, had made the area a battle-ground, even over breakfast.

In spite of my fellow delegates’ outrage with regard to the use of nonsense words, the use of such words to enhance or measure decoding skills is not new. Where the skill of decoding is required to be isolated from other reading skills, such an approach has been employed in a variety of situations. For example, a scheme known as ‘Toe by Toe’ has been regularly used in prisons for over twenty five years to help adults, usually with learning difficulties, to learn to decode. This scheme uses ‘Polynons’ (nonsense words) such as ‘plettonsig’ and ‘affidsad’ in order to help learners develop skills of word-building and syllable division, and to “*focus learners’ attention solely on decoding, and avoid guessing based on any other ‘cue’* ” (DCSF, 2007b: 89).

Bolduc (2008) collated a review of literature, focusing on studies which considered the effects of music instruction on emergent literacy capacities among preschool children (a focus somewhat different to my own), and he included a number of studies which used pseudowords (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 (perceptive melodic and rhythmic abilities), phonological awareness, and “pseudoword” recognition abilities). There was therefore an established research tradition here with regard to nonsense words, to which I was happy to contribute.

## 

## 2.10 What is it, to be a reader?

The national argument centres on definitions of what it is to read, or to be a reader. For example, the answer to the question ‘*Can a child read without being able to decode?*’ depends upon one’s definition of the word ‘read’. Other, less definition-dependent, questions might be asked. Can a child find ways to engage with texts without being able to decode? What reading skills must a child develop to be able to enjoy a book, or a webpage, a magazine, or a piece of music? Children are faced with the interpretation of symbols from a very early age, in a variety of contexts. In mathematics, for example “*the first obstacle faced by young children is that numerals are abstract concepts ... Numerals themselves bear no relation to the concrete objects that can be touched and seen*” (Hansen, 2011: 25). Graphemes and decontextualised music notation (for example, quavers or rests) are also, in themselves, devoid of meaning, until meaning is ascribed to them. In numerals and musical notation the ascribed conventions of meaning-making are consistent. In the decoding of graphemes, especially the reading of English (Goswami, 2005), they are not.

This is not to say that Goswami’s research challenged or threatened my intended study. Indeed, the opposite was true. Goswami’s evidence is that, since children who are learning to read English are faced with an orthography which is inconsistent and various, so too they need multiple strategies in order to engage with it. If children’s decoding of text is shown to be improved by their learning to read music, it is possible that this is a result of an auditory development, which is as essential a part of reading as visual cues. “*Phonological awareness refers to the whole spectrum, from primitive awareness of speech sounds and rhythms; to rhyme awareness and sound similarities; and at the highest level, awareness of syllables and phonemes*” (Copple et al., 2000: 84). It may be that engagement in the reading of music stimulates such phonological awareness. Burns et al.’s (1999) research showed a correlation between children’s auditory capabilities and their ability to decode at an early developmental level. “*Children who have a better degree of phonological awareness when they enter school are better equipped to learn to read*” (Burns et al., 1999: 46). The participants in my study had already entered Year 1 via an Early Years setting of course, but their reading and phonological awareness would still have been developing, and this study contributed to that process.

## 2.11 The notion of transfer

All of this hinges on this notion of ‘transfer’: one skill being able to impact upon another (potentially related) skill. This is contested ground, because the development of one skill (or body of knowledge, or area of understanding) causally enabling the development of another has yet to be demonstrated in any discipline. “*Despite a century’s worth of research, arguments surrounding the question of whether transfer occurs have made little progress toward resolution*” (Barnett and Ceci, 2002: 612). Dettermann and Sternberg (1993) are nothing short of dismissive:

First, most studies fail to find transfer. Second, those studies claiming transfer can only be said to have found transfer by the most generous of criteria and would not meet the classical definition of transfer, defined by Detterman [1993: 4] as "the degree to which a behaviour will be repeated in a new situation". ... In short, from studies that claim to show transfer and that don't show transfer, there is no evidence to contradict Thorndike's general conclusions: Transfer is rare, and its likelihood of occurrence is directly related to the similarity between two situations.

(Dettermann and Sternberg, 1993: 15)

It is of interest that Detterman and Sternberg refer to ‘behaviour’, rather than learning, or knowledge. This reference to similarity of situation is a recurring theme which shall be returned to and explored throughout this literature review’s section on transfer.

Historically there are two philosophical stances on transfer. The first goes back as far as Plato’s philosophical dialogue *Meno* (Scott, 2006) in which he (Plato) explored a rigidly-held binary world view, asking whether continuity (memory, learning, skills) was inherently a human condition, or whether it was somehow located within the world, external to our mental processes. Plato argued that continuity was carried within individuals, and that new information or experiences were actively afforded continuity by the person experiencing them. Continuity, therefore, was found not to be a precondition of nature but a human attribute. This confirmation of human agency parallels Piaget’s (1977) later concept of accommodation (Plato, not considering children’s development, had no need of Piaget’s concept of assimilation). Plato, then, had a concretely formulated notion of knowledge transfer, any case against it being perceived by him as unarguable.

The second is iterated by Thurman (1984, in Beach 2004), who, by contrast, takes a social constructivist approach to transfer, seeing it (transfer) as the result of social contexts, and interactions between individuals; communities of individuals; and communities of communities. Learning and knowledge can be seen as developed by influences external to the human mind, rather than an internal product of the mind. For example, cultural practices and attitudes are developed and maintained corporately by communities, not in isolation by individuals. Writing in a similar period, Bruner (1986) famously observed that:

Most learning in most settings is a communal activity, a sharing of the culture. It is not just that the child must make his knowledge his own, but that he must make it his own in a community of those who share his sense of belonging to a culture.

(Bruner, 1986: 127)

I acknowledge the truth of Bruner’s philosophy in regard to learning and education (indeed it has informed much of the organisation of my teaching for two decades), but Bruner, in his emphasis of the word ‘most’, acknowledges that not all knowledge is socially constructed. For example, the knowledge that one needs to sleep regularly is not learnt through discussion and social interaction; neither is the knowledge that touching a hot thing hurts. But with regard to my study, it was true that if there was a transfer of knowledge, skills or understanding between the two forms of reading that I was looking at, the vehicle for that transfer was the participation in a communal series of lessons or “*literacy events*” (Barton et al., 2000). Unlike the knowledge that hot things hurt, you cannot learn everything there is to know about literacy in a single event. Barton et al. (2000) consider that *“... the notion of events stresses the situated nature of literacy; that it always exists in a social context*” (p.8).

Perhaps this is why the concept of transfer has never been demonstrated, in any discipline. There is a methodological tension here. My positionality as a teacher reveals a constructivist view, but a quantitative study of the size of my own cannot hope to take diverse social factors into consideration. To do so would be to accept a range of variables which would make the quest for quantitative clarity impossible and invalid.

Beach (2004) also acknowledges the difficulties of defining the notion of transfer, as a result of his belief that learning tends to be context-dependent. This necessity for context makes transfer from one field of knowledge to another hard to prove, because it is likely that the contexts of learning will not be directly comparable:

As a construct in educational psychology, [transfer] refers to the appearance of a person carrying the product of learning from one task, problem, situation, or institution to another. It is here that the metaphor begins to break down. Transfer is distinguished from run-of-the-mill learning by virtue of its distinct tasks and situations, yet it does not include the genesis of tasks and situations as part of the process.

(Beach, 2004: 101)

In other words, if knowledge, skills and understandings are learnt, they are learnt in a context. The skilled teacher will give his or her students reasons for learning a fact, skill or attitude. S/he will convey purpose and provide appropriate resources. It is a poor teacher who provides identical learning environments for differing learning intentions. Therefore, if transfer is to exist, it must negotiate journeys both from cognitive positions, and physical ones. For purpose to be given to reading music there must be an instrument to play, and (ideally) an audience to appreciate. The same resources are not required when learning to decode text (although if we are to accommodate the idea of learning being a social construct, an audience is similarly an effective stimulus to learning).

This awareness of the importance of context when considering transfer is not a new notion. The same view was maintained by Shweder (1980):

The everyday mind accomplishes a very difficult task. It looks out at [the] behavioural world of complex, context-dependent interaction events and insubstantial correlations among events, yet it perceives continuities, neat clusters, and simple irregularities.

(Shweder, 1980: 77)

More recently, Hallam (2010) has written on more generic aspects of the impact of music to enhance intellectual, social and personal aspects of children’s development. Her study was larger and wider than my own, but it is interesting that she also spent considerable time exploring and justifying the notion of transfer. She defined transfer, stating that “*transfer between tasks is a function of the degree to which the tasks share cognitive processes*” (p.6), and it is worth noting that Hallam referred to transfer between tasks, rather than between aspects of knowledge. She referred to “*low road transfer [which] depends on automated skills and is relatively spontaneous and automatic ... and high road transfer [which] requires reflection and conscious processing*” (Hallam, 2010: 6). This terminology echoed that used by Gromko (2005), and is drawn from Schunk (2004, see table 2.5, below).

Much has been made in recent years of ‘transferable skills’. Gone are the days when one had a job-for-life. For example, presently, apprentice plumbers at my local post-compulsory college learn technical, mathematical, scientific and health and safety aspects of the trade (as one would expect), but also contained within this apprenticeship programme are modules in English, customer liaison, accounting, first aid, and business administration, none of which is directly relevant to the ability to fit a gas boiler or a tap. These are skills and knowledge which can be easily transferred from one context to another, and are consciously included in the design of the programme to make the programme attractive to prospective students in terms of content and future employability. Transfer is an institutionally recognised phenomenon in current post-compulsory education, and demonstrates a move in policy away from ‘training’ to a broader concern for ‘education’.

Another case in point, highly topical at present as a result of the recent White Paper (DfE, 2010), is the terminology surrounding Initial Teacher Education (ITE). Many of my colleagues involved in the development of the next generation of teachers become highly exercised at the mention of the perceived misplaced acronym of ITT (Initial Teacher Training). Skinner’s famous line that “*education is what survives when what has been learnt has been forgotten*” (Skinner, 1964, May 21st) speaks of this need to educate, rather than train; to contextualise rather than to compartmentalise; to teach in a cross-curricular way, rather than with a linear, subject-based approach; to allow students to construct their own knowledge based upon their previous diverse experiences, rather than to receive knowledge based upon the experiences of their teacher.

## 2.12 Philosophical positions on the concept of transfer

Philosophical positions are important in terms of the driving of education policy. I have already stated that this study was informed by American research, and the reason for this geographical focus may be found in the interest that American theorists took, at the beginning of the 20th century, with regard to transfer. This philosophical legacy clearly remains there. Philosophical positions taken by American policy-makers have had a profound influence on the experiences of children in American schools. Thorndike argued that since (in his view) transfer is an empirical response (Thorndike and Woodworth, 1901), the structuring of learning tasks should be built around elements of diverse tasks which are identical, or clearly linked. Thorndike reportedly held a rigid view of learning and teaching, seeing education as functionalist; a means to an end; a vehicle for the production of productive members of society. Dewey (1916), by contrast, argued directly against Thorndike, as did Judd (1908), both holding that transfer should be acknowledged for its role in meaning-making, and flexibility of learning. They valued process over product, and principles over facts, and they each criticised the policy that this cognitive process of transfer; this aspect of human development; should be shackled by its use as a means to organise a curriculum, and related education procedures. For Judd and Dewey, transfer should be a mode of exploration and invention; a vehicle for the imagination of what might be, based on our knowledge of what is; a reason for weakening curriculum boundaries, not tightening them. It is interesting that in the public argument that these writers maintained, the existence of transfer is assumed, acknowledged, and taken as read by all parties. Thorndike’s argument won the day, and the effects of that functionalist philosophical underpinning of American education continues to this day, explaining why there is an established and vibrant body of writing in American journals on the subject of transfer, and why there continue to be experiments and research on the influence of one skill (for example, in this case, the reading of music) on another (the decoding of text). This is a body of research to which I am happy to contribute, and to import, applying it to an English context.

This is not to say that I support Thorndike’s functionalist views – far from it. But the presentation of Thorndike’s argument does explain the high geographical incidence of American research on transfer.

## 2.13 Taxonomies of transfer

Barnett and Ceci (2002), again writing in America, have developed a taxonomy (glossary, p.246) of transfer (Table 2.3). They ask two interesting and fundamental questions, the first being “*What is being transferred*” (p.621), and the second being “*What and where is learning transferred from and to*?” (p.623). This second question is clumsy, being something of a hybrid, but this taxonomy builds on the analysis of two other American academics, Salomon and Perkins (1989), who demonstrate what they perceive to be happening in transfer by separating the ‘how’ of transfer (the cognitive mechanisms), from the ‘what’ of transfer (the kinds of knowledge and skills that are being transferred). Barnett and Ceci’s (2002) second question, once unpicked, was most relevant for my purposes, because I have found their identification of six ‘contexts’, to which they apply the perspective of ‘near’ and ‘far’ transfer (Table 2.3), to be helpful.

**Table 2.3: Barnett and Ceci’s taxonomy for near & far transfer**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Context: When and where transferred from and to | | | | | |
|  | Near Far | | | | |
| **Knowledge domain** | Mouse vs. rat | Biology vs. botany | Biology vs. economics | Science vs. history | Science vs. art |
| **Physical context** | Same room at school | Different room at school | School vs. Research lab. | School vs. home | School vs. The beach |
| **Temporal context** | Same session | Next day | Weeks later | Months later | Years later |
| **Functional context** | Both clearly academic | Both academic, but one non-evaluative | Academic vs. Filling in tax forms [functional] | Academic vs. informal questionnaire | Academic vs. at play |
| **Social context** | Both individual | Individual vs. pair | Individual vs. small group | Individual vs. large group | Individual vs. society |
| **Modality** | Both written, same format | Both written, multiple choice vs. essay | Book learning vs. oral exam | Lecture vs. wine tasting | Lecture vs. wood carving |

Source: Barnett and Ceci (2002: 621)

It must be said that this is a taxonomy which brings with it some problems. On a technical level, one may wish to question the vocabulary of phrases such as ‘biology *versus* botany’, or to question the phrasal validity of ‘Lecture versus wine tasting’ – the combative vocabulary suggesting that as development in one area benefits, development in the other must as a corollary be disadvantaged. This is an interpretation I am not sure Barnett and Ceci intended. On a cultural level, the figure suggests that there might, in some way, be a tension between the physical contexts of school and home. This observation is sadly and frequently true of course, where school literacies and attitudes can stand in stark contrast to a child’s (or a community’s) home experience and culture, in which case the use of the word ‘versus’ can sometimes be appropriate, but, returning to Barnett and Ceci’s classifications, whether the same is true, and to an identified greater extent, between the school and the beach is arguable.

On a cognitive level, one might also question the entry under ‘Functional contexts’ which gives, as an example of far transfer, “*Academic vs. at play*”. Such a phrase suggests that this taxonomy is not aimed at Early Years or Key Stage 1 learning environments, in which the organisation of purposeful play is a seriously academic business. Although tasks in an Early Years setting may be predominantly kinaesthetic; learning objectives fluid; choice of activity often, but regularly, in the control of the child; social interactions encouraged, planned and manipulated; and assessment opportunities intentionally observational (although currently just as summative and target-led as in any other phase of learning), the idea that there would be a far transfer between academic aspects of learning and play would be anathema to any Early Years practitioner.

On a social level, Barnett and Ceci (2002) demonstrate a commitment to a psychology of individual, isolated learning, rather than accommodating a social constructivist view. I struggled to complete Table 2.4, being an application of Barnett and Ceci’s (2002) taxonomy to my own research programme, because the social context near-transfer classifications of ‘Both small group’ or ‘Both large group’ are not to be found in Barnett and Ceci’s original, and so I had to add them to be able to make analysis of my research design.

However, I make reference to this taxonomy because the point is well made by Barnett and Ceci (2002) that transfer is not an homogenous concept, applicable to all and every learning situation. The reason that transfer is so notoriously difficult to prove is that one is rarely comparing like with like. There are (according to Barnett and Ceci) six contextual variables alone, each with degrees of distance. I have mentioned Barnett and Ceci because for my research model I was building on Gromko (2005), who made analysis of children’s reading scores after a series of lessons in the reading of music. Gromko refers to Barnett and Ceci’s (2002) “*near transfer hypothesis*” (Gromko, 2005: 201), by which she means that the contexts of ‘when, how and where’ the two modes of reading are conducted were very similar. With regard to my own research, there are contextual variants which had to be taken into account methodologically. To this end, I adapted Gromko’s method, the changes from which I will explain, analyse and justify in a later section. But for now, were I to apply Barnett and Ceci’s taxonomy for transfer to my research design, in order to demonstrate the ‘nearness of transfer’, the table would appear as follows.

**Table 2.4: Analysis of the distance of transfer contained in this research design**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Context: When and where transferred from and to | | | | | |
|  | Near Far | | | | |
| Knowledge domain |  | Interpretation of symbols, albeit in different formats |  |  |  |
| Physical context |  | Different room at school |  |  |  |
| Temporal context | During the same six-week period |  |  |  |  |
| Functional context | Both academic |  |  |  |  |
| Social context | Both individual, with the use of small and large groups |  |  |  |  |
| Modality |  | Both written, different format |  |  |  |

This adaptation of Barnett and Ceci’s (2002) taxonomy shows that, under these six classifications, no aspect of my research design required children to make large cognitive leaps in terms of transfer. Returning to Shweder (1980), I have been careful through this literature review to identify “*continuities, neat clusters, and simple irregularities*” (p.77) between the reading of music and the reading of texts, while at the same time offering children “*complex, context-dependent interaction events”* (Shweder, 1980: 77) in the reading of music, as a result of which measurable progress might be made in the children’s decoding of text. Thus, transfer, if not proved, is a most plausible explanation.

The terminology of near- and far-transfer, together with the clumsily-presented notion of ‘When and where transferred from and to’ helped me to make analysis of what was trying to be achieved in my study. Bransford and Schwartz (1999), again writing in America, but preceding Barnett and Ceci (2002), suggested that, if we are to consider the possibilities of transfer, in addition to a consideration of “*the replicative ‘knowing that’, and the applicative ‘knowing how’, we should add the associative ‘knowing with’* ” (Bransford and Schwartz, 1999: 61). Here, Bransford and Schwartz drew heavily on Broudy (1977), who considered that, far from being elusive and non-provable, transfer is essential for any learning to take place, be it new or consolidatory. We can replicate without learning (for example, we might follow another driver to a destination, but without internalising the route). We can apply without learning (for example, as a method of calculation we might correctly insert an additional zero into the tens column during the workings of a two-digit multiplication, but without knowing why). Broudy considers that, for learning to take place, it must occur as a result of an interpretive function, interpreted through the lens of a learner transferring prior and relevant experience to a new situation. Broudy (1977) also emphasises that “*much of the knowledge that supports ‘knowing with’ is tacit and may be unavailable for recall*” (p.11).

## 2.14 Reading skills that the children brought with them into the study

This concept of tacit knowledge can be applied to my own study. At the beginning of the programme the participating children were familiar with the concept and mechanics of reading texts. They knew that written symbols in both texts and music carry meaning, even though initially not one of them knew what meaning the music symbols conveyed or represented. They knew that the symbols in texts are sequential, and the children therefore anticipated (correctly) that this would also be true of written music. They knew that in English, texts are read from left to right, and from top to bottom; that symbols can be written as well as read; and that the symbols hold conservation of meaning (i.e. the same symbol will mean the same thing on each encounter, and that given the same sequence, successful interpretation of that sequence will make the same sound, even in a different book or setting. This is universally true of standard music symbols, but is less true, as we have seen, in the orthography of written English). The children knew that when the symbols stopped, the story (or tune) stopped also.

All these things the children brought with them, and so Broudy (1977) would argue that it is inevitable that their knowledge of how texts work would enhance their learning of the reading of music, and in such a situation (Broudy would say), transfer from one genre of reading to another cannot be avoided. It would be interesting (although ethically prohibitive, and rightly so) to see whether a cohort of children of the same sample size and age, but amongst whom no child had been exposed to books or texts, would be able to interpret simple written musical notation as easily as the children in this cohort did. I suspect they would struggle.

But here the analogy breaks down. I am not investigating whether knowledge, skills and understanding held by developmental readers can be transferred from one field in which they are experienced to another field in which they are not, indeed I am seeking to demonstrate the opposite phenomenon, and I am encouraged by Broudy. In the light of the enthusiastic way in which Detterman and Sternberg (1993) take a hatchet to the concept of transfer, it would be easy to feel that, within the field of educational psychology, I am mistakenly trying to go against the laws of nature. It might be likened to a person trying to develop a counter-argument against the second law of thermodynamics which states that in nature, energy can only be lost, and therefore heat can only pass from a hotter body to a cooler one. In thermodynamics, the transfer is, and can only be, in one direction (entropy), but at least such transfer is acknowledged within the field of science. Does such inflexibility apply to the field of education? I am trying to demonstrate that the process of transfer from the experience of internalising new learning can be of measurable benefit to an already developing skill set, if the degree of transfer can be identified and maintained as being ‘near’, which I have demonstrated is the case, by the use of Barnett and Ceci’s (2002) taxonomy of near and far transfer.

This notion of near and far transfer was taken up by Schunk (2004), who presented another taxonomy (Table 2.5), simpler in design than that of Barnett and Ceci (2002), drawing on the same vocabulary of near and far transfer, but adding new vocabulary to describe or to categorise the relationship between chosen learning situations, and between the nature of what is being transferred or learnt.

**Table 2.5: Schunk’s taxonomy of diverse aspects of transfer**

|  |  |
| --- | --- |
| Type of transfer | Characteristics |
| Near | Overlap between situations. Original and transfer contexts are similar. |
| Far | Little overlap between situations. Original and transfer settings are dissimilar. |
| Positive | What is learned in one context enhances learning in a different setting. |
| Negative | What is learned in one context hinders or delays learning in a different setting. |
| Vertical | Knowledge of a previous topic is essential to acquire new knowledge. |
| Horizontal | Knowledge of a previous topic is not essential but helpful to learn a new topic. |
| Literal | Intact knowledge transfers to new task. |
| Figural | Use some aspect of general knowledge to think or learn about a problem. |
| Low Road | Transfer of well-established skills in almost automatic fashion. |
| High Road | Transfer involves abstraction so conscious formulations of connections between contexts are needed. |

Source: (Schunk, 2004: 220)

Using the vocabulary provided by Schunk (2004), my methodology section (chapter 3) shows that I was creating near-transfer conditions, intending to generate positive and horizontal transfer opportunities. I was struck, however, by the inclusion of ‘negative’ transfer in Schunk’s taxonomy – the idea that what is learned in one context might hinder or delay learning in a different setting. Presumably this might happen when presenting children with conflicting information in seemingly similar formats (for example, metric and imperial measurements), or when presenting identical information which becomes incongruous in different formats (for example, the reading of analogue and digital clocks). The reaction of the children to the intervention, and their progress within it, showed that neither of these negative transfer conditions pertained in my research, but this was an aspect of the research that I had not considered when applying for ethical approval, because my experience of teaching children to read music has never presented to me the notion that this genre of reading was incompatible with the reading of texts. Happily, the data analysis of my research demonstrated that such concerns of negative impact were not applicable in this case.

A further approach to transfer was given by Cousin (2010). She spoke of “*liminal*” aspects of learning, a term she used to refer to classificational features of learning, in which an individual is “*betwixt and between*” conscious (supra-liminal) and unconscious (subliminal) learning. Cousin suggested, in rather Piagetian terms, that for an individual to respond to a stimulus, new or otherwise, it is necessary to “*classify to make sense [of it], but to make connections between diverse classifications to comprehend it*” (Cousin, 8th July 2010). In other words, no learning can be learnt in isolation. The processes of conscious learning are interdependent on previous experience, knowledge, skills and understandings. To this end, it can be argued that the processes involved in developing reading skills – skills of decoding and classification (is it a letter or a number?; a letter or a squiggle?; a letter or a note?; a vowel or a consonant?; a note or a rest?; a crotchet or a quaver?) may be strengthened by the contemporaneous learning of two contrasting reading systems.

## 2.15 Children’s articulation of their experiences of learning to read

In my study I intended not only to teach children to read music, but also to talk to them about their reading experiences. Care was needed here of course. If I had engaged children in conversation about the research, in addition to their involvement in an intervention, I had to be able to identify whether any development of the reading skills was a result of the children’s learning of the reading of music, or of the conversations I had with them. The conversations required them to identify and articulate aspects of reading of which they might not have been previously aware, thus lifting their understanding from the “liminal” to the supraliminal, by helping them to contextualise and develop the learning of one reading code by making conscious connections to another. To ensure that the interviews did not impinge on the data, the interviews occurred after the post-intervention tests were completed.

## 2.16 The assessment of music and of literacy

It is not only the curriculum content or means of learning which hold interesting parallels between the two modes of reading being looked at in this study. If one considers the ways in which music reading and music learning are traditionally assessed, one discovers further insights into relationships between music and literacy. Hansen et al. (2007) presented an analysis of “*assessments of auditory and visual processes important in music reading and music learning”* (p.149), in which were presented 20 discrete aspects of such assessment. Seven are auditory; two are binaural (the ability to discriminate and maintain a musical part in association with other, different, parts); one relates to timbre recognition; and ten are visual. Hansen et al. present such an analysis to highlight exclusively “*skills* [which] *represent discrete tasks decoding both text and music”* (p.146). Not all 20 are relevant to my research, because Hansen et al.’s book aims to cover a broader range of aspects of reading than my own, more focused, study, but nine of the skills are directly relevant, and I have adapted Hansen’s table to present those nine, as follows.

**Table 2.6: Assessments of auditory and visual processes important in music reading and music learning**

|  |  |
| --- | --- |
| **Auditory association**  Match the auditory sample to the sound source through kinaesthetic response | **Visual analysis**  To visually identify specific note types or durations |
| **Auditory blending**  Identify sound patterns in a melody or rhythm | **Visual blending**  To visually locate patterns of melody or rhythm in music |
| **Auditory discrimination**  The identification of diverse elements of music\* | **Visual – special functions**  To demonstrate consistency in responding to music symbols in different music literature |
| **Auditory memory**  To recall and perform a song (or melody) from memory | **Visual associations**  To locate, identify and respond to dynamic markings in music |
| **Visual memory**  To recall and perform a song or melody by looking at the music |
| \* For “*Elements of music*”, see National Curriculum (DfEE, 1999: 124). Hansen et al. do not use the phrase, but they list many of the various components | |

Source: Adapted from Hansen et al. (2007: 149)

Analyses of assessment opportunities and intentions offer a means of clarifying what a teacher is trying to achieve – indeed experienced teachers often begin by determining what is to be assessed, and how, before designing teaching plans and learning environments to facilitate the fair demonstration of the learning. Hansen et al.’s (2007) categorisation of assessment demonstrates effectively aspects of text-reading which might usefully be developed through the teaching of music reading, and a useful comparison can be made between this categorisation and current Early Years and Primary documentation.

### 2.16.1 Auditory skills (i): Auditory association

Both the Early Years Foundation Stage (DCSF, 2007a) and strand 5 of the Primary Framework for the teaching of mathematics and literacy (DfES, 2006a) require children between the ages of 40 and 60 months to be taught to “*Link sounds to letters, naming and sounding the letters of the alphabet*”, and to “*Use a pencil and hold it effectively to form recognisable letters*” (DfES, 2006a: 48). This parallels Hansen et al.’s ‘auditory association’, being a kinaesthetic response to an auditory stimulus. The EYFS (DCSF, 2007a) suggests that phonic teaching should be “*multi-sensory in order to capture [children’s] interests, sustain motivation and reinforce learning*” (p.53).

### 2.16.2 Auditory skills (ii): Auditory blending

Similarly, the vocabulary of ‘auditory blending’ speaks of the identification and synthesis of separate sounds, which Hansen considers an essential component of learning to read music, and is also a central tenet of the systematic synthetic phonics pedagogy found in ‘Letters and Sounds’ (DfES, 2007a). “*Beginner readers should be taught to apply the highly important skill of blending (synthesising) phonemes in the order in which they occur*” (DfES, 2007a: 10). This skill of blending is prevalent in the guidance for the teaching of ‘Letters and Sounds’ (DfES, 2007a), indeed the word ‘blending’ appears 28 times in the 28-page document. Children are encouraged to participate in “*sound talk. For example,* [children are encouraged to] *sound talk a word in an instruction (e.g. Give yourselves a p-a-t on the back*” (p.55). Often robotic chopping arm movements are employed to reinforce the separation, or segmentation, of a word’s component sounds, in an attempt to make the experience multi-sensory.

### 2.16.3 Auditory skills (iii): Auditory discrimination

Hansen et al.’s (2007) ‘auditory discrimination’ relates to the identification of features of specific sounds (high/low, soft/loud, long/short), and this echoes the studies by Butzlaff (2000) and by Lamb et al. (1993), mentioned earlier. The Primary Framework (DfES, 2006a, strand 5) invites children to “*explore and experiment with sounds, words and texts*” (p.48). The first three aspects of phase one of the Letters and Sounds (DfES, 2007b) documentation are listed as follows:

Aspect 1: General sound discrimination – environmental sounds

Aspect 2: General sound discrimination – instrumental sounds

Aspect 3: General sound discrimination – body percussion

(DfES, 2007b: 4)

I acknowledge that phase one was designed for younger children than those in my study, but the pedagogic emphasis shown by the DfES (2007) on the pre-reading development of auditory discriminatory faculties is shown by Hansen et al. (2007) to be well-placed, and highlights that connections made in this thesis between music development and reading development are strong, topical, and relevant to current practice in the teaching of early reading.

### 2.16.4 Auditory skills (iv): Auditory memory

The need to consider ‘auditory memory’ relates to sequencing and structure (beginning, middle and end; verse/chorus; form). Hansen et al.’s categorisation of music reading thus has in mind reading at a textual or whole-piece level, which goes beyond the scope of this thesis, but it is mentioned and described in order to contrast with ‘visual memory’, considered next.

### 2.16.5 Visual skills (i): Visual memory

The majority of Hansen et al.’s assessment features relate to visual aspects of music reading, and, although I have shown that the features of only three of their seven auditory assessments apply directly to recognised aspects of textual reading (as defined and identified within current documentation), six of the 11 features of visual assessable aspects survive the migration. Hansen et al.’s ‘visual memory’ contrasts with ‘auditory memory’ in that auditory memory relates to the “*big shapes*” (Barrs, 1991: 6) that are found in a text, relating to aspects of whole-text awareness, such as setting, characterisation, plot, inference, intertextuality, and children’s preferences in the process of text selection. Hansen et al.’s ‘visual memory’, by contrast, relates to the “*small shapes*” (Barrs, 1991: 6), for which children develop skills of decoding individual graphemes within a word or syllable (synthetic phonics).

However, Barrs’ ‘small shape’ definition of grapheme/phoneme correlation was vague, and has been superseded. Brooks (2003) offered more comprehensive definitions:

**Synthetic phonics** refers to an approach to the teaching of reading in which the phonemes associated with particular graphemes are pronounced in isolation and blended together (synthesized). For example, children are taught to take a single syllable word such as ‘cat’ apart into its three letters, pronounce a phoneme for each letter in turn /k, æ, t/, and blend the phonemes together to form a word.

**Analytic phonics** refers to an approach to the teaching of reading in which the phonemes associated with particular graphemes are not pronounced in isolation. Children identify (analyse) the common phoneme in a set of words in which each word contains the phoneme under study. For example, teacher and pupils discuss how the following words are alike: *pat, park, push* and *pen*.

(Brooks, 2003, in Torgerson et al., 2006: 13)

Strand 5 of the Primary Framework (DfES, 2006a) expects Year 1 children to have an “*automatic recognition of high frequency words*” (p.48). It was interesting and frustrating to note that many of the participating children in my study, whilst engaged in the pre-test, struggled more with two of the nonsense words that were (carelessly on my part) too similar to standard words they knew, than with those nonsense words which held no discernible similarity to words in the children’s sight vocabulary. Several children naturally applied their visual memory to these two words, an eventuality I was hoping to avoid as I was seeking to test decoding skills exclusively.

### 2.16.6 Visual skills (ii): Visual analysis

If a learner is employing a ‘visual analysis’ in music reading, s/he is identifying specific note types or durations (such as the understanding of the function and meaning of such things as quavers, crotchets, and rests; and possibly the identification of clefs). The children in my study were introduced only to crotchets, quavers and rests, and they learned to recognise, clap and play combinations of these notes, on a stave, being written on the treble clef (). In spite of this symbol featuring prominently at the beginning of every line of music, no child enquired as to what this symbol might be (a note or a rest?), and neither were the children told. Cousin’s (2010) classifications (liminal, supra-liminal and subliminal) were presumably unnecessary for the children, and so the question was not asked. I was relieved by this. The theoretical explanation of what a treble clef is, and how it works, was considered by me to an unnecessary complication, on the grounds that it could have potentially caused negative transfer (Schunk, 2004: 220) by introducing an additional musical concept which was not a prerequisite to the children learning to interpret the notes on the stave.

To support this exploration of visual analysis further, an early aspect of strand 7 of the Primary Framework asks for children to “*know that print carries meaning*” (DfES, 2006a: 23). The EYFS (2007b) expects that children aged 30-50 months should “*ascribe meanings to marks that they see in different places*” (p.59), and by 60 months should “*use their phonic knowledge to write simple regular words and make phonetically* *plausible attempts at more complex words*” (p.60). This is only possible through a conscious analysis of what the marks on the page mean, and how they work. Within the focus of my study, this cognitive process, both in terms of symbol recognition and understanding of symbol function, constituted a very near transfer indeed.

As a brief aside, since I was using this EYFS document to make relevant analysis for my study, I do not think it pedantic to point out that the writers of that document cannot mean ‘phonetically plausible attempts’ when referring to developmental writing. They mean ‘phonically plausible attempts’. The production of phonetically plausible attempts (e.g. fǝ'nƐtIklI 'pl**ɔː**zǝbǝl ǝ'tƐmpts) within the writing of children aged 60 months would clearly be a very ambitious expectation indeed, since phonetics is not taught in Early Years settings.

### 2.16.7 Visual skills (iii): Visual blending

Hansen et al.’s (2007) ‘visual blending’ relates to “*the recognition of patterns of melody or rhythm*” (p.149) and this closely corresponds, in another example of near transfer (Barnett and Ceci, 2002: 621), to the important link between the blending and segmenting of individual sounds, represented by either music notation or graphemes. Compare, for example, definitions of blending and segmenting, taken from DFES (2007) and Hansen et al. (2007) respectively:

Blending means merging individual phonemes together into whole words; segmenting is the reverse process of splitting up whole spoken words into individual phonemes.

(DfES, 2007a: 10)

Visual blending [is to] determine pattern from individual icons/notation; to identify specific icons or symbols embedded in a musical context; to look selectively at a musical pattern, or individual notation symbols, in music. (Hansen et al., 2007: 6)

In presenting their analysis of links between music and literacy, Hansen et al. (2007) were attempting to develop “*a common vocabulary”* (p.4), because they believed there are “*valuable skills in music classes that are an integral part of becoming literate people*” (p.2). This focus on blending is just one example of such a skill, but one which directly invites the support which my research hopes to give.

There are some skills regarding the interpretation of symbols which apply across many academic disciplines. For example, in mathematics, children aged 40-60 months are expected not only to recognise single numerals, but also to recognise that those numerals represent the number of items in a set (DCSF, 2007a: 66). The numeral ‘5’ symbolically represents five objects, irrespective of what those objects (or that combination of objects) might be. The child must learn not only the symbol (‘5’); but also that a given symbol is part of a series of symbols (numerals); that the application of that numeral is consistent across all objects; that numerals have a function (ordering); and that numeric symbols work in combination with each other (place value). So too it is with music notation, and with graphemes. The symbols of individual musical notes and letters must be learnt, both individually, and as part of a series (a word, or a musical phrase). The consistency of a given symbol’s meaning, irrespective of context, must be acknowledged. The function of these symbols must be understood, and the way in which the symbols work in combination with each other must be known (such as reading conventions; and the mechanics of the text or melody).

### 2.16.8 Visual skills (iv): Special functions

It is this conceptual aspect of reading music which Hansen et al. (2007) refer to as ‘visual – special functions’. For example, a child must learn that a quaver in one melody has the same function as a quaver in another, irrespective of whether that melody is faster or slower than the first. The digraph <sh> makes the same sound, irrespective of whether the genre is fiction or non-fiction. These essential visual correspondences of early reading acquisition are shared by music and literacy alike.

### 2.16.9 Visual skills (v): Visual associations

Finally, Hansen et al. refer to ‘visual associations’, which relate to those aspects of reading music which go beyond exclusive interpretation of the notes to be produced. In the early stages of reading music these aspects typically are confined to dynamics (the requirement to become louder or quieter) and tempo (faster or slower). Visual associations when decoding text are rather more complicated. Mallett (2005: 24) spoke of “*bibliographic skills*”, by which she meant the ability to know how a text works. These include not only “*big-shape*” (Barrs, 1991: 6) structural aspects of a text, such as prediction (for example, the knowledge that if mother tells Little Red Riding Hood not to go through the forest, then through the forest she is most likely to go), but also “small-shape” aspects. For example, if a child is stuck on a word, s/he may be able to use pictures surrounding the text for assistance, making plausible suggestions about the written text from information not supplied within the written text. Hansen et al. demonstrated that, in both codes, the ability to glean information from a source external to the decoding of the symbols (notes/graphemes) can have an influence on the production or the interpretation of those symbols. This is one area in which the reading of text does not have an equivalent in the reading of music. A helpful picture of Granny in the tune ‘My Granny moves quite slowly’ would not have aided the children’s reading of that melody. This was mentioned by the children in post-test interviews (section 4.25).

Hansen et al. (2007) made no mention of aspects of assessment which are peculiar to music, and therefore isolated from the reading of written texts. Perhaps, given the title of their work, it was not in their interests to do so. Additionally, they only mentioned linear aspects of reading music, which therefore have a physical resemblance to the reading of texts (for example, individual melodies are read horizontally from left to right, and in sequential rows from the top of a page to the bottom). There was no mention by Hansen et al. of the reading of chords or harmonic aspects of reading music, which require vertical interpretation. Neither was consideration given to the reading of music with more than one part, in which case one’s partners’ parts in the piece, which appear on the same page, must be acknowledged, but ignored, being played exclusively by the partner. In reading texts, words come individually, linearly, and sequentially, but in music, notes can come in synchronous clusters, by one or more players, and it is common for whole sections to be repeated. Hansen et al. did not analyse such aspects of the reading of music in relation to the reading of written texts, because it would not have been helpful or relevant to their argument to do so. However, this feature of their analysis does not negatively affect the relevance of their work to my study, because vertical proficiencies were not required in my randomised controlled trial. The recognition of linear sequences of individual notes, together with the rhythmic implications of crotchets, quavers and rests, was the full symbolic range that the children in my study were invited to learn and interpret.

## 2.17 Physiological links between music and other forms of learning

The identification of physiological and cognitive links between the learning of music and other forms of learning is not new. There is a variety of frameworks from which teachers of music might draw were they to seek philosophical support for their methods. These include the Dalcroze eurhythmics; the Kodaly sequence, and the Orff-Schulwerk. Each of these musicians sought to link aspects of music with other areas of learning, and I will briefly look at each of these in turn, in an attempt to find either methodological or theoretical support for my own study.

Dalcroze (in Campbell, 2008) developed a movement-based pedagogy, whereby a response of movement is “*the means to musical understanding, and those who practise Dalcroze allow their students to make the brain-body connection to the music they hear*” (Campbell, 2008: 133). Dalcroze was not interested so much in the functionality of music (notation and formal structure), as the emotional and expressive effects it brought to, and engendered within, those who sought to engage with music. For Dalcroze, the student has to be able to do six things – *“pay attention, turn attention to concentration, remember, reproduce a performance, change, and automate*” (Caldwell, 1993: 27). Although Dalcroze was principally interested in music education, he acknowledged it to be obvious that these skills had a broad general application in teaching and learning.

**Figure 2.2: Kodaly tonic sol-fa hand sequence**

By contrast, the Kodaly sequence (sometimes referred to as ‘the Hungarian method’) is “*a well ordered and logical system of materials and techniques that lead students to the development of performance skills, hearing and notational literacy*” (Campbell, 2008: 133). Campbell observes that, like Dalcroze, Kodaly was adamant that regular involvement in musical activities was essential to human well-being, and he advocated daily (or at least very regular) instruction in traditional, folk and national music. This instruction included both the performance and the transcription of songs and melodies, and children were taught hand movements (Figure 2:2, from Campbell, 2008: 134) to coincide with the tonic sol-fa notation system. This enabled the melodic structure of music, and particularly the intervals between individual notes, to be identified physically as well as aurally, and the physical aspect of describing the melody through prescribed hand movements was the precursor to more formal written notation.



Whereas Dalcroze and Kodaly were each interested in the teaching of music, Orff, through observation of children’s music preferences and habits, made analysis of children’s learning. He categorised four linear stages – “*imitation, exploration, literacy, and improvisation*” (Campbell, 2008: 134), and it is interesting for the purposes of my study to note that the reading and writing of music is deemed by Orff an essential function of children’s music compositional processes. This contrasts starkly with current views about the place of notation (Mills, 2009; Hallam, 2010; Odam, 1995; Glover and Young, 1999; Jones and Robson, 2008) identified in chapter 1 (section 1.4). Orff favoured games which teach structure (question and answer improvisation, ABA and rondo forms).

These three proponents of music education emphasised musical participation over theoretical understanding (Gault, 2005: 7), but each considered that music education had an influence on other aspects of personal development.

Similarly, Thompson and Johnston (2000) conducted research on the reading abilities of “*reading-disabled*” (p.63) children using the reading of non-words as an indicator of analytic and synthetic phonic ability. The aim of their research was clearly different from my own, but as an empirical test method it is worthy of consideration as it offers “*means and standard deviations for descriptor variables of the participant groups*” (p.71). Gromko (2005) and Thompson and Johnston (2000) each employed the use of non-words as a valid test method, and this informed my own study. My research sought to contain as many quantitative data as possible, in order to provide valid and reliable data for analysis, and to provide an informed basis of discussion for interviews.

## 2.18 Relevant literature with regard to dyslexia

Any number of approaches might have been taken in the design of this thesis, and in the early design stages, when direction was being explored with peers and colleagues, I did consider focusing on the experiences of children for whom reading acquisition was difficult, and children with dyslexia was a focus which was often suggested. Indeed, as Welch et al. (2012) point out, “*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). However, I elected not to pursue this line, firstly because children with dyslexia are usually identified to be as such at an age some years after the age phase that I was most interested in (5-6 year olds). Secondly, Ganschow et al. (1994) identify the difficulties that children with dyslexia can experience with the formalised study of music, in particular, music notation. Knowing that I intended to use notation as a core intervention, I was uncomfortable with the idea of intentionally subjecting children to something from which they might not take satisfaction. It was possible that this could have been justified ethically, but I also feared a high attrition rate, which I could not risk. Nonetheless, interesting and relevant work has been undertaken with regard to dyslexia and music.

For example, Forgeard et al. (2008) explore what it is to be dyslexic, and whether it is possible to remediate some of the language deficits experienced by children with dyslexia. They cite a number of studies (Breier et al., 2001; Bryant and Bradley, 1985; Snowling 1987) which show “*the core deficit of these children* [with dyslexia] *to be phonological*” (p.383). Forgeard et al. (2008) also refer to Mody et al. (1997), who posited that dyslexia could be “*a deficit specific to the processing of speech sounds, rather than a general auditory processing problem*” (Forgeard et al.2008: 383). In a series of four studies, exploring either rhythmic or melodic discrimination in children, Forgeard et al. confirmed that a strong relationship exists between auditory musical discrimination abilities and language related skills in children. However, the children in Forgeard’s studies with dyslexia “*appear to have deficits in both pitch and rhythm processing … suggesting that children with dyslexia may have a more global form of musical impairment than has been appreciated in the literature so far*” (p.388). Note how the vocabulary and focus chosen for identifying and isolating any deficit that dyslexic children may have is here taken from that of music, demonstrating a potential close connection between music and reading, or at least “*between musical discrimination and phonemic awareness, which in turn predicts reading abilities*” (p.388). This strengthens the case for a hypothesis which explores connections between aspects of music and the decoding of text.

Thomson and Goswami (2008) take a neurological approach, but maintaining a focus on the link between language and music. Building on the work of Hill (2001) they note that “*most children with a diagnosis of specific language impairment also had a diagnosis of developmental coordination disorder, namely movement difficulties out of proportion with their general development and intelligence”* (Thomson and Goswami, 2008: 120)*.* This was demonstrated by the way in which 10-year-old children with dyslexia, and those with other specific language impairments, were less accurate when tapping in time with a metronome compared with children without impairments, particularly in their reaction times when the speed of the metronome changed. In a supplementary study, Corriveau and Goswami (2009) “*proposed that the* [dyslexic] *children might have subtle neural impairments that affected both motor and language impairments*” (in Thomson and Goswami, 2008: 120). Furthermore, “*Children who were particularly inconsistent in tapping to a particular rate showed the poorest literacy and phonological development*” (p.128).

Returning to auditory processing considerations, Hämäläinen et al. (2005), when exploring the wave-length features of phonemes, reveal that “*low sensitivity to amplitude modulated (AM) sounds is reported to be associated with dyslexia. An important aspect of amplitude modulation cycles are* [sic] *the rise and fall times within the sound*” (Hämäläinen et al., 2005:32). This sensitivity to the modulations of sounds is described by Goswami et al. (2002) and also by Thomson and Goswami (2008: 121) as “*amplitude envelope information in language learning*”, and they place particular importance on the successful readers’ perception of rise time within the onset of syllables.

Similarly Huss et al. (2011) speak of the “*metrical structure*” (p.674) of the organisation of the phonology of language, and they report that amongst the participating children in their study, those children identified with dyslexia were less successful in the perception of musical meter (glossary, p.245), than were children without dyslexia. Again drawing on the vocabulary of Thomson and Goswami (2008), Huss et al.’s study (2011) showed that “*individual differences in the perception of amplitude envelope rise time are linked to musical metrical sensitivity [and that] ... even the simplest metrical task was performed significantly more poorly by the children with dyslexia*” (p674).

Again, I emphasise that this study was not focused on dyslexia, but it is interesting to note that amongst studies exploring reasons behind perceived difficulties faced by children with dyslexia, and with other language impairments, the place and vocabulary of music is shown to be a rich seam, and is proving to be helpful in our understanding of processes of reading acquisition.

## 2.19 An analysis of three studies: Long (2007), Standley and Hughes (1997) and Gromko (2005)

My study, like those of Long (2007), Standley and Hughes (1997) and Gromko (2005), was built upon a quantitative foundation. Measurability was a central tenet, with the attendant notional corollary that development is, by definition, progressive, perhaps linear, or hierarchic, or spiral (Bruner, 1960). Rees (1994, appendix 4) offered a reading continuum, being the presentation of five identifiable reading ‘phases’ (role play; experimental; early; transitional; and independent) which show reading development to be (in Rees’ view) inherently linear. Similarly, the National Curriculum (DfEE, 1999) offered a rigidly linear series of level descriptors, not only to inform the assessment of children’s reading, but also to direct a teacher’s pedagogy of reading (the unspoken yet insistent message being that if level 2C is deemed to be achieved by a child, then, unquestionably, level 2B will be the next desirable destination, arrived at through the acquisition of a designated skill-set). This was relevant to my study because mine was a study attempting to show correlation and impact between one form of reading and another. Such an enterprise would be futile if there were no demonstrable parallel between known developmental aspects of reading acquisition in each.

Can such a linear approach to reading acquisition be assumed in both forms of reading? The level descriptors for music contained within the National Curriculum (DfEE, 1999) were not as well-defined as those for reading. Reading, as one designated area of learning amongst three within the English curriculum (the other two being Writing, and Speaking & Listening), was discretely described. By contrast, a series of single generic level descriptors applied to music, all aspects of National Curriculum Music being covered in each. The hierarchical level descriptors in all subjects demonstrated that the assumption was held within this National Curriculum document that new learning builds upon existing learning in a linear fashion.

However, despite a diligent search on my part, no equivalent to Rees’ (1994) continuum for the reading of music was to be found. In the light of the widely-held opinion (demonstrated in chapter 1) that it is inappropriate for children of Early Years or Primary age to learn to read and write music unless there is a compelling drive to do so on the part of the child (Mills, 2009; Hallam and Creech, 2010; Odam, 1995; Glover and Young, 1999), this is unsurprising. Practical graded instrumental examinations are readily available to those who want them (and it is worthy of note that these are not age-prescriptive. There is no expectation that one should be a grade 3 instrumentalist by a certain age, or even that if one has succeeded at grade 3, then the next automatic target is grade 4). Attempts have been made to describe musical

|  |
| --- |
| **Figure 2.3: Swanwick and Tillman’s (1986) sequence of early musical development** |
|  |
| Source: Swanwick (1988: 68) |

development, but these are generic. For example, Swanwick (1988: 68; Figure 2:3) developed a sequence of early music development, which he was at pains to present as a spiral, but there was no mention in his work of the reading of music reading or of music literacy. Indeed the developmental social aspects of music were presented as a central tenet (anticipating Standley and Hughes’ (1997) work (section 2.19.2). As such, I am unable to offer within this literature review an established delineated progression of skills, functions and features of identifiable progression in the reading of music.

### 2.19.1 Long (2007)

The first of the papers I have found to be sufficiently relevant to my field of study to warrant detailed analysis is Long (2007). Although the focus (the effect of a music intervention on the temporal organisation of reading skills) is not closely related to my own, the intervention was similar (being rhythm-based), and the study was undertaken as a thesis in successful pursuit of a doctorate. As such, I expected the bibliography to be exhaustive; the methodology tight; and I hoped her method of data analysis would be informative. Long’s work contained a variety of small-scale experiments (*n* = 24, glossary, p.245) with children aged 8-10, using targeted groups of identified ability, and she focused her attention upon the comprehension skills of these groups. Numeric data were produced through the interpretation of both reading-comprehension-age tests and reading accuracy tests (p.184), using the Neale Analysis of Reading Ability (p.231). Long found that the children involved in the interventions showed consistently improved progress compared with the control groups, especially those children who were identified as less able:

School children with below average baseline scores in reading made gains in comprehension but not reading accuracy following participation in the music intervention. Findings ... suggest that improvements in reading comprehension may be associated with improved reading fluency*.*

(Long, 2007: 250)

Long did not calculate an effect size (*d,* glossary, p.245) in any of the experiments, preferring to calculate statistical significance (*p,* glossary, p.246).

Long’s work informed my own thesis in so much as she employed methods which, on reflection, I would prefer to do differently. For example, Long’s use of reading-age scores as an assessment tool worried me, because it potentially invited unwelcome individual parental interest. What parent or carer would not want to know ‘how their child is doing’. Since, in my study, the parents had an awareness of the procedures (having given informed written consent), they therefore had a stake in the study, and would perhaps feel that they had a right to know the individual results. For example, of the parents consulted about the 2011 piloted Year 1 phonics screening check, “*99% wanted to receive information on their child’s performance on the Check... and 96% wanted information about what they could do to support their child’s phonic ability*” (Coldwell et al.2011: 2). Long did not disclose whether the school who allowed her access to the children ordinarily used reading-age tests as an assessment tool. Reading-age assessment devices are potentially contentious. Not only are they a blunt tool, but they also suggest norms to be achieved. Children therefore risk being labelled as under-achieving. This is a factor I was keen to avoid. I did not want to present my research as a normative indicator of national expectations, because potentially, laudable parental interest might lead to extraneous influences and additional variables. Instead I used an ipsative (glossary, p.245) assessment tool, specifically designed for the analysis I wanted to undertake.

Another feature of Long’s work that I have chosen not to replicate is her method of presenting series of mean scores using prose, making (at least for me) any immediate cursory visual interpretation impossible:

... a highly statistically significant effect was found in School ‘E’ (p < .001) with a mean pre-test score of 38.15, and 65.83 as a mean post-test score. School ‘F’ produced ... a mean pre-test score of 6.93, and a mean post-test score of 28.79.

(Long, 2007: 243)

I have presented summaries of my data using tables and figures, and have reserved prose for analysis.

### 2.19.2 Standley and Hughes (1997)

The second paper that I have found helpful to review is Standley and Hughes (1997). This study “*evaluated effects of music sessions designed to enhance the pre-reading ... skills of 24 children aged 4-5 ... using two groups of matched participants*” (p.79). The design was a series of two RCTs, the first focused on pre-writing, the second on pre-reading. Each produced pre- and post-test numeric data, and “*the result[s] demonstrated that music significantly enhanced print concepts*” (Standley and Hughes, 1997: 79). I refer to this study because it contained parallels to my own. It is of immediate interest that Standley and Hughes reiterated the paucity of literature in the field, stating that “*as yet, little research exists which focuses specifically on music therapy effects on oral and written language skills of children involved in Early Intervention programmes*” (p.80).

For the purposes of this thesis, Standley and Hughes’ (1997) opening statement needs unpicking a little. The opening paragraphs of their article make clear that the children involved in their study were part of a wider therapeutic intervention programme which sought to address serious behavioural problems in preschool children. As such, the additional intervention was designed to enhance pre-literacy skills in order to empower children to be able to function more effectively and socially within preschool settings. The participating children were working at a pre-phonics stage in their reading, although one of the four areas of pre-writing focus does make reference to the alphabet. The areas of reading on which Standley and Hughes’ study focused were assessed through a Print Concept Checklist adapted from Clay (1985), which included generic aspects of reading (recognising the front of a book, where to start reading, directionality of text, holding the book the right way up), and also more advanced mechanical aspects of text recognition (recognition of a capital letter, a sentence, a paragraph, how to read following the text with a finger).

Standley and Hughes (1997) are overt in their disclosure of the research design that the purpose of the study was “*to evaluate effects of concentrated music sessions ... which were specifically designed to enhance pre-reading and writing skills of children aged 3-5 who were enrolled in inclusive Early Intervention programmes*” (p.80). The music activities were undertaken in a print-rich environment, and included echo games (clapping, marching; vocal activities); echo reading; echo singing; phonics; use of music and reading props; and plenary discussion of activities. One might therefore expect such a specifically-targeted intervention to produce data demonstrating a positive outcome in relation to the null hypothesis, and this was the case. The rationale, the assessment tool, intervention, size and age focus of Standley and Hughes’ work were each therefore stylistically different from my own, but the overall methodology and anticipated outcomes were similar.

Their study informed my own in a variety of ways. Firstly, (in contrast to Gromko, 2005; see section 2.19.3) Standley and Hughes were very diligent in arranging closely-matched groups (in their case by cohort size, gender, age and race). I did the same (in my case by cohort size, age and reading ability). Secondly, they adopted the method of awarding one point for each instance in which a child offered a correct response to a symbol, stimulus or artefact, and awarding no mark for an incorrect response. This reduced the opportunity for subjective judgements being made by testers, and contributed substantially to the validity, replicability and uniformity of the assessments. Thirdly, their assessment tool produced data on a number of categories of aspects of reading, and these could therefore be independently analysed. Again, I copied this data collection design (though to an intentionally lesser extent). Fourthly, summaries of the data were presented clearly and unambiguously in chart and table form; calculations were made, and statistics presented formulaically.

These are positive aspects of Standley and Hughes’ (1997) research design, which I sought to replicate. However, I also observed features of their work which I consciously and assiduously avoided. First amongst these was the wide plurality of intervention activities. Even within their RCT which focused on reading, the activities which the children undertook were wide-ranging in task and structure, each activity seeking to advance generic aspects of the children’s pre-reading development. In such a design it is impossible to ascertain which aspect of the intervention influenced which aspect of reading. My research design used only two activities (the clapping of written sequential rhythms, and the learning of the recorder). In hindsight, even this was one activity-type too many, a feature of the research design which I shall critique in chapter 5 (section 5.7).

The second feature is the very wide plurality of assessment features. Standley and Hughes’ adaptation of Clay’s (1985) Print Concept Checklist assessed 25 separate skills or attributes of reading. This is too many to analyse, especially with a cohort of 24 children. In truth Standley and Hughes did not attempt to analyse each feature individually, but even if the features are clustered, the data categories were disparate, and may or may not have sat comfortably together. Either way, for the purposes of a doctoral thesis, this wide-ranging assessment tool would need justification. My assessment tool focused on only one aspect of reading (phonemic decoding) which lent itself easily to four categories of analysis (single-letter grapheme recognition; cluster and digraph recognition; non-word recognition; and a combination of all three). A fifth category emerged during the data analysis (see chapter 4, section 4.2).

### 2.19.3 Gromko (2005)

The third paper which I have found helpful, and indeed have chosen to adapt, is Gromko (2005), and I am indebted to her for providing theoretical and methodological starting points. The comments which follow constitute a severe critique of her work, but I do not make any critical remarks from a position of superiority, or from a desire to undermine her work. Gromko’s participant groupings, intended audience, and her purpose for writing, were each very different from my own. My purpose was to establish a methodology which would survive the rigour of a doctoral viva: Gromko’s was not. I have learnt a great deal about the garnering and presentation of data from this analysis of Gromko’s work, and for that I am grateful.

Gromko’s article is a quantitative piece, focusing on the effect of music instruction on phonemic awareness in beginning readers, and when I first read it I was dismayed to think that she had pre-empted my intended study. However an analysis of it revealed some aspects of procedure and method which I have chosen not to replicate. Gromko acknowledged the limitations of the research, and at no point did she claim her study was definitive. She did not claim causality; neither did she imply that anyone replicating the study would achieve the same results. In her discussion she offered alternative explanations for the results, so there was no attempt on her part to claim analysis of the data which was not warranted.

I will centre this critique of Gromko’s (2005) article on the four pillars of research - repeatability, replicability, reliability, and validity (Wellington, 2000: 32). I will demonstrate that in terms of both repeatability and replicability, Gromko is sound, but there are substantial methodological features which might leave the study open to criticism with regard to reliability and validity.

Gromko (2004, 2005) referred to Barnett and Ceci’s (2002) Near-Transfer Hypothesis (explained above, section 2.13) and attempted to demonstrate transfer of reading skills as a result of a music-based intervention. In Table 2.7 I reproduce Gromko’s (2005: 204) pre-test and post-test means for DIBELS **(**Dynamic Indicators of Basic Early Literacy Skills) subtests by group. The figures in brackets show standard deviations (glossary, p.246). The means are shown graphically in Figure 2.4.

**Table 2.7: Pre-test and post-test means for DIBELS subtests by group**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Letter naming fluency (LNF) | | Phoneme-segmentation fluency (PSF) | | Nonsense word fluency (NWF) | |
|  | Pre | Post | Pre | Post | Pre | Post |
| Treatment | 33.42 (15.48) | 42.63 (15.22) | 18.61 (16.26) | 44.72 (16.94) | 11.30 (17.09) | 21.16 (18.53) |
| Control | 36.27 (18.87) | 44.10 (15.63) | 25.83 (14.73) | 41.55 (14.50) | 20.61 (23.90) | 35.82 (18.93) |

Source: Gromko (2005)

**Figure 2.4: Pre-test and post-test means for DIBELS subtests by group**

|  |
| --- |
|  |
| Source: Adapted from Gromko (2005) |

Figure 2.4 visually demonstrates significant and unparalleled mean gains by children in the treatment group in the area of phoneme segmentation (glossary, p.245). The figure shows an interesting divergence of starting positions between the two groups. On each test the treatment group’s starting position is weaker than that of the control group. This is most starkly evidenced in the nonsense-word fluency results, and calls into question the equivalence of the groups at the starting point, a factor which I took pains to control. Gromko cannot be criticised for these variants in starting position, because hers was not a randomised controlled trial, but a two-group pre-test/post-test study using unmatched groups. As such, differences in starting position between the two available samples were almost inevitable.

Gromko did not calculate any effect sizes, perhaps because her groups were unmatched, but such calculations would have been helpful to her argument, since she does make reference to the effect size of a piece of research by the National Reading Panel (undated, in Gromko, 2005: 200). She identifies mean gains, which she, like Long (2007), presents in prose, but which I present here in Table 2.8. The mean gains are shown graphically in Figure 2.5.

**Table 2.8: Mean gains (and s.d’s) in Gromko’s study**

|  |  |  |
| --- | --- | --- |
|  | Treatment | Control |
| Letter Naming Fluency | 9.21 (10.38) | 7.83 (12.61) |
| Phoneme Segmentation Fluency | 26.12 (16.05) | 15.72 (13.83) |
| Nonsense Word Fluency | 9.86 (9.79) | 15.42 (19.60) |

Source: Gromko (2005: 205-6)

**Figure 2.5: Mean gains in Gromko’s study**

|  |
| --- |
|  |
| Source: Adapted from Gromko (2005: 205-6) |

Gromko considered that the mean gains in both Letter-Naming Fluency and Nonsense Word Fluency at the treatment school and at the control school are “*not significantly different*”, but that “*the mean gains in phoneme segmentation fluency [are] significantly different*” (p.205). The calculation of effect size might have been helpful to Gromko here. In my own research, the effect sizes of the intervention of these three aspects of reading development were calculated, and so, in order that my own research might be compared with Gromko’s, I calculated the effect sizes of Gromko’s study, based upon her figures. Formulae for calculating effect size are presented in chapter 4 (section 4.4.2). Table 2.9 contains the calculated effect sizes, using the formula employed in my own study, for reasons of consistency.

**Table 2.9: Effect sizes for Gromko’s study**

|  |  |
| --- | --- |
|  | Cohen’s *d* |
| Letter Naming Fluency | 0.09 |
| Phoneme Segmentation Fluency | 0.66 |
| Nonsense Word Fluency | -0.30 |

Source: Calculated from Gromko (2005: 205-6)

Gromko’s (2005) effect size values vary widely. The extent to which the findings from my research in these same three categories compared with those of Gromko is explored in chapter 4.

Gromko’s stated aim was to “*determine whether music instruction was related to significant gains in the development of young children’s phonemic awareness, particularly their phoneme-segmentation fluency*” (Gromko, 2005: 203, abstract), and within the confines of the rubric of her research she did that. The data showed that in a given situation, music instruction had a greater influence on phoneme-segmentation fluency than on letter-naming fluency, or nonsense-word fluency.

A second way in which my study differed from Gromko’s is that she collected no evidence of the improvement in participating children’s ability to read music as a result of the intervention. Gromko arranged for children to be taught standard notation, as did I, but her research made no attempt to gather data on aspects of music reading development. However, I was seeking to demonstrate transfer, and therefore I needed to provide evidence that skills in music reading were present and developing, from which decoding skills might be transferred. I provided this evidence through analysis of the pieces of music the participating children were asked to learn (chapter 3, section 3.14) and also through interviews with research assistants (chapter 4, section 4.29ff).

Clarification is needed here, however. I was looking to demonstrate whether or not learning to read music during developmental stages of learning to read text enhances children’s ability to decode texts. I was not trying to show whether children who have a higher (or lower) than average propensity to successfully learn to read music also have a higher (or lower) than average propensity to successfully learn to decode text. Although that in itself would make an interesting study, such an approach would have constituted an unhelpful aside from my own narrowly chosen focus. Mine was a cohort study, measured in mean scores, not individual ones.

Gromko did not claim that her research design was exhaustively robust. She wrote: *“Although readers should interpret these results with the limitations of this study in mind, I believe that these results have sufficient credibility to serve as the basis for continued enquiry into near-transfer effects of music instruction”* (Gromko, 2005: 205). My study accepted that invitation. Mine was an attempt to take Gromko’s aims and focus, and endow them with a tighter and more robust design which contained fewer variables, in order to provide continued enquiry into the near-transfer hypothesis to which she subscribed.

# Chapter 3 Methods and Methodology

## 3.1 Introduction

The methods chosen for this study were informed by preparatory work undertaken during Part 1 of the EdD programme. Not all of the six assignments which I produced during the first stage of the programme had this final study directly in mind, but three did, and it is useful to here identify the ways in which this study was underpinned by these three pieces of previous work.

The first relevant submission explored whether any knowledge can be value-free. This question is particularly pertinent when considering a methodology which includes quantitative paradigms, because quantitative methods tend to celebrate experimental research: that is, they attempt to demonstrate that *if x, then y*. Although this study has at its core a randomised controlled trial, and no apology is made for that, this first submission taught me that there is a degree of subjectivity in all designs of research, and as a result it may be considered impossible to ‘prove’ anything. This first prior submission also taught me that a piece of research which has such a linear approach, and which makes use of exclusively numeric data, can, if one is not careful, be insensitive to the experience of participants, because participants might be seen as providers of data rather than partners in an enterprise. This is an accusation I was keen to avoid in this study, hence my chosen use of mixed methods (glossary, p.245) from both a design point of view (section 3.2, below), and also from an ethical perspective (section 3.3).

A second submission, entitled ‘What types of research question lend themselves to condensed or longitudinal studies?’, was written exclusively with the needs of this study in mind, and was an attempt to identify both how long, chronologically, the intervention of this study might need to be, and also whether the length of a study had an implication on the title, and hence the focus, of that study. This assignment drew most strongly on the work of Brice-Heath (1983) and Hannon (1987) as examples of longitudinal studies, and Tobin and McInnes (2008) as an example of a condensed study. The latter was particularly helpful, as Tobin and McInnes showed how, with effective use of mixed but complementary methods, “compelling evidence” (p.3) could be procured, the quantitative data being analysed against a backdrop of insight gleaned from qualitative approaches. I sought to replicate the ‘feel’ of Tobin and McInnes’ work, and believe I have done so. With regard to the phraseology of the research question however, my analysis of these three pieces of research, together with a number of other longitudinal and condensed studies, showed no reliable pattern.

The third submission was the research proposal, which of course gave an opportunity to explore in some detail the method, length, size, scope, ethics and feasibility of the final intended study. This was a helpful opportunity to define and defend the enterprise, and the successful negotiation of this proposal gave me confidence to consider that what was to come was a worthwhile and properly considered study. The intended title at that time was ‘How do reading skills pertaining to the decoding of text compare with those pertaining to the decoding of music notation?’, and although, through the writing of this proposal, the title and focus evolved and narrowed, as did the method to some extent, the proposal anticipated the vast majority of what was outplayed in the study itself.

## 3.2 Design

The research design used four distinct methods. Although independent, each is interconnected in an attempt to recreate the levels of methodological synthesis achieved by Tobin and McInnes (2008). Figure 3.1 identifies the interconnectedness of the four approaches.

**Figure 3.1: The four selected methods**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | The RCT, designed to measure degrees of fluency in the recognition of diverse phonemic stimuli at given assessment points. |  | Interviews with research assistants, designed to identify the presence of improved fluency in the recognition of music-reading, as a result of the intervention. | |  |  |  | |  | The research question |  | |  |  |  | | A comparison between the participating children’s SATs results in reading, with those of the previous year, to give additional evidence of an effect of a possible intervention. |  | Interviews with children, designed to identify children’s perspectives of learning to read, and of learning to read music, and also to confirm that participation in the study was a positive experience for them, conducted ethically. | |

Firstly, a matched pairs randomised controlled trial (RCT), was devised, in order to provide transparent and measurable data to confirm or not confirm the hypothesis. There are a number of designs within the family of RCTs. Some closely match the groups at pre-test; some (often large-scale) do not; and some do not use a pre-test at all (Torgerson et al., 2006). Since mine was a small-scale study, I chose to match the groups as closely as possible in order to make the pre-test scores and the literacy teaching as equal as possible at pupil level, rendering the identification of effect size transparent. This RCT mirrored the focus, but not the design, of Gromko (2005), and constituted the controlling aim of the research design.

The research was led by a hypothesis that there is a possible correlation between children’s ability to read music and their ability to decode text during the early stages of reading development - hence the research question ‘Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text?’ As such, the experiment was “*confirmatory* [in that it sought] *to support, or not to support, a null hypothesis*” (Cohen et al., 2007: 272). I was mindful that for an experiment to be successful, “*the independent variables are isolated and controlled carefully*” (Cohen et al., 2007: 272).

Secondly, analysis was made of the host school’s results of related internal tracking tests in reading, which gave an indication of the children’s termly progress, measured as National Curriculum levels. In this section, I made comparison between the scores of the 55 Year 1 children who participated in the intervention, and the scores of the 53 Year 1 children in the previous year.

Thirdly, to contextualise the research, I conducted a series of semi-structured interviews with a large proportion of the participating children, in order to see whether their experiences of engaging in the project could be articulated in order to contextualise, confirm or counter any patterns found in the data.

Fourthly, I conducted semi-structured interviews with three of the five research assistants who taught the recorder to the children, in order to ascertain whether, in their opinion, from their experience of the teaching, they were aware of whether degrees of progress had been made by the children in the reading of music.

I took these four disparate approaches because any writing on the subject of research design emphasises the importance of either triangulation or mixed methods, which are not necessarily the same thing. Triangulation is reductive, and attempts to arrive at a single conclusion, calculation or position from a plurality of angles or approaches. The results of a research study which has successful triangulation procedures, particularly if that research is quantitative in nature, tend to be tidy, unambiguous and narrow, and answer specific (often closed) questions. By contrast, a research study which employs mixed methods does so in an attempt to “*to produce a fully rounded, more authentic, portrayal of the social world*” (Coffey and Atkinson, 1996: 14). The use of mixed methods can reveal not only whether or not an effect has taken place, but also what the implications of that effect might be for a participant or participants. Mixed methods can reveal, for example, what the experience of engaging in the production of that effect was like, and whether or not participants would be happy for that experience to be repeated or to continue. “*The more we examine our data from different viewpoints, the more we may reveal – or construct – their complexity*” (Lankshear and Knobel, 2004: 364). Triangulation focuses a question in order to reveal or prove a truth or a phenomenon. Mixed methods explore a question from a variety of approaches, perhaps in turn leaving some questions partially unanswered, or even revealing new, previously unexpected questions.

Each of the four methods I undertook addresses one aspect or another of the title question ‘Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text?’, and all the data that I collected to inform an exploration of that question were acquired during the 2010-11 school year (although a small proportion of those data related to the previous 2009-10 school year).

## 3.3 Ethics

There is more to ethics in a piece of research than just the securing of ethical approval (appendix 29). This section gives examples of some of the decisions, attitudes, procedures and policies I adopted prior to, and during, the study, in order to make overt that it was planned and conducted in an ethical fashion.

I did not consider the participant children to be merely a source of data, but wanted them, and the research assistants, to feel that they were partners in an enterprise. To this end I ensured that the children knew the following:

* what the project was about, explicitly - how it was to be conducted, and for how long
* that participation was not compulsory
* that they had to opt in (appendices 6 and 7)
* that they had the right to withdraw, and without giving a reason (appendix 5)
* that the data would be kept anonymously, and securely

I was careful to conduct the study in an ethical fashion. For example, I visited the school immediately prior to the study, introducing myself and the research assistants to the children, in order to tell the children about the project, and to put the children at ease. Secondly, I sent a reminder to four children who had not brought in permission slips, but although it would have been in my interests to pursue these children diligently, I did not badger them beyond that reminder. Their decision not to participate was respected. In spite of that, the research assistants and host teachers ensured that these children did not feel isolated on the testing days, and they were allowed to participate in the tests if they wanted to, even though their data was not used. All did want to.

Furthermore, the testing procedures, outlined in section 3.17, demonstrate that we had the children’s self-esteem at heart at all times. No child was coerced or forced to take the tests, or to go further in a test than they wanted to.

The post-intervention interviews that were conducted with the children (section 4.23ff) showed that the children felt overwhelmingly positive about their experiences, had enjoyed the project, including the testing. These responses suggest that they felt safe with me, and the research assistants, and with the procedures of the programme.

## 3.4 The selection of school

I was successful in securing the cooperation of a two-form-entry school in North West Kent, from which the head teacher and two Year 1 staff (who shall be referred to respectively as Mr Barker, Mr Simms and Mrs Barrs) enthusiastically responded to my invitation to participate. This school was chosen for three reasons. Firstly, it is a school with which I have maintained close professional links. I worked there as a teacher between 1991 and 2002, and was on the board of governors for some years during that period. I continue to have professional links with the school, as they host many of the student teachers from the university at which I work. Secondly, it is a school which holds a commitment to the teaching of music. All children learn the recorder in Year 2, and all children are offered subsidised instrumental tuition (trumpet, flute or violin) for one academic year during their time in Key Stage 2. Thirdly, it is a school of diverse ethnic intake, a factor which I considered would provide the research with participants representing a wide range of early reading experience and ability. The immediate catchment area is a formerly council-owned estate, now generally privately-owned, but no means affluent. The school also attracts children from a new-build estate, in a commuter-belt area between one and three miles from the school. 13.9% of children were eligible for, and were claiming, free school meals in January 2011, which compares with the 2011 national average of 19.2% (DfE, 2011b, table 3b).

I had considered undertaking the research programme in more than one school, and might have done so, but I was dissuaded from this because, although such an enterprise would have perhaps doubled the number of participants, it would not have enhanced the validity of the data by the same proportion, because this approach would have introduced a number of additional variables, and I would not have been comparing like with like between each project. In effect, I would have been doing the programme twice, rather than doubling the number of participants.

I presented the intended study to Mr Barker, Mr Simms and Mrs Barrs, and in discussion with them (together with five research assistants drawn from the student body of the university in which I work), agreed the research programme.

## 3.5 Selection of age range

I chose to conduct the research with Year 1 children, for both educational and logistic reasons. Kalof et al. (2008) refer to such a predetermined approach as “*purposive sampling [whereby] the theory, research questions and data analysis processes determine the sample*” (Kalof et al., 2008: 44). The choice of Year 1 children was taken consciously because I wanted to focus upon participants who were in specific formative stages of their reading. The time of year was therefore important. Had my opportunity for research fallen at the end of the academic year, I would have chosen a Reception class. 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 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’, ‘phonically’ must be meant).

These developmental skills tied in closely with the focus of my research. By contrast, the expectation for children by the end of Year 1 (for whom phase six of Letters and Sounds (DfES, 2007b) 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, 2007b: 168)

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

## 3.6 Enlisting participants

I introduced the project to all 60 Year 1 children in the school at the very beginning of the academic year by visiting the school, playing the recorder to them, and explaining what I intended to do, and why I wanted to do it. I was careful at this early stage to avoid explaining to the children any features of the reading of music, but I did show them a very simple manuscript (crochet rest quaver crochet) so that they were able to make an informed decision about what participation would involve. A pamphlet explaining the project was also given to their parents/carers (appendix 5), and permission was sought from them (appendix 6). A simplified version of the permission slip was also designed for the children (appendix 7).

## 3.7 Pairing and randomisation methods

Half the children from each class were to have recorder lessons during the first half of the Autumn term in the school year 2010-11, and the other half during the second half of that term. An identical method of allocation was employed for each of the two classes. The children were organised into ‘strata’, which Kolaf et al. (2008: 207) describe simply as “*groupings of a population of interest*”, based upon the reading ability groups with which the host teachers had furnished me. From these strata I organised the children into pairs. The pairings were arranged from lists given to me by the host teachers who, from information taken from end-of-Foundation Stage assessments had both allocated each of the children into four reading groups (lower; lower middle; upper middle; and higher). These groups each helpfully held an even number of children, and so I simply paired the children in the order in which they appeared on my lists. As such, a child designated in the least-able group was paired with another in the same class from the same group. The same was true for children from those deemed to be the most able. For each pair I threw a die (actually, a virtual die from a software programme), and if it came down as odd, the first-named of the pair was placed in group 1, and the second in group 2. If the die came down even, the opposite allocation was made. In this way the two groups would have a similar range of reading ability, but the allocation of the children within pairs was randomised.

I had considered pairing the children according to the data gleaned from my initial test results, and not from the groupings given to me by the host teachers, but I decided against this because the teaching which these children would receive would be most similar, at least in the early stages of their Year 1 education, in the groups that the host teacher had allocated. I also considered it rather arrogant to suggest that my baseline data, gleaned as they were from a single afternoon’s testing of children I had previously met only briefly, and which focused on a very narrow decoding aspect of early reading, should be more valid, informed or compelling than those groupings given to me by the host teachers.

Table 3:1 shows the pairings of children, contained within their classes, and the given recorder groups.

**Table 3.1: Original matched pairs of children within classes and recorder groups**

Class 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Recorder group 1 | 10 | 2 | 4 | 5 | 6 | 12 | 25 | 13 | 14 | 15 | 19 | 22 | 29 | 27 | 23 |
| Recorder group 2 | 1 | 3 | 16 | 18 | 11 | 7 | 8 | 9 | 20 | 17 | 21 | 24 | 26 | 28 | 30 |

Class 2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Recorder group 1 | 36 | 32 | 33 | 34 | 53 | 42 | 38 | 44 | 59 | 46 | 50 | 49 | 60 | 54 | 57 |
| Recorder group 2 | 31 | 41 | 51 | 45 | 35 | 37 | 43 | 39 | 40 | 56 | 47 | 48 | 52 | 55 | 58 |

I had intended to include all the Year 1 children in the RCT, but permission slips were not received from five of them. The children for whom permission slips were not received were numbers 12, 32, 37, 55 and 60. As such, in the interests of keeping the groups of identical construct, and also from an ethical perspective, the use of the data of children with whom these five were paired (7, 41, 42, 54 and 52) were not allowable, and did not contribute to the RCT in any way. The number of children in the RCT was therefore 50. However, in the over-time comparison with the previous year group I was able to use data from the five children whose pairs were lost through non-receipt of permission slips; thus the number of children from this year group in that comparison was 55.

## 3.8 Absentees at pre-test

A complication arose from two children (29 and 30) who presented permission slips, but who were absent at the pre-testing. Clearly, from an ethical point of view they were eligible to be included, but their missing data were somewhat problematic. Oppenheim’s (1992) observation that the best way of dealing with missing data is to not have any is intentionally ironic, but unhelpful. Kolaf et al. (2008: 66) acknowledge that “*missing data are a constant issue in research*”. However, most writing on the subject of missing data in social research is based on the problem of incomplete questionnaires or interview answers as a result of participants’ reluctance, refusal or inability to answer one or more question amongst several others. Gray (2009: 456) lists four categories of missing data (‘*not applicable’; ‘refused’; ‘did not know’;* and *‘forgot to answer’*), none of which address the situation in my research which arose as a result of these two children (both from one class) being absent during the pre-intervention testing. Changing the pairings so that they were together was out of the question, since the pairing of the children was done in advance, and the validity of the experiment would be lost if these pairings were changed.

Gyimah (2001) was helpful. He makes the distinction between “*case missing (or unit non-response scenarios) ... and item missing scenarios*” (p.3) – the former being where “*a unit is eligible but no response is obtained, emanat[ing] from inaccessibility*” (p.3), and the second being a missed item (for example a response to an individual question) amongst a successfully-gleaned series of other items. The distinction is necessary, because different techniques are used to deal with each. In my research then, I had cases missing, for which eventuality Gyimah (2001) notes that “[case-missing or] *case non-response is usually rectified by population or sample weight adjustments where [participants] are differently weighted to retain the overall sample fraction*” (p.4), in other words, to attribute to the missing cases the mean of the data from other participating cases. This is strictly “*only applicable to monotone patterns of data*” (Little and Rubin, 1989: 295), or groupings that internally are homogenously constituted, i.e. the members of a group are similar in terms of aspects which might affect significant variables. In the case of my research, this might apply closely to children’s prior ability in grapheme recognition and blending skills, but could also include such reported features as age and gender.

Gyimah demonstrates, through citing a range of examples (Vach, 1994; Dodge, 1985; Madow et al., 1983), that “*by correcting for missing data, we significantly increase the internal and external validity of our findings*” (Gyimah, 2001: 3). It is therefore better, in Gyimah’s view, to make active adjustments for the missing data, than either to analyse the remaining data regardless of the omissions (a course of action which, in my research, was not an option statistically, because this would have made the group sizes unbalanced, and have automatically skewed the findings), or to delete the missing case(s) and any associated paired cases, which will deny the researcher otherwise valid and usable data from the member of the paired case(s) who did participate.

Duffy (2006), a writer in the field of quantitative medical research, highlights that the underlying reasons for data being missing are important in determining what to do about them. She suggests that a random example of non-participation (due for example to absence, death or other cause of attrition) is less problematic than a systematic example of non-participation which is intentional on the part of the participant, and perhaps selective. She writes “*A random pattern refers to values missing accidentally*” (Duffy, 2006: 274). There was no active non-participation from the two children who missed the pre-test – they merely were on holiday at the beginning of term when we started the project, and they signed up enthusiastically on their return, which was before the first weekly session of the intervention. Duffy confirms that “*if a large amount of data is missing from a small or moderately sized dataset ...*” (by which she identifies a proportion of >5%) “...*serious problems can arise unless the researcher takes steps to handle the problem ... often by imputation*” (p.273). This echoes Gyimah’s call to be active in addressing the situation, rather than being reductive.

Duffy (2006) goes on to define imputation as “*the process of estimating missing data based on valid values of other variables or cases in the sample*” (p.274), and she holds that common methods of imputation include “*prior knowledge ...*” (in which the researcher, knowing the field or the participant well, makes an educated guess at the missing figure) and “... *mean or median substitution*” in which the mean for the sample is applied. “*This* [latter] *commonly-done procedure is considered a conservative approach because the distribution mean as a whole does not change and the researcher does not have to guess at missing values*” (p.274). Tabachnick and Fidell (2007) prefer, if the research design allows, the imputation of a group mean rather than the total sample mean, because such an approach maintains the integrity of the data against which variables are to be measured and analysed. This last approach was entirely feasible in my own research situation, since the class means were easily identified, and so this is the approach I took. The imputed values are therefore notional, but it is better to include plausible and neutral notional scores, than to lose actual post-test data, together with the actual pre- and post-test data of the children with whom these two were randomly paired. Calculations of imputed scores for children 29 and 30 are presented in appendix 8.

## 3.9 The phases of the study

The quantitative element of the study was originally intended to have three phases. The first two, using Cohen et al.’s (2007: 278-282) diagrammatic style, are represented in table 3.2

**Table 3.2 Phases 1 and 2 of the study**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phase 1 Phase 2  (RCT) (follow-up) | | | | |
| Time 1 |  | Time 2 |  | Time 3 |
| RO1 | Χ | O2 |  | O3 |
| RO4 |  | O5 | χ | O6 |

The graphic conventions used by Cohen et al. in table 3.2 follow those from Campbell and Stanley (1963), in that horizontal entries refer to the same persons; left to right order signifies temporal sequence; vertical entries are synchronous; R denotes randomization; O is a test or measurement event; and χ denotes an intervention.

The RCT (phase 1) analysis used data drawn from the pre-test and the first post-test, but there was an additional round of testing of all available participating children after the second intervention (phase 2, follow-up, *n* = 42). These additional data cannot be said to been gleaned from a second matched pairs RCT, since the starting positions of the two groups of children at the beginning of the second intervention were different from each other, and so the groups were no longer closely matched.

In this respect, this second phase of my study was more closely comparable to that of Gromko (2005) than was the first (see section 3.17, below). The variables were still fewer than in her study, because I was able to retain parity of group size, and the literacy and music tuition input that the paired children received remained closely matched, but the two phases cannot be considered comparable, not only because the paired children were no longer closely matched in terms of reading ability, but also because, towards the end of the study, severe weather conditions forced the school to close for two consecutive Tuesdays, reducing the taught element by a third. There was also a high proportion of absenteeism during the early weeks of phase 2, which was not the case in phase 1 (which enjoyed an attendance rate of over 96%, and no child missed more than one session). Additionally, the pieces that were composed on a weekly basis during phase 1 were tailored to the needs and progress of the first intervention group (section 3.14). Although the pieces were the same for the second intervention group, this tailoring was not present.

Consequently, although it was important for ethical reasons to undertake the second phase of this study, the quantitative data it produced lacked the validity of the first, and so are not analysed in chapter 4. The qualitative data gleaned from interviews with the children during this second phase were still used, as insightful data were produced through the children’s articulation of their experiences and preferences of learning to read music. There was no methodological tension here, since the two data types were not mutually dependent. Valid qualitative data have been kept. Invalid quantitative data have been rejected. Methodologically, this is only responsible approach.

The third phase compared data drawn not from the intervention, but from the school’s termly SATs (glossary, p.246) records, in which I compared the progress of the participant children during the intervention period with that of the previous year’s cohort for the equivalent period.

The fourth and fifth phases were qualitative in nature, being interpretations of interviews undertaken with participant children (phase 4) and with research assistants (phase 5) immediately after the intervention. The numbers of children participating in the various aspects of the study are shown in Figure 3.2.

**Figure 3.2: The phases of the study**

|  |
| --- |
| Phase 1, The RCT  Original intention, n = 60 children (full sample)  30 Placed into matched pairs 30  55 permission slips received, 5 permission slips not received  Attrition = 10 children (the 5 without permission, and their pairs)  25 n = 50 25 |
| Phase 2, The follow-up  Four children were absent during the last round of testing  Therefore, attrition from phase 1, = eight children (these four, and their pairs)  21 n = 42 21 |
| Phase 3, The comparison of the intervention group’s Year 1 SAT results  with those of the previous year’s Year 1 cohort  Original intention, n = 113 children (full sample, 60 intervention + 53 from the previous year)   |  |  |  | | --- | --- | --- | | 55 permission slips received,  5 permission slips not received |  | Attrition = 5 children |   n = 108 |
| Phase 4, Interviews with the children  Original intention, n = 60   |  |  |  | | --- | --- | --- | | 55 permission slips received,  5 permission slips not received |  | Attrition = 5 children |   All 25 children in the first intervention were interviewed in groups after phase 1.  11 children in the second intervention were interviewed individually after phase 2  n = 36 |
| Phase 5, Interviews with research assistants  At the end of the intervention, three of the five research assistants were willing and available to be interviewed about the children’s progress in reading music, clapping rhythms, playing the instrument, and the children’s attitudes to the intervention and to the testing. |

## Phases 1 and 2. The Randomised Controlled Trial, and follow-up

## 3.10 Power calculation

A power calculation is “*an estimate of the ability of the test to separate the effect size from random variation*” (Gorard, 2001: 14). Ideally, such a calculation is made in advance of an RCT, to establish either how many participants would be needed in an RCT to have a given probability (usually 80%) of detecting a given effect size (say 0.5) with a confidence interval of (usually) 95%; for these values it has been established that the required sample size is a minimum of 64 participants in each arm.

However, in educational research it is often the case that the sample size is predetermined for practical reasons, as in my study. In these circumstances a power calculation can still be performed, but retrospectively, in order to determine the achieved power of the experiment and (by carrying out a sample size calculation) compare its sample size with the sample size that would have been required to achieve the same power level with statistical significance.

The statistical power of a given investigation is inextricably linked with statistical significance, the effect size and the sample size of the investigation. Indeed, these four “*are so related that any one of them is a function of the other three*” (Cohen, 1988: 14). A variety of online software packages are available which will calculate the statistical power of a research design. I used the statistical power calculator developed by Schoenfeld (2012) which demonstrated that given a statistical significance of 0.05; together with the post-test mean of the control group as a percentile of that of the intervention group (58%), and with the pooled post-test standard deviation (16.93), the study was shown to have only a 27% probability of achieving statistical significance (appendix 9), well below the desirable power level of 80%. Thus, the study was unlikely to have achieved statistical significance. By changing the factors imputed (appendix 10), Schoenfeld’s calculator showed that for the study to achieve a power level of 80%, 278 participants would have been required (139 in each arm), which was over five times the size of my available sample. Not only was this significantly beyond the resources I had at my disposal, but such a research design would require a nine-form entry Primary school (of which there is not one example in England), and if the study were to be conducted across several schools, the required number would be even higher as a result of this additional variable. It must be acknowledged therefore that due to its size, the power of the study is very low, but it was beyond my power to organise it otherwise.

This is not to say a study which lacks statistical significance lacks worth, or is discredited. Where an effect has been demonstrated, a statistically non-significant study contributes to the field in which it focuses in two ways. Firstly, it shows that the focus is valid, with potential for statistical significance if replicated with more ambitious parameters, and is therefore worthy of further investigation. Secondly, it contributes to existing data from similar studies, allowing “*research syntheses*” (Hodges, 1984: 61), which is the combination of the results of multiple independent research studies, none of which might achieve statistical significance, but combined, may reveal patterns and consistent outcomes. Such research synthesis is now known as ‘*meta-analysis’* (Torgerson, 2003: 73). If meta-analyses are to be successful and valid, the selected trials must be homogenous in terms of method, focus, and size (Torgerson, 2003: 82), but if these conditions do apply, meta-analysis can allow the identification and validation of small, and possibly educationally important effect sizes. Hodges was also an advocate of synthesising studies of similar focus and quality. When selecting studies for inclusion in such a synthesis, he warned against censoring non-significant studies:

If the quantitative result of a study is [recognised] only when the mean difference is statistically significant, the observed mean difference, variance, and effect size are biased estimators of the corresponding population parameters.

(Hodges, 1984: 61)

More recently, Gerber and Malhotra (2009) examined the range of articles published in two leading science journals seeking evidence “*for the presence of publication bias due to reliance on the 0.05 significance level*” (p.313). They expressed the same view as Hodges, but were more forthright:

Our findings raise the possibility that the results reported in the leading political science journals may be misleading due to publication bias ... [which has] been found to be ubiquitous in several academic disciplines.

(Gerber and Malhotra, 2009: 313 & 314)

They go on to celebrate the place of the smaller, less powerful (yet effectively designed) study:

Publication bias might be avoided if greater attention was given to the research design rather than the estimates produced by a particular study. Since an innovation in research design can improve all subsequent work, there should be strong encouragement for innovative studies that show how a causal effect or important quantity can be estimated in a convincing fashion. For such studies, it may sometimes be the case that data is scarce or measurement difficult and, consequently, the power of the particular study is weak. Nevertheless, since a strong design can point the way to subsequent innovations, there are good reasons to publish the work even if the results fall short of guidelines conventionally employed in evaluating hypotheses.

(Gerber and Malhotra, 2009: 323)

This celebration of the small but effectively designed study is welcome in my case, since I have been careful to undertake a thorough audit (chapter 4, section 4.16). Nonetheless a power level of only 27% must be acknowledged. The implications of this power calculation for the statistical findings are considered in chapter 4.

## 3.11 The size of the research programme

In the light of this calculation, the rationale for my choice of sample size must be explored. I was originally influenced by a range of methodologists. My trawl of books and articles on quantitative methods and methodology (Brown and Dowling, 1998; Hoy, 2010; Kalof et al, 2008; Gray, 2009) largely revealed silence on the matter of sample size. Each of these writers was informative about types of sample and population, but none offered any insight as to the numeric size that a piece of research should be, suggesting (falsely, I suspected) that the question is either unanswerable, or irrelevant.

Torgerson et al. (2003) and Cohen et al. (2007) are specific. Torgerson et al. (2003) analysed a number of RCTs, critiquing each methodologically. They ask many auditing questions (which were each applied to my study, see chapter 4, section 4.16), but they also give consideration of cohort size, both in terms of sample size, and the justification of sample size. The identification of sample size is important because if a given intervention produces a large effect size, this could be statistically non-significant because of a small sample. Equally, a large sample which produces a small effect can be statistically significant.

Attrition rates are also given weight by Torgerson et al., but this was not relevant to my study as none of the participating children who started the programme were lost during it.

The sample size in my RCT was 50. The sample sizes analysed in Torgerson et al.’s review varied between 24 and 1,115. Many contained fewer than 70. Of a sample size of 24, Torgerson et al. (2003: 243) write:

Clearly this sample size is low and would not have had very much statistical power to detect even one standard deviation difference in improvement (which requires at least 32 participants in a two-group comparison).

Cohen et al (2007: 101) confirm this figure:

A sample size of thirty is held by many to be the minimum number of cases if researchers plan to use some form of statistical analysis in their data … Typically an anticipated minimum of thirty cases per variable should be used as a ‘rule of thumb’.

In my research design, I was very careful to ensure a single variable (namely, the intervention of the recorder lessons). Group size, ability spread, pre-test mean scores, pair-matching, intervention procedures, testing procedures, and the quality, quantity and level of literacy teaching were all consistent for each pair of children for the duration of the RCT. As such, there was only one identifiable independent variable (the intervention itself), and so my participant number exceeded the minimum required in terms of Cohen et al.’s (2007) advice.

Hutchison and Styles (2010) consider that the use of baseline data can also reduce a sample size, whilst retaining statistical integrity. “*Since the use of baseline data can make our analysis more sensitive, it is possible that it can reduce the sample size required for the experiment*” (Hutchison and Styles, 2010: 28).

I acknowledge that I was very much working at the bottom end of what is allowable in terms of sample size in an RCT, but to do otherwise was not within my power or budget.

## 3.12 Research assistants

Ten research assistants were enlisted to assist with the study, five to do the pre-testing and the teaching, and a separate group of five to do the post-testing. The first group of five were all final-year students who, as part of their Initial Teacher Training (ITT) programme, were required to produce a research project as their dissertation. Permission was gained, from the Dean of the School of Education at the university in which I work, for the students to have access to the anonymised raw data gleaned from the two phases of the RCT for use in their dissertations. As such, the five students were pleased to participate, as they were attracted by being involved with a project far bigger than they could have arranged independently.

This did not violate any ethical protocols, because my introductory pamphlet to parents/carers (appendix 5) was careful to identify that “*The only people who will have access to the results which have your name on (known as ‘the raw data’) will be Mr Barker, Mrs Barrs, Mr Simms, and me. Once I have removed all of your names on the data, other researchers whom I work with will be interested to know the results.”*

It was important that the intervention was delivered in a uniform way by all research assistants. The procedures and protocols of the lessons are presented in section 3.13, and those for the testing in section 3.17. These procedures and protocols were agreed and rehearsed by everyone involved in delivering the programme (including the host teachers) prior to, and throughout, the intervention. Prior to the intervention, this was done through three meetings, clearly setting out aims, objectives and methods. During the intervention period, the assistants and I would meet on a weekly basis (usually, but not always, in the school on the afternoons of the intervention) to monitor the progress of the intervention. At these meetings, individual children were not discussed, but everyone was given opportunities to articulate their experiences and any concerns, which gave a sense of confidence, unity and direction to the intervention. It was clear that throughout the programme, everyone involved knew what was happening, what was expected, and what the procedural norms were.

The post-testers (five in number) were friends of the original five research assistants, and were also students on the ITT programme. They were enlisted without enticement or inducement, but were willing to assist, under my direction (see testing procedure 3.16, below). Again, there was a great deal of rehearsal, and clarification of procedures and protocols, before the testing events, and although I did not participate in any of the post-testing, I was present in case clarification of any part of the process was sought by the testers.

My own involvement in the teaching and testing was intentionally limited. I assisted with the pre-testing, but did not involve myself in post-testing in any capacity. I also was required to assist in two of the teaching afternoons during phase 2 (the data for which were not used), on both of which occasions research assistant numbers were short due to diverse taught commitments at the university.

In order to maintain as much validity for the programme as possible, I am at pains to report that the post-testing of the children was administered ‘blind’: that is, by a different body of research assistants from those who did the teaching. I did not want the people who were involved in the post-testing to have any information about the decoding ability of the children, or to have knowledge of which children had been involved in which intervention. Also, I did not want any of the post-testers to have developed any sort of pedagogic or emotional relationship with any of the children. This is important because:

The relationship (positive to negative) between the assessor and the testee exerts an influence on the assessment. This takes on increasing significance in teacher assessment where the students know the teachers personally and professionally – and vice versa – and where the assessment situation involves face-to-face contact between teacher and student.

(Cohen et al., 2007: 160)

## 3.13 The format and content of the music teaching

The children undergoing the intervention were taught on a half-class-cohort basis (i.e. each afternoon of the intervention, 15 of the 30 children from class 1 would receive their music lessons, followed by 15 of the 30 children from class 2). The lessons took place in the school canteen on six consecutive Tuesday afternoons. The research assistants were unable to begin the recorder lessons until the canteen had been swept and cleared (by 1.15). Playtime was at 2pm for 15 minutes, and, with regard to the second group, the host teacher wanted the children back in the classroom by 3pm in order to prepare for home time. Thus, in order to ensure that the lessons were of uniform length the children were collected from their classrooms at either 1.15 or 2.15, and were returned at either 2pm or 3pm. The canteen was at the opposite end of the school from the children’s classroom, and there was inevitably some time spent collecting recorders and getting organised, so the lessons themselves began at 25 minutes past the hour, and finished at 5 minutes to the hour, allowing a lesson time of 30 minutes. These timings were consistent throughout the intervention.

The total ‘contact’ teaching time for the intervention was therefore 6 x 30 minutes = 3 hours. However, the children were practising, and talking to their peers and parents/carers, throughout the six weeks of the programme.

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. The combinations began with five crotchets, which would be clapped repeatedly, each repeat interspersed with a pause of three beats so that a feeling of four beats in a bar was established. Then, one by one, crotchets would be replaced by rests or quavers, and then with rests and quavers.

Following this activity, the children would clap 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 would be told that this was the rhythm of the piece. They would then learn to sing or chant the new piece (the title of the first four pieces reflected the rhythm), and then chant and clap the piece at the same time. The piece was therefore heavily reinforced before the children would play the recorders. This section of the lesson would take 10-12 minutes.

In the second section the children would play the piece they had learnt and practised from the week before. Again, this was done as a whole cohort. This tended to be fairly cacophonous, since the children did not uniformly cover the holes with their fingers, but the research assistants were asked to model, praise and reward correct rhythm and fingering, rather than the physical accuracy of fingering, since the intervention was a comparison of reading activity, not of fine motor skills. The children would then return to the new piece, with more clapping and singing to remind them of what they had learnt in section 1. This section of the lessons took 5-7 minutes.

Thirdly, the children were placed in groups of 3 or 4 (each research assistant now taking one small group), and the children would be 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. Children did not have the same instructor each week. This was a conscious decision, because not all research assistants were available every week, and I did not want some children to have consistency of recorder teacher, and others not. It usually took 2-3 minutes to reorganise and refocus the children, as the music stands and groups were moved as far apart as possible, and so this part of the lesson took 10-12 minutes.

## 3.14 Analysis of the pieces

I have made below an analysis of the six pieces that were learnt. The first piece was deemed by all research assistants to be somewhat over-ambitious, and so there is actually a reduction in complexity between piece 1 and piece 2, but thereafter the increasing complexity is gradual and identifiable. I had originally planned the six pieces in advance, and produced a small booklet, but after week one it became clear that my expectation of progress was inappropriate, and so in communication with the research assistants, I wrote new pieces on a weekly basis, based upon the perceived progress of the children in the previous week.

I have analysed here the progression. Over the six weeks the pieces develop more in terms of rhythmic and structural complexity than melodic. I confined the children to 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).

### 3.14.1 Piece 1: B A G spells Bag

**Figure 3.3, Piece 1**

|  |
| --- |
| 1, BAG spells Bag |

This piece introduces the children to the three notes mainly used in the intervention (B, A, and G). There are only crotchets and rests (glossary, p.246) in this piece. The children were comfortable learning the first two lines. Their difficulty arose in the second two lines, where the melody in inverted. To their credit the assistants caught on to this difficulty immediately, and the children were asked to learn and play only the first two lines. The title of this piece was not an attempt to merge the two reading codes, but merely an attempt to give the children something to call it. Effectively, bar 4 constitutes a coda (being different in structure from the previous three bars), which was learnt by most, but not all children. This feature informed the structure of the next piece.

### 3.14.2 Piece 2: One, two, three, You and me, (Dad’s in the kitchen cooking my tea)

**Figure 3.4: Piece 2**

|  |
| --- |
| 2, One, two, three, You and me |

The number of named notes is identical to piece 1, and the structure (four bars of repeated Bs, followed by repeated As and repeated Gs) is also the same. However, intentionally, in this piece there is no coda, and therefore only three phrases. The children were comfortable with this structure, even though it feels unfinished to a trained European ear (which has been taught to expect a symmetrical number of meters). The piece is more complex rhythmically, including as it does crotchets, quavers (glossary, p.246) and rests, but because of its more simple melodic structure compared with piece 1, the children were, as a group, more successful in learning and performing it. The rhythm of each phrase is reflected in the title.

### 3.14.3 Piece 3: My granny walks quite slowly

**Figure 3.5: Piece 3**

|  |
| --- |
| 3, My granny moves quite slowly |

Piece 3 is no more complex rhythmically than is piece 2, but melodically it contains one note change in each phrase. Each new phrase begins on the same note (B) to enable the children to easily regroup after each note change. Again, the rhythm of the first three phrases reflects the title. The rhythm of the opening notes of each phrase (‘My granny walks quite’) is identical to a section of piece 2 (‘Dad’s in the kitchen’). I experimented with a coda again in this piece, which the children coped with much better than in piece 1, showing evidence that the children were developing in their ability to respond to minor changes in melodic and rhythmic structure.

### 3.14.4 Piece 4: One, two, How do you do

**Figure 3.6: Piece 4**

|  |
| --- |
| 4, One, two, How do you do |

This is the first piece that does not start on a B (the easiest note for the children to play, involving the thumb and only one finger). In the first three pieces the melody has always fallen from the opening B, and almost always (with only one exception in bar 4 of piece 3) to an adjacent note. This piece starts on G (a tricky, three fingered note) and rises with a jump, to B in the second phrase. To allow for this melodic initiative, there are no changes in note within each phrase itself. Again, each of the first three phrases is rhythmically identical to the others, and the use of coda is retained, again, successfully on the part of the children. As in the two previous pieces, the title of the piece reflects the rhythm.

### 3.14.5 Piece 5: A tune which starts on an A

**Figure 3.7: Piece 5**

|  |
| --- |
| 5, A tune that starts on an A |

I was keen to entitle this piece in a way that did not give the children rhythmic clues, and I was pleased to see that this did not negatively affect their ability to learn it, demonstrating that the children were definitely reading the rhythm of the score as opposed to the possibility of them merely reproducing the taught rhythms contained in previous titles. This piece (as the title suggests) begins on an A, and again has a rising motif on the first change of note, then immediately falling by an interval of a third to the G. Melodically this is more complex than previous pieces, having all three learnt notes in each phrase. Additionally, the structure is different from previous pieces, with phrases 1, 2, and 4 being rhythmically and melodically identical, but phrase 3 being unique (in effect, a ‘middle 2’). So encouraged was I by the successful way that the children were able to learn this piece I decided to introduce a fourth note for the final lesson.

### 3.14.6 Piece 6: How low can you go?

**Figure 3.8: Piece 6**

|  |
| --- |
| 6, How low can you go |

Again, the title gives no rhythmic clues. Because of the new melodic and kinaesthetic complexity (the new note D involves three fingers of each hand along the full length of the recorder), rhythmically, the piece begins as did piece 1, with only crotchets and rests. However, an error was made on my part here, because I did not make sufficient rhythmic similarity between each of the phrases. As such, although the children picked up the reading of the new note very quickly, identifying it as ‘low’ on the score, and reproducing this with the fingers having to go ‘low’ on the recorder, the assistants were obliged, for a second time, to teach only the first two lines, which was successful. The new individual note was learnt and recognised well (which was the main object of the exercise), but the piece was structurally too hard for them. In my defense, I reasoned that since the children had made such obvious progress in the previous lessons, I wanted to extend their reading of music in this last lesson, but my ambition outstretched the children’s ability on this occasion. The fact that as an experienced teacher I was tempted to take this approach shows the extent of the progress that the children were making.

The above analysis shows that the reading of music gave children planned incremental opportunities to interpret and internalise repetitive symbols and to give them kinaesthetic opportunities to demonstrate their understanding of those symbols, much in the same way as *Letters and Sounds* (DfES, 2007b), and *Jolly Phonics* (2011) employ repetition of a narrow selection of graphemes in the early stages of reading acquisition. In this intervention 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. It is my contention that allowing children opportunities to interpret a variety of incremental symbol combinations in this way helped them to decode text, and my RCT in music reading serves to illustrate that point. I do not claim to have proved a connection, but I do claim to have contributed to the body of evidence by successful illustration of the effect such an intervention can have.

## 3.15 The activities of the control group during the intervention

During the intervention of the RCT (phase 1), in the first half-term of the programme, the 15 children in Mr Simm’s class who constituted half of the control group were involved in art work. The control group in Mrs Barr’s class had PE. As such, during the intervention, although the control children had consistent activities, each class had different activities from the other. It was important to report this in order to demonstrate that the children in the control group were not receiving either additional literacy or additional music input throughout this intervention period.

## 3.16 The testing materials

The participating children were each pre- and post-tested with regard to their recognition of the 26 letters of the alphabet; and their ability to decode a selection of consonant clusters and digraphs, and a set of nonsense words.

The tests were designed with three influences at work. The focus came from Gromko (2005), whose comparative study analysed data on three aspects of early reading – recognition of single graphemes; clusters and digraphs; and nonsense words. I wanted to have something against which to compare both the method and data of my own work, and so I employed these categories for my own study. The second influence was the (then proposed) decoding test for six-year-olds, commissioned by the DfE and first piloted in Primary schools during 2010-11. This test uses nonsense words, and although, because of its narrow focus, it has been heavily criticised as a test of general reading (Reedy et al., 2011), as a tool for exclusively measuring decoding skills, such an approach is useful. I also considered that the topical nature of this method might make my own research more relevant to the current initiatives which are at work in Primary and Early Years education, and thus might attract an interested audience within the field. Because of the contentious nature of this focus, I repeat that this research was not aiming to contribute to the debate about what is the “*best method for teaching reading*” (DfE, 2010: 11). I was merely trying to identify links between two codes of symbol interpretation, and to demonstrate whether one can enhance another. My motives were cognitive, not pedagogic.

The third influence was the work, explored in chapter 2, of Goswami (2005), Seymour et al. (2003) and Dehaene (2009) with regard to the cumulative structure of the nonsense words in terms of orthography and syllabic complexity.

The tests were in three sections. The first (single-letter grapheme recognition) asked children to identify the 26 alphabet graphemes, presented in mixed order (maximum score, 26). The second asked them to identify four consonant clusters (<bl, dr, sp, str>); four consonant digraphs (<ch, sh, th, ph>); and one suffix containing a vowel letter and a consonant digraph (<ing>), (maximum score, 9). The third required the identification of three sets of 15 nonsense words, each set containing three examples of five word-categories, the structure of which built in complexity (appendices 1-3). For example, the first three words of the first set had a CVC (consonant-vowel-consonant) structure. The first three words of the second set had a CVCCVC structure, and in the third set the first words had a prefix-CVC structure (maximum score, 45). The sections were always administered in the same order. This progressive structure was identical for the pre- and the post-tests, as were the clusters, digraphs and suffix, but the nonsense words were different on each test.

## 3.17 The testing procedure

The research assistants undertaking the tests were given clear instructions about which responses were acceptable as a correct response, 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 (for the implications of this approach to non-attempted items, see chapter 4, section 4.3).

When decoding single-letter graphemes, 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 /mǝ/ (‘muh’), not /ǝm/ (‘mmm’). I 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 /mǝ/ or /ǝm/. 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 initial and 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, it became clear that some children articulated <th> as /f/, as a result of prevalent local diction. I had anticipated this, and asked the assistants to ask such children who recognised <th> but articulated /f/ to name a word which started with <th>. If the children said (for example) /fruː/ (‘froo’, local pronunciation of ‘through’), a mark was awarded. If they said /fɒg/ (‘fog’), it wasn’t. 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’*?*”*

The nonsense words were divided into three sections of increasing complexity (appendices 1-3). The ground rules applied in these sections more than in any other, and if it became clear that a child was struggling, even before a section was completed, the test was terminated. Some children refused to read the nonsense words at all, because they were ‘not real words’, and this was respected. Interestingly, this response to nonsense words parallels findings with regard to pseudowords contained in Caldwell et al.’s (2011) evaluation of the 2011 pilot of the proposed Phonics Screening Check for 6-year-olds. The evaluation 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)

It is not true that as many as 12% of participants in my study refused to read the nonsense words or were confused by them, because the format and content of the tests were carefully explained to the children prior to their experiencing them. However, research assistants did observe that all children noticed that the nonsense words were not real words (some children liked this, thinking the words to be funny), but a small proportion refused to read them as a result.

This attitude to nonsense words from children in the developmental stages of reading was analysed by Barr et al. (1996), who categorised young readers into three groups

1. Logographic readers, who read few, if any words in isolation
2. Phonetic cue readers, who can read real words, but not any nonsense words
3. Phonological decoders, who can read real words, and can decode nonsense words

(adapted from Barr et al, 1996: 412)

This categorization suggests that the children who refused to read the nonsense words were, by Barr et al.’s (1996) description, phonetic cue readers.

The testing procedures and conditions were the same for every child on every occasion. They were conducted in the same communal area of the school (with which each child was familiar), at the same time of day (Tuesday afternoons). The children were tested in the same order each time (class register order), and so each child was tested within ten minutes of the time of day they had been tested previously. Even the weather was similar (each testing day was dry and with little wind, although obviously the external temperature decreased as winter approached). All tests were conducted individually, and each took approximately ten minutes. This was another conscious decision, since the testers had 60 children to get through on each afternoon, and there were five testers on each occasion. Therefore, on average, each tester had twelve children to get through, and they had two hours in which to complete the task, so it was important that they kept to time. The children who scored more highly tended to decode more quickly (but not always). It was not the case that the children who scored least well had less or more time, or vice versa.

# Phase 3. The interpretation of existing data from the participating school’s internal tracking records

## 3.18 The design of Phase 3

For a piece of research to be considered valid, it is obviously not necessary for the data, quantitative or otherwise, to be gleaned exclusively through the planning and outworking of an RCT. A great many studies are informed by the comparison of one or more sets of existing data. For example, Butzlaff’s (2000) work has been shown to be an analysis of a number of other studies. An RCT is often referred to as a ‘true’ experiment, but clearly it is not always practicable for such an experiment to take place, since one may not always be in a position to do so, or perhaps one is not always seeking to compare like with like.

At the end of the 2010-11 school year, that is at the very end of the academic year in which the intervention had taken place, I arranged to have sight of anonymous versions of a set of data drawn from the internal termly tracking reading records of the outgoing (2010-11) Year 1s (55 of whom had undergone the intervention during the autumn term), together with the equivalent records of the previous Year 1 (2009-10). Each cohort had shared the same teachers, and so although it cannot be claimed that the teaching received by these two cohorts of children was identical in content and quantity, it is likely that it was similar in coverage and quality. These records tracked each pupil’s progress through a series of teacher assessments through the allocation of National Curriculum (DfEE, 1999) level descriptors, and were therefore quantitative in nature, easily converted into numeric values. Although such a research approach cannot be said to be of the same ‘gold standard’ to which my RCT originally aspired, it was nonetheless a valid approach, and had the advantage of having 108 participants, compared with the 50 participating in the RCT.

These data did not focus upon phonemic aspects of children’s reading, as the RCT in my study did, because the data did not furnish that level of detail, but they instead took a more generic approach to reading development. Nonetheless, such a study would still be relevant to the title of my programme.

Such a research design is referred to as quasi-experimental, because it contains many features of a true experiment, but may also include additional variables, often (as in this case) chronological in nature. This form of quasi-experiment involved me making comparison between data from the two cohorts of participants, each of whom had different experiences (intervention and non-intervention) but at different times, and is known as a ‘pre-test/post-test non-equivalent group, with variable situational factors design’. The pre- and post-tests were the SATs tests in reading undertaken by the children at the beginning of each term. Using the same graphic conventions as were used in table 3.2, in section 3.9 (Campbell and Stanley, 1963, in Cohen et al., 2007: 275) this design can be represented diagrammatically as:

O3 - χ – O4

--------------------------------

O1 - O2

The dashed line represents non-randomisation of group assignment. Campbell and Stanley (1963) describe such chronologically separate quasi-experiments as “*containing the ‘who and to whom’ of measurement, but lacking control over the ‘when and to whom’ of exposure*” (p17).

Gromko’s study, by which I have been influenced, had a similar design, although her situational variable was not chronological, but physical, the control and intervention groups being in different schools. Her model would therefore be represented as follows:

O1 - χ – O2

-------------

O3 - O4

At first glance this model looks closely like an RCT, but lacks the feature of matched groups and, much more importantly, randomisation.

# Phase 4: Semi-structured interviews with the participant children

## 3.19 Use of mixed methods

A piece of research which has only numeric data as its source may be unambiguous and compelling, but modern research methods tend to involve mixed methods, or triangulation. Cohen et al. (2007: 141) observe that:

The use of multiple methods contrasts with the ubiquitous but generally more vulnerable single-method approach that characterises so much of research in the social sciences ... The more the methods contrast with each other, the greater the researcher’s confidence.

I had no wish for my research to be labelled ‘vulnerable’, and so I sought to broaden my design. It is one thing to identify that children’s textual decoding skills may or may not be improved by the synchronous learning of the reading of music, but if the learning experience of the children is shown to be negative, or if the process of the learning of music might be more or less effective by introducing or avoiding given aspects of the experience, exclusively numeric data will not add to that discussion. As such I sought ways in which my research might be enhanced or contextualised by some quantitative data.

I thus considered structured or semi-structured interviews with the participating children, and these were carried out with children at the end of their series of lessons. Fraenkel and Wallen (1993, 2003) are helpful in considering important aspects of the structure of interviews, and the categorisation of questions that might be asked, identifying features of interview questions, and the variety of question styles that provoke effective responses from questionnaires. They identify six types of question:

* Background or demographic (routine enquiries about respondents, e.g. age, education, etc.)
* Knowledge questions (trying to ascertain what the respondents consider to be factual information as opposed to opinion or belief. For example, in my case, questions about names of musical symbols, nonsense words, interpretation of rhythms through clapping activities)
* Experience or behaviour questions (What have you found easy or hard? How much did you practice?)
* Opinion or value questions (respondents’ opinions, or attitudes)
* Feelings questions (Which piece did you like best? Why?)
* Sensory questions (what would I see/hear, etc.)

(Fraenkel and Wallen, 1993: 386)

I have reproduced my questions below, and have annotated them according to these six categories. The background question has no direct bearing on the data, but I considered such a question to be an important courtesy, even with children who were five or six years old. Getting the participants engaged in the interview was an important first step, and this background question served both to put the interviewee at ease, and to focus the conversation, making it seem relevant to them as well as to the researcher.

Exploration of the questions showed one background question; one knowledge question; two experience or behaviour questions; four opinion or values questions; four feelings questions; and one sensory question. As such my selection of questions contained the full range of question types that Fraenkel and Wallen (2003) consider necessary to glean cogent and assessable data. Intentionally, there was a proportional preference in my questions in favour of opinion, value and feelings questions, in order to provide contrast with the hard quantitative data gleaned from the testing of the children. For example, I had little need of knowledge questions, since I had intentionally eliminated anything except the children’s knowledge in the test formats. The questions I selected are shown in figure 3.9.

**Figure 3.9: Interview questions for children**

|  |  |
| --- | --- |
| What is your name, and whose class are you in?  Could read music before we started teaching you the recorder?  Can you read music now?  Can you read words in books?  Look at this piece of music. What can you see?  Tell me how easy it was, learning to read music?  Why?  Do you enjoy reading music?  How about books and writing – do you enjoy reading those?  Is it easier to read music or words?  Why?  What are the differences between reading music and reading words?  Do you think that learning to read music has helped you to read words and letters? | (Background)  (Experience)  (Knowledge, opinion)  (Knowledge, opinion)  (Sensory)  (Experience)  (Opinion)  (Feelings)  (Feelings)  (Opinion, value)  (Feelings)  (Opinion, knowledge)  (Opinion, feelings) |

Some of these questions were harder to answer than others of course, but I wanted to cater for a range of linguistic and cognitive ability. After the first round of lessons (October 2010) I interviewed all the children who had received the recorder lessons, and I did this in groups of four or five. The analysis in Chapter 4 will show that these interviews were only partially successful in identifying important features of the children’s experiences, or patterns in their learning. I abandoned the structured questions for the interviews with the children after the second phase of recorder lessons, which were conducted with individual children, and were therefore much less formal. These were generally shorter, but gave greater insight.

Nutbrown and Hannon (2003) would not have been surprised that a less formal approach led to greater fluency of response. Concerns that they have with regard to the interviewing of young children are that:

* Interviews would have to be brief, consisting of a few short, clear questions
* The children may be overwhelmed by a situation [they have not experienced before]
* Children may tell the researcher what they think the researcher wants to hear
* It is more possible than when interviewing adults that the children may not say anything at all.

(Nutbrown and Hannon, 2003: 4-5)

The intentional formality of my early interview technique was honourably (but to some extent misguidedly) intended to provide children with equality of experience and opportunity so that they would be led to talk about the same things, and so that comparison between the interviews could be made. It was clear straight away however that equality of response was not forthcoming.

It is unrealistic, of course, to expect young children to provide comprehensive reports [of their experiences], but what may well be possible is for young children to tell us something about who and what is most salient in *their* literacy experiences.

(Nutbrown and Hannon, 2003: 205)

I discovered that what is most salient for child A contrasts heavily with that for child B, and again for children C, D, and E, even within a semi-structured group discussion. As such, the less formal one-to-one discussions allowed children to expand to some extent on their thoughts, because they were not in competition with a range of other salient points, and I was able to give time to explore such salient thoughts before they were superseded or forgotten.

I used a large degree of positive reinforcement at the end of the interviews. Baumrind (1964) considers it beholden on researchers to be openly aware of a positive indebtedness to their participants for their services. The children who took part in my intervention were only partially successful. At the end of the project, they were not fluent readers of music, just as they were not yet fluent readers of texts or graphemes. There is therefore a risk of a feeling, if not of failure, but of a job only half done. Furthermore, the children may have been thoroughly enjoying the six-week experience, which then came to a crashing end, and they may have felt disappointment in this. “*If the research involves subjects in a failure or loss experience, for example*”, write Cohen et al. (2007: 62) “*researchers must ensure that the subjects do not leave the situation with less ... self-esteem ... than when they arrived.*”

# Phase 5: Structured interviews with research assistants

## 3.20 Rationale

All five of the research assistants were invited to be interviewed. One did not reply, one declined, and three agreed.

I considered it important that I should have evidence that the participating children had made progress not only in their reading of letters, etc., but also in their reading of music. The title statement seeks to explore whether reading skills that are “developed through the reading of music can be transferred to the early decoding of text”. Such an enterprise would fall at the first hurdle if the existence of reading skills that are developed through the reading of music cannot be shown to be present.

To this end, following the conclusion of the intervention, I held structured interviews with three of the research assistants (two together, one separately) asking them the questions shown in Figure 3.10.

**Figure 3.10: Interview questions for research assistants**

|  |
| --- |
| Think back to the very first lesson that the children received. What were your initial impressions of the children’s level of recognition of the three music symbols (crotchets, quavers and rests) during the first lesson?  As the children continued with their lessons, can you think of any evidence that would suggest that it was clear that their understanding of reading rhythms was improving? (If the opposite is true, please say so, and why you think that).  As the children continued with their lessons, can you think of any evidence that would suggest that it was clear that their understanding of reading the notes on the stave was improving (B, A, G, D)? (Again, if the opposite is true, please say so, and why you think that).  How easy or difficult was it for the children to transfer their knowledge of reading music to the playing of the recorder, and what make you say that?  How easy or difficult was it for the children to transfer their knowledge of reading music to the clapping of rhythms, and what makes you say that? |

These questions were designed to clarify the following:

* the level of knowledge of music notation that the children brought with them to the programme
* the children’s ability to read rhythms
* the children’s ability to read notes on a stave
* the extent to which the clapping, and the recorder lessons, assisted or obstructed the children in their demonstration of their understanding of the reading of music.

Although no quantitative data can be presented here, the transcripts of these interviews (appendices 11 and 12) are analysed in chapter 4, and show that the responses of the research assistants in the two interviews were consistent, and they each confirmed that no child brought any perceivable knowledge of the reading of music to the programme. They also show that the children became successful and independent in the clapping of rhythms, and in their recognition of individual notes on the stave, but that the logistics of playing the recorder appeared problematic. However, in terms of the reading of music, the majority of children made good progress, some very good, and it was considered by all research assistants that all children made at least some progress.

## 3.21 Conclusion

I have, in this methodology chapter, described, analysed and anticipated as many aspects of the four identified methods as possible, in order to make the data analysis and results chapter which follows as transparent as possible. These methods, as with those of Long (2007); Standley and Hughes (1997) and Gromko (2005), are not beyond criticism, but they combined to produce as robust and comprehensive a research programme as was possible within the time and budget available to me. Gromko (2005: 208) hoped that “*future research can and should put the near-transfer hypothesis to additional tests.*” In spite of the above limitations of resources, I have done this, and I feel confident that these methods were sufficiently well-considered to have produced credible results.

# Chapter 4 Data Analysis (Findings)

## 4.1 Introduction

This chapter is structured into five sections. The first outlines and analyses the data analysis methods. In particular, a rationale is given for the selection and range of the sub-sets chosen for analysis, and for the treatment of non-attempted items in the tests. The interpretation of the significance of an effect size within educational research is also given consideration.

The second section reports on the RCT. In this section the equivalence of the pre-test samples is demonstrated; and pre- and post-test data are presented. In addition to calculations made from the overall scores on the tests, four sub-categories are explored, three of which were informed by those used by Gromko (2005). For each of these five categories, the mean pre- and post-test scores, effect size, and statistical significance are presented and interpreted. The results are also compared with Gromko’s (2005). It would have been possible to analyse many more than four sub-categories, but to do so would have risked type I errors, given the size and calculated power of the study. This second section concludes with a systematic audit (Torgerson et al., 2003: 241-242) – a series of checks designed to ascertain the health and validity of a given RCT. The data gleaned from phase 2 of the study have not been analysed, for reasons of invalidity, identified in chapter 3.

The third section explores phase 3 of the study: the comparison of the intervention group’s Year 1 SAT results with that of the previous year’s Year 1 cohort. Again, this was not an RCT, and the focus of the SATs data was not specific to phonemic skills of decoding, but comparison could be made of progress in general aspects of reading encompassed by national tests between the two cohorts of children. In this study, National Curriculum (DfEE, 1999) levels were converted into numeric scores, and *t*-tests (glossary, p.246) on the gains were calculated.

The fourth section of this chapter is an analysis of the interviews with 36 of the 55 children in the 2010-11 cohort. Five categories of interest presented themselves in these interviews, in some cases during the interviews themselves, and in some cases during the analysis. These interviews gave a clear picture of different aspects of the study that the children found either easiest, or the most challenging, and served to illuminate and contextualise the findings from the numeric data. They also informed a critique of the study, identifying diverse aspects in method that, were the study to be repeated, might usefully be adapted, amplified or removed.

The fifth and final section presents findings from interviews with three of the research assistants, which, together with the analysis of the pieces (section 3.14) provided evidence of the children’s development in their ability to read music. This section is important because this study asks whether reading skills that are developed through the reading of music can be transferred to the early decoding of text. The analysis of the collected numeric data clearly identified measurable development in the children’s early decoding of text, but could not, on its own, demonstrate development in the reading of music. If the study was unable to show that such development in the reading of music was present, the premise of the whole study would founder.

# Section 1: Datasets and forms of analysis

## 4.2 Categories of test data used for statistical analysis

The RCT (phase 1) contained analyses of the overall mean scores, and of four sub-sets. Originally it was my intention to divide the analyses into four categories:

* Overall test scores
* Single-letter grapheme recognition
* Cluster and digraph recognition
* Nonsense word recognition.

These last three categories were taken from Gromko (2005), a study that I adapted for the purposes of this study. However, in my study, the data of single-letter grapheme recognition presented an analysis problem. Howitt and Cramer (2008) consider that it is important to study the distribution shape of each variable, because “*statistical techniques are at their most powerful when the distributions of the variables involved are normally distributed*” (p.46). Frequency histograms for the overall data for the pre- and post-tests (Figure 4.1) showed a steep frequency distribution at both assessment points, with the distribution skewed somewhat to the left, showing that overall the tests had the capability of allowing measurement of a good deal of progress for the children. This was confirmed by the observation that only one child scored a maximum score on one occasion during the two rounds of testing. This distribution feature also allowed meaningful statistical testing.

**Figure 4.1: Distribution of total scores, pre and post**

|  |  |
| --- | --- |
|  |  |

By contrast, the distribution of the data with regard to recognition of individual letters (Figure 4.2) showed a pronounced skewing to the right at both points, showing that this aspect of reading was at an advanced stage in the children’s development, and as such there was little or no room for measurable improvement as a result of the intervention.

**Figure 4.2: Distribution of single-letter grapheme recognition scores, pre and post**

|  |  |
| --- | --- |
|  |  |

Given such a pronounced ceiling in the progress that children could make in the recognition of single-letter graphemes, I included a fifth category of analysis – the overall scores, but with the data on individual letter recognition removed. The ethics of this is explored below (Section 4.10).

## 4.3 Non-attempted items

Appendices 13 and 14 respectively show the success (or otherwise) of the children’s pre- and post-test attempts. Appendix 15 shows the points at which each of the 50 children stopped engaging with both of these tests, either as a conscious decision on their part, or as an imposed decision on the part of the testers if it became clear that, given the incremental nature of the tests, a child was not going to be able to score additionally as a result of subsequent involvement. The appendices show that there was no pattern to be found in the points of withdrawal – that is to say that there were no discernible components of either test which proved to be the undoing of a significant proportion of the children.

Non-attempted items are problematic if it is shown that some items within a test are demonstrably and unexpectedly more difficult than others (Schrader, 1984; Kaplan and Saccuzzo, 2009). Such a situation is problematic because sudden or disproportionately hard items are shown by these writers to be an influential factor which might cause participants to withdraw either partially or completely, even when (unbeknown to the participant) there is still a good likelihood that subsequent items might be met with success, and thus positively influence the scores. If participants are denied a likely opportunity to score (for example, if the majority of participants failed to score on a range of items, but were able to go on to score on an item which a given participant who failed those same items chose not to engage with), this must be factored in to the results of those participants. However, appendices 13-15 reveal that no such pattern arose in this study, and as such, there was no need to attribute notional scores to non-attempted items (indeed, without such a pattern, it would be impossible to do so). In practice, therefore, all non-attempted items were attributed a score of zero, and data from all 50 children for all sections of the testing were used in the statistical analysis of the RCT.

## 4.4 Forms of statistical analysis used

I used five forms of statistical analysis:

* *t*-tests of pre-test differences to establish equivalence of groups
* effect sizes
* analyses of variance (anova) (glossary, p.245)
* *t*-tests of post-test differences
* confidence intervals around the effect sizes.

The last four were all, in their different ways, attempts to establish the significance of the results. These five forms were used in the following ways:

### 4.4.1 *t*-tests of pre-test differences

These were used to investigate whether the two groups’ mean scores differed significantly at the pre-test stage (appendices 16-20). Because there was no hypothesis identifying the likely direction of difference, two-tailed *p*-values were identified and presented.

### 4.4.2 Effect sizes

There is a range of possible approaches when determining an effect size, of which ‘Cohen’s *d*’ (Cohen, 1988: 20; Cohen et al., 2007: 521) is the established and most commonly-used method (and the one used here). This is an approach which specifies an effect size in units or proportions of standard deviations, the formula for which is:

Mean gain of group 1 – mean gain of group 2

*d* = *s*

Where the two post-test standard deviations (s.d’s) are very similar, pooling them by simply using their arithmetic mean gives a result very similar to the more complex formula shown below, and is therefore satisfactory.

However, in situations in which the s.d’s are widely different (e.g. if they differ by more than 5% of the larger standard deviation), *s* is better calculated as follows (where *n1* and *n2* are the sample sizes of the intervention and control groups respectively, and *s1* and *s2* are their post-test s.d’s):

s = \sqrt{\frac{(n_1-1)s^2_1 + (n_2-1)s^2_2}{n_1+n_2} },

(Hartung et al., 2008)

In my study, the pairs of standard deviations which were sufficiently different to warrant the use of the more complicated formula occurred in single-letter recognition (s.d’s of 2.26 and 1.46); and cluster and digraph recognition (2.97 and 2.74). The smaller post-test standard deviations of the other three categories fell within 5% of the larger value. However, to be consistent, all effect sizes were calculated using the more complex formula.

### 4.4.3 Analyses of variance

“*Where one [or more] groups of participants are studied at different time points ... a two-way mixed analysis of variance will be particularly appropriate*” (Howitt and Cramer, 2008: 196 & 197). This variant of anova (also known in older texts as ‘two-way analysis of variance with repeated measures on one factor’) was used here. Appendices 21-25 contain the anova results produced by SPSS from the RCT data. The anova results of interest were not those for the main factor, time, but for the interaction of time and group membership – this was to investigate whether or not one of the two groups made significantly more progress than the other.

### 4.4.4 *t*-tests of post-test differences

These were used to investigate whether the two groups’ mean scores differed significantly at the post-test stage (appendices 21-25). Here, because there was a hypothesis identifying the likely direction of difference (in favour of the intervention group), one-tailed *p*-values were identified and presented.

### 4.4.5 Confidence intervals around effect sizes

Confidence intervals can serve as an additional test of significance:

Although there has been a large increase in the use of confidence intervals and of effect sizes in recent years, [Howell, 2011] goes a step beyond either confidence intervals or effect sizes by discussing how we can place a confidence interval on an effect size. [This] can be [most easily] done with available software.

(Howell, 2011: 1)

The software I used was an online calculator provided by ClinTools (2012), and the results are presented in section 4.13.

## 4.5 Effect sizes and statistical significances

Levels of effect size can be interpreted and categorised. Cohen et al. (2007: 521) list the following descriptors:

|  |  |
| --- | --- |
| 0-0.2  0.21-0.50  0.51-1  >1 | = weak effect  = modest effect  = moderate effect  = strong effect |

These are subjective descriptors, placed within arbitrary boundaries, and as such are not universally welcomed (even Cohen was reluctant to categorise them). Cohen first published his categories of effect size in 1969 (Cohen, 1969: 23) where he described an effect size of 0.2 as “*small*”; an effect size of 0.5 as “*medium*”; and an effect size of 0.8 as “*grossly perceptible and therefore large*”. In that early publication he did acknowledge the danger of using terms like 'small', 'medium' and 'large' out of context, particularly in educational settings. Glass et al*.* (1981) argued that “*the effectiveness of a particular intervention can only be interpreted in relation to other interventions that seek to produce the same effect*” (Glass et al, 1981, in Coe, 2002: 3). This approach to effect size was applied to my own work in its stated comparison to that of Gromko (2005). Glass et al. also made the point that, although statistical significance is clearly important as a statistical construct, in educational fields, if it can be demonstrated that an intervention that is easily administered and free of significant cost can lead to an effect size of even 0.1, it could be deemed to be educationally significant, if not statistically so.

More recently, Hutchison and Styles (2010) were also rather dismissive of Cohen et al.’s descriptive terminology, which they (Hutchison and Styles) consider should only be used “*when one cannot think of anything else*” (p.41), and they went on to even question whether “*the definitions carry over to education*” (p.41).

Wiliam (2008: 254) calculated that “*0.3 standard deviations is approximately the average progress in achievement made by a pupil in a year*”, and in an educational setting he posited “*an effect size of 0.1 or even 0.05 as being substantively useful*” (p.254). Wiliam’s descriptors of levels of effect size are in stark contrast to those of Cohen et al. (2007), because an effect size of 0.2 would be described as weak using the Cohen conventions, but according to Wiliam could be considered meaningful in an educational context.

Similarly, Torgerson (2003) cited Weiner (1994) who “*showed an effect [size] of phonemic awareness training ... of about 0.3 amongst 30 pupils, which is considered* [to be] *a reasonable effect size in educational research*” (Torgerson, 2003: 59). Torgerson’s point here, however, was that, even though a clear effect was shown in a particular study, the size of the sample did not allow the findings to be statistically significant.

The same will be shown to be true for this study. To anticipate somewhat, here the effect size from total scores will be shown to have been 0.29 (modest), but the anova on total scores from the RCT will show that the small sample resulted in a value of *p* = 0.084; and the *t*-test of the difference in total post-test scores resulted in a *t*-score with a *p*-value of 0.29, both non-significant.

However, a calculation of a statistical significance >0.05 does not at a stroke discount the validity of a piece of research. As Davies and Crombie (2009: 1) put it, *“Non-significance does not mean ‘no effect’. Small studies will often report non-significance even when there are important, real effects”*. It may jeopardise one’s chances of publishing that research, or from making generalisations, but an effect size is an effect size, and is of interest in its own right, particularly if it can be shown to concur with other pieces of research with the same focus. For example, a calculated *p* level of 0.167 can still represent a strong effect, albeit in a small sample, just as a *p* level of 0.01 might represent a tiny effect amongst a huge sample.

Gromko’s (2005) research was published in an internationally recognised peer-reviewed journal, and showed statistical significance in the data regarding phoneme segmentation fluency (*p* = 0.01; p.205), yet the other two areas of her study (letter-naming fluency and nonsense word fluency) demonstrated non-significant effects. The title of her article (“*The effect of music instruction on phonemic awareness in beginning readers*”) might well have been very different had the data shown statistical significance in one or both of the other areas. Below, I have replicated Gromko’s (2005) focus of breaking down the data into the same three respective areas, and I have compared her findings with mine.

# Section 2: Phase 1 – The randomised controlled trial

## 4.6 Pre-test equivalence of samples

A valid RCT requires that features which are important and relevant to the research hypothesis, and are held by the constituent members of the two randomly assigned groups, are closely matched prior to the intervention. Table 4.1 shows mean pre-test scores for the two groups, together with *t*- and *p*-values, generated through SPSS (appendices 16-20). This table demonstrates that this need for close matching of means was met for all of the areas of reading that this research programme aimed to focus upon. The two samples were also intentionally equivalent in size (25 in each arm). All pre-test *t*- and *p*-values were statistically non-significant.

**Table 4.1: Mean pre-test scores, *t*-test values, and statistical significances**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *n* | Mean Overall | Mean of Overall Scores  Minus Single-Letter Graphemes | Mean single-letter recognition | Mean Cluster & Digraph  recognition | Mean Nonsense word  recognition |
| Intervention  group | 25 | 33.96 | 10.60 | 23.36 | 3.16 | 7.12 |
| Control  group | 25 | 33.64 | 11.12 | 22.48 | 3.36 | 8.00 |
| *t*-value |  | 0.080 | 1.086 | -0.114 | -0.259 | -0.294 |
| *p*-value |  | 0.936 | 0.283 | 0.886 | 0.797 | 0.770 |

*t*- and *p*-values (two-tailed) were taken from SPSS outputs, appendices 16-20

## 4.7 Presentation of SPSS data

In the following five presentations of data, the corresponding appendices (21-25) each contain SPSS outputs which present tables for the following:

* Descriptive Statistics (showing paired-samples’ pre- and post-test overall mean scores, together with standard deviations, from which an effect size was calculated)
* Multivariate tests (identifying values for statistical significance and intervention effects)

The mean scores of the two test events are presented below as line graphs for ease of interpretation only. There is no intention to suggest that the children’s progress was linear, or that it was measured on a continual basis.

## 4.8 Overall (total) mean scores

**Table 4.2: Overall (total) mean scores**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 25 | 33.96 | (13.47) | 45.20 | (16.59) | 11.24 | *d* = 0.29 |
| Control | 25 | 33.64 | (14.63) | 40.12 | (17.26) | 6.48 |

**Figure 4.3: Overall (total) mean scores**

|  |
| --- |
|  |

By Cohen et al.’s (2007: 521) categorisation, this effect size is “*modest*”. The anova results (appendix 21) show that the effect of time (factor 1) was highly significant (F(1,48) = 43.144, p<0.001), as would be expected, but that the interaction of time and group was not statistically significant (F(1,48) = 3.113, *p* = 0.084). The test of the mean post-mean scores was also non-significant (*t* = 1.061; *df* = 48; one-tailed *p* = 0.147).

## 4.9 Single-letter grapheme recognition mean scores

**Table 4.3: Single-letter grapheme recognition mean scores**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 25 | 23.36 | (1.75) | 24.16 | (1.46) | 0.80 | *d* = -0.45 |
| Control | 25 | 22.48 | (3.65) | 24.12 | (2.26) | 1.64 |

**Figure 4.4: Single-letter grapheme recognition mean scores**

|  |
| --- |
|  |

Cohen et al.’s (2007: 521) categories hold for both positive and negative values. The anova results (appendix 22) show that the effect of time (factor 1) was highly significant (F(1,48) = 20.797, *p* <0.001), but that the interaction of time and group was not statistically significant (F(1,48) = 2.465, *p* = 0.123). The test of the mean post-mean scores was also non-significant (*t* = 0.074; *df* = 48; one-tailed *p* = 0.471).

Gromko (2005) found the effect size of her intervention on letter-naming fluency to be to be *d* = 0.09 (see chapter 2, 2.19.3), which was a much smaller effect than those in the other two categories of her study. Because of the ceiling, in my study the effect of the intervention on children’s ability to identify single graphemes is much more pronounced even than in Gromko’s, to the extent that it has produced a negative result.

A calculated effect size of *d* = -0.45 for single-letter grapheme recognition, demonstrates that the control group made significantly greater progress than the intervention group. An explanation for this can be found in the data of two children (one from each class) who each scored only eleven marks in the pre-test, all of which were gained exclusively from recognition of single-letter graphemes (the focus of this analysis), but who had both been randomly allocated to the control group by the casting of the die. 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 of this area at pre-test was very close to the maximum. The ‘catch-up’ that these two children enjoyed explains the greater average progress of the control group. Because their scores produced a distorted result, the next section presents overall scores, excluding the single-letter recognition scores.

## 4.10 Overall (total) mean scores, excluding single-letter grapheme recognition

**Table 4.4: Overall (total) mean scores, excluding single-letter grapheme recognition**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

**Figure 4.5: Overall (total) mean scores, excluding single-letter grapheme recognition**

|  |
| --- |
|  |

Again, by Cohen et al.’s (2007: 521) categorisation, this effect size is “*modest*”. The anova results (appendix 23) show the effect of time (factor 1) 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 test of the mean post-mean scores was non-significant (*t* = 1.089; *df* = 48; one-tailed *p* = 0.181).

In this effect size calculation, I have purposely excluded specific data (single-letter grapheme recognition), and in doing so have revealed a stronger effect size than that demonstrated in the data from overall scores. In acknowledging this, I might stand accused of manipulating the data; however, there is established 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)

## 4.11 Cluster and digraph recognition mean scores

**Table 4.5: Cluster and digraph recognition mean scores**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 25 | 3.16 | (2.44) | 5.52 | (2.74) | 2.36 | *d* = 0.49 |
| Control | 25 | 3.36 | (2.99) | 4.36 | (2.97) | 1.00 |

**Figure 4.6: Cluster and digraph recognition mean scores**

|  |
| --- |
|  |

Again, by Cohen et al.’s (2007: 521) categorisation, this effect size is “*modest*”, although it falls at the very upper end of Cohen’s categorisation. The anova results (appendix 24) show the effect of time (factor 1) was highly significant (F(1,48) = 20.794, *p* <0.001), but that the interaction of time and group was not statistically significant (F(1,48) = 3.407, *p* = 0.071). The test of the mean post-mean scores was also non-significant (*t* = 1.435; *df* = 48; one-tailed *p* = 0.079).

Although the analysis of children’s recognition of clusters and digraphs has been combined (in order for comparison with Gromko’s (2005) findings to be made), it is interesting that the interpretation and decoding of clusters was clearly more problematic than that of digraphs for these participating children. Appendices 13 and 14 (showing the children’s correct and incorrect answers) reveal that in the pre-test, 33 of the 50 children were unable to recognise and sound any of the four clusters. This number was reduced to fifteen children in the post-test. Of the eighteen children who moved from a score of 0 at pre-test to >0 at post-test, twelve were in the intervention group. This imbalance of progression contributed to a positive effect size larger than any other category in this research programme. By Wiliam’s (2008: 254) reckoning (“*0.3 standard deviations is approximately the average progress in achievement made by a pupil in a year*”), the effect size of 0.49 represents about 1.5 year’s progress, an astonishing amount of progress over a six-week period.

**4.12 Nonsense word recognition mean scores**

**Table 4.6: Nonsense word recognition mean scores**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 25 | 7.12 | (10.07) | 15.64 | (14.43) | 8.52 | *d* = 0.37 |
| Control | 25 | 8.00 | (11.09) | 11.44 | (13.73) | 3.44 |

**Figure 4.7: Nonsense word recognition mean scores**

|  |
| --- |
|  |

By Cohen et al.’s (2007: 521) categorisation, this effect size is “*modest*”, but again, is stronger than that for total scores. The anova results (appendix 25) show the effect of time (factor 1) was highly significant (F(1,48) = 25.495, *p* <0.001), and, as with the result of the tests of overall scores excluding single-letter grapheme recognition fluency, the interaction of time and group was statistically significant (F(1,48) = 4.600, *p* = 0.037). The test of the mean post-mean scores was non-significant (*t* = 1.055; *df* = 48; one-tailed *p* = 0.149).

Further analysis of the nonsense-word recognition data revealed some interesting additional patterns.

**Table 4.7: Further analysis of Nonsense word recognition scores**

|  |  |  |
| --- | --- | --- |
|  | 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 |

The most noticeable difference between the two groups was the pre-test / post-test reduction in the number of children unable to recognise a single nonsense word. Of the 7 children who moved from a score of 0 in the pre-test to >0 in the post-test, 6 were from the intervention group. Numbers are not significantly different between the recognition of 1-9 and 10-19 words, but again there were significant differences in the increases between 20-29 and >30. Of the 6 children who moved from a score of <30 at pre-test to 30 or more at post-test, 5 were from the intervention group. It was true that the participant numbers were small in some categories, but in all situations where there was an apparently significant 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 pleasing that in all categories other than single-letter recognition, demonstrable and distinct effect sizes were demonstrated, albeit of modest size (by Cohen et al.’s, 2007 classification). Wiliam (2008: 254) would describe all of 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.

## 4.13 Presentation of confidence intervals around effect size

A confidence interval demonstrates the bounds within which a researcher can be (usually 95%) certain that the results would fall, were a study to be repeated:

A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data ... The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter. ... Confidence intervals are more informative than the simple results of hypothesis tests (where we decide "reject” or “don’t reject” [the null hypothesis]) since they provide a range of plausible values for the unknown parameter. ... If the confidence interval includes 0 we can say that there is no significant difference between the means of two populations, at a given level of confidence.

(Easton and McColl, 2012: 1)

It is common to calculate confidence intervals around calculated means, but, if an effect size is positive, it is now also possible to calculate confidence intervals around it, using a complex formula, which involves the calculation of a non-central *t-*value (Howell, 2011: 5). However, Howell also recommends online calculators of confidence intervals, and I took his advice, using that provided by ClinTools (2012). The results are shown in Table 4.8 and Figure 4.8.

**Table 4.8: Confidence intervals around effect sizes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Total | Total less single-letters | Clusters & Digraphs | Nonsense |
| Higher | 0.85 | 0.91 | 1.05 | 0.93 |
| Actual | 0.29 | 0.35 | 0.49 | 0.37 |
| Lower | -0.27 | -0.21 | -0.08 | -0.19 |

**Figure 4.8: Confidence intervals around effect sizes**

|  |
| --- |
|  |

The range of these intervals is wide, as a result of the small sample size and consequent low power of the study. Most significantly however, each of the calculated confidence intervals of effect size straddles 0 at the lower end, showing that it cannot be claimed, to a 95% level of confidence, that were my study to be repeated, an effect would be present.

## 4.14 Comparison of this study with that of Gromko (2005)

Although the focus of this study intentionally matched that of Gromko (2005), it is significantly different in design. Gromko’s was an unmatched groups study, two levels weaker than an RCT. As such, it should not be expected that the results would correspond closely, and indeed they do not, other than the observation that in both studies the intervention had a more profound effect on cluster and digraph recognition than on other identified areas. Torgerson (2003) would not have been surprised at this lack of synergy through meta-analysis, since the two research designs are not sufficiently homogenous for useful comparison. Nonetheless, Gromko’s was the inspiration for my own study, and it would seem churlish to make no comparison at all.

**Table 4.9: Comparison of this study with that of Gromko (2005), by calculated effect sizes**

|  |  |  |
| --- | --- | --- |
| Focus | Effect size | |
| This study | Gromko (2005) |
| Overall mean scores  Individual grapheme recognition scores  Overall, excluding single-letter grapheme recognition  Mean scores, cluster and digraph recognition  Mean scores, nonsense words only | 0.29  -0.45  0.35  0.49  0.37 | N/A  0.09  N/A  0.66  -0.30 |

**Figure 4.9: Comparison of this study with that of Gromko (2005), by calculated effect sizes**

|  |
| --- |
|  |

It was not possible to compare the two studies by the calculation of confidence intervals of effect size using the online calculator employed in my own study (ClinTools, 2012), since that tool makes analysis of post-test scores (together with s.d’s and sample sizes), and holds matched pre-test groups as a prerequisite, which were not present in Gromko’s study.

## 4.15 Overall scores by gender

As an interesting aside, it was also possible to use the data to analyse whether the effects were specific in terms of gender. These data did not have the validity of the RCT, as the groups were not randomly constituted, neither were they identical in size. Pre-test scores were not equivalent. However, given the same formula as was used above (Hartung et al., 2008), Tables 4.10 and 4.11 show that there was a negligible difference in the effect sizes for girls and boys.

**Table 4.10: Overall results, girls**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 10 | 28.64 | (6.22) | 39.09 | (11.11) | 10.45 | *d* = 0.30 |
| Control | 13 | 35.67 | (13.63) | 42.00 | (15.97) | 6.33 |

**Table 4.11: Overall results, boys**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 15 | 36.07 | (15.99) | 46.86 | (19.32) | 10.79 | *d* = 0.29 |
| Control | 12 | 29.62 | (14.70) | 35.31 | (16.88) | 5.69 |

The result shown by the boys’ data was identical to that shown by the overall data: the effect size for the girls was marginally higher, but only as a result of rounding. There was therefore no specificity with regard to gender in these overall findings.

## 4.16 A systematic audit of the quality of the RCT

An RCT is not effective or reliable simply by virtue of being labelled such, but by including attributes worthy of the name. A primary consideration is the prevention of bias, and Torgerson (2003) considered that, although the size of an RCT is important in terms of providing a representative sample, and by the calculation of statistical significance, the quality of the method must be maintained. She identified reasons for which one might be critical of RCTs, and which might cause them to fall below the “*gold-standard*” (Torgerson, 2003: 52). Irrespective of how big an RCT is, unless the component features which she identified as important are present, it could still contain a fatal flaw. The avoidance of such fatal flaws is judged by systematically auditing the internal validity of the study. Such a systematic audit is provided here.

**Figure 4.10: Characteristics of a valid RCT**

|  |  |
| --- | --- |
| Characteristics | Present in this programme? |
| 1, Was the study population adequately described? (i.e. were the important characteristics of the randomised [participants] described e.g. age, gender)  2, Was the minimum important difference described (i.e. was the smallest educationally important effect size described)?  3, Was the target sample size adequately determined?  4, Was intention to treat analysis used? (i.e. were all [participants] who were randomised included in the follow-up and analysis).  5, Was the unit of randomisation described (i.e. individual [participants] or groups of [participants])?  6, Were the participants allocated using random number tables, coin flip, computer generation?  7, Was the randomisation process concealed from the investigators (i.e. were the researchers who were recruiting [participants] to the trial blinded to the [participant’s] allocation until after that [participant] had been included in the trial)?  8, Were follow-up measures administered blindly (i.e. were the researchers who administered the outcome measures blind to treatment allocation)?  9, Was estimated effect on primary and secondary outcome measures stated?  10, Was precision of effect size estimated (confidence intervals)  11, Were summary data presented in sufficient detail to permit alternative analyses or replication?  12, Was the discussion of the study findings consistent with the data? | Yes  Yes  No. The sample size was fixed externally (*n* = 50).  Yes, except the children for whom permission was not received, and their pairs  Yes  Yes. A software-generated throw of a die was used.  No – unavoidably, I as the investigator had to know the allocation, and perform it. See below.  Yes.  Yes.  Yes.  Yes.  Yes (although for the purposes of a doctoral thesis, it is for others to judge the consistency). |

Torgerson et al. (2003) reviewed a range of RCTs (their focus, in contrast to mine, was RCTs which looked at adult literacy and numeracy). I have used Torgerson et al.’s (2003: 241-242) audit as a check-list for my own work (Figure 4.10). Torgerson et al. were focused upon adult participants, and so I have substituted the word ‘participant’ for the word ‘adult’, where appropriate.

Only two of Torgerson et al.’s (2003) characteristics resulted in a negative answer. Characteristic 3 asks for the target sample size to be determined prior to the RCT through a power calculation. This was not possible, since I was confined to a ceiling of 60 participants, of whom, in the event, only 50 could be used in the RCT calculations.

Characteristic 7 requires that the randomisation process be concealed from the investigators. This was impossible, since I was both the organiser and the investigator. However, the allocations were made even before the study had started. I had not met the children in any capacity, and as such the participants were completely unknown to me. The only information I had about the children was the reading group to which each had been allocated, and this baseline information had been asked for so that, firstly, the balance of reading ability between the control and intervention groups was as closely matched as possible, and secondly, that the reading instruction of each group was similar throughout the duration of the study.

A related question, implied, but not quite stated, in Torgerson et al.’s (2003) audit is whether the allocation of participants to groups was maintained throughout. It was. The allocation schedule was therefore not subverted.

I could therefore approach with a fair degree of confidence the analysis of data that the programme provided. This breakdown of Torgerson et al.’s (2003) characteristics found in effective RCTs shows that I have done everything possible to minimise the existence of bias and of additional variables. I have acknowledged that the size of the sample was situated very much at the lower end of what is deemed acceptable for such a trial, but, having given due diligence to the constitution of the groups, and the methodological restraints incumbent upon a successful RCT, I could be confident that the data which this RCT produced were valid and reliable in relation to the participating cohort (*n* = 50). The statistical power of the programme was not sufficient to allow generalisation, but it was sufficient to calculate and identify effect sizes in a narrow range of foci; to compare findings with those of Gromko (2005); and to illuminate the initial hypothesis.

# 4.17 Section 3: Analysis of phase 3 – the school’s internal tracking records

As an additional comparison, I took the school’s internal tracking of progress levels in Reading achieved at the beginning of the autumn term, and again at the beginning of the spring term, by the 55 children involved in the intervention, and compared them to the previous year’s cohort. These internal tracking mechanisms therefore fell at the very beginning, and then shortly after the end, of the period during which the intervention took place.

Brown and Dowling’s (1998: 38) assertion that in comparative studies “*two groups are actually matched in terms of their characteristics and experiences”* refers to the need for the two groups to be demonstrably matched at pre-test. The results (appendix 26), based upon National Curriculum (1999) levels, reveal the following.

**Table 4.12: Mean pre-test scores, phase 3**

|  |  |
| --- | --- |
| Mean pre-test scores  (2010-11 cohort) | Mean pre-test scores  (2009-10 cohort) |
| 4.40 (55 children) | 4.26 (53 children) |

The groups were therefore closely, if not exactly, matched, both in size and pre-test ability. Moreover, the pre-test *t*-value (SPSS output, appendix 27) was non-significant (*t* = 0.744; *df* = 106; two-tailed *p* = 0.459).

## 4.18 Absentees

The five children in the 2010-11 cohort for whom permission was not obtained were again omitted from this comparison, but on this occasion there was no need for their pairs to be removed, since they were not being matched to each other. The data for five of the 58 children in the 2009-10 cohort who missed either round of testing were not taken into account, since their data was incomplete. Hence, the population of the 2009-10 cohort was 53 children.

## 4.19 National Curriculum levels

It is necessary to give an overview of what these changes in level represent. The National Curriculum (DfEE, 1999, Key Stages 1 & 2) presented a seemingly clear and unambiguous expectation of pupil progression in the Primary phases.

**Table 4.13: National Curriculum Attainment Targets**

|  |  |  |  |
| --- | --- | --- | --- |
| Range of levels within which the great majority of pupils are expected to work | | Expected attainment for the majority of pupils at the end of the key stage | |
| Key Stage 1 | 1 - 3 | At age 7 | 2 |
| Key Stage 2 | 2 - 5 | At age 11 | 4 |
| Key Stage 3 | 3 - 7 | At age 14 | 5/6 |

(Source: Appendix 1 of National Curriculum, 1999)

Although there is no baseline figure with regard to the expected attainment for the majority of pupils at the beginning of Key Stage 1, it can be seen that during the four years of the Key Stage 2 phase, it is anticipated (and expected) that pupils move up two full levels. It must be observed that this is an ambiguous instrument of measurement. It is somewhat blunt because the differences in knowledge, skills and understanding described under 2c in Reading are very different and simplified from those described under 2a, and so a broad range of progress and ability is accommodated within a single level descriptor. A child whose progress during the Key Stage 2 phase allowed an assessment profile to show a move from 2c - 4a (a possible but unlikely improvement of eight sublevels) would be monitored, in terms of levels of National Curriculum progress, under the same categories as a child who moved from 2a – 4c (an improvement of only four sublevels). Were we to assume however that progress for level 2 to level 4 averaged an improvement of six sublevels (from 2c - 4c, or from 2a - 4a), this would require an average annual improvement of 1.5 sublevels per year. Each term then, the child would need to progress by half of a sub-level, which is a nonsense concept in terms of the assessment of a single child (there is no such assessment category of 2.5a or 2.5b), but it is relevant and assessable in terms of the mean progression of a cohort.

## 4.20 Imputation of numeric values

The children’s reading levels were tracked at the start of each term, and I compared the tracking data of the 2010-11 cohort with the previous year’s cohort, for whom no intervention was administered. The following numeric values were attributed to the sublevels so they could be imputed onto SPSS.

**Table 4.14: Numeric values allocated to National Curriculum sub-levels**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Level | Wc\* | Wb | Wa | 1c | 1b | 1a | 2c |
| Value | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

(\*W = ‘working towards level 1’)

## 4.21 Presentation of Phase 3 data

Looking at the assessed levels of the cohorts (appendix 26) it was clear that over their respective Year 1 autumn terms, each cohort achieved a mean improvement in excess of this notional 0.5 sublevel/term expectation, but the 2010-11 cohort’s progress was the stronger of the two by a proportion of 17% (intervention group’s mean gain = 1.02 sublevels; control group’s mean gain = 0.85 sublevel). An effect size calculation and the SPSS anova output (appendix 28) reveal the following.

**Table 4.15: Phase 3, pre- and post-test mean scores, and effect size**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | *n* | Pre-test | | Post-test | | Gain | Effect size |
|  |  | mean | (s.d.) | mean | (s.d.) |  |
| Intervention | 55 | 4.40 | 0.94 | 5.42 | (1.13) | 1.02 | *d* = 0.14 |
| Control | 53 | 4.26 | (0.96) | 5.11 | (1.34) | 0.85 |

The anova results (appendix 28) show the effect of time (factor 1) was very highly significant (F(1,106) = 208.798; *p* <0.001), but that the interaction of time and group was not statistically significant (F(1,106) = 3.113; *p* = 1.713). The test of the mean post-mean scores was also non-significant (*t* = 1.279; *df* = 106; one-tailed *p* = 0.102).

By Cohen et al.’s (2007: 521) categorisation, this effect size is “weak”. However, I return to Wiliam’s (2008: 254) observation that “*0.3 standard deviations is approximately the average progress in achievement made by a pupil in a year”*, ... “*and* [referring to an educational setting] *an effect size of 0.1 or even 0.05 [can be] substantively useful*”. This places the effect size of these parallel cohorts into a positive light. Once again, the result is statistically non-significant, but an effect is shown.

**Table 4.16: 95% confidence interval of effect size, phase 3**

|  |  |  |
| --- | --- | --- |
| Lower | Actual | Higher |
| -0.13 | 0.14 | 0.63 |

Table 4.16 identifies the upper and lower ranges of the 95% confidence interval, using ClinTools (2012). These mirrored closely those of the RCT. The values of the confidence interval straddle 0, and so, as in phase 1, it was not possible to be 95% certain that, were my study to be repeated, an effect would be present.

## 4.22 Comparison of degrees of progress, by National Curriculum sublevels

**Table 4.17: Tally of degrees of progress, by National Curriculum sublevel**

|  |  |  |  |
| --- | --- | --- | --- |
| Year 1, 2010-11 | | Year 1, 2009-10 | |
| Progress  (in sublevels) | Number of children | Progress  (in sublevels) | Number of children |
| 0 | 9 | 0 | 18 |
| +1 | 38 | +1 | 26 |
| +2 | 8 | +2 | 8 |
| +3 | 0 | +3 | 1 |

Table 4.17 (summarising appendix 26) demonstrates that, by comparison with the children from the previous year’s cohort, the number of children whose progress during the intervention period was not sufficient to take them up into the next sublevel band was lower by 9, yet the number of children who progressed by one sub-level was higher by 12. The number of children who progressed by two sub-levels remained unchanged. Across the two cohorts, only one child (from the 2009-10 cohort) progressed by 3 sublevels. This table does not give an indication as to whether the intervention benefited children whose prior learning was stronger or weaker than the norm, but it does give an indication that the intervention probably had the effect of allowing a number of children to progress by a single National Curriculum sub-level who, without the intervention, might not have done so.

# Section 4: Analysis of semi-structured interviews with participating children

## 4.23 Rationale

These interviews were conducted on 19th October 2010, and 14th December 2010. The October interviews were conducted in line with the format described in Chapter 3 (section 3.18) and although helpful and usable data were collected in both rounds of interviews, I was aware that a group mentality often took the conversations in directions which detracted from the research focus, often with competitive claims about reading experiences which were both unbelievable and unhelpful. As such, the second round of interviews was conducted individually, and although the interviews were much shorter than in the October, a greater quality of data was collected, as answers could be developed without risk of interruption, or the temptation for exaggeration.

All children’s names have been changed in the reporting of these interviews.

The interview data that were collected present a variety of patterns, and questions for further thought. These can be categorised into five areas, which are presented here, in no particular order, other than the order in which I became aware of them when analysing the transcripts.

1. Children’s attitude to the testing
2. The children’s attitude to learning to read music, and the respective levels of difficulty in learning to read music and/or text
3. The respective levels of difficulty associated with learning to read music through the clapping of rhythms, or through the playing of an instrument
4. Children’s perceptions of themselves as readers of texts, and as readers of music
5. The lack of correlation between children’s given reading levels, and their ability to articulate their thoughts about reading acquisition.

## 4.24 Children’s attitude to the testing

It was clear from the transcripts that the vast majority of the children thoroughly enjoyed the testing. I had considered that this might be a dry and somewhat unwelcome experience for them: a functional necessity, driven by the needs of the research project. However, when conducting the individual interviews, the following reactions were recorded

(intervewer = I)

I: Hello Natalie [child 3]. Thank you for agreeing to talk to me [about your recorder playing]. Would you like to do the talk or the test first?

Natalie: (clearly) Test.

I: Test first. Do you like the test? .... (Natalie, nods to me, willingly).

I: Hello Meheni [Child 52]. I’d like to talk to you about your recorder lessons. Would you like to do the test first, or the talk first?

Meheni: Test first.

I: Test first. Do you like the tests then (Meheni nods, smiling)

It may have been that the children wanted to get the tests out of the way, but that was not at all the impression I got. I became careful to change the order in which I asked the question, and the way in which the session was introduced to the children in case the children responded to the first or last option they heard.

I: Hello Alice [child 47]. What I’d like us to do is one of these tests. You’ve seen these before?   
Alice: Yes.

I: Then I’d like to talk to you about your recorder lessons. Do you want to do the talk first, or the test?

Alice: The test.

I: The test. Do you like the test then? .... Alice has nodded.

I: Hello, Arthur [child 20] ... [some introductions]. What would you like to do first, the test, or the chat?

Arthur: (loudly and definitely) Test!

I: The test first. Do you like the test?

Arthur: Yes!

I was pleased that the children had given this reaction, because, mindful of Cohen et al.’s (2007: 62) caution that “*researchers must ensure that the subjects do not leave the situation with less ... self-esteem ... than when they arrived*”, I had worried that since the test is designed to take children to the limit of their decoding skills, they would resist the experience. Of the eleven children I asked about their attitude to the test, only one (child 39) expressed a preference to do the talk first.

I: [some introductions] ... so Colin [child 39], do you want to do the test first, or the chat first?

Colin: (quietly, and without confidence) Chat first.

I: The chat first ... You’re not very keen on the tests then?

Colin: (no answer, shakes his head)

In the interests of maintaining Colin’s self-esteem, I arranged for him to be given a choice as to whether to do the test at all, but he was willing to do so. In both rounds of testing he demonstrated a knowledge of single grapheme/phoneme correspondences (23/26 in the pre-test, rising to 25/26 in the post-test). Additionally, in the post-test he recognised three of the four digraphs, but he did not recognise a single cluster, neither was he able to decode a single nonsense word in either of the tests. As such, in spite of much positive reinforcement from the testers, the experience must have lacked any sense of progress or satisfaction for him.

## 4.25 The children’s attitude to learning to read music, and the respective levels of difficulty in learning to read music and/or text

In the main, the interviews showed that children found the learning of reading music to be straightforward. The scope of the music that they were asked to learn to read was restricted to four named notes (B, A, G, and D) and three types of symbol for recognising rhythm (crotchets, introduced to the children as ‘ones’; quavers, known as ‘quickies’; and rests, known as ‘rests’). Nonetheless, just as there are conventions that have to be learnt with the decoding of words (reading from left to right, spaces between words, correct orientation of graphemes), so to there are with the reading of music, conventions that were not to be found in the children’s experience of decoding texts (interpretation of the treble clef, rhythm demands, and the need to maintain an agreed pace).

By contrast, there are also universal consistencies which make the reading of music less problematic than in the decoding of text; a level of consistency which the decoding of texts cannot boast. A quaver is always a quaver, and, when playing the recorder, the ‘A’ above middle C is always played in the same way, and makes the same sound. Such consistency cannot be said for many of the graphemes that the children are encountering in their reading of text (for example, the grapheme ‘c’ in the words ‘circus’, and ‘cello’).

I was interested to see how threatened the children, as a cohort, would be with regard to the reading of music, since this was a code none of them had experienced before (for evidence, see section 4.29ff). Some children were indeed apprehensive.

I: What was the hardest or easiest bit of learning to read music, Shefki?

Shefki [child 25]: The hardest bit, when you first came into our class and we didn’t know what the bit of music on the front ....

Simon [child 29]: ... (interrupting) Yea, we thought, and we thought, we thought it was about colouring.

Shefki: Yea, we thought when we first started and we saw a quickie, we thought it was just scribble.

Simon: And when we first see the, saw the ... when we first saw the first little funny thing.

Shefki: What we forgot was that we didn’t know what it was.

These two children were clearly not sure what they were looking at when I first showed them some isolated notes on my initial visit, and most of the cohort shared their bemusement. However, when the children were asked in the post-intervention interviews how easy or hard they found the reading of music, and why, the answers they gave were interesting, and disclosed children’s clear perception of their own reading acquisition. Some children considered that books were easier to read than music, simply because they (the children) did so much more of it, at school and at home.

I: So is it easier reading music, or is it easier reading words?

Meheni [child 52]: Easier reading words.

I: Why is that?

Meheni: Because ... (pointing to the notes) ...I don’t really read them much.

I: Do you mean not often in a day?

Meheni: Yes. Not often in a day.

I: Which is easier?

Edward [child 54]: Reading books.

I: Why?

Edward: Because we learn that much more.

I: Oh what, you mean you read more at school.

Edward: Yes, we read at school, lots of times.

I: Do you think people should teach children to read music as much as they do books.

Edward: Noooo! It would take a year to do that.

It is interesting that when asked about the notes that the children could remember, they tended to reply about the types of notes (crotchets, quavers and rests) rather than the names of notes (B, A, or G).

I: Tell me about learning to read music. Was it easy or hard?

Janet [child 37]: Well, when I started it, it was a bit tricky ... When I first learnt it, I think it was going to be easy, but it started off a little hard.

I: What notes can you remember?

Janet: The one, the quiet one, and the quickie.

I: Colin, do you remember any of the notes? The names of any of the notes that we learnt?

Colin [child 39]: Quickies

I: Quickies. Good.

Colin: One.

I: Good

Colin: (short pause) Rests.

I: Tell me about learning to read music. Was it easy or hard?

Meheni [child 52]: It’s quite easy, playing the recorder. Words, I keep in my head so that I don’t forget them.

I: Which words are these?

Meheni: Quickie, one ...[pause] ... rest!

I: And what tunes can you remember best?

Shannon [child 9]: Erm ... can’t remember

I: OK. What notes can you remember?

Shannon: Quickie, one, and rest.

I: Good. Quickie, one, rest. And do you remember any of the names of the notes?  
 Shannon: Er ... No

I: So it’s very much the rhythm side that you remember.

(Shannon nods)

Not all children responded primarily to the types of note. Some had a mathematical or graphical approach to decoding the letter names of the notes, though not all lines of thought were sequential.

I: How hard is it, learning to read music?

Ian [child 49]: It’s easy peasy.

I: Why is it easy peasy, Ian?

Ian: Um, well, we know what notes they are by the look of them, because one’s on the line, one’s just below the line, these on the line, the very first one, B is there [points], and G is the very bottom one, and when you get old you can read better.

Some children considered that music was easier to read than words because of the physicality of producing the sound.

I: You thought reading music was easier Isaac. Why’s that?

Isaac [child 59]: Because it’s quicker, and um, you just nudge this ... You have a finger in your head and then you point to it, and then, and then you, then you do that [demonstrates the finger movement of playing a recorder], and you easily learn very quick.

I: And Kiki, you also thought it was easier to read music?

Kiki [child 53]: Because we get to blow ... blow but not talk.

I: Ah ha! To blow but not talk. So can you read without talking?

Kiki: No! (laughs)

I: Elizabeth, is it easier to read music, or is it easier to read words?

Elizabeth [child 17]: Read, um ... [thinks] music.

I: What makes it easier?

Elizabeth: You just have to put your fingers on the holes.

I: Michael?

Michael [child 57]: Well you clap notes, but you don’t clap words.

Others considered books to be easier, because of being able to make connections with the content.

I: So what other differences are there between reading music and reading texts?

Raymond [child 24]: Well that [points to a piece of music] is not so easier than that [points to a book], and that [the book] is .. it’s got little bit colour in and that [the music] is not.

I: So there are more pictures? Books have pictures, and music doesn’t?

Raymond: Yes, so I don’t know what that is [pointing to the music].

I: Oh, I see. Are you saying that the pictures help you to read?

Raymond: Yes.

I: Should I have put some pictures on the music? Would that have helped you?

Raymond: [Thinks] ...Yes. [unconvinced]

I: Is it easier to read books, or music?

Anthony [child 33]: Books

I: Why is that Anthony?

Anthony: Because when you read it, when you read it, it’s got funny stuff in.

I: Funny stuff. There’s no funny stuff in music then!?

Anthony: No!! (laughs).

These excerpts from transcripts show that the children were not at all threatened by being asked to read simple, but formal, music notation at an early age. The excerpts also show the range of strategies that the children brought with them to make sense of the “scribble” (child 25).

## 4.26 The respective levels of difficulty associated with learning to read music through the clapping of rhythms, or through the playing of an instrument

With regard to the logistics of learning to read music, it became apparent through talking to the children that one of the reasons that so many of the children can remember the names of the notes in terms of crotchets, quavers and rests, and not in terms of B, A and G, is because the actual learning of the instrument, the physicality of matching symbols to combinations of holes and finger movements, does not serve to enhance the learning experience, but rather, to get in the way. I had anticipated that the kinaesthetic aspects of learning an instrument would have been a potential reason for success, but the opposite seems to be true. The crotchets, quavers and rests are easily responded to by clapping or chanting the names (“all together ... and, one, quickie, one, rest, one”). These three types of note can be combined to make rhythmic phrases of increasing length and complexity, and several times during the course of the programme I had cause to wonder whether the results of my research may have been equally successful, or perhaps even more so, if I had simply taught the children rhythmic notation, rather than melodic; using drums, rather than recorders. Certainly, the logistics of the enterprise would have been considerably easier. Again, the transcript excerpts told a story.

I: What was the hardest or easiest bit of learning to read music, Alice?

Alice [child 37]: The clapping.

I: The clapping, was that easy or hard?

Alice: Yes, easy.

I: What about the playing of the recorder?

Alice: That was hard, because you had to put four of them [fingers] down.

I: What made it hard, Isaac?

Isaac [child 59]: The quickies.

I: The quickies? Why was that?

Isaac: You start with a B, and when you go to a lower note and then you go up, you start with a G then you go up to a B straight away, after the G. It’s quite difficult.

I: Why is it difficult?

Isaac: Because you don’t have time to put two fingers up.

I: Oh I see. So it’s the finger movement that’s the hard bit. Fiona?

Fiona [child 36]: It’s a little bit hard for me because every time when you’re blowing the flute I can’t join them.

I: So it’s the blowing that you find difficult.

Fiona: I can’t put my fingers on.

And here, Shefki and Simon display forensic logic for children who have not long been five years old.

Shefki [child 25]: Reading books is easier than reading music, because you don’t have to play an instrument while you are doing it!

Simon [child 29] Yea, you don’t have to play music, you just have to read.

Clearly, the dual activities of interpreting symbols, whilst manipulating an instrument, are problematic for children of this age. When asked, Courtney [child 2, who found all aspects of reading texts difficult, yet was successful in the clapping activities], and children 3, 9, 17, 20, 30, 39, 49, 50, 57, 58, were unanimous in their verdict that the clapping of rhythm symbols was easy, but the playing of the recorder was hard. Perhaps it is the fine motor skills that were the obstacle here.

It should not be assumed, however, that just because the playing of a recorder was hard, the children did not enjoy doing this. Meheni [child 52] observed that “on note one [B] I put one finger down, the two, I put two fingers down, and three I put three fingers down. It’s easy”. Alice [child 37] “liked playing the recorder best”, and Natalie [child 3] not only considered that she liked playing the recorder but she disclosed that she could “play a song where there’s not the notes”, meaning she was able to play a tune which she had herself composed from memory.

I: Really, Natalie? Go on then.

Natalie: Right. This is my song (she plays to the group from memory, simply and inaccurately, but there is a definite structure).

I: Do you know what that is called?

Natalie: Yes (but she does not say; she just looks proud)

The primary function of this research programme was to endeavour to demonstrate whether the early reading experiences of children could be enhanced through the transfer of reading skills from one code to another. No piece of research is so comprehensive that it answers every possible related question, and so as a suggestion for further research it would be interesting to know whether similar successful numeric results to those presented above are achieved with only the teaching of rhythmic notation. The physicality of playing a particular instrument was not central to my hypothesis. The responses of the children participating in this randomised controlled trial, together with their ready recollection of rhythmic vocabulary and their proficiency at accurately reproducing simple written rhythms through clapping, contrasted with their rather laboured ability to recognise note names by letter. This suggests that children in the early months of Year 1 are receptive to the interpretation of simple formal rhythmic symbols, but not necessarily to the interpretation of notes on a stave. The reproduction of the patterns of those notes named by letter, through the medium of a recorder, appeared not only surplus to the requirements of the study, it may have actually impaired the measured effect size of the intervention. I can as yet offer no evidence to support or counter that argument of course, other than the analysis of these parallel conversations.

The chance conversation with Natalie [child 3] also raised questions about compositional skills. This was an area outside of the scope of this thesis of course, but it would be of interest to consider whether children’s ability to compose and to generate ideas within fiction, non-fiction, poetry or multi-modal genres would be enhanced were they (the children) to be given opportunities to compose and generate musical pieces, performances, either instrumentally or through song, performance poetry or dance. Such a question draws upon the observations held in the Henley report (DfE, 2011a), where it is overtly assumed that “*the benefits of a quality music education are those of ... improved academic attainment in areas such as ... literacy and language* (DfE, 2011a: 42). Spandel and Stiggens (1997) put it beautifully:

In order to [produce] any kind of performance well, we have to get inside it, see it from the inside out, truly get in touch with all the details and particulars. Surely, this is also true of writing.

(Spandel and Stiggens 1997: xi)

Whether such a benefit could be identified is another matter. The measurability of compositional aspects of text generation (as opposed to transcriptional aspects) would be methodologically problematic to pursue, but if, as Hansen et al. (2007) demonstrate in the field of reading, there are linear relationships between music and literacy, this connection to compositional aspects of writing would be a further interesting area for research.

## 4.27 Children’s perceptions of themselves as readers of texts, and as readers of music

It was very interesting to discover how, and whether, the children in this study perceived themselves as either readers of texts, or as readers of music, or both. As a first example, consider the experience of Courtney [child 2], who scored only on single-letter graphemes in the first round of testing (23/26), yet recognised all four of the digraphs in the second. In both rounds of testing she was unable to decode any of the nonsense words.

I: I know you can all read music, but are you good readers of books?

All (except Courtney [child 2]: Yes / yes / I am.

I: Courtney, you said no. You can read music, but you can’t read books?

Courtney: I just can’t. I can read little ones.

I: You read little ones.

Courtney: Yeah

I: If I showed you a piece of music, there we are, there’s a piece of music there ... what do you know on that piece of paper?

Courtney: I know that (she indicates the taught hand movement for a rest).

I: Yes. That’s right. Courtney, can you see a rest on the paper.

(Courtney points, correctly)

I: Shall we all clap a rest together (we do). Well done. So how easy was it then, learning to read music?

All (including Courtney): Easy/ quite easy/ good/ Yeah.

Courtney was willing, even amongst a group of three other children who each declared themselves to be good readers of books, to disclose that she “just can’t” read texts, except little ones. This showed some confidence and self-awareness. Yet when asked to confirm whether she was a reader of music, she was able to make a positive, and genuine, affirmation. It may be that this programme made some, none or a great deal of difference to Courtney’s ability to decode text, but if nothing else it gave her an opportunity to say, amongst her peers, that she was, with them, a reader of music. Cohen et al. (2007), with their insistence on ethical procedures which protect self-esteem, would approve.

Meheni was another whose perception of herself as a reader of texts was not confident.

I: I know you can read music, but are you a good reader of books?

Meheni [child 52]: I’m not a very good reader. All the boys ...

I: (interrupting) ...What stage are you on?

Meheni: Three.

I: Sorry. All the boys?

Meheni: All the boys are on stage six or seven.

I: And are you a good reader of music, Meheni?

Meheni: Yes.

I: OK. So you are a good reader of music, but not yet a good reader of words. Will you be a good reader of words?

Meheni: Yes.

I: I think you will. You read some beautiful stuff to me today.

After six lessons, Meheni (who has learnt to read music as successfully as any of her peers) considered herself to be a good reader of music, and by comparison to her peers, she was right. Yet she did not consider herself to be a good reader of words or books. Elizabeth [child 17], who scored an identical total to Meheni in the pre-test (34), but whose score improved greatly by comparison to Meheni in the post-test, in spite of being in the same recorder group (43 in the post-test, compared with Meheni’s 35), had no such agonies of self-doubt.

I: Can you clap this rhythm for me, Elizabeth? [she does, very accurately] ... and are you a good reader of words, Elizabeth?

Elizabeth: (confidently) Yes!

I: ... and of music?

Elizabeth: (confidently) Yes!

By contrast, a number of children could read aspects of music, but were reluctant to claim so. Michael [child 57] was a case in point.

I: If I showed you some music, like this ... are you able to read that?

All: Let’s see / um / yes / I can

I: Michael, can you read that?

Michael: (reading the title of the tune) 1, 2, How do you do?

I: (laughs) That’s the title. Can you read these bits (pointing to the music)

Michael: No

I: Can you clap the rhythm?

Michael: Yes.

I: Yes? Come on then. (Michael does, correctly, if jerkily). Excellent. Now you could do that couldn’t you? Now you just told me that you couldn’t read music, but you just did, didn’t you?

Michael nods, but appears unconvinced.

Clearly, Michael could read aspects of music, but did not yet identify himself as a reader of music. This was a child who scored very highly in the tests (totals of 61, and 54 – unusually he scored less well in the post-test than the pre-). When asked to read from the page of music placed in front of him, he applied himself directly to the familiar – the writing, the title. He generally read appropriate texts with fluency and confidence, yet did not do so with music. This is because, once again, for Michael to have a sense of success, he felt the need to be able to interpret the music sufficiently well to be able to produce the product – the tune from the recorder – yet the individual letter names of the notes were not automatic to him. See how he steered the conversation away from his perceived lack of success at reading music to his successful decoding of texts.

I: How hard is it, learning to read music?

All (except Michael): Easy / quite easy / quite hard

Michael: I’m on stage 6 soon

Did he mis-hear the question, or was he returning to aspects of reading where he could be ahead of the field? It was not possible to tell, but it was interesting to compare the responses of weaker readers of texts, who found themselves able to confidently assert themselves to be readers of music (Courtney, Meheni), with stronger readers of texts, who either had different standards about the level of engagement one had to have before claiming proficiency (Michael), or who were willing to rejoice in their success at whatever level (Elizabeth).

## 4.28 The lack of correlation between children’s given reading levels, and their ability to articulate their thoughts about reading acquisition

|  |
| --- |
| wordle3 |
| **Figure 4.11: Wordle;** from group interviews. The most commonly-used words are presented most prominently. The word ‘stage’ is shown to be used often. |

In the participating children’s experience of this research programme, there was of course no standard against which they could judge or identify their own progress. There were no ‘stage 3’ or ‘stage 6 or 7’ readers of music for them to compare themselves against. This experience of numbered stages of reading, through which children must progress, was keenly felt. The closest I came to the use of a Likert scale (Likert, 1932), was when the children were asked to consider whether some of the pieces I had taught them were harder than others.

I: Some books are harder than others. Have some of the pieces I have asked you to play been harder than others?

Mavis [child 14]: The first one was really easy – the Granny.

Alan [child 27]: I liked the one with the dad – that was really easy for me.

Brian [child 21]: At school, I’m on stage 6.

I: There’s not stages in reading music though, is there? If this was ... I mean stage 6 or 8 ... how many stages are there? (pause ... nobody knows...) Let’s say there’s six. So a stage six book is going to be really quite hard. A stage 1 book is going to be really easy.

Keith [child 4]: I got stage 2, but I’ve got Spot the dog in my book.

Brian: I know, but even though my stage 6 book is hard I can still read it because I am really good at reading.

I: So if you were putting this piece of music into a stage, what stage would it be? Stage 1? Stage 3? Stage 5?

Mavis: Stage 5.

Alan: No. I think stage 1.

Brian: Stage 5.

Keith: Stage 5.

Once again, the responses suggested that I was right not to employ a Likert scale, because these children could only categorise the level of difficulty at one end of the spectrum or the other. However, although the question had been unsuccessful in terms of promoting reflection about reading music, it had been raised because this particular group of four children possessed a level of self awareness about the process of reading acquisition that I had not seen in other groups. Encouraged by this, I raised a big question.

I: Do you think that learning to read music has helped you to read words?

Mavis / Alan / Brian: Yes / No / Um ...

(long pause)

Keith: Yes, because if we do the music, then we learn the recorder, then we learn the clapping, we know what the, ... what the, ... what the words do and then we know, in the books we know, what letter they are.

It was a hard question I know, but although Keith’s articulation was elementary, a connection had clearly been made between the two codes of reading. Keith was not an able reader of texts, indeed the groupings of children given to me at the start of the programme placed him in the least able set in his class. In the pre-test he scored 23, accumulated from 22 of the 26 alphabet graphemes, and a single nonsense word. He failed to recognise and sound a single cluster or digraph. At post-test his overall score had risen to 28, having gained recognition of two more letters, one cluster, one digraph, and one additional nonsense word. This was not a child who had any degree of autonomy in the reading process, yet he was confident enough to attempt to articulate his thoughts about a complex issue (the subject of a doctoral thesis, no less) within the vocabulary he possessed, in front of his apparently more able peers.

Sometimes, connections between the reading of music and the reading of texts were made unexpectedly, and not by my direction of the conversation.

Claire [child 34] ... and when you are grown up you can help your children.

I: Oh right. And does your mum read the music to you?

Claire: Yes, but she can’t read the music. I can.

I: Oh right. So you’ve been teaching your mum have you?

Claire: Yes.

Stuart [child 49]: I haven’t taught my mum. My mum can’t spell anything.

I: Spelling’s very hard, isn’t it?

Stuart: I can’t spell lots of words.

I: But you said that this was a B, and A, and G. Stuart. Is that spelling?

Stuart: No

I: Is there spelling in music?

Stuart / Claire: No / No! (laughs)

Meheni [child 52] was another child whose responses feature heavily in this presentation of data, able to talk freely about her reading acquisition experiences. She too was not an able reader, placed in the lower middle reading group, yet through the medium of the interviews she was able to identify her perceived strengths and weaknesses.

It was intriguing to see young children grappling with concepts which caused them to speak at the very limits of what their understanding and vocabulary allow. In addition to Keith’s laudable attempts at articulating his theoretical connections, other children found the interview process a scaffolded opportunity to express previously un-thought thoughts.

I: Is there anything on this page that you’d be able to tell me what it was?

Alex [child 6]: You remember the ones that were that, and that, and that (he points) ... that’s a quickie, and the one like that, that isn’t a quickie, you just slow down on that (he claps), and a rest, you have to go like this (he demonstrates) .... (pause, then proudly, realising his mastery of the three music symbols) ... I know everything.

I: Alisha, you thought that learning to read music helped you to read words. How does that work?

Alisha [child 22]: Yeah. Because, um, because sometimes pieces of paper have words on them. In music you can see the words, and then you just have to clap the words. And the easiest way to read is to keep on learning and then, then, first learn easy words, and second learn hard words.

I: Are there easy notes and hard notes?

Alisha: Yes, some notes are hard, and some are easy.

I: What’s a hard note?

Alisha: Errr, ... quickies.

I: Quickies are hard are they? Someone else said that today. Is it because ....

Raymond [child 24]: (interrupts) Yea. When you do a quickie, when you do a quickie and then another, another vvv... another rest (he claps) it’s a little bit harder, so when, if, if, if you do it faster then it moves, on the on the, the ..

I: On the recorder?

Raymond: Yes, on that.

Table 4.18 identifies the children who have featured in this presentation of data, and shows that with the exception of child 52, between two and six of their comments have each been represented. The allocated identification number is taken from the ranked scores in the pre-test (weak to strong). As such the weakest child in Mr Simms’ class was allocated number 1, and the strongest, number 30. Similarly the weakest in Mrs Barr’s class was allocated number 31, and the strongest, 60. The table shows that a spread of decoding ability was contained amongst the children who responded most relevantly.

**Table 4.18: Tally of children whose responses feature in the presentation of data**

|  |  |  |  |
| --- | --- | --- | --- |
| ID number (Mr Simm’s class) | Number of references in transcript entries | ID number (Mrs Barr’s class) | Number of references in transcript entries |
| 2 | 5 | 33 | 3 |
| 3\* | 3 | 36 | 2 |
| 9\* | 3 | 37 | 2 |
| 17\* | 4 | 39\* | 5 |
| 20\* | 2 | 47\* | 5 |
| 24 | 6 | 49 | 2 |
| 25 | 4 | 52\* | 12 |
| 28\* | 6 | 53 | 2 |
| 29 | 3 | 54 | 4 |
| Interviewed individually\* |  | 57 | 6 |
|  |  | 59 | 4 |

This table, and these excerpts from transcripts, have shown that participating children in this project, from a spread of reading proficiency, have the ability to articulate their understanding of the reading process, and to make plausible (and on several occasions, insightful) observations about what it is to be a reader of texts, and what it is to be a reader of music. They have compared the two experiences, expressed preferences, and, on a few occasions, have drawn parallels between the two codes of reading. This observation has surprised me. One might imagine that it would be the most able readers who have the most insight into the processes involved in learning to read, but the identification of the allocated numbers of children who, at interview, articulated something deemed by me to be sufficiently articulate or valuable to be included in this presentation of data demonstrates that it was not the most able who were the most represented.

# 4.29 Section 5: Interviews with research assistants

The premise of this research study is to explore whether reading skills which are developed through the reading of music can be transferred to the early reading of text. No formal pre- or post-testing of music reading was undertaken, because during my initial presentation of the programme to the children, where the three music symbols (crochet rest quaver) were shown but not explained, no child could suggest what they might be. From an ethical point of view, it seemed counterproductive to try to test children in an area where they apparently had no prior experience – such tests could easily discourage children from participation and have a detrimental effect on self-esteem.

Instead, three of the research assistants were interviewed in a semi-structured way once the intervention was complete; to ascertain whether there was evidence that progress in music reading had been made. Five questions formed the basic structure of the interviews, and these, together with a summary of responses, are presented as follows.

(J, E & G = Research Assistants 1,2 &3 respectively: I = interviewer)

4.30 Question 1, Think back to the very first lessons that the children received. What were your initial impressions of the children’s level of recognition of the three music symbols (crotchets, quavers and rests) during the first lessons?

J: I don’t think any of them could.

E: Maybe one or two said something like they had seen a recorder before, and might have done it at home, but, I don’t know, not really.

G: Yeah, I can’t remember any one in particular being able to.

J: No, not the symbols, no. I didn’t get the impression that any of them knew any of the symbols before they started.

4.31 Question 2, As the children continued with their lessons, can you think of any evidence that would suggest that it was clear that their understanding of reading rhythms was improving? (If the opposite is true, please say so, and why you think that).

G: In terms of, like, rhythmically, I think they grasped that a lot more than, note reading, in my opinion, I think.

E: Yeah, and the idea of rests. I think the crotchets and stuff they could get.

J: With the clapping, definitely, being able to follow the quickies. And I think they enjoyed that, especially the notes having their names, the quickies and the ones.

G: At the very beginning they couldn’t do that at all, but I’d say it got better, wouldn’t you?

E: Yeah, I would.

J: I would say that by the end they could all do it. I don’t think there were any children who couldn’t. Not that I noticed anyway.

G: Definitely. Yeah. I would say that improved more than note reading.

I: What makes you say that?

G: Because when some of them were sort of blowing aimlessly into a recorder, they still pretty much kept the rhythm, whereas I noticed more that they weren’t really playing the right notes, but sometimes, like, the no fingers on any holes playing, they could keep the rhythm.

4.32 Question 3, As the children continued with their lessons, can you think of any evidence that would suggest that it was clear that their understanding of reading the notes on the stave was improving (B, A, G)? (Again, if the opposite is true, please say so, and why you think that).

J: I would say that most of them could do it. We would put the notes up on the board, then point to different notes, and most of them got it right every time by the end of it. It’s hard to tell when you’ve got a whole group, because you never quite know if they are just following everyone else, but the general impression was that they all got it, because they would all call out the right note.

I: To be clear, did they bring any recognition with them of recognising the notes B, A, or G to the programme, on lesson 1?

G & E, together: No (emphatic).

I: Unequivocally, no?

G: No.

E: No.

I; So by lesson 2? 3? 4? 5? Are they quite accurately recognising the notes?  
 E: Yeah, like, in the warm-up we’d have the notes for B, A, G written on the board, and then they would have to tell us, and we would keep pointing to a different one, and they would nearly always get the right one.

4.33 Question 4, How easy or difficult was it for the children to transfer their knowledge of reading music to the playing of the recorder, and what make you say that?

G: I think that at the beginning of each lesson when we went through the clapping and refreshed everything we would always go over the notes and the fingering of them, and I can remember that they could always pretty much tell us what was what, but then when it came to actually playing it they all got a bit too excited and just went for anything didn’t they? I think that some of it was because their fingers were too short. I know that sounds silly, but I think that some of them struggled really hard to cover some of the holes.

E: I think that they know that one different note meant that you had to change fingers, yeah, each note you had to do something different.

J: Some could play, they could follow the music, and they could play the recognisable tunes, yes, definitely, definitely. To be honest with you I’m not sure how many, because it’s not something that I think we were looking for.

4.34 Question 5, How easy or difficult was it for the children to transfer their knowledge of reading music to the clapping of rhythms, and what makes you say that?

E: Well, at the start, they couldn’t read them.

G: Yeah, I would say that none of them could read any of it.

I: And what proportion of them by the end, were clapping correctly ... roughly ...

E; Nearly all. That was probably the most successful part of it I think, reading the rhythms of the music

G: Yeah, I would say about 85% of them. Well, actually they could all do it.

J: They certainly enjoyed doing the clapping, and doing the different combinations of notes.

It was clear from these exchanges that the participating children, as a cohort, brought with them no discernible prior learning of music reading to the programme, but all made identifiable progress in the reading and clapping of rhythms, and many were successful in recognising the notes on the staff. Many, but by no means all were successful in transferring this new reading skill to the playing of the recorder. This conformed to the responses of the children, who found these same categories to be either accessible or difficult.

## 4.35 Summary and conclusions

This study has presented clear, identifiable and measurable evidence that, in the case of a given cohort of 55 Year 1 children, reading skills that were developed through the reading of music were transferred to the decoding of texts at grapheme and single-word level. With the exception of single-grapheme recognition, every indicator, in every phase of the study, disclosed a positive effect, although the lack of the study’s power has resulted in wide confidence intervals of effect size, each of which straddled 0. The possible reasons for the single-grapheme-recognition anomaly have already been rehearsed (in section 4.2). Not all of the five categories of focus in the RCT can claim statistical significance, but two can, and consistently positive effect sizes have been calculated. The same is true for phase 3, in which a small but statistically significant effect was disclosed. From comments made by the children and the research assistants, it is possible that, had the study been chronologically longer, a stronger effect might have been shown to be present.

Interviews with the children showed that they understood and could reproduce the taught vocabulary, and enjoyed the programme. Generally they found the reading of exclusively rhythmic phrases easier than melodic ones, but could recognise and interpret both. For many, however, the kinaesthetic process of accurately converting their knowledge of rhythmic and melodic manuscripts to the playing of the recorder was problematic. Nonetheless, this inability to consistently make pleasing music did not invalidate the programme, since this was a cognitive study, not a physical one. It is clear from the interviews with research assistants, and from the analysis of the progressive nature of the pieces the children were asked to learn (which were tailored to their needs), that considerable progress was made in the recognition and interpretation of formal music notation. The intervention was organised in such a way that this new skill (for new it clearly was, for all participants) was the only significant variable for the duration of the programme, and so there can be no other explanation for the consistent and measurable progress of the participating children’s enhanced development in the recognition of clusters, digraphs and nonsense words across almost all aspects of the programme. The study lacks power and, as a consequence, statistical significance, but it has disclosed measurable and consistently positive effects.

# Chapter 5: Conclusion

This chapter will recap on where I started, consider what has changed, and as a consequence, identify what has been learned. The limitations of the programme will be identified. In recapping, I first return to Hart’s (1998) view that a doctoral thesis should serve two purposes - “*to develop the skills of the researcher, [and to have] a public dimension”* (Hart, 1998: 27). This thesis has been successful on both counts, and I shall explore each of these in turn.

## 5.1 The context of the study

This thesis has been eleven years in the making, and it has grown with unexpected rigour throughout that time. Initially, whilst I was a none-too-academic Key Stage 1 Primary school teacher and music specialist, the research question lurked quietly and undefined as a series of semi-connected, semi-articulated musings in my mind. Then it germinated and found thwarted direction as the intended subject for a final MA dissertation, vetoed by tutors (I now see, rightly) on the grounds that it was too big a question to be accommodated and answered satisfactorily at that level. Finally, I embarked upon a doctorate programme, and ultimately organised a full-blown randomised controlled trial, which actively involved one head teacher, one deputy-head teacher, two host teachers, five research assistants, five independent testers, three academic tutors, 55 children, 24 interviews, more reading than I thought was possible, and the production of a doctoral thesis. I arranged and did all this in order to satisfy the insistent question which over the eleven years had grown and changed in my consciousness beyond the level of an interested, professional, theoretical whimsy, becoming instead a demanding Tantalus, hungry to be satisfied. At the start of these eleven years, I had not anticipated any aspect of what has subsequently happened, and I am still surprised by the scale and speed of the development.

The title of my study is “Can reading skills that are developed through the reading of music be transferred to benefit the early decoding of text?”, and I have demonstrated that they can. I have also demonstrated that the children enjoyed the experience, and could articulate their attitudes, frustrations and enthusiasms, and those who were interviewed were happy to participate. Some children were able to make insightful comparative analyses on what it is to read, and what it is to be a reader of each genre.

It would be folly then to focus merely on the calculations of effect size. It is true that the calculations are eye-catching headlines, constituting easily-remembered sound bites. Certainly, the production of the data for these calculations was the most labour-intensive aspect of the study, but that should not blind us to the patterns in data which were gleaned in a less physically-demanding way. Calculations do not tell the whole story. I was intentionally careful to employ mixed methods, so that my study should not become so reductive that it produced no more than a list of effect sizes. Mathematical equations are solid, unequivocal, and have the appearance of permanence, but they may not always reveal the things of most significance. One is reminded of Adams’ (1979) iconic moment in science-fiction literature:

The answer to the great question ... [said Deep Thought]

(Yes ...?!)

Is ...

(Yes ...?!!)

Is ...

(Yes ...?!!!)

Forty-Two.

...

...

It was a long time before anyone spoke.

(Adams, 1979: 128)

It would be an example of a misplaced opportunity if, after all the effort on the part of all the people involved in this study, the answer to my great question was ‘*d* = 0.29’. The answer to my question is very effectively illustrated, but not confined by, that statement. This is a study of shared experiences in which children’s learning of the reading of music was shown to contribute to their reading development as a cohort. It was as a cohort that they learnt, and as a cohort that they engaged in “*literacy events*” (Barton et al., 2000). It is not the case that I have proved that in every cohort these skills will transfer, but I have shown that in one situation, with a given body of participant children, analysed as a cohort, these skills were developed, and did transfer, and the experience for them was a positive one.

It cannot be claimed that what I have done constitutes an entirely original contribution to human knowledge. Several scholars and educationalists had already considered similar questions (principally Gromko, 2005; Standley and Hughes, 1997; Butzlaff, 2000; and Hansen et al., 2007), but these writers asked their similar questions in America. Their research questions were therefore formulated in response to prevailing American educational culture and conditions. Their data was collected in American schools or in American Early Years settings. Their theoretical underpinning was drawn from what appears to be a tightly-knit community of American researchers in the field, and their findings were all published in American academic journals or by American publishing houses. Their interventions were conducted within the ethos and conventions of individual American schools, and their tests of reading or grapheme recognition were conducted by predominantly American researchers, with participating children using (predominantly) American accents. Many of the tests that these researchers employed as assessment tools were standard American tests, designed to reveal aspects of reading of interest to American institutions and education authorities.

What I have done is to recreate and adapt this existing American research, identifying and improving the research designs of others where I considered improvements might be made. I have presented and summarised the (sparse) published writing that currently exists on the subject of the relationship between music and the development of early reading, and the identified effect that the one can have on the other, and I have applied all this to a British setting. Whilst I have described and taken into account the current phonics-centric ethos which currently pertains throughout British Primary and Early Years education, I have consciously and diligently refused to support or deny the insistent current clarion call that the teaching of phonics is the ‘best way’ to teach early reading. I merely report that in the case of 50 children in a school in North West Kent, the formal teaching of the reading of music had a measurably positive effect on their phonemic decoding of text.

## 5.2 The skills of the researcher

With regard to the skills of the researcher, through the generation of this thesis I have developed on a number of fronts, not least amongst which is a developing understanding of quantitative methods, data analysis and statistical terminology, and the limitations of the study come into relief here. It was unfortunate that I could not afford a larger-scale quantitative approach (in terms of time, and volunteer personnel). In hindsight, a larger cohort would have made the analysis more valid and statistically more significant. In spite of Cohen et al’s (2007: 101) advice that “*A sample size of thirty is held by many to be the minimum number of cases if researchers plan to use some form of statistical analysis in their data*”, I feel disappointingly misinformed by this. I am now aware, for example, that “*to have an 80% chance of detecting an effect size of 0.5 (half a standard deviation) significant at the 5% level, a minimum of 64 subjects in each arm of a two-arm trial is required*” (Greg Brooks, personal communication 2011). Hutchison and Styles (2010: 17) offer similar advice. Before appreciating this I had felt that my sample of *n* = 50 was more than sufficient to produce compelling data and incontrovertible analysis, and although it is true that my actual analysis of the available data identified a calculated effect with regard to a range of aspects of decoding, because of the small sample size these calculations do not all reach statistical significance.

As far as the logistics of the project were concerned, I wish at the outset I had been more aware of the implications that missing data would have on the sample size of the cohort, and the ethics of dealing with missing data. In the first few days of the programme, when permission slips were being sent and received, I was insufficiently mindful that, since the design of an honest RCT demands random allocation of pairs, one missing participant will consequently, and always, result in two. I did realise at the beginning that, if one is to have groups which are “*well-matched at base-line*” (Torgerson et al., 2003: 245) the numbers must match as well as the characteristics of the constituent members, but from an ethical perspective, my initial assumption was that, having randomly allocated the children into control and intervention groups, the pairs could subsequently be rearranged to minimise the effect of ineligible children.

In this I was shown to be substantially in error. The majority of published work on missing data (for example, Gyimah, 2001; Vach, 1994) refers to large-scale studies where a single example of missing data represents less than 1% of whole. This is a further, if indirect, reference to the need for a cohort of perhaps 128 or more (two arms of 64 participants in each). In a potential cohort of 60, a single child who declines to take part represents 1.66% of the cohort, and four wanting to take part but failing to return permission slips constitutes a further 6.66%. Combined they make over 8% of the cohort, and when removed with their pairs result in almost a 17% reduction in available participants. Thus, I inadvertently found that the attrition rate was running away from me in the first two weeks of the programme. A more experienced researcher might have paid more attention to this. Neither Gyimah nor Vach have anything to say on a 17% level of missing data, other than to suggest that “*you generate unbiased data with the view of making valid inferences and conclusions ... if necessary making pairwise deletions”* (Gyimah, 2001: 4). ‘Pairwise deletions’ based upon the original randomised allocations were accordingly made from all examples of numeric data, and it hurt to do so. Fortunately, no additional losses were suffered as the study subsequently progressed.

In my defense, the size of my research design was not only limited by time (mine, and that of research assistants) and by cohort availability, but also by the size of existing research programmes that influenced my thinking. I established in Chapter 2 that there are few research studies in the field with which to compare mine, but those that have been published have not been populated by as many as 128 participants. Gromko’s (2005) was the largest study (*n* = 103), but these participants were taken from two separate schools, with groups of unequal size, thus engendering multiple variables. Standley and Hughes (1997, *n* = 24), was smaller than my own (but with engineered well-matched pairs). Each of these writers produced work deemed to be considered worthy of publication in a peer-reviewed journal. Long (2007, *n* = 24) produced a PhD thesis deemed worthy of the name. My work is therefore couched amongst other independently-produced published pieces of similar size. Had a selection of larger examples of relevant research studies been forthcoming in my review of literature, I might well have been swayed to attempt a larger-scale study (although the logistics of that for an independent researcher are very daunting, possibly even prohibitive). Having said that, even if I had managed to secure permission slips from all of the children, or had somehow secured twice as many participants, this would still have fallen a long way short of the 278 projected by Schoenfeld, (2012) required for the study to have achieved statistical significance.

## 5.3 Possible adjustments to the tests

The study also held limitations with regard to the shape of the distribution curves produced on the frequency diagrams, particularly with regard to the recognition of nonsense words, and single-letter grapheme recognition. Not much could be done about the distributions for single-letter grapheme recognition because of the advanced level of previous knowledge that the children brought with them to this aspect of the programme. However, in hindsight, for the nonsense words it might have been helpful to have one or more sections of the tests in which the children were asked to read words such as ‘red’ and ‘egg’, in order to identify those children who recognised high-frequency words by sight, but who could not yet freely blend and segment nonsense words. Perhaps the nonsense words being assessed could have been interspersed amongst un-assessed common words of identical syllabic complexity and orthographic depth (Seymour et al., 2003). This might have helped those (few) children who, having taken one look at the nonsense words, announced that they could or would not read them because “They are not real words.” It is clear from this immediate reaction that whole-word recognition is a common way for children to interpret and blend multiple graphemes, since such children must have ‘read’ or engaged with the word in some way in order to be able to pronounce judgement upon it as ‘not being real’.

This research is focused on decoding skills, not whole-word recognition, but any research that involves itself with early reading development must acknowledge that whole-word recognition is a legitimate method used by many children to read high-frequency words; ‘*tricky words*’ (DfES, 2007b: 194); or those words which have a passing physical resemblance to the object they represent (such as ‘aeroplane’). However, methodologically, the inclusion of recognisable high-frequency words before or during the reading of the nonsense words might have served to put children at their ease, and thus perform more comfortably in the tests. This might have produced a more ‘bell-like’ curve on the frequency diagrams.

Similarly, when asking children to decode the clusters it would have been useful to determine whether or not the letters that made up those clusters (dr, fl) were letters known to those children who struggled on the single-grapheme recognition section of the test. It is obviously unfair to expect a child who has failed to recognise either ‘d’ or ‘r’ to decode the cluster ‘dr’. It would be better to create, during the test, four possible clusters based upon known skills of single-letter grapheme recognition.

## 5.4 The public dimension

In terms of a public dimension (Hart 1998), the findings of my study did not conform closely to those of Gromko (2005), despite sharing the focus. There was some symmetry with regard to cluster and digraph recognition, in that this area enjoyed the most noticeable effect size in both studies, but there was less correlation in all other areas. Gromko would consider this an unsurprising outcome because she herself had anticipated that “*the development of aural perception would lead to significant gains in the development of young children's phonemic awareness, particularly in their phoneme-segmentation fluency, which is an aural skill*” (Gromko, 2005: 205). This observation that the ability to blend and segment individual sounds is aural in its origin is shared by Campbell (2008); Kodaly (in Campbell, 2008); and Hansen et al. (2007).

Gromko’s is the only published study closely related to my own, but although this closeness of resemblance in terms of focus was coincidental and happily found, the divergence in method, particularly in the use of an RCT in this study, was intentional on my part. However, I was also pleased to explore the notion of near-transfer hypothesis, which Gromko mentioned, but did not expand upon.

The public dimension has been, and is, on-going. I presented my research proposal at the UKLA international conference in Greenwich in 2009, and although this final thesis contains many procedural and methodological changes since then, some are as a result of that presentation. I have been approached by two publishing houses with regard to publishing the findings. The first is UKLA, who have invited a submission to be considered for their peer-reviewed *Journal of Research in Reading*. The second is *Sage*, who in the light of the Henley report (DfE, 2011) have expressed an interest in producing a book on the relationship between Music and Literacy, similar to that of Hansen et al. (2007), but for a British audience. It has yet to be decided whether this is to be an individual or collaborative enterprise. I would also like to publish a comparison between my own work and Gromko’s (2005) in the American *Journal of Research in Music Education* (in which Gromko has been regularly published), and to make an exploration of Barnett and Ceci’s (2002) “*near-transfer hypothesis*” (Gromko, 2005: 201) in UKLA’s other peer-reviewed journal *Literacy*. However, I have not yet explored these last two possibilities, and these are not therefore commissioned submissions. Nonetheless, I feel they would each be worthwhile articles, of interest to the readership of those specific journals.

## 5.5 The limitations of the study

I have already explored, to some extent, the limitations of this study. It is clear that from a statistical significance point of view the study does not allow generalisation to be made. But whilst considering the public dimension, it is helpful to return to Gromko (2005), who, when considering the limitations of her own study, wrote:

Although readers should interpret these results with the limitations of this study in mind, I believe that these results have sufficient credibility to serve as the basis for continued inquiry into near-transfer effects of music instruction. … A rigorous test of the hypothesis that children may benefit in a second symbol system (e.g., language) from the exercise of a first one (e.g., music) will require a particular kind of music instruction, one that involves active music-making and the association of sound with developmentally appropriate symbols. The implication for schools is that music instruction, while valuable for liberating the artistic and musical potential of every child, may significantly enhance children's language literacy as well.

(Gromko 2005: 205)

I have done exactly what Gromko considered might be a next step in the exploration of the field. I have taken up her challenge and have responded to it in the way she suggested, and this is why I believe the editors of the *Journal of Research in Music Education,* and, I would hope, Gromko herself, would be interested in my findings. Mine has been a rigorous test, identifying the measurable benefits on a second symbol system (phoneme recognition) from the exercise of a first (the reading of simple formal notation).

## 5.6 Other factors which may have contributed to the effect

Given the calculated effect size, it is necessary to consider what factors, other than the intervention, may have contributed to the effect. I was as diligent as possible in controlling possibly confounding variables, but inevitably, over a six-week period, extraneous variables will have an influence. The children may have had additional reading experiences at home (although the parents were asked not to do this, see appendix 5), but, even so, it is as likely that this might affect a child in the control group as in the intervention. The attendance records of the control and intervention group, reported in chapter 3 (section 3.8), were very good, and are not noticeably different, so there is no implication there.

Gromko cited three possible alternative explanations for the effect in her study – the systematic differences of literacy provision enjoyed by the separate control and intervention groups (not a factor present in my study); the difference in starting scores between the two groups (again, not applicable to my study); and the Hawthorne effect, defined as “*the initial improvement in performance following any newly introduced change*” (Wellington, 2000: 197). This last suggestion is a possibility which I was unable to check against, since the intervention in itself was somewhat invasive for the host teachers, and I felt unable to ask them to allow an additional change to their routine. It may be that the effect is due simply to the intervention group spending time in a half-class setting, and then in small groups with the five very enthusiastic research assistants. This possibility could only have been catered for by enlisting further research assistants to enthuse the control group in whatever activities they were undertaking in the classroom in the absence of the intervention group, but even then it would be impossible to say whether that in itself would constitute the introduction of a further intervention and additional variable requiring very close control of the alternative intervention. An even more rigorous design to cope with this would have been to recruit not only a no treatment control group (as I did), but also a Hawthorn Effect control group (as just outlined) – again, unrealistically demanding.

## 5.7 Further research

This study has raised as many questions as it has 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 consider whether children’s reading of text can be enhanced by giving them focused and guided opportunities to articulate their individual thoughts on the reading acquisition process. It seemed to me, from the interviews that were conducted with the children, that several connections and realisations were made through the articulation of their experiences of learning to read. This need not be a piece of research peculiar to the literacy/music connection, but would be of interest to anyone seeking to undertake research in the field of early reading.

One might also 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 again the interviews with the children tended to show that, generally, they found the fine motor skills of playing an instrument problematic. I am reminded of Shefki [child 25] who spoke for the many when he observed “*Reading books is easier than reading music, because you don’t have to play an instrument while you are doing it.*” For the majority of the children, the physical understanding of how to play the recorder was evident, and the connection had been made between the contours of the notation and the mechanics of the instrument ( “*on note 1 [B] I put one finger down, the two, I put two fingers down, and three I put three fingers down. It’s easy*” (Meheni, child 52), 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, I have applied two interventions, the clapping of rhythms and the playing of an instrument, where one intervention might have been more effective.

Approaches very different from my own might also have been considered. Potential for exploring the implication of connections between music and literacy for children with dyslexia was been rehearsed in section 2.18. Alternatively, studies have been made of reading development which take a neurological approach. For example, 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 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, and (if a budget allowed) taking brain scans at the same time.

One might also explore the use of background music on children’s reading. Ziv and Goshen (2006), explored the effect on children’s perception of a story read to them, when happy, sad, and no music was played in the background. It would be interesting to discover whether the heightened effects of background music revealed by Ziv and Goshen were maintained when children of various ages were reading independently.

## And finally ...

The formulation of this thesis was of considerable benefit to everyone who participated in it. The children enjoyed participating, and they benefited both musically, and also (demonstrably) in their early reading skills. Research assistants developed their understanding of research procedures and ethics for their own research projects; as did I. But, most importantly for me, I have at last satisfactorily and, I am pleased to report, positively answered the question that has exercised me for over a decade. It has been a long journey, and I am grateful to all those who have supported and contributed, because I could not have done this on my own.

Mark Betteney

Bexleyheath

August, 2012

# References

Adams, D. (1979) *The Hitchhiker’s Guide to the Galaxy* Basingstoke and Oxford, England: Heinemann

Atterbury, B. W. (1985) ‘Musical differences in learning-disabled and normal-achieving readers, aged seven, eight and nine’ *Psychology of Music*, 13, 114-123 DOI: 10.1177/0305735685132005

Barnett, S. M., and Ceci, S.J. (2002) ‘When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer’ *Psychological Bulletin* Vol 128, No 4, 612-637 DOI: [10.1037/0033-2909.128.4.612](http://psycnet.apa.org/doi/10.1037/0033-2909.128.4.612)

Barr, R., Kamil, M.L., Mosenthal, P.B. and Pearson, P.D. (1996) *Handbook of Reading Research, Volume II*  Mahwah, New Jersey: Routledge

Barrs, M. (1991) *The Reading Book* London, England: CLEP

Barton, D., Hamilton, M., and Ivanic, R. (2000) (Eds.), *Situated literacies: Reading and writing in context* New York: Routledge.

Baumrind, D. (1964) ‘Some thoughts on ethics of research: After reading Milgram's "Behavioral Study of Obedience."’ *American Psychologist*, Vol 19(6), pp421-423. DOI: [10.1037/h0040128](http://psycnet.apa.org/doi/10.1037/h0040128)

Beach, K. (2004) ‘Consequential Transitions: A Sociocultural Expedition beyond Transfer in Education’ *Review of Research in Education* Vol 24 pp101-139

Retrieved from http://www.jstor.org/stable/1167268

Bentley, A. (1966) *Measures of Musical Ability* London, England: Harrap

Besson M., and Friederici A. (2005) ‘Part II: Language and Music—A Comparison’ *Annals of the New York Acadamy of Sciences*

DOI: 10.1196/annals.1360.061

Bolduc, J. (2008) The Effects of Music Instruction on Emergent Literacy Capacities among Preschool Children: A Literature Review *Early Childhood Research and Practice* Vol 10 No 1, no page numbers available.

Retrieved from <http://ecrp.uiuc.edu/v10n1/bolduc.html>

Bradley, L. and Bryant, P. E. (1978) ‘Difficulties in auditory organisation as a possible cause for reading backwardness’ *Nature* 271, 746-747

DOI: 10.1038/271746a0

Bransford, J.D., and Schwartz, D.L (1999) *Review of Research in Education* Vol. 24 pp. 61-100 Retrieved from http://www.jstor.org/stable/1167267

Breier, J. I., Gray, L., Fletcher, J. M., Diehl, R. L., Klass, P. and Foorman, B. R.,(2001). Perception of voice and tone onset time continua in children with dyslexia with and without attention deficit/hyperactivity disorder *Journal of Experimental Child Psychology,* 80(3), 245-270.

DOI :10.1006/jecp.2001.2630

Brice Heath S (1983) *Ways With Words: Language, life, and work in communities and classrooms* Cambridge, England: Cambridge University Press

Broudy, H. S. (1977). Types of knowledge and purposes of education. In Anderson, R.C., Spiro, R.J., and Montague, W.E. (Eds.), *Schooling and the acquisition of knowledge* (pp. 1-17). Hillsdale, New Jersey: Erlbaum.

Brown, A. and Dowling, P. (1998) *Doing Research/Reading Research: A Mode of interrogation for Education* London, England: Falmer

Bruner, J. S. (1960) *The Process of Education* London, England: Oxford University Press

Bruner, J. S. (1986) *Actual Minds: Possible Worlds* London, England: Harvard University Press

Bryant, P. E., and Bradley, L. (1985). *Children’s readingproblems* Oxford: Basil Blackwell.

Burns, S.M., Griffin, P., and Snow, C. E. (1999) *Starting out right: a guide to promoting children’s reading success* Washington DC: National Academy Press

Butzlaff, R. (2000) ‘Can Music Be Used to Teach Reading’ *Journal of Aesthetic Education* vol. 34, 167-178 Retrieved from http://www.jstor.org/stable/3333642

Central Advisory Council for Education (1967) *Children and their Families (The Plowden Report)* London, England: HMSO

Caldwell, T. (1993) ‘A Dalcroze perspective on skills for learning’ *Music Educators Journal* Vol 79, No 7 pp 27-29

Retrieved from http://www.jstor.org/pss/3398612

Campbell, P. S. (2008) (Ed) *Musician and Teacher: An orientation to Music Education* New York: Norton

Campbell, D. T. and Stanley, J. (1963) *Experimental and Quasi-Experimental designs for Research on Teaching* Boston MA: Houghton Mifflin

Clay, M.M. (1985) *The Early Detection of Reading Difficulties* Auckland, New Zealand: Heinemann

ClinTools Software(2012) *Effect Size Generator* Retrieved from http://www.clintools.com/victims/resources/software/effectsize/effect\_size\_generator.html

Coe, R. (2002) ‘*It's the Effect Size, Stupid: What effect size is and why it is important*’ Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, September 2002

Coffey, A. and Atkinson, P. (1996) *Making sense of Qualitative Data Analysis: Complementary Strategies* Thousand Oaks, CA: Sage

Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd Ed) Hillsdale, New Jersey: Lawrence Erlbaum Associates:

Cohen, J. (1969) *Statistical Power Analysis for the Behavioural Sciences* Hillsdale, New Jersey: Lawrence Erlbaum Associates

Cohen, L., Manion, L., and Morrison, K. (2007) *Research methods in Education (6th Edition)* London, England and New York: RouteledgeFalmer

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, England: Centre for Education and Inclusion Research Retrieved from https://www.education.gov.uk/publications/eOrderingDownload/DFE-RB159.pdf

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

Corriveau, K.H., and Goswami, U. (2009) ‘Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat’ *Cortex* 45 (1) 119-130

DOI:10.1016/j.cortex.2007.09.008

Cousin, G. (2010) *Global Citizenship as a Graduate Attribute* Keynote lecture, Metropolitan University Learning and Teaching conference, 8th July 2010, London, England

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

DCSF (2009) *Independent Review of the Primary Curriculum: Final Report* London, England: HMSO

DCSF (2008) *Practice Guidance for the Early Years Foundation Stage* London, England: HMSO

DCSF (2007a) *The Early Years Foundation Stage* London, England: HMSO

DCSF (2007b) *What works for pupils with literacy problems? The effectiveness of intervention schemes* London, England: HMSO

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

DES (1985) *Curriculum Matters 4: Music from 5 to 16* London, England: HMSO

Detterman, D. L. (1993) The case for the prosecution: Transfer as epiphenomenon, in D. K. Detterman and R. J. Sternberg (1993) (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, New Jersey: Ablex.

Detterman, D. K., and Sternberg, R. J. (Eds.). (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, New Jersey: Ablex.

Dewey, J. (1916) *Democracy and Education. An introduction to the philosophy of education (1966 edn.)*, New York: Free Press

DfE (2011a) *Music Education in England* London, England: HMSO

DfE (2011b) *Statistical First Release, Schools, Pupils, and their Characteristics, SFR 12* National Statistics

Retrieved from http://www.education.gov.uk/rsgateway/DB/SFR/s001012/sfr12-2011.pdf

DfE (2010) *‘The Importance of Teaching - The Schools White Paper’* London, England: HMSO

DfEE (1999) *The National Curriculum: Handbook for Primary Teachers in England and Wales.* London, England: DfEE Publications

DfEE (1998) *The National Literacy Strategy* London, England: DfEE Publications

DfES (2007a) *Letters and Sounds: Principles and Practice of High Quality Phonics – notes of guidance for practitioners and teachers* London, England: DfES Publications

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

DfES (2006a) *Primary Framework for literacy and mathematics* London, England: DfES Publications

DfES (2006b) *Independent Review of the Teaching of Early Reading: Final Report* London, England: DfES Publications

Dodge, Y. (1985) *Analysis of Experiments with Missing Data*. New York: Willey Series in Probability and Mathematical Statistics

Douglas,S. and Willats, P. (1994) ‘The relationship between Musical Ability and Literacy Skills’ *Journal of Research in Reading*, 17 (2) 99-107

DOI: 10.1111/j.1467-9817.1994.tb00057.x

Duffy, M.E. (2006) ‘Handling Missing Data: A Commonly Encountered Problem In Quantitative Research’ *Clinical Nurse Specialist* 20(6), pp 273-276

Retrieved from http://ovidsp.tx.ovid.com.eresources.shef.ac.uk/sp-3.4.1b/ovidweb.cgi?&S=MEHDFPEJAKDDJKCBNCBLOCGCLOJOAA00&Link+Set=S.sh.15.16.19.23.27.31.35.39.43.46%7c5%7csl\_10

Easton J.E. and McColl J.H. (2012) Statistics Glossary vol 1.1, STEPS Retrieved from http://www.stats.gla.ac.uk/steps/glossary/confidence\_intervals.html

Ericson, L., and Scjuliebo, M. F. (1998). The phonological awareness handbook for kindergarten and primary teachers. Newark, DE: International Reading Association

Fetzer, L. (1994). Facilitating print awareness and literacy development with familiar children's songs. (Unpublished doctoral dissertation, East Texas State University [now Texas A & M University], Commerce, TX, 1994).

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., and Winner, E. (2008) ‘The relation between music and phonological processing in normal-reading children and children with dyslexia’ *Music perception* 25 (4) 383-390

DOI: 10.1525/mp.2008.25.4.383

Fountas, L, and Pinnell, G. (1999). Voiceon word matters: Good first teaching for all children. Portsmouth, NH: Heinemann.

Frankel, J.R. and Wallen, N.E. (2003) *How to design and evaluate research in education (5th edition)* Singapore: McGraw-Hill

Frankel, J.R. and Wallen, N.E. (1993) *How to design and evaluate research in education (2nd edition)* Singapore: McGraw-Hill

Ganschow, L., Lloyd-Jones, J., and Miles, T.R. (1994) ‘Dyslexia and Music notation’ *Annals of Dyslexia* 44 (1) 185-202

DOI: 10.1007/BF02648161

Gault, B. (2005) ‘Music Learning through all the channels’ *General Music Today* Vol 19 Issue 1, 7-9 DOI: 10.1177/10483713050190010103

Gerber, A. and Malhotra, N. (2008) ‘Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals’ *Quarterly Journal of Political Science* 3: 313–326 DOI: 10.1561/100.00008024

Glover, J. and Young, S. (1999) *Primary Music: later years* London, England: Falmer Press

Gorard, S. (2001) *Quantitative Methods in Educational Research: The role of the Numbers Made Easy* London, England: Continuum

Goswami, U. (2008) ‘Reading, complexity and the brain’ *Literacy* 42(2) 67-74

DOI: DOI: 10.1111/j.1741-4369.2008.00484.x

Goswami, U. (2005) ‘Synthetic Phonics and Learning to Read: A Cross-language Perspective’ *Educational Psychology in Practice* vol 21, Issue 4, pp 273 - 282

DOI: 10.1080/02667360500344823

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) 10911-10916

DOI: 10.1073/pnas.122368599

Goswami, U, (1995) ‘Phonological development and reading by analogy: what is analogy, and what is it not?’ *Journal of Research in Reading*, Vol 18 (2) pp139-145

DOI: 10.1111/j.1467-9817.1995.tb00080.x

Gray, D.E. (2009) *Doing Research in the Real World* (2nd Ed) London, England: Sage

Gromko, J. (2005) The effect of Music Instruction on Phonemic Awareness in Beginning Readers *Journal of Research in Music Education* Vol 53 (3) 199 – 209

DOI: 10.1177/002242940505300302

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

Retrieved from http://www.jstor.org/stable/3345521

Gyimah, S. O-M. (2001) "Missing Data in Quantitative Social Research," PSC Discussion Papers Series: Vol. 15: Iss. 14, Article 1. pp 1-28  
Retrieved from: <http://ir.lib.uwo.ca/pscpapers/vol15/iss14/1>

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* Vol 28 Issue 3 pp269-289 DOI: 10.1177/025576141037065

Hämäläinen, J., Leppänen, P.H.T., Torppa, M., Müller, K., and Lyytinen, H. (2005)‘Detection of sound rise time by adults with dyslexia’ *Brain and Language* 94 (1) 32-42

DOI: 10.1016.j.bandl.2004.11.005

Hannon P (1987) ‘A study of the effects of parental involvement in the teaching of reading on children’s reading test performance’ *British Journal of Education Psychology*, 57, (1) 56-72 DOI: 10.1111/j.2044-8279.1987.tb03061.x

Hansen, A. (2011) *Children’s Errors in Mathematics* (2nd edition) Exeter, England: Learning Matters

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

Hansen, D. and Bernstorf, E. (2002) ‘Linking music learning to reading instruction’ *Music Educators Journal*; 88, (5) 17 – 25

Retrieved from http://www.jstor.org/stable/3399821

Hart, C. (1998) *Doing a literature review*. London, England: Sage

Hartung J., Knapp G., and Sinha B.K. (2008). *Statistical Meta-Analysis with Application*. Hoboken, New Jersey: Wiley.

Hill, E.L. (2001) ‘Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments’ *International Journal*

*of Language and Communication Disorders* 36 (2), 149–171

DOI: 10.1080/13682820010019874

Hodges, L. (1984) ‘Estimation of Effect Size under Nonrandom Sampling: The Effects of Censoring Studies Yielding Statistically Insignificant Mean Differences’ *Journal of Educational and Behavioral Statistics* 9 (1) 61-85

DOI: 10.3102/10769986009001061

Howell, D. C. (2011) *Confidence intervals on effect size* Vermont: University of Vermont, Retrieved from http://www.uvm.edu/~dhowell/methods8/Supplements/Confidence%20Intervals%20on%20Effect%20Size.pdf

Howitt, D. and Cramer, D. (2008) *Introduction to SPSS in Psychology (4th Edition)* Harlow, England: Pearson Education,

Hoy, W.K. (2010) *Quantitative Research in Education: A primer* Thousand Oaks, California: Sage

Huss, M., Verney, J.P., Fosker, T., Mead N., and Goswami, U. (2011) ‘Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology’ *Cortex* 47 (6) 674-689

DOI: 10.1016/j.cortex.2010.07.010

Hutchison, D. and Styles, B. (2010). *A Guide to Running Randomised Controlled Trials for Educational Researchers* Slough, England: NFER.

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

Jones, P. and Robson, C. (2008) *Teaching Music in Primary Schools* Exeter, England: LearningMatters

Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review*, 36, 28–42

Kalof, L., Dan, A., and Dietz, T. (2008) *Essentials of Social Research* Maidenhead, England: Open University Press

Kaplan, R.M., and Saccuzzo, D.P. (2009) *Psychological Testing: Principles, Applications and Issues* (7th ed.) Belmont, Canada: Wadsworth CENGAGE Learning

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

**DOI:** 10.1080/0144341930130103

Lanksheare,C. and Knobel, M. (2004) *A Handbook for Teacher Research From Design to Implementation* Maidenhead, England: Open University Press

Likert, R (1932) ‘A technique for the measurement of attitudes’ Archives of Psychology, No 140, in Brown, A. and Dowling, P. (1998) *Doing Research/Reading Research: A Mode of interrogation for Education* London, England: Falmer

Little, R., and Rubin, D. (1989) ‘Analysis of Social Science Data with Missing Values’ *Sociological Methods and Research* 18 (2-3) 292-326

DOI: 10.1177/0049124189018002004

Long, M. (2007) *The effect of a music intervention on the temporal organisation of reading skills* Unpublished PhD, Institute of Education, University of London, England

Madow, W. G., Nissellson H. and Olkin I. (eds., 1983) *Incomplete Data in Sample Surveys* New York: Academic Press

Mallett, M. (2005) *The Primary English Encyclopaedia* London, England: David Fulton

McMahon, O. (1979) ‘The relationship of music discrimination training to reading and associated auditory skills’, in Bolduc, J. (2005)’Pitch processing and phonological awareness’ *Psychomusicology* Vol 19 pp3-14 Retrieved from http://ojs.vre.upei.ca/index.php/psychomusicology/article/viewFile/444/593

Meek, M. (1994) *Learning to Read* London, England: Bodley Head

Miller, A., and Coen, D. (1994). The case for music in the schools. Phi Delta Kappan, 75, 459-461. Retrieved from http://www.jstor.org/pss/20405141

Mills, J. (2009) *Music in the Primary School* (3rd ed.) Oxford, England: Oxford University Press

Mody, M., Studdert-Kennedy, M., and Brady, S. (1997) ‘Speech perception deficits in poor readers: Auditory processing or phonological coding?’ *Journal of Experimental Child Psychology,* 64 (2), 199-231.

DOI: 10.1006/jecp.1996.2343

Moreno, S., Marques, C., Santos, A., Santos, M. Castro, S.L., and Besson, M. (2009) ‘Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity’ *Cerebral Cortex* 19, 712--723

doi:10.1093/cercor/bhn120

Nutbrown, C. and Hannon, P. (2003) ‘Children’s perspectives on family literacy: Methodological issues, findings and implications for practice’*Journal of Early*

*Childhood Literacy* 3(2) 115–145 DOI: 10.1177/14687984030032001

Odam, G. (1995) *The Sounding Symbol: Music Education in Action* Cheltenham, England: Nelson Thornes

OFSTED (2011) *Getting them reading early: guidance and training for inspectors* Manchester: Office of Standards in Education, Children’s Services and Skills

retrieved from http://www.ofsted.gov.uk/resources/getting-them-reading-early

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

DOI: 10.1017/S014271640000984X

Oppenheim, A.N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement* (2nd Edition) London, England: Pinter

Ouvry, M (2004) *Sounds like Playing* London, England: British Association for Early Childhood Education

Page, N. (1995) Music as a way of knowing New York: Stenhouse

Parker, D. H. H. (1970) ‘Musical perception and backwardness in reading’ *Educational Research* 12, (3), 244-246 **DOI:** 10.1080/0013188700120312

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

DOI: 10.1111/1467-9817.00159

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. (A. Rosin, Trans). New York: The Viking Press.

Reedy, D. (2011) BBC News Bulletin, 19.02.2011, retrieved from <http://www.bbc.co.uk/news/education-12509477>

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

Rees, D (1994) *First steps: Reading developmental continuum,* Longman Cheshire, Melbourne, Australia: Education Department of Western Australia

Salomon, G. and Perkins, D. N. (1989) ‘Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon’ *Educational Psychologist* Vol 24 (2), 113-142 **DOI:**10.1207/s15326985ep2402\_1

Schrader, W.B. (1984) *Three studies of SAT-Verbal Item Types*

College Board Report No. 84-7: ETS RR No 64-33 Retrieved from http://professionals.collegeboard.com/profdownload/pdf/RR%2084-7.PDF

Schunk, D. H. (2004). *Learning theories: An educational perspective (4th ed.)*. Upper Saddle River, NEW JERSEY: Merrill/Prentice Hall.

Scott, D. (2006) *Plato’s Meno* Cambridge, England: Cambridge University Press

Seymour, P., Aro, M. and Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143-74.

DOI: 10.1348/000712603321661859

Schoenfeld, D. (2012) *Statistical considerations for a parallel trial where the outcome is a measurement* MGH Mallinckrodt General Clinical Research Center

Retrieved from http://hedwig.mgh.harvard.edu/sample\_size/js/js\_parallel\_quant.html

Shweder, R. A. (1980) ‘Rethinking Culture and personality theory: From genesis and typography to hermeneutics and dynamics (Part 3)’ *Ethos* 8 (1) 60-94

Retrieved from http://www.jstor.org/stable/640137

Skinner, B.F. (1964) *New Scientist* May 21st

Smith, J. (2000) ‘Singing and songwriting support early literacy instruction’ The Reading Teacher, 53 (8) 646-649. Retrieved from http://proquest.umi.com.eresources.shef.ac.uk/pqdweb?index=0&did=54165097&SrchMode=1&sid=2&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1314356371&clientId=29199

Snowling, M. (1987). *Dyslexia: A cognitive developmental perspective*. Oxford: Basil Blackwell.

Spandel, V. and Stiggins, R. (1997) *Creating Writers: Linking writing assessment and instruction* (2nd Ed) New York: Longman

Spychiger, M., Patry, J., Lauper, G., Zimmerman, E. and Weber, E. (1993) ‘Does more music teaching lead to a better social climate?’ In R. Olechowski and G. Svik (eds), *Experimental Research in Teaching and Learning*. Bern, Switzerland: Peter Lang.

Standley, J.M., and Hughes, J.E. (1997) ‘Evaluation of an Early Intervention Music Curriculum for Enhancing Prereading/Writing Skills’ *Music Therapy Perspectives* Vol 15 (2), 79-85. Accessed through the British Library

Swanwick, K. (1988) *Music, Mind, and Education* New York: Routledge

Swanwick, K. and Tillman, J. (1986) ‘The sequence of musical development’ *British Journal of Music Education* 3 (3) 305-39, DOI:10.1017/S0265051700000814

Tabachnick, B. and Fidell, L. (2007) Using Multivariate Statistics. 5th ed. Boston: Pearson Education, Inc

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

DOI: 10.1016/j.tins.2006.06.003

Thompson G.B. and Johnston R.S. (2000) ‘Are nonword and other phonological deficits indicative of a failed reading process?’ *Reading and Writing: An Interdisciplinary Journal* 12: 63–97, Kluwer Academic Publishers. Retrieved from http://www.springerlink.com/content/h341238g00n0987w/

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) 120–129

DOI: 10.1016/jphysparis.2008.03.007

Thorndike, E.L., and Woodworth, R.S., (1901) The efficiency of one mental function upon the efficiency of other functions *Psychological Review* 8 (3), 247-251

Retrieved from http://ovidsp.tx.ovid.com.eresources.shef.ac.uk/sp-3.4.1b/ovidweb.cgi?&S=AGHOFPPHHLDDIKFJNCBLDFLBOOBDAA00&Complete+Reference=S.sh.15.16.21.26%7c2%7c1

Tobin, R. & McInnes, A. (2008) ‘Accommodating differences: variations in differentiated literacy instruction in Grade 2/3 classrooms’ *Literacy*  42 (1), 3-9

DOI: 10.1111/j.1467-9345.2008.00470.x

Torgerson, C.J., Brooks, G., and Hall, J. (2006) *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling.* London, England: Department for Education and Skills Research (Report No. 711)

Torgerson, C.J., Porthouse, J., and Brooks, G. (2003) ‘**A systematic review and meta‐analysis of randomised controlled trials evaluating interventions in adult literacy and numeracy’** *Journal of Research in Reading*, Volume 26 (3), 234–255

DOI: 10.1111/1467-9817.00200

Torgerson, C.J. (2003) *Systematic Reviews* London, England and New York: Continuum

Vach, W. (1994) *Logistic Regression with Missing Values in the Covariates*. New York: Sprnger-Verlag

Wallace, K. (2009) *The Tale of a Tadpole* London, England: Dorling Kindersley

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: England: International Music Research Centre

Wellington, J. (2000) *Educational Research: Contemporary Issues and Practical Approaches* London, England & New York: Continuum

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

**DOI:** 10.1080/09695940802417426

Yopp, H. K., and Yopp, R. H. (1997). Ooples and boo-noo-noos; Songs and activities for phonemic awareness. New York: Harcourt Brace

Yung,B. (Ed) (1997) *Celestial Airs of Antiquity: music of the seven-string zither of China* A-R Editions Inc: USA

Ziv, N., and Goshen, M. (2006) ‘The effect of ‘sad’ and ‘happy’ background music on the interpretation of a story in 5 to 6 year old children’ *British Journal of Music Education* 23(3) 303-314

DOI: 10:1017/S0265051706007078

## Glossary

Anova: Analysis of Variance: A statistical test showing whether or not the mean scores of a number of groups are significantly different. The use of anova avoids the need for multiple *t*-tests (see ‘*t*-test’)

Attrition: the loss of participants during the course of an investigation.

Blending: the act of combining individual phonemes to form a word or a syllable. See also ‘Segmenting’, and ‘Sounding-out (for reading)’

Coda, spoken: the final consonant phoneme or consonant cluster (if any) of a syllable. See also Onset, and Rime.

Crotchet: the symbol used in music to denote a single beat

d = : see ‘Effect size’

Effect size: a measure of the difference in gains of two independent groups. Represented as ‘*d* =’ and expressed in standard deviation units.

Grapheme: a symbol (letter or letters) which denotes an individual speech sound (phoneme). In English, unusually, some graphemes can represent sequences of two phonemes, e.g, the single-letter grapheme <u> in the two-phoneme sequence /ju**ː**/ (as in ‘union’).

Ipsative: a method of assessment which measures change from a given starting point.

Mixed methods: The use of a variety of research types

Musical meter: the structure of a given melody or individual rhythm, often identified by the number of notes

*n* = : the number of participants in a given piece of research, or one of the groups within it.

Onset: that part of a syllable (if any) falling before, but not including, the vowel sound. See also Rime, and Coda

Orthographies: the writing systems of different languages

*p* = : see ‘Statistical significance’

Phoneme: The smallest meaningful unit of sound in a word or syllable.

Phonological processing: the way in which the brain responds to the auditory stimuli of speech sounds. In many theories of reading, phonological processing is also held to mediate the recognition and/or comprehension of written words.

Quaver: a note whose duration is half of that of a crotchet

Rest: a note which denotes a period of silence

Rime: that part of a syllable which falls after the onset. The rime begins with, and may consist solely of, the syllable’s vowel sound

SATs: Standard Assessment Tests (sometimes ‘Tasks’), taken at the end of an academic year

Segmenting: the act of breaking down a spoken word or syllable into its phonemes, also known as ‘sounding out’ (for spelling), Or the act of breaking down a written word or syllable into its graphemes. See also ‘Blending’, and ‘Sounding out’

Semantically: making use of a reader’s knowledge of the world to predict or make sense of a word or phrase, rather than decoding

Sounding out (for reading): the act of pronouncing a phoneme for each grapheme of a written word. Conceptually, follows segmenting, and precedes blending

Standard deviation: a calculation of the variance of a group’s data from the mean.

Statistical significance: a statistical assessment of whether a difference between two (or more) sets of data shows a real effect, rather than just chance. Represented as ‘*p* =’

Syntactically: predicting or making sense of a word through knowledge of the grammar or structure of a sentence or phrase

Synthetic phonics: the identification of the smallest units of sound in a word or syllable, and the blending or segmenting of these sounds to learn to read and write

Systematic synthetic phonics: a way of teaching synthetic phonics, introducing the various phonemes and graphemes in a prescribed order

Taxonomy: A list, or a classification

*t*-test: shows the extent to which the means of two groups are statistically different from each other. See also ‘Anova’

## Appendix 1: Nonsense words (pre-test)

Group 3

|  |  |
| --- | --- |
| Hig, reb, tov  ost, arn, ent  shib, thun, quop  fluz, grad, scug  hibe, vome, yake | c-v-c  v-c-c  Digraph-v-c  Cluster-v-c  Split digraph |

Group 4

|  |  |
| --- | --- |
| nugfim, fetzum, jumdop  fiss, gudd, wobb  leab, waig, joam  tremp, drint, sculp  chish, shoth, phash | c-v-c- x2  Double final consonants  Medial vowel digraphs  Cluster, beginning and end  Digraphs, beginning and end |

Group 5

|  |  |
| --- | --- |
| remar, disfug, prejox  zining, yogful, hitate  luttle, gimming, vesser  ginkly, fornter, duntness  Buntingham, slantering, preflopation | c-v-c with prefix  c-v-c with suffix  Double medial consonants  two syllables, separate medial consonants  Multi-syllable |

## Appendix 2: Nonsense words (post-test)

Group 3

|  |  |
| --- | --- |
| Hos, riv, taz  ist, orn, ent  shog, thid, quat  flam, grun, scir  kibe, wome, vake | c-v-c  v-c-c  Digraph-v-c  Cluster-v-c  Split digraph |

Group 4

|  |  |
| --- | --- |
| nagfot, fitzud, jimdap  liss, fudd, tobb  weab, baig, moam  slemp, brint, frulp  thish, photh, chash | c-v-c- x2  Double final consonants  Medial vowel digraphs  Cluster, beginning and end  Digraphs, beginning and end |

Group 5

|  |  |
| --- | --- |
| demar, risfug, pregox  zinate, yoging, hitful  muttle, timming, nesser  minkly, pornter, tuntness  Lartingham, blintering, trendigation | c-v-c with prefix  c-v-c with suffix  Double medial consonants  two syllables, separate medial consonants  Multi-syllable |

## Appendix 3: Nonsense words (second post-test)

Group 3

|  |  |
| --- | --- |
| nep, sut, paz  ast, ulp, eld  thog, phid, shab  trun, blat, stug  dite, sape, voke | c-v-c  v-c-c  Digraph-v-c  Cluster-v-c  Split digraph |

Group 4

|  |  |
| --- | --- |
| limfit, natzum, sugron  biss, tudd, wobb  seab, laig, foap  bremp, sluft, frint  thash, phuth, chash | c-v-c- x2  Double final consonants  Medial vowel digraphs  Clusters, beginning and end  Digraphs, beginning and end |

Group 5

|  |  |
| --- | --- |
| defug, ungox, premot  hinate, zoging, yitful  nottle, mitter, pugged  timply, honter, darful  Mortingham, flantering, plundigation | c-v-c with prefix  c-v-c with suffix  Double medial consonants  two syllables, separate medial consonants  Multi-syllable |

## Appendix 4: Overview of Reading Developmental Continuum (Rees, 1994)



## Appendix 5: Explanatory leaflet

**A study of the effects of learning to read music on children’s reading of words and texts, by Mark Betteney.**

**What is the project about?**

I would like to discover whether aspects of children’s early reading (the decoding of text) can be improved by teaching them to read simple music notation. To find this out I would like to deliver a short program of instrumental tuition (six to eight weekly recorder lessons). Before and after this series of lessons I will assess each child’s reading in different ways, and record the progress. At the end of the series of lessons I will also talk to the children in simple terms about their understanding of the reading of music.

Your school is one of only three schools in England chosen to take part in the project, but if the results show that the reading of music does improve the reading of a good proportion of participating children (or even just a few), it is very likely that the project will be repeated in other schools to the benefit of many children. If the results show that the reading of music does not improve children’s reading, then it will not have been a waste of time, because the children will have learnt a new skill, and I will have found an answer to a question that has been nagging at me for several years.

You may have some questions about the project. I try to answer some of these on the next page.

**What do I have to do?**

If you are a child in Mrs Barrs or Mr Simms class, I would like to you come to a short series of recorder lessons at school which will happen on Tuesdays. The lessons are quite short, but it would be helpful if during the six week period you would practise your recorder for a few minutes every day. I will provide you with a recorder and a booklet of music, both of which you can keep at the end of the project.

If you are an adult, please encourage your child to practise for a few minutes regularly, and try to ensure your child brings his or her recorder to school on the correct days. It might be tempting for you to read with and to your child more often during the project than you usually do. I would ask you, throughout the project, not to change any routines that you have, or the results will be due to parental input rather than music development.

**Do I have to pay for anything?**

No. Everything needed is provided free of charge.

**How will the project work?**

The children will engage individually in three different reading activities which will give me a series of reading scores. The class will then be divided in half, by drawing names out of a hat. A first half of the class will have the recorder lessons, in groups of between six and eight, after which all of the children will engage in a second round of reading activities. The second half of the class will then have the same recorder lessons that the first half of the class had. All of the children will engage in a third round of reading. I will also talk to the children individually about their knowledge of music reading at the start and end of the project.

**Do I have to do this?**

No. Mr Barker, Mr Simms, and Mrs Barrs have kindly agreed that the two classes can take part, but no one is forced to join in, and if you start the programme and find you don’t like it, you can stop at any point, just by telling me. I won’t be annoyed or disappointed – I will just be grateful that you joined in with the project in the first place.

**What will happen to the reading assessment results?**

I will look at all the different reading scores and then write about the changes in reading skills. Before I do this, I will change your name on the results, so that no one who reads it will know it is you. The only people who will have access to the results which have your name on (known as ‘the raw data’) will be Mr Barker, Mr Simms, Mrs Barrs, and me. Once I have removed all of your names on the data, other researchers whom I work with will be interested to know the results. If the results are very exciting (by which I mean if all of the children who take part make good progress in their reading, or if one or two children make enormous progress) the results may be published, but if that happens your real names will not be mentioned.

**Will the data be kept securely?**

Yes. The results will be kept by me for no more than three years. Hard copies of the assessment scores will be kept in a locked filing cabinet in my office at the University of Greenwich. A digital version will be kept in a password-protected folder on my computer, which itself is username and password protected.

## Appendix 6: Permission slip (parent/carer)

Thank you very much for reading about the project. If both you and your child are happy for your child to take part, please fill in the two forms accompanying this leaflet (one for you, and one for your child). If you have any questions about the project please contact me at ([m.w.betteney@gre.ac.uk](mailto:m.w.betteney@gre.ac.uk)). I look forward to an interesting project.

A study of the effects of learning to read music on children’s reading of words and texts.

I give permission for my child to take part in this study.

I understand the following

* there is no charge for participation,
* a recorder will be provided.
* my child may withdraw from the project at any time.
* information and data will be securely stored
* results will be anonymous, and may be published
* my child may keep the recorder at the end of the project

Child’s name …………………………………………………

Signature of parent/carer …………………………………...

Date …………………………………………………………..

## Appendix 7: Permission slip (child)

Yes please! I would like to take part!



recorder

****

**……………………………………………………………….**

**Child’s name**

## Appendix 8: Calculation of imputed data for children 29 and 30

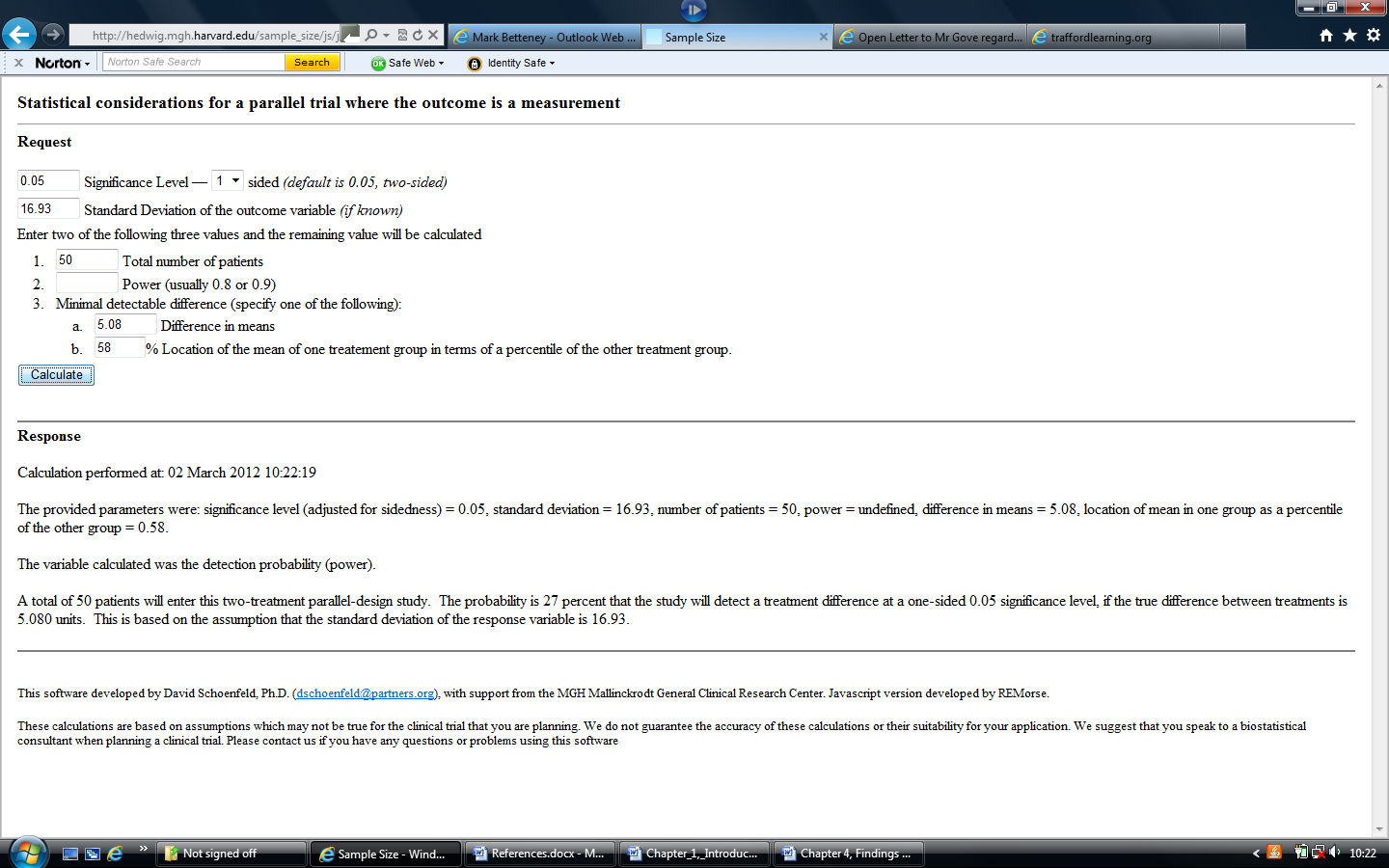
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Child 29 | Total | Single-letter grapheme | Digraph recognition | Cluster recognition | Nonsense word recognition group 3 | Nonsense word recognition group 4 | Nonsense word recognition group 5 |
| Post score | 68 | 24 | 5 | 4 | 12 | 12 | 11 |
| Class difference  Post- / pre- |  | 96% | 76% | 47% | 60% | 34% | 19% |
| Calculation |  | 24 x 0.96 = 23.04 | 5 x 0.76  = 3.8 | 4 x 0.47  = 3.04 | 12 x 0.60  = 9.12 | 12 x 0.34  = 4.08 | 11 x 0.19  = 2.09 |
| Imputed pre-score | 45 | 23 | 4 | 3 | 9 | 4 | 2 |

Imputed scores for child 29

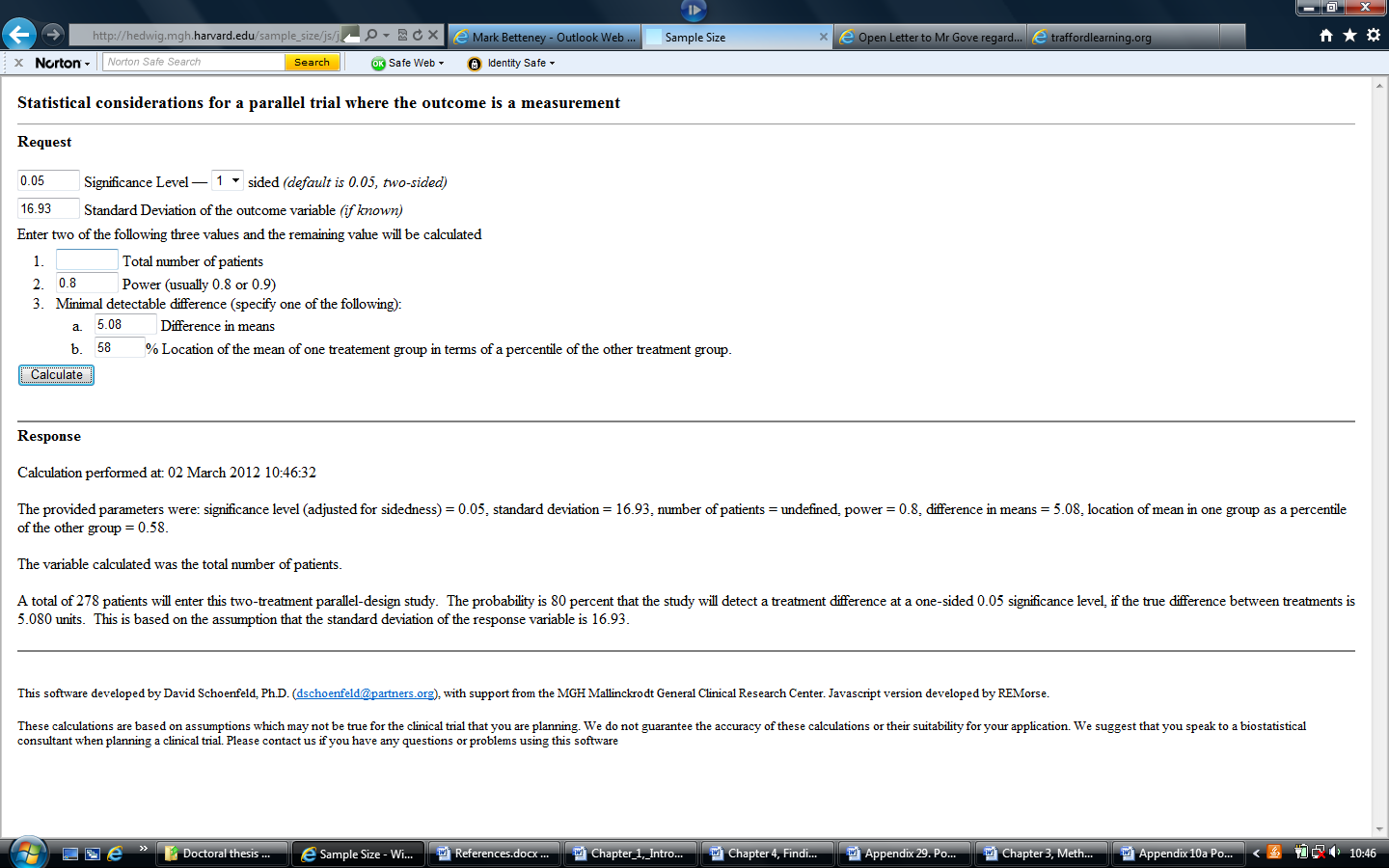
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Child 30 | Total | Single-letter grapheme | Digraph recognition | Cluster recognition | Nonsense word recognition group 3 | Nonsense word recognition group 4 | Nonsense word recognition group 5 |
| Post score | 61 | 23 | 5 | 4 | 15 | 11 | 2 |
| Class difference  Post- / pre- |  | 96% | 76% | 47% | 60% | 34% | 19% |
| Calculation |  | 23 x 0.96 = 22.08 | 5 x 0.76  = 3.8 | 4 x 0.47  = 3.04 | 15 x 0.60  = 9.00 | 11 x 0.34%  = 3.74 | 2 x 0.19  = 0.38 |
| Imputed pre-score | 42 | 22 | 4 | 3 | 9 | 4 | 0 |

Imputed scores for child 30

## Appendix 9: Power calculation (i)

Source: Schoenfeld (2012)

## Appendix 10: Power calculation (ii)



Source: Schoenfeld (2012)

## Appendix 11: Interview with research assistants G and E

M = Me

G = Research assistant 1

E = Research assistant 2

M: Well it’s a while since we did the research, and in fact the data is quite clear, and exciting in terms of their progress – I’ve had to work out effect sizes and all sorts, and it’s very, very good. But as I say, my tutor is looking for evidence that the children made progress in their ability to read music. So if you think back to the very first lessons the children had, perhaps the very first one, um, your initial impressions of the children’s ability to read music before we started – could any of them read music at all?

E: Barely

G: Yeah, I can’t remember any one in particular being able to

E: Maybe one or two said something like they had seen a recorder before, and might have done it at home, but, I don’t know, not really

G: No

E: I don’t remember, when we showed them the notes, I don’t remember them

M: Because I showed them the notes when we did that initial visit, I showed them what the notes were, but I was careful not to say ‘this is a rest’, or ‘this is a quaver’. There wasn’t, there wasn’t a, I’m trying not to feed questions to you, there wasn’t a, you weren’t aware of children who could say ‘that’s a crotchet, that’s a rest, I’ve seen this before’

G: No, I didn’t. In any of the groups I did there was no one who was really confident, no one who was evidently better at it than anyone else.

M: Right. And, right at the end, was it clear to you that children were reading music, or were they ... what they were asked to play was in some way an approximation of what was on the paper, or were they still floundering?

E: Mmmm ... both. Some and some, do you think?

G: There were some children that just blew (laughs)

E: anything

G: into their recorder weren’t there?

E: Yeah

G: and in terms of, like, rhythmically, I think they grasped that a lot more than, like , note reading, in my opinion I think

E: Yeah, and the idea of rests. I think the crotchets and stuff they could get

M: Ok. So let’s explore that a bit more. So if we take away the recorders for a second, just the clapping of the rests, the quavers and the crotchets, so at the very beginning they couldn’t do that at all

E: Yeah

G: Yeah

M: Is that right?

E: Yes

M: By the end, were they quite ... to what extent were they independent and confident, by the end?

E: Mostly, in terms of the clapping, I would say

G: Yeah, at the beginning, when we did the clapping and all that, they all did it it, didn’t they?

E: Yeah, it was quite easy by the end, wasn’t it?

G: Yeah,

E: There was still the odd few that didn’t watch, that just didn’t follow

M: When you say at the beginning, do you mean at the beginning of the programme, or the beginning of the lessons

E: The lessons

G: Yeah, I was talking about the beginning of the lessons, were we did the warm up, the quickies, and all

E: Yes

G: The ones, and rests

M: Because we started off just having five notes, and we were changing the rests, and by lesson 4, 5, and 6 of the first intervention, were those getting longer?

E: Yeah, yeah

M: OK, and the children’s confidence in clapping?

E: Yeah

G: I would say that got better

E: Yeah, I would

G: Definite. Yeah. I would say that improved more than note reading

M: What makes you say that?

G: Because when some of them were sort of blowing aimlessly into a recorder, they still pretty much kept the rhythm, whereas I noticed more that they weren’t really playing the right notes, but sometimes, like, if any notes, you know, the no fingers on any holes, but they could kind of keep the rhythm. Do you, did you think that as well?

E: Yeah, they could follow it, couldn’t they

G: Yeah, so I didn’t know why really, I think, think that some of it was because their fingers were too short. I know that sounds silly, but I think that some of them struggle really hard to cover some of the holes

M: Yes I wonder whether the recorder was actually a poor choice, in the end. Those ocarinas might have been better, or just sticking to the clapping. Were you aware when the children were playing or just blowing away while it was screeching and sounding dreadful that the fingering was roughly correct even if they weren’t anywhere near the holes.

E: Mmmmm, sometimes (not convinced)

G: Yeah, ...I would, I, I would think so. I do remember thinking that if only they could have been better at reaching the holes, some of them, it would have been better

M: So what I’m getting at is that even if the sound itself may not have been a very satisfying experience, could you tell that what they were reading on the stave was being, at least exercised on the recorder in terms of fingering if not

E: Yes I think that they know that one note meant that you had to change fingers, yeah, each note you had to do something different. You didn’t normally just sit on the same note. Sometimes you did, must most of the time you had to change

M: And did that develop over the six weeks, or was that much the same all the way through?

G: I think that at the beginning of each lesson when we went through the clapping and refreshed everything we would always go over the notes wouldn’t we and like the fingering of them and stuff, and I can remember that they could always pretty much tell us what was what, but then when it came to actually playing it they all got a bit too excited and just went for anything didn’t they? I think if they actually ... if we were doing one on one lessons and it was all a bit calmer, I think they would have ... but I think if they were, just like if they wanted tl play anything, but I think if actually asked them, and got them to show you ...

E: Yeah, they could do it, they needed a bit more time maybe

G: Like A, B, and Cs and things, they all knew what they were didn’t they, when we talked about it all together

M; So was that the case at the beginning, in the first couple of lessons

E: They were probably doing anything they want then

M: So they wouldn’t ... I’m thinking not in terms of rhythm, crotchets and quavers, but in terms of B, A and G, the notes on the stave, did they recognise, did they bring any recognition with them to the programme, on lesson 1

E: what of B, A & G was to look at

M: Yes

G: No (emphatic)

E: No (emphatic)

M; Unequivocally, no

G: No

E: No

M; So by lesson 2? 3? 4? 5? Are they quite accurately recognising the notes?  
  
E: Yeah, like as you said, in the warm-up we’d have B, A, G written on the board, and then they would have to tell us, and we would keep pointing to a different one, and they would nearly always get the right one, because there was only three to remember

M: Do you remember any particular children who were successful. I know remembering their names would be difficult

G: Er ... I don’t think I could

E: Who was that girl you had in the first group? Eleeah. Eleeah?

G: I know. It sounds familiar ... It’s hard because we didn’t even use names in our project

M: OK. So we are saying that at the beginning of the programme they knew nothing of the rhythmic names of notes, quavers, crotchets, rests. Some of them you said maybe had seen notes before, and knew they were notes

G; Yeah, I think a few of them like yeah recognised it was sheet music or you know, but I would say that none of them could

E: have picked it up and started playing

M: And what proportion of them by the end, I’m thinking really of the first intervention, up to half term , what proportion of them by the end were clapping correctly ... roughly ...

E; Nearly all

G: Yeah, I would say about 85% of them They could all do it, I just think some of them messed around a bit

M: OK. And recognising the notes. If you pointed to a note at the start of the intervention, they couldn’t tell you if it was a B, and A or a G

E: No

M: But by, well how long, week 3, 4, 5 ...

G: Yeah, by week 3 easily

M: What sort of proportion, say by the end of the intervention, what sort of proportion would be able to tell you that’s an A, that’s a G

E: I’d say it was quite high

G: Yeah, again, about 90%

E: Yeah, I would

M: OK. That’s fine. One of the questions I put on the bottom here – I’ve said how easy or difficult was it for the children to transfer their knowledge of reading music to their playing of the recorder, and you’ve said just the physicality of the thing ... how easy or difficult was it for the children to transfer their knowledge of reading music to the clapping of rhythms, so you’ve said that 85% got that

G: Yes. That was probably the most successful part of it I think, reading the rhythms of the music

## Appendix 12: Interview with research assistant J

M = me

J = Research assistant 3

M: OK, J, you’ve given me this piece of paper, just for the sake of the tape, tell me what you’re showing me here.

Overall progress of children in intervention 1 (n=50)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Very good progress | | Progress | | Not obvious progress | |
| Femi  Tiffany  Paige  Joshua  Ben  Charlotte  Ellie-May  Cameron  Millie  Elliot | 25  out  13  15  44  2  14  38  3  50 | Lily  Louis  Priyanka | 23  21  5 | Noah  Xander  Christopher  Aaliyah  Adam  Kira | 10  6  out  22  59  53 |

J: Right, what we’ve got here is how the children progressed within their recorder lessons, so we’ve got the first group that were having the lessons.

M: before or after half term.

J: Before half term, yes, so we’ve got a group of children who progressed very well, with their lessons with the actual playing of the recorder, which I would say, is down to them, who could read, who could follow the music, and who could follow the notes that they were reading on the page, therefore they were able to play the notes.

M: OK, and this is something you developed for yourself, just for your own research project.

J: Yes, yes, just out of interest really. At the time when we did it we weren’t too sure whether it was going to be needed within our research. As it happens we didn’t use it, but it was something that we thought might be useful

M: And so the first group, very good, group 2 …

J: ... Did progress, but not as well as this group here, the first group, and the third group did not, did not really progress. They were reading some of the notes but not relating it to playing notes on the recorder. They learnt a little bit but not as much.

M: So this is very much based on their recorder playing as opposed to their clapping of notes.

J: Yes, it is.

M: Yes OK.

J: And I know from the first group that there were children who were taking the recorders home, practicing at home. Obviously that will have a huge influence.

M: It does. OK That’s brilliant, it’s very helpful. Um. The reason I’m here is because my tutor is keen to have evidence of exactly the sort of thing that you have just shown me, now this (the page) is to do with their recorder playing. Would you say that this group three here, these seven children here made no progress in their recorder playing?

J: I would say they made some, but only a small amount of progress. It was also, I think, a behaviour thing as well. There were a couple of them that were doing it but only because they were told to do it, not because they were truly interested in it, so I think that ... there was a little girl called [name changed], and she just .. I think the whole thing was just beyond her, really.

M: Ok, returning to the questions, the same questions I asked to the other research assistants, thinking back to the very first lessons, probably the very first lesson, I’m thinking of the children’s recognition of the rhythm symbols, the rests and crotchets and quavers, what were your initial impressions of the children’s recognition of those in the very first lesson

J:I think some of them were very capable, they were able to follow the rhythms quite ..

M: did they bring any knowledge of those rhythms with them to the programme

J: Certainly I would say some of them did. Not all of them though, I think, some of them struggled

M: So you think some of them had read music before

J: I don’t know that it would be reading music, maybe following the rhythms. Are you talking about when we clapped and they followed?

M: Well, when we wrote the rhythms.

J: OH, right.

M: When we wrote those rhythms up, did any of them know any of those symbols.

J: Oh no, Not the symbols, no. I didn’t get the impression that any of them knew any of the symbols before they started.

M: No. And by the end of the intervention, at the end of the programme, were they much more capable, confident, or were they still struggling?

J: Oh no, I would say that by the end they could all do it. I don’t think there were any children who couldn’t. Not that I noticed anyway.

M: with the clapping.

J: With the clapping, definitely, being able to follow the quickies. And I think they enjoyed that, especially the notes having their names, the quickies and the ones.

M: What sort of things were they doing that gave you that impression that they had made this progress?

J: Well they were doing it correctly. They were following it. I think as the rhythms got a little bit more complicated a couple of them wavered a bit, just getting the quickies and rests a bit confused.

M: What made them more complicated.

J: IT was she way we had a one, a rest, a quickie. If it was one, rest, one, rest, one, rest, then they’d be fine, but if it was one, rest, quickie, one, rest, one, that kind of thing, because .. more of a combination, then they did, but I mean, it was the combinations. But then there were those that could cope with that, and I wouldn’t have thought that it could have been done any other way, because there are always going to be some children that pick it up quicker than others.

M: And then when we think about the reading of the notes, A, B, and G, did any of the children, on the first lesson, were there any that had experienced anything like this before.

J: No. I don’t think any of them could.

M: And the same question, by the end, you said that their clapping of rhythms got very much more confident, the recognition of those notes on the stave ...

J: I would say that most of them could do it. We would put the three notes upon the board, then point to different notes, and most of them got it right every time by the end of it.

M: That’s brilliant, so when you say ‘most’...

J: Probably, one or two from this group (pointing to the children who made little progress). It’s hard to tell when you’ve got a whole group, because you never quite know if they are just following everyone else, but the general impression was that they all got it, because they would all call out the right note, but when you came down do doing individual or small group, when we were doing the actual recorder lessons, that was when we were able to pick out that some of the children couldn’t then translate the reading of the music to the playing.

M: You are anticipating the next question there – so the playing of the recorder itself. ASt the beginning, how many of them had played a recorder before?

J: I don’t think any of them had. No.

M: And how successful were they in recognising the notes and converting that to the recorder

J: B y the end of the lessons? The children in the first group (referring to the page of children who made progress), so we’ve got ten of them who could play, they could follow the music, and they could play the recognisable tunes, yes, definitely, definitely. I would say so. And then the three in the middle group I would say got it some of the time, but the seven in the lower group.

M: so these ones that you’ve put on the lower group, and the ones in the higher group got all aspects so these groups are true across the board really, with the clapping and the recognition of GAB and so on .. Do you feel that there were any children who were struggling to play the recorder, but who could read the music ... where the physicality of the thing was getting in the way .. or was it generally if they couldn’t read it, they couldn’t play it.

J: To be honest with you I’m not sure, because it’s not something that I think we were looking for, so I couldn’t say. They certainly enjoyed doing the clapping, and doing the different combinations of notes.

M: G and E were of the opinion that they were struggling to cover the holes, perhaps their fingers were just not big enough, so I asked them the question were there some children that you could tell were reading the music, the right fingers were going down, but they were not covering the holes, they couldn’t be sure, and you are saying the same thing.

J: Yes.

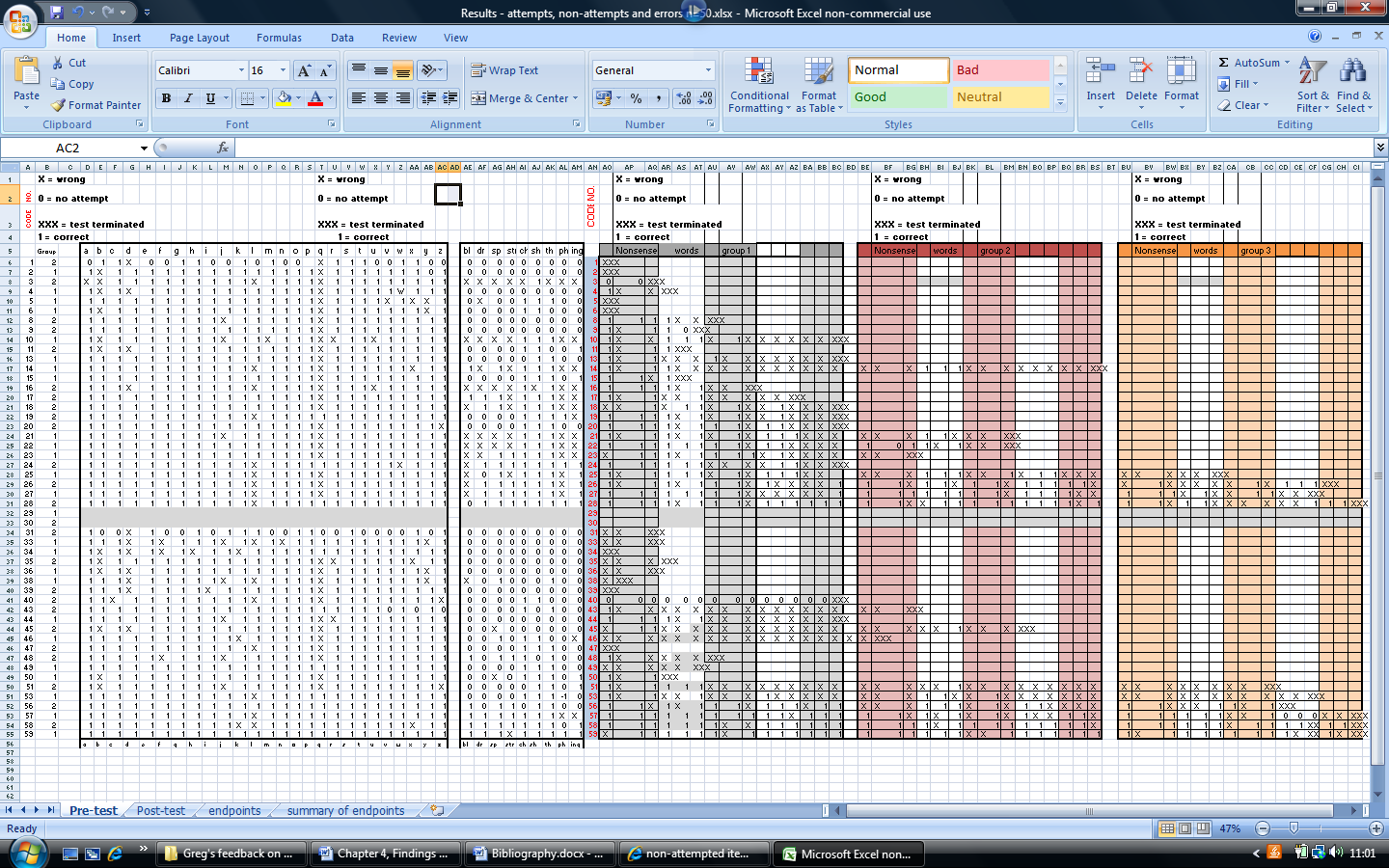
M: But what I wanted was evidence that in your view the children generally had made progress in their music reading throughout the programme, and you are happy to confirm that in the main.

J: Oh definitely, definitely.

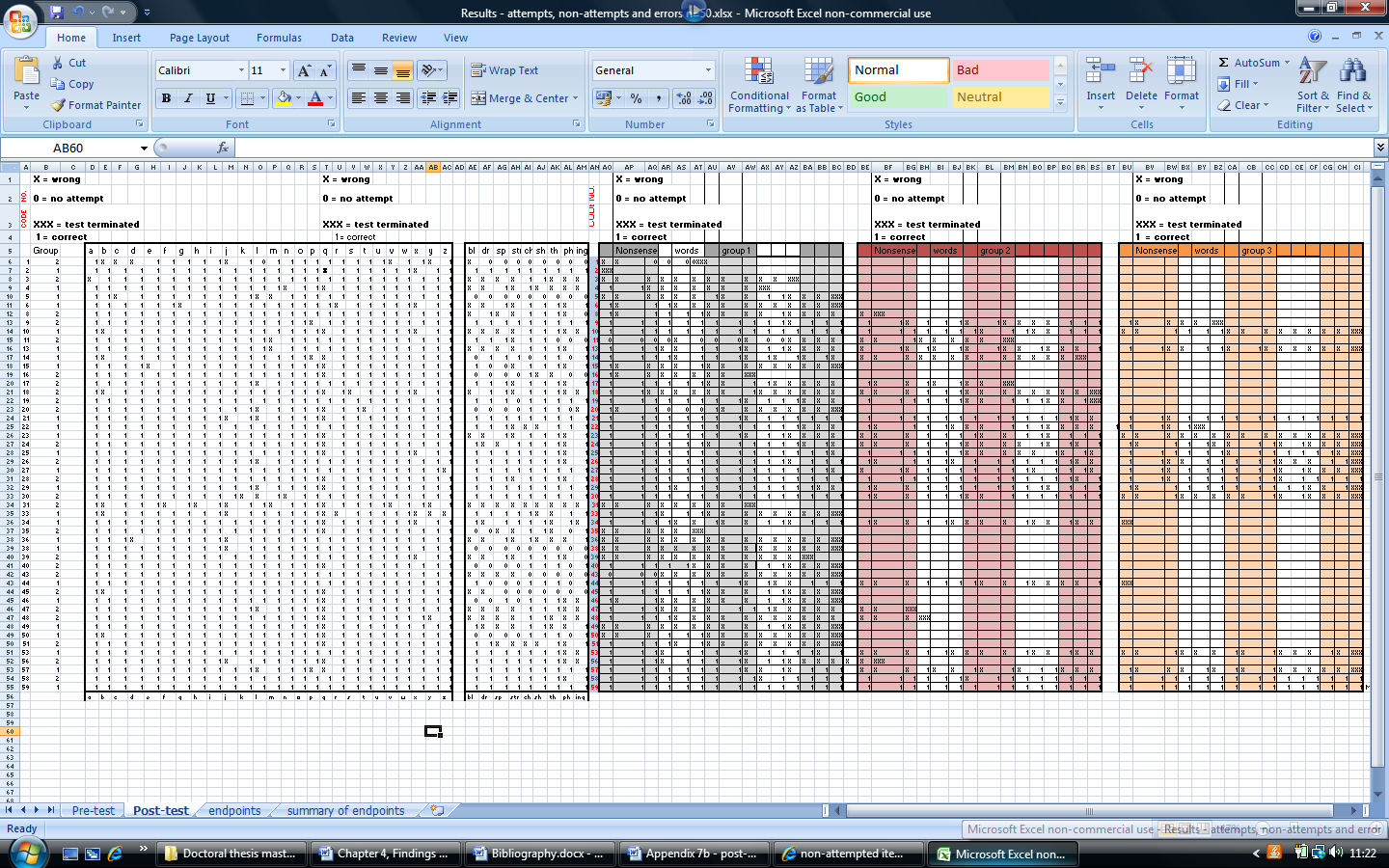
M: And even these ones here (children in group 3 of the list), identified as the least able as it were in this particular area, they made some progress in your view.

J: Oh yes, certainly I would say they made some progress. I mean, they all made progress.

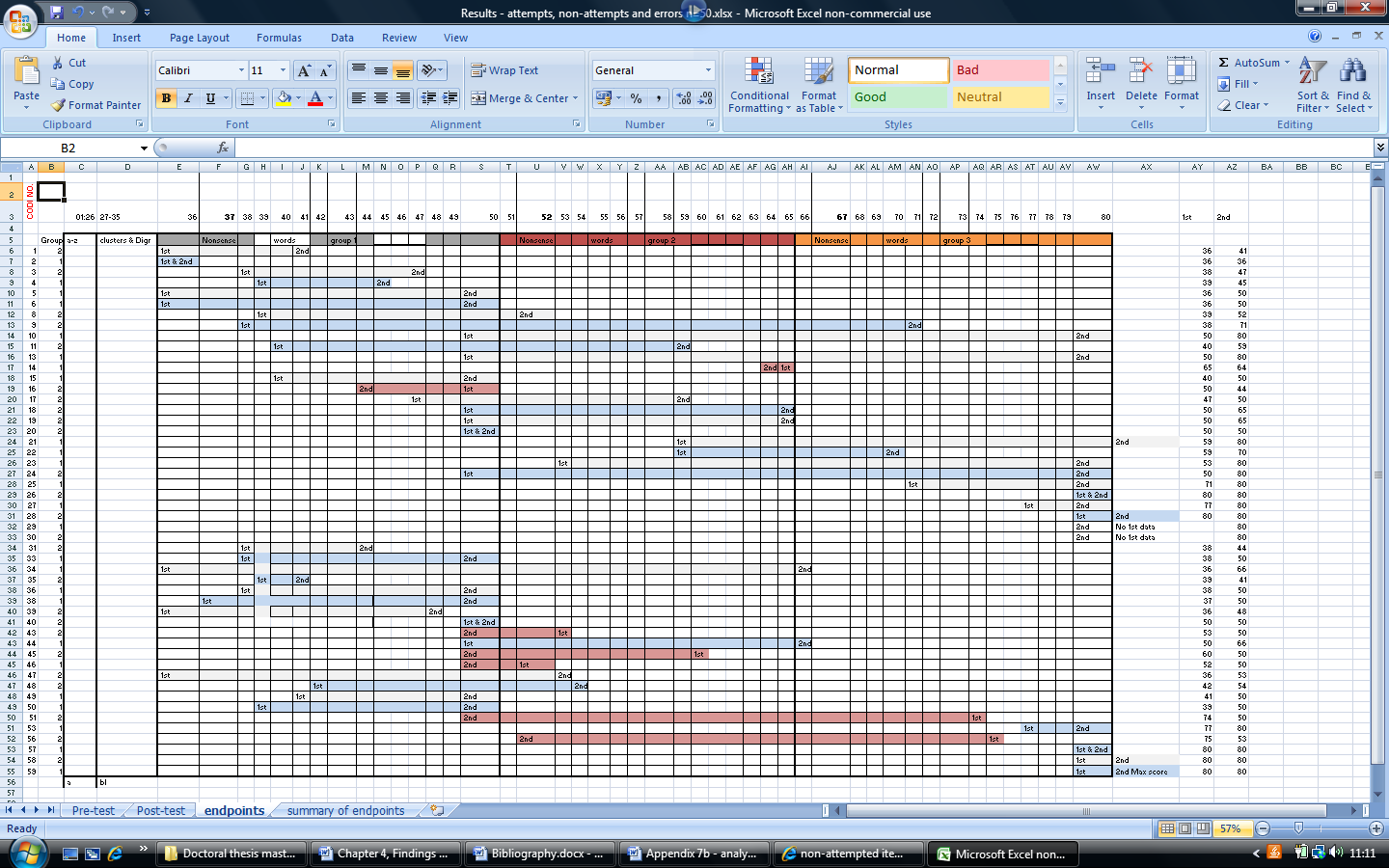
## Appendix 13: Pre-test attempts



## Appendix 14: Post-test attempts



## Appendix 15: Summary of pre- and post-test endpoints



## Appendix 16: RCT, pre-test *t*-test, overall scores

From SPSS outputs

| **Paired Samples Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | Total1 | 33.96 | 25 | 13.474 | 2.695 |
| Group1 | .00 | 25 | .000 | .000 |
| Pair 2 | Total1 | 33.64 | 25 | 14.634 | 2.927 |
| Group2 | ,00 | 25 | ,000 | ,000 |

| **Independent Samples Test** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| TotalPretest | Equal variances assumed | .051 | .822 | .080 | 48 | .936 | .320 | 3.978 | -7.679 | 8.319 |
| Equal variances not assumed |  |  | .080 | 47.676 | .936 | .320 | 3.978 | -7.681 | 8.321 |

| **Paired Samples Test** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | Total1 - Group1 | 33.960 | 13.474 | 2.695 | 28.398 | 39.522 | 12.602 | 24 | .000 |
| Pair 2 | Total1 - Group2 | 33.640 | 14.634 | 2.927 | 27.599 | 39.681 | 11.494 | 24 | .000 |

## Appendix 17: RCT, pre-test *t*-test, single letter grapheme recognition

From SPSS outputs

| **Paired Samples Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | SingleLetter1 | 23.36 | 25 | 1.753 | .351 |
| Group1 | .00 | 25 | .000 | .000 |
| Pair 2 | SingleLetter1 | 22.48 | 25 | 3.653 | .731 |
| Group2 | ,00 | 25 | ,000 | ,000 |

| **Paired Samples Test** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | SingleLetter1 - Group1 | 23.360 | 1.753 | .351 | 22.636 | 24.084 | 66.625 | 24 | .000 |
| Pair 2 | SingleLetter1 - Group2 | 22.480 | 3.653 | .731 | 20.972 | 23.988 | 30.770 | 24 | .000 |

| **Independent Samples Test** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| SingleLetter1 | Equal variances assumed | .930 | .340 | 1.086 | 48 | .283 | .880 | .810 | -.749 | 2.509 |
| Equal variances not assumed |  |  | 1.086 | 34.499 | .285 | .880 | .810 | -.766 | 2.526 |

## Appendix 18: RCT, pre-test *t*-test, overall scores less single-letter grapheme recognition

From SPSS outputs

| **Paired Samples Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | TotalLessSing1 | 10.60 | 25 | 12.560 | 2.512 |
| Group1 | .00 | 25 | .000 | .000 |
| Pair 2 | TotalLessSing1 | 11.12 | 25 | 12.991 | 2.598 |
| Group2 | ,00 | 25 | ,000 | ,000 |

| **Paired Samples Test** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | TotalLessSing1 - Group1 | 10.600 | 12.560 | 2.512 | 5.416 | 15.784 | 4.220 | 24 | .000 |
| Pair 2 | TotalLessSing1 - Group2 | 11.120 | 12.991 | 2.598 | 5.757 | 16.483 | 4.280 | 24 | .000 |

| **Independent Samples Test** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| TotalLessSing1 | Equal variances assumed | .010 | .919 | -.144 | 48 | .886 | -.520 | 3.614 | -7.786 | 6.746 |
| Equal variances not assumed |  |  | -.144 | 47.945 | .886 | -.520 | 3.614 | -7.787 | 6.747 |

## Appendix 19: RCT, pre-test *t*-test, clusters and digraphs

From SPSS outputs

| **Paired Samples Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | ClustDigraph1 | 3.16 | 25 | 2.444 | .489 |
| Group1 | .00 | 25 | .000 | .000 |
| Pair 2 | ClustDigraph1 | 3.36 | 25 | 2.998 | .600 |
| Group2 | ,00 | 25 | ,000 | ,000 |

| **Independent Samples Test** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| ClustDigraph1 | Equal variances assumed | 2.097 | .154 | -.259 | 48 | .797 | -.200 | .774 | -1.756 | 1.356 |
| Equal variances not assumed |  |  | -.259 | 46.125 | .797 | -.200 | .774 | -1.757 | 1.357 |

| **Paired Samples Test** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | ClustDigraph1 - Group1 | 3.160 | 2.444 | .489 | 2.151 | 4.169 | 6.465 | 24 | .000 |
| Pair 2 | ClustDigraph1 - Group2 | 3.360 | 2.998 | .600 | 2.122 | 4.598 | 5.603 | 24 | .000 |

## Appendix 20: RCT, pre-test *t*-test, nonsense words

From SPSS outputs

| Paired Samples Statistics | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | N | Std. Deviation | Std. Error Mean |
| Pair 1 | NonsenseWord1 | 7.12 | 25 | 10.072 | 2.014 |
| Group1 | .00 | 25 | .000 | .000 |
| Pair 2 | NonsenseWord1 | 8.00 | 25 | 11.094 | 2.219 |
| Group2 | ,00 | 25 | ,000 | ,000 |

| Paired Samples Test | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Pair 1 | NonsenseWord1 - Group1 | 7.120 | 10.072 | 2.014 | 2.963 | 11.277 | 3.535 | 24 | .002 |
| Pair 2 | NonsenseWord1 - Group2 | 8.000 | 11.094 | 2.219 | 3.421 | 12.579 | 3.605 | 24 | .001 |

| Independent Samples Test | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| NonsenseWord1 | Equal variances assumed | .165 | .686 | -.294 | 48 | .770 | -.880 | 2.997 | -6.906 | 5.146 |
| Equal variances not assumed |  |  | -.294 | 47.558 | .770 | -.880 | 2.997 | -6.907 | 5.147 |

## Appendix 21: RCT anova and unrelated t-tests, overall scores

| **Descriptive Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | Group | | Mean | Std. Deviation | N |
| Total1 | dimension1 | 1 | 33.96 | 13.474 | 25 |
| 2 | 33.64 | 14.634 | 25 |
| Total | 33.80 | 13.923 | 50 |
| Total2 | dimension1 | 1 | 45.20 | 16.593 | 25 |
| 2 | 40.12 | 17.256 | 25 |
| Total | 42.66 | 16.950 | 50 |

|  | | **Independent Samples Test** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Total2 | Equal variances assumed | | .000 | .989 | 1.061 | 48 | .294 | .147 | 5.080 | 4.788 | -4.547 | 14.707 |
| Equal variances not assumed | |  |  | 1.061 | 47.927 | .294 | .147 | 5.080 | 4.788 | -4.547 | 14.707 |

| **Multivariate Testsb** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .473 | 43.144a | 1.000 | 48.000 | .000 | .473 |
| Wilks' Lambda | .527 | 43.144a | 1.000 | 48.000 | .000 | .473 |
| Hotelling's Trace | .899 | 43.144a | 1.000 | 48.000 | .000 | .473 |
| Roy's Largest Root | .899 | 43.144a | 1.000 | 48.000 | .000 | .473 |
| factor1 \* Group | Pillai's Trace | .061 | 3.113a | 1.000 | 48.000 | .084 | .061 |
| Wilks' Lambda | .939 | 3.113a | 1.000 | 48.000 | .084 | .061 |
| Hotelling's Trace | .065 | 3.113a | 1.000 | 48.000 | .084 | .061 |
| Roy's Largest Root | .065 | 3.113a | 1.000 | 48.000 | .084 | .061 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group  Within Subjects Design: factor1 | | | | | | | |

## Appendix 22, RCT anova and unrelated t-tests, single-letter graphemes

| **Multivariate Testsb** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .302 | 20.797a | 1.000 | 48.000 | .000 | .302 |
| Wilks' Lambda | .698 | 20.797a | 1.000 | 48.000 | .000 | .302 |
| Hotelling's Trace | .433 | 20.797a | 1.000 | 48.000 | .000 | .302 |
| Roy's Largest Root | .433 | 20.797a | 1.000 | 48.000 | .000 | .302 |
| factor1 \* Group | Pillai's Trace | .049 | 2.465a | 1.000 | 48.000 | .123 | .049 |
| Wilks' Lambda | .951 | 2.465a | 1.000 | 48.000 | .123 | .049 |
| Hotelling's Trace | .051 | 2.465a | 1.000 | 48.000 | .123 | .049 |
| Roy's Largest Root | .051 | 2.465a | 1.000 | 48.000 | .123 | .049 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group Within Subjects Design: factor1 | | | | | | | |

| **Descriptive Statistics** | | | | |
| --- | --- | --- | --- | --- |
|  | Group | Mean | Std. Deviation | N |
| SingleLetter1 | Intervention | 23.36 | 1.753 | 25 |
| Control | 22.48 | 3.653 | 25 |
| Total | 22.92 | 2.870 | 50 |
| Single2 | Intervention | 24.16 | 1.463 | 25 |
| Control | 24.12 | 2.261 | 25 |
| Total | 24.14 | 1.885 | 50 |

|  | Independent Samples Test | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Single2 | | Equal variances assumed | 1.276 | .264 | .074 | 48 | .941 | .471 | .040 | .539 | -1.043 | 1.123 |
| Equal variances not assumed |  |  | .074 | 41.102 | .941 | .471 | .040 | .539 | -1.047 | 1.127 |

## Appendix 23: RCT anova and unrelated t-tests, overall scores minus single-letter graphemes

| Descriptive Statistics | | | | |
| --- | --- | --- | --- | --- |
|  | Group | Mean | Std. Deviation | N |
| TotalLessSing1 | Intervention | 10.60 | 12.560 | 25 |
| Control | 11.12 | 12.991 | 25 |
| Total | 10.86 | 12.649 | 50 |
| TotalLessSing2 | Intervention | 21.04 | 16.306 | 25 |
| Control | 16.00 | 16.432 | 25 |
| Total | 18.52 | 16.400 | 50 |

|  | Independent Samples Test | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| TotalLessSing2 | | Equal variances assumed | .000 | .987 | 1.089 | 48 | .282 | .181 | 5.040 | 4.630 | -4.269 | 14.349 |
| Equal variances not assumed |  |  | 1.089 | 47.997 | .282 | .181 | 5.040 | 4.630 | -4.269 | 14.349 |

| **Multivariate Testsb** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .409 | 33.159a | 1.000 | 48.000 | .000 | .409 |
| Wilks' Lambda | .591 | 33.159a | 1.000 | 48.000 | .000 | .409 |
| Hotelling's Trace | .691 | 33.159a | 1.000 | 48.000 | .000 | .409 |
| Roy's Largest Root | .691 | 33.159a | 1.000 | 48.000 | .000 | .409 |
| factor1 \* Group | Pillai's Trace | .083 | 4.368a | 1.000 | 48.000 | .042 | .083 |
| Wilks' Lambda | .917 | 4.368a | 1.000 | 48.000 | .042 | .083 |
| Hotelling's Trace | .091 | 4.368a | 1.000 | 48.000 | .042 | .083 |
| Roy's Largest Root | .091 | 4.368a | 1.000 | 48.000 | .042 | .083 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group  Within Subjects Design: factor1 | | | | | | | |

## Appendix 24: RCT anova and unrelated t-tests, Clusters and Digraphs

| Descriptive Statistics | | | | |
| --- | --- | --- | --- | --- |
|  | Group | Mean | Std. Deviation | N |
| ClustDi1 | Intervention | 3.16 | 2.444 | 25 |
| Control | 3.36 | 2.998 | 25 |
| Total | 3.26 | 2.709 | 50 |
| ClustDi2 | Intervention | 5.52 | 2.740 | 25 |
| Control | 4.36 | 2.970 | 25 |
| Total | 4.94 | 2.888 | 50 |

|  | | Independent Samples Test | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| ClustDi2 | Equal variances assumed | | .361 | .551 | 1.435 | 48 | .158 | .079 | 1.160 | .808 | -.465 | 2.785 |
| Equal variances not assumed | |  |  | 1.435 | 47.692 | .158 | .079 | 1.160 | .808 | -.465 | 2.785 |

| Multivariate Testsb | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .302 | 20.794a | 1.000 | 48.000 | .000 | .302 |
| Wilks' Lambda | .698 | 20.794a | 1.000 | 48.000 | .000 | .302 |
| Hotelling's Trace | .433 | 20.794a | 1.000 | 48.000 | .000 | .302 |
| Roy's Largest Root | .433 | 20.794a | 1.000 | 48.000 | .000 | .302 |
| factor1 \* Group | Pillai's Trace | .066 | 3.407a | 1.000 | 48.000 | .071 | .066 |
| Wilks' Lambda | .934 | 3.407a | 1.000 | 48.000 | .071 | .066 |
| Hotelling's Trace | .071 | 3.407a | 1.000 | 48.000 | .071 | .066 |
| Roy's Largest Root | .071 | 3.407a | 1.000 | 48.000 | .071 | .066 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group  Within Subjects Design: factor1 | | | | | | | |

## Appendix 25: RCT anova and unrelated t-tests, Nonsense words

| Descriptive Statistics | | | | |
| --- | --- | --- | --- | --- |
|  | Group | Mean | Std. Deviation | N |
| Nonsesnse1 | Intervention | 7.12 | 10.072 | 25 |
| Control | 8.00 | 11.094 | 25 |
| Total | 7.56 | 10.496 | 50 |
| Nonsense2 | Intervention | 15.64 | 14.425 | 25 |
| Control | 11.44 | 13.730 | 25 |
| Total | 13.54 | 14.098 | 50 |

|  | | Independent Samples Test | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| Nonsense2 | Equal variances assumed | | .298 | .587 | 1.055 | 48 | .297 | .149 | 4.200 | 3.983 | -3.808 | 12.208 |
| Equal variances not assumed | |  |  | 1.055 | 47.883 | .297 | .149 | 4.200 | 3.983 | -3.809 | 12.209 |

| Multivariate Testsb | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .347 | 25.495a | 1.000 | 48.000 | .000 | .347 |
| Wilks' Lambda | .653 | 25.495a | 1.000 | 48.000 | .000 | .347 |
| Hotelling's Trace | .531 | 25.495a | 1.000 | 48.000 | .000 | .347 |
| Roy's Largest Root | .531 | 25.495a | 1.000 | 48.000 | .000 | .347 |
| factor1 \* Group | Pillai's Trace | .087 | 4.600a | 1.000 | 48.000 | .037 | .087 |
| Wilks' Lambda | .913 | 4.600a | 1.000 | 48.000 | .037 | .087 |
| Hotelling's Trace | .096 | 4.600a | 1.000 | 48.000 | .037 | .087 |
| Roy's Largest Root | .096 | 4.600a | 1.000 | 48.000 | .037 | .087 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group  Within Subjects Design: factor1 | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year 1 2010-2011 | | | |  | Year 1 2009-2010 | | | |
| PIN | Autumn | Spring | Progress |  | PIN | Autumn | Spring | Progress |
| 1 | 1b | 1a | 1 |  | 1 | 1b | 2c | 2 |
| 2 | 1b | 1b | 0 |  | 2 | 1c | 1c | 0 |
| 3 | 1b | 2c | 2 |  | 3 | 1c | 1b | 1 |
| 4 | 1b | 1a | 1 |  | 4 | 1c | 1c | 0 |
| 5 | 1b | 1a | 1 |  | 5 | 1b | 1a | 1 |
| 6 | 1c | 1b | 1 |  | 6 | 1b | 1a | 1 |
| 7 | 1b | 1b | 0 |  | 7 | 1b | 1b | 0 |
| 8 | 1b | 1a | 1 |  | 8 | 1c | 1b | 1 |
| 9 | 1b | 1a | 1 |  | 9 | 1c | 1c | 0 |
| 10 | 1c | 1c | 0 |  | 10 | 1c | 1b | 1 |
| 11 | 1b | 2c | 2 |  | 11 | 1c | 1c | 0 |
| 12 |  |  |  |  | 12 | 2c | 2b | 1 |
| 13 | 1c | 1c | 0 |  | 13 | 1b | 1a | 1 |
| 14 | Wa | 1c | 1 |  | 14 |  |  |  |
| 15 | 1a | 2c | 1 |  | 15 | Wb | 1c | 2 |
| 16 | 1b | 1a | 1 |  | 16 | 1b | 2c | 2 |
| 17 | Wa | 1c | 1 |  | 17 | 1c | 1b | 1 |
| 18 | 1c | 1b | 1 |  | 18 | 1b | 2c | 2 |
| 19 | 1b | 1a | 1 |  | 19 | 1c | 1c | 0 |
| 20 | 1c | 1c | 0 |  | 20 | 1b | 2c | 2 |
| 21 | 1c | 1b | 1 |  | 21 | 1c | 1b | 1 |
| 22 | 1c | 1b | 1 |  | 22 | 1b | 1b | 0 |
| 23 | 1c | 1a | 1 |  | 23 | 1b | 2c | 2 |
| 24 | 1c | 1b | 1 |  | 24 | 1c | 1b | 1 |
| 25 | Wa | 1c | 1 |  | 25 | 1c | 1c | 0 |
| 26 | 1c | 1c | 0 |  | 26 | 1c | 1b | 1 |
| 27 | Wc | Wb | 1 |  | 27 | 1b | 2c | 2 |
| 28 | 1c | 1b | 1 |  | 28 | Wc | Wc | 0 |
| 29 | 1c | 1b | 1 |  | 29 | 1c | 1c | 0 |
| 30 | 1c | 1b | 1 |  | 30 | 1b | 1a | 1 |
| 31 | 1b | 1a | 1 |  | 31 | 1b | 1a | 1 |
| 32 |  |  |  |  | 32 | 1c | 1b | 1 |
| 33 | 1c | 1b | 1 |  | 33 | Wc | Wc | 0 |
| 34 | 1b | 1b | 0 |  | 34 | 1c | 1b | 1 |
| 35 | 1b | 2c | 2 |  | 35 | 1c | 1c | 0 |
| 36 | 1c | 1a | 1 |  | 36 | 1b | 1b | 0 |
| 37 |  |  |  |  | 37 | 1b | 1a | 1 |
| 38 | 1c | 1b | 1 |  | 38 |  | 1c | - |
| 39 | 1b | 1a | 1 |  | 39 | 1c | 2c | 3 |
| 40 | 1b | 2c | 2 |  | 40 | 1b | 1a | 1 |
| 41 | 1b | 1a | 1 |  | 41 | 1c | 1b | 1 |
| 42 | 1a | 2b | 2 |  | 42 | 1c | 1b | 1 |
| 43 | 1c | 1b | 1 |  | 43 | 1c | 1b | 1 |
| 44 | 1b | 1a | 1 |  | 44 | Wa | 1c | 1 |
| 45 | 1c | 1b | 1 |  | 45 | 1c | 1c | 0 |
| 46 | 1c | 1b | 1 |  | 46 | 1c | 1b | 1 |
| 47 | 1b | 1a | 1 |  | 47 | 1b | 1b | 0 |
| 48 | 1b | 1b | 0 |  | 48 |  | 1c | - |
| 49 | 1a | 2c | 1 |  | 49 | 1c | 1b | 1 |
| 50 | 1c | 1b | 1 |  | 50 |  | 1b | - |
| 51 | 1c | 1b | 1 |  | 51 | 1c | 1b | 1 |
| 52 | 1b | 2c | 2 |  | 52 | 1c | 1c | 0 |
| 53 | 1b | 1a | 1 |  | 53 | 1c | 1b | 1 |
| 54 | 1a | 2c | 1 |  | 54 | 1b | 1a | 1 |
| 55 |  |  |  |  | 55 | 1b | 2c | 2 |
| 56 | 1b | 1a | 1 |  | 56 |  |  |  |
| 57 | Wb | 1c | 2 |  | 57 | 1c | 1c | 0 |
| 58 | Wa | Wa | 0 |  | 58 | 1b | 1b | 0 |
| 59 | 1c | 1a | 2 |  | 59 |  |  |  |
| 60 |  |  |  |  | 60 |  |  |  |
| Mean pre-test score | | | 4.40 |  | Mean pre-test score | | | 4.26 |
| Mean post-test score | | | 5.42 |  | Mean post-test score | | | 5.11 |
| Mean Gain | | | 1.02 |  | Mean Gain | | | 0.85 |
| Post-test Standard Deviation | | | 1.13 |  | Post-test Standard Deviation | | | 1.34 |

## Appendix 26. Year 1 SATs results in reading: comparison between the participating cohort with the previous year’s cohort.

## Appendix 27: Phase 3, pre-test *t*-test

| Independent Samples Test | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| pretest | Equal variances assumed | .385 | .536 | .744 | 106 | .459 | .136 | .183 | -.226 | .498 |
| Equal variances not assumed |  |  | .743 | 105.515 | .459 | .136 | .183 | -.227 | .498 |

## Appendix 28: Phase 3 SPSS outputs

| Descriptive Statistics | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | Group | | Mean | Std. Deviation | N |
| pretest | dimension1 | 1 | 4.40 | .935 | 55 |
| 2 | 4.26 | .964 | 53 |
| Total | 4.33 | .947 | 108 |
| posttest | dimension1 | 1 | 5.42 | 1.134 | 55 |
| 2 | 5.11 | 1.340 | 53 |
| Total | 5.27 | 1.243 | 108 |

|  | | Independent Samples Test | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
| F | Sig. | t | df | Sig. (2-tailed) | Sig. (1-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| posttest | Equal variances assumed | | .044 | .835 | 1.279 | 106 | .204 | .102 | .305 | .238 | -.168 | .778 |
| Equal variances not assumed | |  |  | 1.275 | 101.828 | .205 | .102 | .305 | .239 | -.170 | .779 |

| Multivariate Testsb | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
| factor1 | Pillai's Trace | .663 | 208.798a | 1.000 | 106.000 | .000 | .663 |
| Wilks' Lambda | .337 | 208.798a | 1.000 | 106.000 | .000 | .663 |
| Hotelling's Trace | 1.970 | 208.798a | 1.000 | 106.000 | .000 | .663 |
| Roy's Largest Root | 1.970 | 208.798a | 1.000 | 106.000 | .000 | .663 |
| factor1 \* Group | Pillai's Trace | .016 | 1.713a | 1.000 | 106.000 | .193 | .016 |
| Wilks' Lambda | .984 | 1.713a | 1.000 | 106.000 | .193 | .016 |
| Hotelling's Trace | .016 | 1.713a | 1.000 | 106.000 | .193 | .016 |
| Roy's Largest Root | .016 | 1.713a | 1.000 | 106.000 | .193 | .016 |
| a. Exact statistic | | | | | | | |
| b. Design: Intercept + Group  Within Subjects Design: factor1 | | | | | | | |