

**Can a non-phonics based intervention scheme enable children who are falling behind in literacy to make better progress than normal classroom teaching? A pilot-feasibility study.**

Louise Amanda Davis

Registration Number 090194933

Submitted in part fulfilment of the award of Doctor of Education

(Ed.D.)

Department of Education

University of Sheffield

February 2018

For Chris, who has provided support and encouragement throughout this educational roller-coaster.

Also, for Nathanael and Beatrice who have encouraged me continually.

## **Acknowledgements**

I would like to thank the following people who offered me their help, support and advice.

Jane Brzozwski

Terry Corrigan

Patrick O'Connor

Janis Lawrenson

Paul Lawrenson

Pete Morgan

Val Morgan

Vicky Morten

Jason White

Thanks to the children who helped me with this project without whom this task would have been impossible.

A special thanks to Professor Greg Brooks, whose patience has no bounds; without his guidance and help this thesis would have remained unwritten - thank you

Louise Davis (2018) Can a non-phonics-based intervention scheme enable children who are falling behind in literacy to make better progress than normal classroom teaching? A pilot-feasibility study.

### **Abstract**

*The aim of this study was to investigate whether a non-phonics-based intervention, Fischer Family Trust (FFT) Wave 3, might help children in Years 1 and 2 who were falling behind in literacy to catch up with their peers. The study consisted of an 'outer' and an 'inner' study. The outer study was a pilot/feasibility study (PFS) which investigated whether the quantitative approach used in the inner study would be suitable for scaling up to investigate the research questions. The inner study was a quantitative evaluation of FFT Wave 3. It consisted mainly of a randomised control trial (RCT) using standardised literacy tests, supplemented by a number of non-standardised assessments.*

*A small (n=24) two-armed RCT lasting 10 weeks was undertaken in a state primary school in Derbyshire. Standardised tests of reading comprehension (primary outcome), of spelling and of oral word and sentence reading (secondary outcomes), were administered to both groups pre and post, and after the control group had received the intervention. At pre- and post-tests both groups answered an attitudes questionnaire, and the intervention group provided one-sentence writing samples, and their reading ages were estimated (exploratory outcomes).*

*Results for the primary and secondary outcomes fell into a confusing pattern, and were inconclusive, and results for the exploratory outcome of attitudes to reading were null. Results for the exploratory outcomes of reading ages and writing showed statistically significant gains, but could not be considered definitive because no parallel data were gathered from the control group. Thus the inner, quantitative study failed to show conclusively whether the FFT Wave 3 intervention had real impact. On the other hand, the PFS successfully showed that, with adjustments, a quantitative, mainly RCT, approach could be a suitable method for assessing a non-phonics-based intervention.*

# Contents

|   |    |
|---|----|
| Acknowledgements.....   | 2  |
| Abstract.....   | 3  |
| List of Tables.....   | 11 |
| List of Figures .....   | 14 |
| 1 Nature and origin of the research .....   | 15 |
| 1.1 Nature of the research .....  | 15 |
| 1.2 Structure of this thesis .....  | 15 |
| 1.3 The importance of literacy.....   | 16 |
| 1.4 Scale of the problem.....   | 17 |
| 1.5 The cost of poor literacy.....  | 20 |
| 1.6 Literacy levels in the UK .....   | 23 |
| 1.7 The influence of the Rose report.....   | 24 |
| 1.8 Limitations of phonics .....  | 24 |
| 1.9 Bad experience with phonics .....   | 27 |
| 1.10 Principal and subsidiary research questions and hypotheses .....                               | 28 |
| 1.10.1 Pilot/feasibility study research question and hypotheses .....                               | 29 |
| 1.10.2 Quantitative study research questions and hypotheses .....                                   | 29 |
| 1.10.3 Summary.....   | 31 |
| 2 Positionality and insider status.....   | 32 |
| 2.1 Introduction .....  | 32 |
| 2.2 What is the researcher's positionality?.....  | 32 |
| 2.2.1 A definition of positivism.....   | 33 |
| 2.2.2 Positivism in educational research, or the qualitative versus<br>quantitative dichotomy ..... | 38 |
| 2.2.3 Quantitative versus qualitative approaches.....   | 39 |
| 2.2.4 Randomised control trials: a useful quantitative tool .....                                   | 40 |
| 2.2.5 Lessons from randomised control trials in healthcare research ....                            | 42 |

|       |   |    |
|-------|---|----|
| 2.2.6 | Summary .....   | 44 |
| 2.3   | The dilemma of insider research.....  | 44 |
| 2.4   | Conclusion .....  | 47 |
| 3     | Pilot/feasibility studies literature review .....   | 49 |
| 3.1   | Introduction .....  | 49 |
| 3.2   | Definitions of the terms ‘pilot studies’, ‘feasibility studies’ and ‘pilot/feasibility studies’ ..... | 49 |
| 3.3   | Why do a pilot/feasibility study? .....   | 52 |
| 3.4   | Pilot/feasibility study consistency .....   | 54 |
| 3.5   | Review of educational pilot/feasibility studies .....   | 56 |
| 3.5.1 | Introduction .....  | 56 |
| 3.5.2 | Education Endowment Foundation.....   | 57 |
| 3.5.3 | Other education pilot/feasibility studies .....   | 61 |
| 3.5.4 | Summary of pilot/feasibility studies .....  | 68 |
| 3.6   | Conclusion .....  | 69 |
| 4     | Why investigate alternatives to phonics?.....   | 71 |
| 4.1   | Introduction .....  | 71 |
| 4.2   | Why phonics or not phonics?.....  | 71 |
| 4.3   | What is phonics? .....  | 72 |
| 4.4   | Theories of reading.....  | 73 |
| 4.4.1 | The simple view of reading.....   | 73 |
| 4.4.2 | Dual-route models .....   | 74 |
| 4.4.3 | Summary .....   | 76 |
| 4.5   | Jean Chall: The Great Debate .....  | 76 |
| 4.6   | Phonics in the National Curriculum.....   | 80 |
| 4.7   | Experimental research supporting the use of phonics in the teaching of beginning reading.....         | 96 |

|       |  |     |
|-------|--|-----|
| 4.7.1 | How effective is systematic phonics teaching for (mainly) normally-developing children of primary age? .....                             | 96  |
| 4.7.2 | Summary of data supporting the use of systematic phonics for primary-age children .....  | 101 |
| 4.7.3 | Do systematic phonics interventions help falling-behind readers to become more skilled at reading? .....                                 | 103 |
| 4.7.4 | Summary of the effectiveness of systematic phonics instruction for children who are at risk of falling behind .....                      | 107 |
| 4.8   | The opposition to phonics .....  | 107 |
| 4.8.1 | 'A monstrous regimen of phonics' (Davis, 2012) .....   | 107 |
| 4.8.2 | Carbo (1987) .....   | 108 |
| 4.8.3 | Rhetoric of the phonics vs whole-word debate .....   | 109 |
| 4.8.4 | Stephen Krashen – an advocate for whole-word approaches .....  | 111 |
| 4.9   | What is the evidence that systematic NON-phonics-based instruction benefits children learning to read and write? How strong is it? ..... | 115 |
| 4.9.1 | Reading Recovery .....   | 115 |
| 4.9.2 | Other teacher-taught schemes .....   | 117 |
| 4.9.3 | Tutored schemes .....  | 118 |
| 4.10  | Summary of relevant non-phonics-based experimental data .....  | 119 |
| 4.11  | Conclusion .....   | 119 |
| 5     | Review of the effect of reading on spelling, and of research on writing and attitudes to reading .....                                   | 121 |
| 5.1   | Introduction .....   | 121 |
| 5.2   | Review of reading and spelling links .....   | 121 |
| 5.3   | Spelling theories .....  | 121 |
| 5.4   | What is writing? .....   | 130 |
| 5.5   | Emergent literacy .....  | 131 |
| 5.6   | Beginning writing and reading .....  | 132 |
| 5.7   | The reader in the writer .....   | 134 |

|       |   |     |
|-------|---|-----|
| 5.8   | Attitudes towards reading .....                                     | 136 |
| 5.8.1 | Introduction .....  | 136 |
| 5.8.2 | Motivation.....   | 136 |
| 5.8.3 | Attitudes towards reading surveys.....                              | 136 |
| 5.9   | Conclusion.....   | 140 |
| 6     | Method.....   | 142 |
| 6.1   | Introduction.....   | 142 |
| 6.2   | The school .....  | 142 |
| 6.2.1 | Phonics teaching in the school.....                                 | 143 |
| 6.2.2 | Intervention teaching in the school.....                            | 144 |
| 6.3   | Identification of an appropriate intervention .....                 | 144 |
| 6.3.1 | The criteria used to determine the most suitable intervention ..... | 144 |
| 6.3.2 | Choosing the intervention.....                                      | 146 |
| 6.4   | The intervention: Fischer Family Trust Wave 3.....                  | 147 |
|       | (FFT Wave 3) .....  | 147 |
| 6.4.1 | Reading level determination.....                                    | 148 |
| 6.4.2 | Reading Day .....   | 148 |
| 6.4.3 | Writing Day.....  | 149 |
| 6.4.4 | Previous research on Fischer Family Trust Wave 3.....               | 150 |
| 6.4.5 | Arranging to work with the Fischer Family Trust .....               | 157 |
| 6.5   | Obtaining consent.....  | 157 |
| 6.6   | Arranging permission and selecting the participants.....            | 159 |
| 6.6.1 | Gaining permission from parents .....                               | 159 |
| 6.6.2 | Selection of participants .....                                     | 159 |
| 6.6.3 | Introducing the research to the children .....                      | 159 |
| 6.7   | Measuring the effectiveness of FFT Wave 3.....                      | 160 |
| 6.8   | Power calculations.....   | 160 |
| 6.9   | Assignment to groups using minimisation.....                        | 161 |



|        |   |     |
|--------|---|-----|
| 6.10   | The assessment instruments.....   | 163 |
| 6.10.1 | Choice of standardised reading and spelling tests.....                                      | 163 |
| 6.10.2 | Statistical robustness of the standardised tests .....                                      | 165 |
| 6.10.3 | Regression towards the mean .....   | 165 |
| 6.10.4 | Administration and marking of the standardised tests (primary and secondary outcomes) ..... | 166 |
| 6.10.5 | Scores.....   | 167 |
| 6.10.6 | Analyses of standardised test data .....  | 167 |
| 6.11   | Exploratory outcomes.....   | 167 |
| 6.11.1 | Book band levels.....   | 167 |
| 6.11.2 | Writing assessment.....   | 167 |
| 6.11.3 | Attitudes questionnaire.....  | 168 |
| 6.12   | Timetable.....  | 168 |
| 6.13   | Summary .....   | 169 |
| 7      | Results.....  | 170 |
| 7.1    | Introduction .....  | 170 |
| 7.2    | Primary and secondary outcomes: equivalence of groups at pre-test                           | 170 |
| 7.3    | Description of statistical analyses for Phase 1 (RCT) and Phase 2 (follow-up) data .....    | 172 |
| 7.3.1  | Introduction .....  | 172 |
| 7.3.2  | Phase 1 analysis .....  | 172 |
| 7.3.3  | Phase 2 analysis .....  | 173 |
| 7.4    | Primary outcome: reading comprehension .....  | 174 |
| 7.4.1  | Phase 1 ANOVA.....  | 174 |
| 7.4.2  | Phase 2 <i>t</i> -tests .....   | 175 |
| 7.5    | Secondary outcomes .....  | 175 |
| 7.5.1  | Hodder Oral Word Reading.....   | 175 |
| 7.5.2  | Phase 1 ANOVA.....  | 176 |

|        |  |     |
|--------|--|-----|
| 7.5.3  | Phase 2 t-tests .....  | 177 |
| 7.5.4  | Hodder Oral Sentence Reading .....                               | 177 |
| 7.5.5  | Phase 1 ANOVA.....   | 178 |
| 7.5.6  | Phase 2 <i>t</i> -tests .....                                    | 179 |
| 7.5.7  | Diagnostic Spelling Test.....                                    | 179 |
| 7.5.8  | Phase 1 ANOVA.....   | 180 |
| 7.5.9  | Phase 2 <i>t</i> -tests .....                                    | 181 |
| 7.6    | Primary and secondary outcomes: Non-parametric analysis .....    | 181 |
| 7.6.1  | Summary of primary and secondary outcome results .....           | 181 |
| 7.7    | Exploratory outcomes .....                                       | 182 |
| 7.7.1  | Book Band analysis.....  | 182 |
| 7.7.2  | Writing analysis .....   | 184 |
| 7.7.3  | Summary of writing analysis.....                                 | 187 |
| 7.7.4  | Attitudes towards reading.....                                   | 187 |
| 7.8    | Summary .....  | 190 |
| 8      | Discussion .....   | 191 |
| 8.1    | Introduction .....   | 191 |
| 8.2    | Testing the quality of the randomised control trial.....         | 191 |
| 8.3    | Judging the quality of the exploratory measures.....             | 193 |
| 8.4    | Testing the quality of the pilot/feasibility study .....         | 194 |
| 8.4.1  | Suitability and limitations of the PFS .....                     | 195 |
| 8.5    | Research designs and rigour .....                                | 196 |
| 8.6    | Limitations.....   | 197 |
| 8.7    | Type I and Type II errors.....                                   | 199 |
| 8.8    | What experience has been gained? .....                           | 199 |
| 8.9    | Principal conclusions .....                                      | 200 |
| 8.10   | Principal and subsidiary research questions and hypotheses ..... | 201 |
| 8.10.1 | Quantitative study research questions and hypotheses.....        | 201 |

|        |  |     |
|--------|--|-----|
| 8.10.2 | Pilot/feasibility study research question and hypotheses ..... | 202 |
| 8.10.3 | Conclusion .....   | 203 |
|        | References.....  | 204 |
|        | Appendix 1: Description of Letters and Sounds phases.....      | 217 |
|        | Appendix 2: Ethical Approval .....                             | 218 |
|        | Appendix 3: Reading Record .....                               | 223 |
|        | Appendix 4: Letter for Parents .....                           | 224 |
|        | Appendix 5: Children’s consent sheet.....                      | 227 |
|        | Appendix 6: Minimisation data and results.....                 | 228 |
|        | Appendix 7: Attitudes Questionnaire.....                       | 230 |
|        | Appendix 8: Book level raw data.....                           | 231 |
|        | Appendix 9: Mann-Whitney results .....                         | 232 |

## List of Tables

|  |     |
|--|-----|
| Table 1.1: Percentages of children in England achieving below level 2 in reading in Key Stage 1 National Curriculum teacher assessments, or below level 4 in reading in Key Stage 2 national tests, 2007–15 .....                  | 17  |
| Table 1.2: Percentages of children in England achieving below level 2 in reading in Key Stage 1 National Curriculum teacher assessments, or below level 4 in reading in Key Stage 2 national tests, with age-cohorts aligned ..... | 18  |
| Table 1.3: England PIRLS results 2001-2016 .....   | 18  |
| Table 1.4: Comparison of adult reading levels for the years 2003 and 2011 .....  | 20  |
| Table 3.1: CONSORT statement for pilot and/or feasibility studies in educational research.....   | 55  |
| Table 3.2: Summary of EEF-funded pilot/feasibility studies .....   | 58  |
| Table 3.3: Mean difference in age standardised scores between initial and final assessments and percentages of pupils showing score changes .....  | 61  |
| Table 3.4: Summary of pilot/feasibility studies .....  | 68  |
| Table 4.1: Ehri et al.’s mean effect sizes for normally–developing and at-risk children, by age group and overall .....  | 97  |
| Table 4.2: Effect sizes calculated by Camilli et al. (2006) .....  | 98  |
| Table 4.3: Effect sizes from Ehri et al. compared with $r$ and $r^2$ as calculated by Hammill & Swanson .....  | 98  |
| Table 4.4: Comparison of the findings of Ehri et al. (2001) and Torgerson et al. (2006a) .....   | 102 |
| Table 4.5: Characteristics of five studies identified from Ehri et al. ....  | 104 |
| Table 4.6: Effect sizes of four relevant studies as stated by Ehri et al. and Torgerson et al.....   | 105 |
| Table 4.7: RCTs identified from EEF and What Works Clearinghouse .....   | 106 |
| Table 4.8: Krashen’s (2001) results for SSR studies.....   | 113 |
| Table 4.9: RCTs on Reading Recovery – effect sizes .....   | 116 |
| Table 4.10: Effect sizes from Center et al. (1995).....  | 116 |
| Table 4.11: Overview of the interventions investigated by Pinnell et al. (1994) .....  | 117 |

|  |     |
|--|-----|
| Table 4.12: Post-test mean effect sizes for the four interventions investigated by Pinnell et al.....  | 118 |
| Table 4.13: Results of tutored non-phonics-based schemes .....   | 119 |
| Table 5.1: Comparison of reading attitudes for the years 1998 and 2003 (percentages) .....   | 138 |
| Table 6.1: Summary of results from FFT Wave 3 pilot study (Canning, FFT wave 3: Report of findings from a 10-week pilot, 2004) .....   | 151 |
| Table 6.2: Assessment results pre- and post-intervention in Canning (2009) ....  | 153 |
| Table 6.3: Book band results from FFT Wave 3 main study (Canning, 2009) ....   | 154 |
| Table 6.4: Pre- and post-test average Reading Recovery book bands and reading ages in years and months, gains in book bands, gains in reading accuracy in months of reading age, and ratio gains derived from Canning (2009) ..... | 156 |
| Table 6.5: Timetable of research study .....   | 168 |
| Table 7.1: Results of Shapiro-Wilks and Levene's tests of standardised pre-test data .....   | 171 |
| Table 7.2: Pre-test mean scores, by group (standard deviations in brackets) .....  | 171 |
| Table 7.3: Hodder Group Reading Test mean scores (and standard deviations), by group and assessment points .....   | 174 |
| Table 7.4: Repeated measures ANOVA, Hodder Group Reading Test, Phase 1 .....   | 175 |
| Table 7.5: Dependent <i>t</i> -test results, Hodder Group Reading test, Phase 2 ....   | 175 |
| Table 7.6: Hodder Oral Word Reading Test mean scores (and standard deviations), by group and assessment points .....   | 176 |
| Table 7.7: Repeated measures ANOVA, Hodder Oral Word Reading Test, Phase 1 .....   | 177 |

|  |     |
|--|-----|
| Table 7.8: Dependent <i>t</i> -test results, Hodder Oral Word Reading test, Phase 2  | 177 |
| Table 7.9: Hodder Oral Sentence Reading Test mean scores (and standard deviations), by group and assessment points         | 177 |
| Table 7.10: Repeated measures ANOVA, Hodder Oral Sentence Reading Test, Phase 1  | 178 |
| Table 7.11: Dependent <i>t</i> -test results, Hodder Oral Sentence Reading test, Phase 2                                   | 179 |
| Table 7.12: Diagnostic Spelling Test mean scores (and standard deviations), by group and assessment points                 | 179 |
| Table 7.13: Repeated measures ANOVA, Diagnostic Spelling Test, Phase 1..   | 180 |
| Table 7.14: Dependent <i>t</i> -test results, Diagnostic Spelling test, Phase 2.....                                       | 181 |
| Table 7.15: Book band, national curriculum level and reading age results .   | 183 |
| Table 7.16: Dependent <i>t</i> -test results for book band, NC and reading age analyses                                    | 184 |
| Table 7.17: The first and final sentences written by the intervention group  | 185 |
| Table 7.18: Number of words written by each child in their first and final sentences                                       | 186 |
| Table 7.19: Intervention group questionnaire results   | 188 |
| Table 7.20: Control group questionnaire results  | 189 |
| Table 8.1 Audit of RCT against an adaptation of Torgerson et al.'s (2006b) checklist of characteristics of valid RCTs..... | 192 |
| Table 8.2: Completed CONSORT table to determine the validity of the PFS .  | 195 |
| Table 8.3. EEF PFS questions answered with respect to this research.....   | 203 |

## List of Figures

|             |  |     |
|-------------|--|-----|
| Figure 2.1: | Comparison of deductive and inductive research approaches...                           | 34  |
| Figure 2.2: | Hierarchy of methods.....  | 35  |
| Figure 4.1: | The simple view of reading .....   | 73  |
| Figure 4.2: | Basic dual-route model of reading aloud.....   | 75  |
| Figure 4.3: | Year 1 phonics screening check mark distribution, 2012 & 2013...<br>.....              | 88  |
| Figure 4.4: | National phonics check results for 2012, actual and remodelled                         | 89  |
| Figure 4.5: | Phonics check data 2012-2016.....  | 90  |
| Figure 4.6: | Group allocation in Machin et al. (2016, p. 26).....                                   | 95  |
| Figure 5.1: | Diagram of Uta Frith's literacy stage theory .....                                     | 125 |
| Figure 5.2: | Levels of reading enjoyment, KS2, 2010-2015 .....                                      | 139 |
| Figure 5.3: | Percentages of children with attitudinal statements in 2010 to 2015<br>.....           | 140 |
| Figure 6.1: | Book band levels of children at the start of the Canning (2009)<br>intervention .....  | 154 |
| Figure 6.2: | Book band levels at the end of the intervention .....                                  | 155 |
| Figure 7.1: | Hodder Group Reading Test mean scores, by group and<br>assessment points .....         | 174 |
| Figure 7.2: | Hodder Oral Word Reading Test mean scores, by group and<br>assessment points .....     | 176 |
| Figure 7.3: | Hodder Oral Sentence Reading Test mean scores, by group and<br>assessment points ..... | 178 |
| Figure 7.4: | Diagnostic Spelling Test mean scores, by group and assessment<br>points .....          | 180 |

# **1 Nature and origin of the research**

## **1.1 Nature of the research**

The aim of this study was to investigate whether a non-phonics-based intervention, Fischer Family Trust (FFT) Wave 3, might help children in Years 1 and 2 who were falling behind in literacy to catch up with their peers. The study consisted of an 'outer' and an 'inner' study. The outer study was a pilot/feasibility study (PFS) which investigated whether the quantitative approach used in the inner study would be suitable for scaling up to investigate the research questions. The inner study was a quantitative evaluation of FFT Wave 3. It consisted mainly of a small two-armed randomised control trial (RCT) using standardised tests of literacy, supplemented by a number of non-standardised assessments. In total, 24 children (12 in Year 1 and 12 in Year 2) were selected by their class teachers to take part in this study and randomised into two groups.

## **1.2 Structure of this thesis**

As mentioned, this research project was structured as an outer study and an inner study. The outer study describes the scale of the study and the approach that was used. In this case, due to the number of participants available, it was decided that a pilot/feasibility study should be done. A pilot/feasibility study (PFS) mimics a larger study and checks whether all aspects of a proposed larger study work well together. In addition, there is an understanding that statistically significant results are not expected from a PFS, usually due to the small number of participants.

The inner study investigated the effect of a non-phonics-based intervention scheme for children in Years 1 and 2 who were struggling with literacy. The intervention investigated was Fischer Family Trust Wave 3 (FFT). This is a one-to-one intervention that aims to improve not just a child's reading skills, but also their written skills, by alternating reading days with writing days.

The principal approach used to investigate the effectiveness of the intervention was a RCT. The reasons for using this research design are described in chapter 2, which discusses the researcher's positionality and the application of RCTs in educational research. Within the RCT, standardised tests were used to assess



reading comprehension, spelling, and oral word and oral sentence reading. The primary outcome measure was reading comprehension, since this was testing the children's understanding of the written text, and not just their ability to decode, which was tested using the other reading tests. Those and the spelling results are the secondary outcomes.

In addition, three exploratory outcomes were investigated: book band analysis (which yielded supplementary measures of progress in reading), improvement in writing, and possible changes in attitudes to reading.

The thesis chapter structure is as follows: the rest of this first chapter discusses the importance of literacy skills in the world today, and outlines the principal and subsidiary hypotheses; chapter 2 describes the researcher's positionality and insider research position; chapter 3 reviews the principles behind pilot/feasibility studies and the suitability of using this approach in this study; chapter 4 reviews the arguments for and against the use of phonics; chapter 5 reviews the links between reading, spelling and attitudes; chapter 6 describes the methods used; chapter 7 is the results chapter; and chapter 8 discusses the findings of this research and offers the final conclusions.

### **1.3 The importance of literacy**

The ability to read and write is a key skill and is paramount to the success of any child in his/her academic life and later their adult life. If children find it difficult to read, they will consequently find it difficult to access the rest of the school curriculum. Reading and writing are skills that are used daily throughout life. Not only does low literacy affect children's school lives, it also has an impact on their lives outside school. In the age of social media children could be ostracised due to their inability to read/write, since they would not be able to access Facebook, Snapchat or even possess the skills to send/receive text messages. Neither would they be able to use the internet or books to obtain information. During childhood, these individuals would manage since they would have their parents and teachers to help them.

Some poor readers grow up into what are termed 'functionally illiterate' adults. Due to their poor literacy skills, they find everyday life difficult (National Literacy Trust, 2011a), and may feel isolated and unintelligent. Their resulting poor self-

esteem makes it unlikely that they would ask for help, and their progress stalls. For ‘functionally illiterate adults’, their poor skills make access to everyday information difficult, preventing them from being able to fill out a job application form or read a newspaper with ease. It is essential that the teaching of literacy in all schools is excellent, to prevent children with weak reading and writing becoming functionally illiterate adults.

## 1.4 Scale of the problem

Most children learn to read and spell at least to an adequate level, almost irrespective of how they are taught, but even in Year 1, and certainly by Year 2, some make less than expected progress, and are at risk of falling steadily further behind their peers, unless identified and given additional help. Table 1.1 shows the percentages of children not attaining the nationally expected levels at KS1 and KS2 in recent years. (Data for 2016 and 2017 are not shown because of the radical change in method of reporting the KS1 results.)

**Table 1.1: Percentages of children in England achieving below level 2 in reading in Key Stage 1 National Curriculum teacher assessments, or below level 4 in reading in Key Stage 2 national tests, 2007–15**

| Year | Key Stage 1 | Key Stage 2 |
|------|-------------|-------------|
| 2007 | 16%         | 16%         |
| 2008 | 16%         | 13%         |
| 2009 | 16%         | 14%         |
| 2010 | 15%         | 17%         |
| 2011 | 15%         | 16%         |
| 2012 | 13%         | 13%         |
| 2013 | 11%         | 14%         |
| 2014 | 10%         | 11%         |
| 2015 | 10%         | 11%         |

(Sources: DfE, (2015a, b))

These results show that there are significant numbers of children who are falling behind at KS1 (most recently about 1 in 10) and at the end of KS2 (most recently about 1 in 9). Even if the data are re-arranged to align the KS1 and KS2 results for the same cohorts (see

Table 1.2), recent apparent improvements at KS2 have only begun to remedy the position, and it will be several years before the recent improvements at KS1 (see again Table 1.1) feed through to KS2. This causes concerns, since the increased demands at KS3 mean these children are at risk of falling yet further behind, and once children have fallen too far behind they find it difficult to catch up with their peers without help.

**Table 1.2: Percentages of children in England achieving below level 2 in reading in Key Stage 1 National Curriculum teacher assessments, or below level 4 in reading in Key Stage 2 national tests, with age-cohorts aligned**

| Year | Key Stage 1 | Year | Key Stage 2 |
|------|-------------|------|-------------|
| 2007 | 16%         | 2011 | 16%         |
| 2008 | 16%         | 2012 | 13%         |
| 2009 | 16%         | 2013 | 14%         |
| 2010 | 15%         | 2014 | 11%         |
| 2011 | 15%         | 2015 | 11%         |

(Sources: DfE, 2015a, b)

In addition to national data, the government also has access to international data in the form of the Progress in International Reading Literacy Study (PIRLS) series. These are international studies which have investigated the reading ability of children with respect to the two main purposes of reading: the degree of enjoyment that the reader gains from reading, and the skill of reading to gain knowledge. PIRLS provides data that the government can use to compare the reading ability of 9- to 10-year-olds in England with similar-aged children from other countries, and therefore determine any areas for improvement. To date there have been four PIRLS studies, starting in 2001 and occurring every five years. In the 2016 study (McGrane, et al., 2017) 50 countries took part. The results for England, 2001-16, are shown in Table 1.3.

**Table 1.3: England PIRLS results 2001-2016**

| Year | England's score |
|------|-----------------|
| 2001 | 553             |
| 2006 | 539             |
| 2011 | 552             |
| 2016 | 559             |

In 2001 England were third in the chart, behind The Netherlands and Sweden. By 2016, although England obtained their highest score, they were significantly behind the top scorers of the Russian Federation (581) and Singapore (576). England scored comparatively highly when compared with other European countries, but still behind the Republic of Ireland (567), Finland (566), Poland (565) and Northern Ireland (565). The report states that the improvement in England's average score was mainly due to an improvement in the average scores gained by boys, and the progress of the low-achieving pupils.

With regard to attitudes to reading, the findings with respect to England were that, although English children possessed a great confidence in reading, a greater percentage of them did not enjoy reading compared with other countries, which is concerning as it could have a direct impact on achievement (see section 5.8)

Beyond the school years, there are also data on levels of adult literacy. In England, less than 1% of the population are 'completely illiterate' (National Literacy Trust, 2011b), but this statistic is rarely used. The statistic more usually used is the rate of 'functional illiteracy' (National Literacy Trust, 2011b). A functionally illiterate person is often described as an adult who does not have the literacy skills expected of an 11-year-old (former National Curriculum level 4). Comparing adults' and children's skills levels is misleading, but does provide an understanding of the skills that the adults possess, or lack. The adult skills audit summarised by the National Literacy Trust (2011b) goes on to explain how adult reading levels have changed, using results from the *Skills for Life* surveys of 2003 and 2011 (writing was not assessed). The results are summarised in Table 1.4.

Overall there had been a 1.3% decrease in the percentage of adults with reading skills at Entry level 3 and below between 2003 and 2011, and an increase in the percentage of adults with reading skills at Level 1 and above. Although there had been a 1.6% increase in adults with Entry level 1 or below reading skills this result was not significant. There are eight years between the two *Skills for Life* surveys. The minimal change in the results poses a serious question about the teaching of literacy skills in our schools, and about the lessons that are failing to be learned, and whether phonics is providing the improvement that is required.

**Table 1.4: Comparison of adult reading levels for the years 2003 and 2011**

| Adult reading levels  | Percentage of adults in 2003 | Percentage of adults in 2011 |
|---|------------------------------|------------------------------|
| Level 2 or above<br>(GCSE grades A* - C)                        | 44.2                         | 56.6                         |
| Level 1<br>(GCSE grades D – G)                                  | 39.5                         | 25.8                         |
| Entry level 3<br>(NC expectation of 11-year-olds)               | 10.8                         | 7.8                          |
| Entry level 2   | 2.0                          | 2.1                          |
| Entry level 1 or below<br>(NC expectation of 5- to 7-year-olds) | 3.4                          | 5.0                          |

Source: BIS (2011, p. 5)

In addition, the Leitch report (2006) stated that there needed to be a significant improvement in the UK's basic skills to ensure that it would be able to compete effectively in the global economy. The report stated that 95% of the adult population would need to be functionally literate by the year 2020 to maintain the UK's standing in the higher quartiles of OECD countries. This provided the country with a significant problem, since 70% of the people who will be in the workforce then had already left school (Leitch, 2006). There are national programmes to help people who are in the workforce; the problem arises in helping those who are not. If these people had been identified and helped within schools this might not have been an issue.

## **1.5 The cost of poor literacy**

Evidence on the cost of poor literacy in the UK was provided by an audit by the UK charity Every Child a Chance Trust (2009). They looked at relevant research and identified specific areas where they felt poor literacy skills had financial implications. For example, they stated that children who had poor literacy skills by the end of Year 2 had a 34% chance of receiving a statement of special educational needs by the age of 11. The report went on to state that the costs of providing interventions at primary school are not reflected in their effectiveness. So it was suggested that, due to the fact that the economic benefits of providing interventions are not reflected in sufficient progress, many primary schools may decide not to offer them. The report also stated that there are links between poor

literacy skills and antisocial behaviour, which in turn leads to exclusions and additional costs. The cost of SEN provision (expressed as total lifetime costs) was shown to be £298 million. They estimated that every pound spent trying to overcome literacy difficulties before a child turns seven would generate a 'return on investment' of between £11 and £17 over their lifetime. But if these literacy problems were not dealt with, a public cost of £64,000 per individual over their lifetime could be foreseen, as a result of the cost to the welfare system.

In their interim report (Cree et al., 2012) and the final report of the same title (World Literacy Foundation., 2015), the authors cite the UNESCO definition of functional literacy more explicitly:

[It is] measured by assessing reading [and] writing ... in the various domains of social life which influence individual identity and insertion into society. From this perspective, literacy involves not only reading and writing but also the acquisition of the skills necessary for effective and productive performance within Society.

Cree et al. (2012, p. 3)

The results in Table 1.4 suggest that there are approximately 5.2 million functionally illiterate adults in England (National Literacy Trust, 2011b), 3.5 million of whom, according to the same source, are at the upper end of the Entry level scale and have strengths and weaknesses in different areas. Many can read straightforward texts, and are more proficient at tasks involving subjects that they know about. However, they will struggle with new subjects. With this level of literacy, their ability to get a job or help their children with their homework is very limited

At the first World Literacy Summit, held in Oxford in 2012, the cost of poor literacy to the UK economy was stated to be:

- £81.312 billion lost to the UK economy each year through poor literacy skills
- £58 billion lost by individuals and businesses through lower personal income or business earnings due to low literacy skills
- Approximately £23.312 billion of tax-payers' money spent on benefits and social programmes. Many people with poor literacy skills are unemployed because they do not have the necessary reading and writing skills to apply for a job or indeed perform a job. Lack of money and focus can then

unfortunately cause some to resort to crime, resulting in imprisonment. Some make poor health choices since they struggle to read the appropriate information, and costs for healthcare become higher.

Cree et al. (2012, p. 2)

These figures are based on several key areas and are based on lifetime costs. For example, the report states that poor literacy skills can lead to poor job prospects. Thus further costs are incurred due to the need for a welfare state, which provides financial support for adults who are not working. However, financial penalties are also noted because the jobs that adults with poor literacy do are not as well paid. Health is another area where cost implications are highlighted. As well as there being an increase in the risk of mental illnesses, such as depression, there can be illnesses caused by poor diet, hygiene and personal health. Again this has a financial implication.

The final report *The Economic and Social Cost of Illiteracy: A Snapshot of Illiteracy in a Global Context* (World Literacy Foundation, 2015) reiterated the findings of the interim report, but stated that illiteracy had a global cost of US\$1.2 trillion. A number of recommendations were made. The first was to establish adult and parental literacy strategies. By improving the literacy skills of all adults, they will be able to perform everyday tasks easily. By improving the literacy skills of parents, they can help their children, as well as encouraging their children to enjoy literacy. The second recommendation was to develop strategies by which children with poor literacy skills continue to attend school and appreciate its importance. To achieve the above two recommendations, it is necessary that there are sufficient resources and training, which need to be paid for. The report continued to state that, considering that doing very little to alleviate the problem is costing US\$1.2 trillion, investment should reduce the cost over time.

However, as with all findings that quote such large cost implications, it could be a stratagem to encourage world governments to invest in the area of early literacy education. By emphasising high costs and poor literacy levels, governments are provided with the evidence they need to invest, as it is of long-term benefit to the economy and the country.

## 1.6 Literacy levels in the UK

It is a fact that low literacy levels in the UK are not decreasing as rapidly as would be liked, as shown in Table 1.1 and Table 1.2. Investment is needed, but maybe the way in which reading, spelling and writing are taught in schools also needs to be reappraised.

The failure to acquire good literacy skills at school has long-term effects which affect not just the individual but the whole global economy, as previously discussed in the previous section.

The report *Reading: the next steps: supporting higher standards in school* (DfE, 2015c) stated that poor literacy skills not only affect employment prospects and thus financial security, but also impact on social issues:

For example, adults in England with low literacy levels have twice the odds of reporting low levels of trust as their peers with high literacy, and three times the odds of reporting poor health.

DfE (2015c, p. 7)

The report went on to say that a 2012 study (Programme for International Student Assessment (PISA)) showed that 15-year-olds in the UK were outperformed by those in 17 other countries with respect to literacy. And the gap between our highest and lowest achievers was above average and, more worryingly, two-thirds of the OECD countries had smaller gaps (DfE, 2015c, p. 8).

These figures show that literacy teaching in the UK is failing some children. The report blamed the UK's poor performance on a 'decade of stagnation' with respect to assessments, even though a large amount of money had been invested in education. The report cited the OECD, whose 2012 study showed that 'the UK was the only country whose younger adults did not have better literacy skills than those nearing retirement age' (DfE, 2015c, p. 8).

The findings are concerning. There has been a significant investment made in education with respect to literacy teaching. The benefits from these investments should have been seen, and literacy skills of younger adults and the older generation should at least be equal; ideally young adults should be excelling. The



young of today have far more opportunities than the older generation had, but they also have far more distractions, e.g. social media, smart phones, etc.

## **1.7 The influence of the Rose report**

Government policy on improving literacy levels currently focuses on the teaching of initial literacy in English schools, and this is dominated by a greatly increased emphasis on phonics, especially synthetic phonics. This can be dated largely from the publication of the Rose Report (Rose, 2006), which is discussed further in chapter 4. Here it is only necessary to note that it has had powerful effects. Most obviously, both the former Labour government and especially the coalition government of 2010-15 emphasised synthetic phonics as the sole approach to teaching reading, in the face of the evidence and valid concerns about wider matters – and this push has continued under the two subsequent Conservative administrations.

The Rose Report also led immediately to a burgeoning of phonics-based intervention schemes. Brooks (2007, p. 26) pointed out that there were many more such schemes present in his 2007 edition than in his 2002 edition, where he reviewed just four such interventions (Brooks, 2002). He stated that this increase was due to the influence of the Rose report. Continuing this increase in phonics interventions, in the 4<sup>th</sup> edition (Brooks, 2013) there were 23 phonics-based schemes, compared to just 12 non-phonetic approaches (these numbers include primary and secondary intervention schemes), whilst Brooks (2016) contains 26 phonic approaches compared to 14 non-phonetic approaches (again including primary and secondary schemes). However, phonics schemes, especially those based wholly or mostly on phonics, may suffer from a set of limitations.

## **1.8 Limitations of phonics**

There is a wealth of evidence (see chapter 4) supporting the use of phonics and the advantages that it offers children who are beginning to learn to read and spell. For this reason, phonics has become a mainstay in the national curriculum but, as with all good things, there are limitations to the exclusive use of phonics to teach children to read.

The major limitation of phonics is that it does not teach children how to read. It teaches children how to decode words, that is, work out and identify words that are unfamiliar to them (so far) in print. It does not provide them with comprehension skills. Many parents say their children can read, but do they understand what they are reading? Initially children learn how to decode the words they see on a page, and as time progresses their understanding of what they are reading increases. But the ability to decode words must not be confused with reading. The teacher must be able to differentiate between those children who are reading and those who do not grasp the meaning of the words that they have just 'read' (aloud).

In addition, Davis (2012) is of the opinion that the use of synthetic phonics slows down the reader's overall reading speed, which would impact on the ability of a child to comprehend what they are reading. Davis expresses his distaste for phonics in far stronger terms and is concerned that there is too much emphasis on the phonics approach. He believes that teaching phonics to children who can already 'read for meaning' is a retrograde step. His views suggest that teaching phonics causes the children who can read to start to double-guess what they are reading, and to look just at single words rather than the whole context.

Phonics is a cuing system designed to help beginner readers to decode new words, it is not reading. Phonics provides a child with the skills needed to identify (generate, 'sound out') the correct phonemes from the corresponding grapheme, and blend those sounds to produce a recognisable word, whether this is a previously known word or a new word. Phonics also provides the skill to eliminate similar-looking words and identify the correct word. This allows the reader to determine which word to say. This second skill is only useful if there is meaning in the reading. Phonics is only a time-limited means to an end, and should not be used as a means to determine a child's ability to read.

A further limitation with merely teaching phonics is that English spelling is not completely phonemic. Although there are 44 phonemes in spoken English, there are only 26 letters in our alphabet, which means that the system has to be supplemented with digraphs, trigraphs and even some four-letter graphemes (e.g. <ough> representing /u:/: in *through*). Also, almost every phoneme has more

than one spelling, for example <au, or, aw> representing the phoneme /ɔ:/ as in *fraud, ford, crawl*, and most graphemes have more than one pronunciation, for example <th> can be pronounced /θ, ð, t/, as in *think, this, thyme*. By contrast, Finnish has one grapheme per phoneme and *vice versa*, which makes it much easier to learn to read the language aloud (and it helps that every word in Finnish is stressed on the first syllable) and (for native speakers) with understanding.

When children are taught how to read English, considerable emphasis has to be placed on teaching children the 100 most frequent words, and analyses of those further illustrate the limitations of phonics. Stuart et al. (2003) constructed a database which consisted of the reading material from an inner-city school which in turn consisted of a variety of reading schemes. Using this database they determined the frequency of words. This showed that the 100 most frequent words accounted for 54.1% of all word tokens in their database, and that the 500 most frequent words accounted for 76.3% of all word tokens. The authors went on to state that 49 of the 100 most frequent words are non-decodable, and emphasise the importance of the need to teach these words as sight words.

A later analysis by Solity & Vousden (2009) looked at the importance of the 100 most frequent words in different types of texts, mainly real books and reading scheme books. They discovered that 50% of all the word tokens found in all texts were one of the 100 high frequency words, showing that, during their time in reception through to year 2, all words that children read have a 50% chance of being in the 100 most frequent list. Interestingly, they go on to say that just 39 of the 100 can be decoded, suggesting that it is important that the remaining 61 words are taught as 'sight words', otherwise the majority of the text would be less accessible to the readers. It is not immediately obvious why their count of non-decodable words differs from Stuart et al.'s.

Thus a strictly phonics-only diet would bump hard up against ineluctable facts about English orthography. If children start to fall behind in their reading, intervention schemes are used. Yet the intervention schemes that are currently available tend to be synthetic phonics-based, probably because synthetic phonics is currently the favoured method of early reading tuition encouraged by the government. This fails to provide children who are falling behind with

alternative reading strategies; offering the same approach as a remedial action could produce more problems than it solves.

## **1.9 Bad experience with phonics**

There are several issues surrounding the teaching of children who are falling behind in their literacy skills, and more specifically the apparent necessity to use synthetic phonics at any opportunity.

The author of this research had repeatedly witnessed children ‘sounding out’ every word that was read and then failing to comprehend what was written, mainly because their reading was stilted. Further frustration developed when children were heard to ‘sound out’ high-frequency words that were not decodable, for example, the word *said* – children would say /s-æ-ɪ-d/, then say /sed/, correctly, so the sounding-out was futile. Although two of the phonemes are the same (/s/ and /d/), /æ-ɪ/ do not produce /e/. There appeared to be an absence of the correct teaching of sight words, and a lack of understanding on the children’s part that not all words are decodable.

In addition to this, children would ‘sound out’ increasingly complex words, for example *grandfather* would be /g-r-æ-n-d-f-æ-t-h-e-r/, which is so long that the children would not be able to blend the sounds to form a coherent word; even if they could it is unlikely it would be /'grænfɑːðə/.

The fact that children used ‘sounding out’ at inappropriate times may not be a limitation of phonics, but more a limitation that the education system has put on itself. When the children came across an unknown word that they could not decode, they did not possess any other skills that they could use to work out these more complicated words. This then led to the question whether there were alternative approaches that could provide struggling readers with additional literacy skills and therefore improve their reading.

Brooks (2007) provided details of over 30 interventions for the key stage 1 age group which had been quantitatively evaluated in the UK and shown to be effective. Many of these schemes were entirely phonics-based, and almost all contained some phonics. Since synthetic phonics is now prescribed by the

government as a substantial element of the initial teaching of literacy in England, phonics-based intervention schemes represent giving children who are falling behind 'more of the same'; this research study is based on the hypothesis that an alternative approach to reading tuition may improve children's reading skills by the same degree as, or even more than, a phonics approach. For reasons that will be given in due course, the choice of a scheme for testing this idea fell on FFT Wave 3, which is not phonics-based; the research design of choice was quantitative (mainly an RCT, with supplementary aspects), and was intended as a pilot/feasibility study.

This was the basis of this investigation, to identify and evaluate a non-phonics-based intervention, and this provided the overall research question. Is it possible that the introduction of a complementary strategy, that is an alternative non-phonics-based intervention, would provide children struggling with literacy with the additional skills to improve their reading, spelling and writing? On the basis of theory, data and experience it became apparent that the use of a non-phonics-based reading intervention, instead of providing possible confusion, might help and encourage struggling children to overcome their literacy difficulties. The additional strategies that they would be taught would give them confidence, which might improve their self-esteem and consequently improve their reading skills. This study was designed to test this idea by comparing the progress made by a sample of children who were given a non-phonics-based intervention with that made by an equivalent group of children who were not given the intervention.

### **1.10 Principal and subsidiary research questions and hypotheses**

Due to the nature of this study, as already declared, there were within it what might be called an 'outer study' and an 'inner study'. The outer study was the testing of the methods to determine whether the methodology chosen would be a suitable approach that could be replicated in a larger study to assess the effectiveness of a non-phonics-based literacy intervention; this will be referred to, for brevity, as the pilot/feasibility study, or PFS, even though that label strictly applies to the entire study. The inner study was the quantitative evaluation of the chosen intervention, its aim being to statistically analyse the data and determine the effectiveness of the intervention, with the understanding that the likelihood of

obtaining statistically significant results was low due to the small number of participants in the trial.

As such, there are both 'outer' and 'inner' research questions and hypotheses.

### **1.10.1 Pilot/feasibility study research question and hypotheses**

#### **Research question:**

Is the quantitative evaluation within this pilot/feasibility study suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study?

#### **Null hypothesis:**

The quantitative evaluation within this pilot/feasibility study is not suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study.

#### **Substantive hypothesis:**

The quantitative evaluation within this pilot/feasibility study is suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study.

### **1.10.2 Quantitative study research questions and hypotheses**

Because there are several outcomes (the reasoning behind this is explained later), there is a corresponding plethora of research questions and hypotheses, as follows:

#### **1.10.2.1 Primary outcome**

##### **Research question:**

Can a non-phonics-based intervention significantly improve struggling children's reading comprehension?

**Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's reading comprehension.

**Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's reading comprehension.

**1.10.2.2 Secondary outcomes****Research question:**

Can a non-phonics-based intervention significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling?

**Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling.

**Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling.

**1.10.2.3 Exploratory outcomes****Research question:**

Can a non-phonics-based intervention improve struggling children's progression through book band levels and/or writing ability and/or attitudes to reading and books?

**Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's progression through book band levels or writing ability or attitude to reading and books.

### **Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's progression through book band levels and/or writing ability and/or attitude to reading and books.

#### **1.10.3 Summary**

This chapter has briefly explained the nature of the research and the concept of the outer and inner studies. The chapter then explained the importance of literacy skills in everyday life, and the effect of life-long poor literacy skills on the global economy, and more specifically on the UK economy, highlighting the importance of all children and adults having access to the best teaching available. The effect of the Rose Report (Rose, 2006), and the subsequent governmental influence on the introduction of systematic phonics within the national curriculum are discussed, and how this shift in beginning reading tuition has resulted in an influx of phonics-based interventions. Limitations of and bad experiences with phonics were illustrated.

The final section identifies the research questions and hypotheses. There are a number of these, dependent upon whether they are concern the outer or inner study. The inner study questions are split into primary, secondary and exploratory outcomes.



## **2 Positionality and insider status**

### **2.1 Introduction**

This chapter discusses the positionality of the researcher, and how the use of an RCT approach for the investigation of the inner study complemented her positivist views, as well as being the most appropriate method for this study. The final section of this chapter looks at the issues that need to be considered when doing research as an insider.

### **2.2 What is the researcher's positionality?**

Before any research is undertaken it is important that the researcher is aware of his/her positionality, that is, how the researcher's own values and beliefs impact upon the research questions asked, and how the research is carried out, as well as how the data gathered are interpreted and subsequently presented.

At the start of this research project the researcher regarded herself as strictly positivist, but as the research progressed it became apparent that the black-and-white nature of a purely positivist approach was not possible in a sociological setting, and in reality, difficult to achieve in purely scientific disciplines (traditionally the realm of positivist theories) since researchers' beliefs and values fundamentally influence what questions they ask and subsequently how they answer them.

During the course of this research the researcher's positionality altered, from being a classical positivist to a positivist with post-positivist tendencies. This journey from classical positivist to positivist with post-positivist tendencies will be traced by describing the attributes and flaws of both theories.

It was a natural decision for the researcher to identify herself as a classical positivist, since she had a scientific background and a preference for dealing with facts and figures rather than the more subjective qualities of qualitative methods. More importantly, an empirical approach was the most appropriate approach for the research questions of the 'inner' study.

In any research, the most appropriate method(s) should be used to ensure that the hypotheses are rigorously tested, and the data reported are repeatable,

reproducible and accurate. Some projects will lend themselves to a qualitative approach, such as case study or grounded theory, whereas some will be more appropriately tackled using quantitative approaches such as RCTs.

Within an educational research scenario, a purely classical positivist approach is not possible. Children are individuals, and their personal qualities undoubtedly affect how they progress, as well as external factors such as age, gender, family situation, where they live, or indeed how they are feeling on any given day. This is an issue of concern across the board with social science/educational research, and is often an argument used against the use of quantitative approaches in these areas.

### **2.2.1 A definition of positivism**

Classical positivism is a philosophy that rigidly sticks to the concept that only 'factual' knowledge that has been obtained by direct observation (the senses), which includes measurement, is trustworthy. A researcher in a classical positivist study is purely an observer who collects data and subsequently interprets those data using an objective approach. The findings of the research are therefore observable and quantifiable (Dudovskiy, 2016).

Although a classical positivist approach of observation and recording of data can be seen as far back as Galileo Galilei (1564-1642), Auguste Comte (1798-1857) is viewed by some as the father of classical positivism.

Comte, like his compatriot Saint-Simon (1760-1825), realised how important science was, and how scientific enquiry could be used to study and improve society. Comte believed that, as there are laws that explain many 'scientific' aspects of the world, such as gravitational theory, there must also be 'laws' that influence social phenomena. One would assume that Comte would think that, by discovering these laws, society could be manipulated (Fletcher & Barnes, 2016). He wanted to create a branch of science that would, he hoped, be as influential as other natural sciences. This branch of science was termed social science.

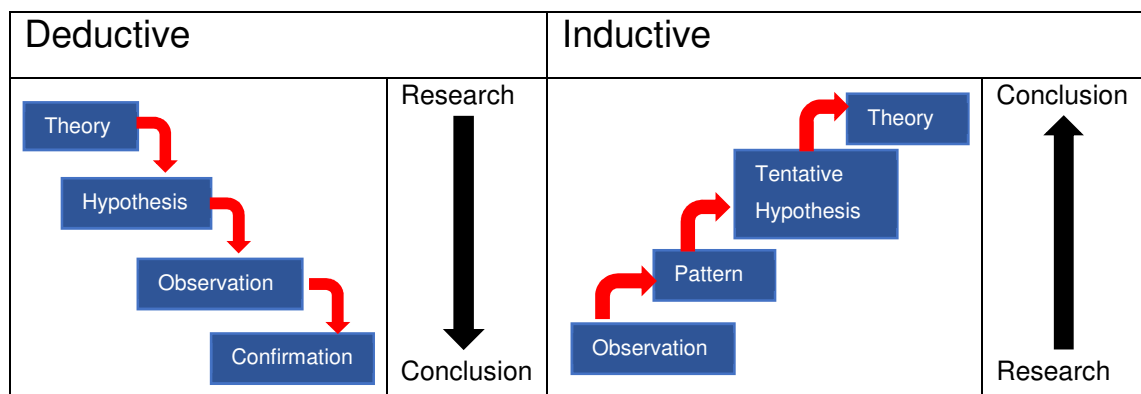
The idea that there are specific 'laws' that govern society has become less popular, and society has realised that human beings are far more complex and

cannot be so easily categorised. Maybe these ‘laws’ referred to by Comte should be termed ‘trends’, unspoken rules that govern how we should live within society.

The term ‘laws’ implies unbreakable qualities, suggesting a ‘Brave New World’ society where individuality is unacceptable, whereas the term ‘trends’ suggests that there can be some variation in accordance with the fact that we are all individuals.

Studies that use a positivist approach usually have a deductive rather than an inductive approach. In this thesis the term ‘deductive’ is used in accordance with the following definition. A deductive approach involves the researcher developing a hypothesis based on theory, and then collecting information and data to prove or disprove the given hypothesis, whereas an inductive approach (which is generally applied to qualitative methods) is much more flexible. There is no need to establish pre-determined theory and then collect data to support that theory. The researcher observes data and collects facts to reach a possible hypothesis and define a theory as per the research problem (see Figure 2.1).

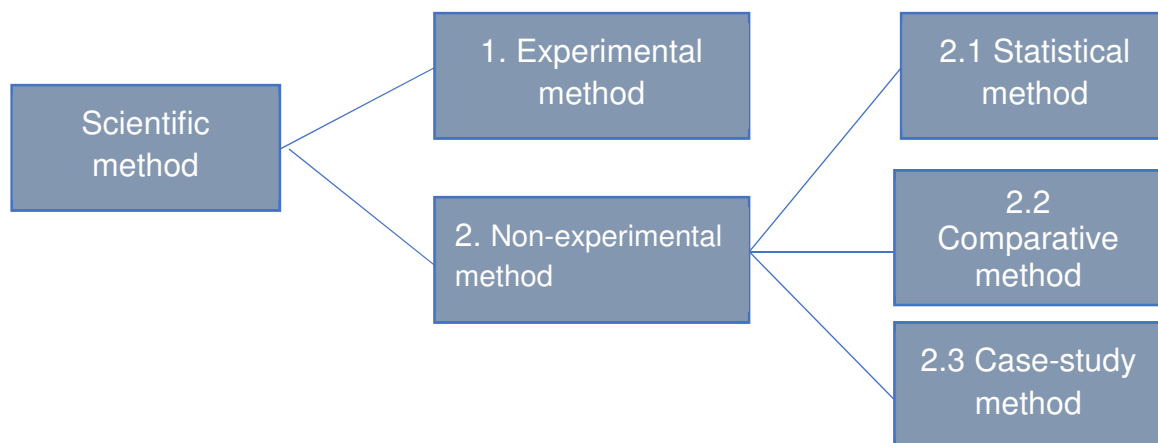
**Figure 2.1: Comparison of deductive and inductive research approaches**



Modified from Trochim, (2006, online)

A positivist approach is seen by most as the pinnacle of all approaches, and this is shown in a hierarchy of methods (Figure 2.2). The use of a scientific framework to investigate social situations should provide evidence that the research has not been influenced by the values and beliefs of the researcher. These interpretations are made based on a rational approach. This provides, in the eyes of the world, undeniable truth and fact.

**Figure 2.2: Hierarchy of methods**



Taken from Jakobsen (2013, online)

Reviewing the approaches used in empirical scientific experiments, the researcher realised that research scientists do not have complete control over their experiments, especially after a chemical reaction begins. Scientists can control external factors such as pressure and temperature, plus the concentrations of the chemicals, which allows them to perform repeatable experiments. But whether the observations that are made are done with complete objectivity is questionable.

The flaws of a positivist approach are regularly highlighted when this approach is applied to a social science scenario. As mentioned previously, one of the main 'flaws' of a positivist approach is the alleged impossibility of controlling all outside influences on the behaviour/outcome of the participants. An example of this concern was experienced whilst this study was being undertaken. It was stated by a senior member of the School of Education that 'it must be remembered that children are not chemicals in a test tube that will all behave in the same manner.' This fact is self-evident, and any educational researcher would be aware of this. What must be remembered is to take into account the plethora of external factors that may affect a child's behaviour at any one time during the study. If necessary, notes should be taken to remind the researcher of incidents that may over- or underestimate the effectiveness of the intervention. If researchers embark on a research project expecting everyone to behave in an identical manner they are opening themselves up to failure. However, this researcher believes that trends can be identified, and their presence shown and proven by the use of statistics and quantitative method.

The problems of classical positivism in relationship to educational research can be summarised as three shortcomings. The first is that classical positivism relies on direct observation and recording of data; however, there are always variables that are not based on direct observation, which include time, space and cause. In educational research, the variables that cannot be directly observed include emotional status, socio-economic factors and parental support.

The second shortcoming is the assumption that interactions that happen between individuals or relationships between people are all based on a known variation of a set of pre-existing variations, whereas in reality the range of interactions between individuals is variable almost on a day-to-day basis, and trying to define them specifically is an impossibility.

Finally, the third shortcoming of classical positivism, when dealing with areas of research in the social sciences, is failing to acknowledge the effect of human nature and its influence on any study undertaken.

As time has progressed it has become apparent that the prescriptive nature of classical positivism cannot be so easily applied to a social science environment. The question to be asked is: how is it possible to build the human element into an approach that is as objective as possible?

Classical positivism, according to Ryan (2006), started to fall from grace for a number of reasons. The first, in her opinion, is that neutral knowledge does not exist. Neutral knowledge is knowledge that is gained and understood without personal feelings/experiences influencing the information gained. Ryan (2006) states that knowledge cannot be neutral since knowledge is influenced by personal experiences, beliefs and understanding. Ryan (2006) suggested that the knowledge we gain, whether it is objective or subjective, public or private, scientific or emotional, is socially constructed. The fact that social construction theories presume knowledge, in this case, is dependent on our perceptions and built on our experiences, means neutral knowledge cannot exist. Ryan (2006. online) stated 'It is not possible to separate ontology from personal experience.'

Bruner (1991) discussed the impact of narrative on knowledge acquirement. He used the term 'narrative' as a description of how knowledge is interpreted by the

individual and is a reflection of a person's experiences and beliefs. This view supported Bruner's stance that knowledge is socially constructed from previous experiences whether it is encoding action-based knowledge, e.g. a baby shaking a rattle (enactment), or encoding knowledge from pictures (iconic), or the final of Bruner's phases where information is encoded from symbols, such as language (symbolic).

The second reason for classical positivism's fall from grace is linked in with the idea of dualistic thinking, the idea that things are either black or white. Researchers realise that it is not possible to separate the polarised views of thinking, whether subjective versus objective, or scientific versus emotional. The human element is so complex that shades of grey are inevitable, and trying to simplify educational research into an either/or scenario is not possible and consequently becoming outdated (Ryan, 2006).

The third reason is down to ethical considerations. The idea that a researcher should see participants purely as numbers and experimental data is no longer acceptable. Thus procedures, techniques, analysis must always be subjected to ethical approval.

The answer to these issues is the theory of post-positivism. Ryan (2006, online) stated that post-positivism has the following characteristics:

- Theory and practice cannot be kept separate. Theory cannot afford to be ignored for the sake of just facts.
- The commitment and motivation of the researcher to the research is paramount.
- The thought that research is only concerned with the correct methods for collating and analysing data is not sufficient.

Ultimately, post-positivist researchers approach their research by acknowledging that things will not be value-free, and that subjectivity is part of research, as is the essence of humanity that will affect all social science research.

As Taylor (2013) stated, post-positivism is a modified scientific approach used in the social sciences; it aims to provide knowledge about social patterns from research, knowledge that can be classed as objective and generalizable.

Consequently, this information can be used to confirm the presence of ‘universal properties/laws amongst pre-defined variables’ (Taylor, 2013, online).

The following sections discuss the importance of a positivist approach in educational research, since randomised control trials are seen as a positivist approach.

### **2.2.2 Positivism in educational research, or the qualitative versus quantitative dichotomy**

Even though many believe that a scientific approach in social science research is unsuitable, there is support for a positivist positionality, just as there are those who would reject this approach. For example, Howe (2009) feels that a positivist approach in educational research should be cast off, as it has been in philosophical research. He believes that trying to treat human beings, or indeed understand them, as if they were simple molecules reacting in a chemical reaction is unacceptable. He suggests that trying to pre-determine a specific outcome is an inadequate method of conducting social science research. Indeed, his thoughts are reflected in the quote by Giddens (1974, p. 13):

Those who are still waiting for a Newton of social science aren't only waiting for a train that won't come in, they're waiting in the wrong station altogether.

Giddens is suggesting that the quantitative/qualitative opposition, or indeed perennial disagreement, is insurmountable.

However, Bredo (2009), in his response to Howe (2009), has a more measured understanding of the issues regarding a positivist stance. He acknowledges that there does exist a quantitative/qualitative dichotomy which has been created by those who, in simplified terms, think their way is the best way. Bredo believes that, instead of creating or indeed fuelling “Methodology Wars”, the most important thing is producing good research. It is imperative to ensure that the most suitable methods are used to investigate the proposed research question. Whether these are qualitative or quantitative methods exclusively, or a combination, is important, but subordinate to ensuring that the most appropriate method is used.

Gorard (2007) further describes these antagonists specifically as “two idealized villain”, those researchers who will either only deal with numbers or those researchers who will not touch numbers (the latter might be called ‘arithmophobes’). In addition, taking such polarised views could affect the efficacy and reliability of any research undertaken. Historically there has been some poor quantitative research based on badly calculated statistics. Understandably, these poor studies have thrown doubt on quantitative approaches, especially with respect to social science research. However, if only purely qualitative research were used this could create problems. The journals would be full of purely subjective studies creating a forum where prejudice and preconceived ideas would be allowed to remain, without having the ability to prove or disprove them.

### **2.2.3 Quantitative versus qualitative approaches**

Within educational research there is a reluctance to embark on quantitative research, as many researchers still believe it is unethical to use quantitative research in a social science setting. The main issue should be identifying the most suitable method for answering a specific research question. As Walsh (2012) so eloquently explains, the quantitative/qualitative dichotomy is false. This so-called division is created by researchers trying to judge qualitative data using the standards applied to quantitative data, and vice versa. As Walsh states:

“... these discussions have often been with academic colleagues who were judging the quality of the student’s work by criteria that were inappropriate given the research question and the approach the student had taken. Most recently there have been a couple of occasions when a student’s qualitative descriptive research was judged by criteria most normally applied to statistical research.”

Walsh (2012, p. 9)

Walsh (2012) continued by saying that the qualitative work was criticised for being non-generalisable and non-representative.

Walsh (2012) obviously states that it is not possible to judge qualitative research by quantitative criteria, or *vice versa*, because the expectations cannot be met – they are incompatible. If these incompatible judgements occur regularly the validity of all research has been unfairly judged but the quantitative/qualitative



debate is perpetuated. Under these conditions neither quantitative nor qualitative approaches would be acceptable or produce the quality of results needed to substantiate a hypothesis. All educational researchers need to understand that qualitative and quantitative approaches are not mutually exclusive but combining quantitative and qualitative methods the quality of the data produced can be significantly improved.

Researchers need to be aware that educational research cannot be purely quantitative. As Johnson & Christensen, (2012) describe it, this incompatibility thesis ('the proposition that one cannot mix quantitative and qualitative research') is a failure:

...to recognize that creative and thoughtful mixing of assumptions, ideas, and methods can be very helpful and offers a third paradigm. The mixing of ideas and approaches has been present throughout history because mixing or combining builds upon what we know and offers new ways to understand and study our world.

Johnson & Christensen, (2012, p. 31)

#### **2.2.4 Randomised control trials: a useful quantitative tool**

Reflecting the post-positivist stance of the researcher it was determined that a quantitative approach, such as an RCT, should be considered for the inner study. Although RCTs are more commonly used in healthcare, where direct comparisons of treatment versus non-treatment groups are usual occurrences, this approach can easily be adapted for use in educational research where interventions are being assessed.

It is often assumed that education researchers started using RCTs after observing their successful use in medical research. This is untrue. The first medical RCT was completed in 1948, and was a study looking at the treatment of tuberculosis with antibiotics (Medical Research Council, 1948); but it was in 1901 that educationalists Thorndike and Woodworth recognised the necessity to have a control group in one of their experiments (Oakley, 1998). Using an RCT allows certainty that, to the best of the researcher's knowledge and ability, all factors which might affect the outcome of a piece of research are levelled. This

means that the chances of these factors influencing the results are eliminated, barring rogue results (which you then further control by using large samples and replication). This therefore increases the chance to show whether, for example, a specific intervention is effective

A number of authors describe the effectiveness of RCTs. For example, Sibbald (1998) states that ‘...randomised control trials are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome.’

West & Spring, (accessed 10/10/16) state that random assignment to treatment and control groups results in groups that contain a mix of known and unknown participant and environmental factors, thus ruling out the risk of bias.

Since randomised control trials produce numerical data that can be analysed by statistical tests this provides numerical evidence of the effectiveness of the intervention (Himmelfarb Health Sciences Library, 2011), which allows for easier comparisons across different studies, allowing for meta-analyses to be undertaken. In addition it is easier to blind/mask than with observational approaches, as it is possible for the researcher to be unaware of which group is being given the intervention. However, this is not so easy in educational research.

As with all approaches, there is a need to identify their drawbacks. Sanson-Fisher et al., (2007) describe the limitations of an RCT’s ability to evaluate population-based interventions based upon both methodology and pragmatic concerns. These include:

- Population availability: the number of available participants may be limited, which could influence the outcome(s) of the study by either under- or over-estimating the overall effect of the intervention;
- Time for follow-up: some follow-ups would need to be undertaken years in the future. Obtaining funding for such long-term studies would be difficult to gain;
- External validity: the selected sample is not a true representation of the whole population;
- Ethics and informed consent: the main issue here is a perennial one, the possibility of giving life-changing help to a certain group but not to another.

This is often dealt with by offering the neediest patients the initial opportunity, which occurs in medicine regularly.

Sanson-Fisher et al. (2007) do conclude by saying that, for individual-oriented studies, RCTs are still the gold standard.

### **2.2.5 Lessons from randomised control trials in healthcare research**

Torgerson et al. (2005) rightly state that the standard of RCTs, indeed any research, should be of the highest quality. To determine the standard of RCTs they compared the standards of non-surgical and non-pharmaceutical healthcare RCTs with those undertaken in educational research. Their findings were conclusive: the healthcare trials were found to have a significantly higher average quality than those in education. They explained that this could be due to the quality of the journals that the healthcare papers were obtained from. But even when they looked in lower-quality healthcare journals the quality of the RCTs was still higher than those RCTs published in educational journals. In addition, they stated that the quality of the healthcare trials was improving over time.

In answer to the need for high-quality RCTs the Medical Research Council (MRC) published a set of guidelines '*A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health*' (2000). Complex interventions, according to this report, can be therapeutic or preventive and what makes them complex is the fact that they contain numerous different facets that could be directly linked to the success or failure of the intervention. Complex interventions are shown to sometimes contain forms of educational interventions where interventions are used to improve the skills of healthcare professionals or test the effectiveness of new treatment guidelines or protocols.

The report (MRC, 2000) continues to state that the expected design for a main study design will be an RCT, and that the purpose of the report is to provide guidance on how to use qualitative (observational studies) as well. RCTs may not provide all the answers, but they do provide reliable evidence.

Bonell et al. (2013) conducted a matched-pair individual allocation randomized trial investigating "Teens and Toddlers" (T&T), an intervention designed for at risk 13/14-year-old girls. It was hoped that, by promoting the girls' awareness of the

responsibilities linked to looking after a child, the risk of teenage pregnancy would be reduced. The intervention consisted of 3-hour weekly sessions over an 18- to 20-week period. The sessions took place in local pre-school nurseries.

Questionnaires were used at three time points (pre-test, post-test (immediately after the intervention) and follow-up (1 year later)) to gather information from the girls. The results for the four primary outcomes showed no significant effect; however, there was an improvement in three of the fourteen secondary outcomes.

The healthcare education report '*Randomised controlled trials for policy interventions: a review of reviews and meta regression*' (Oliver, et al., 2010) compared randomised and non-randomised trials on policy interventions. The main recommendation made by the report, though is that 'policy evaluations should adopt randomised designs wherever possible' (Oliver et al., 2010, p. 80).

As can be seen, RCTs are the first choice in healthcare education research, and the lessons learnt by healthcare professionals with respect to what errors to avoid are there for the taking.

'*A guide to running randomised controlled trials for educational researchers*' (Hutchison & Styles, 2010) offers educational researchers the guidance they need to conduct high-quality randomised control trials, including guidance on design and statistical analyses that could be done. The authors compared and contrasted RCTs done in healthcare with those done in education with respect to population, ethics, risk, randomisation, intention to treat, and finances. The outcomes for both areas were similar, and showed that RCT's are a suitable form of investigation in educational research. This guidance was written to encourage the education research community to embark on this type of research by providing evidence that RCTs are a suitable method to be used.

The interventions used in healthcare education research are comparable with interventions that are used in education. Thus the use of RCTs in education is a practical, but under-used resource.

### **2.2.6 Summary**

In this section the positionality of the researcher has been discussed and its influence on the choice of research method that should be used to investigate the inner study. Evidence has been provided to show that a well-designed RCT can provide reliable and robust results.

Primarily the method used should be suitable to answer the research question. In this study the inner study question lends itself to a quantitative approach, and the principal method is an RCT which could provide evidence of causality, not just correlation.

Although a one group pre-test/post-test design could have been implemented, it would not have provided evidence that the progress made was any different to a group who had not received the intervention. This approach would provide evidence of correlation but not of causation. Many studies do use a single-group pre-test/post-test design. In fact, the great majority of the studies in 'What works for literacy difficulties?' (Brooks, 2007, 2013, 2016) were done in this manner.

The next best approach would be a matched-groups quasi-experiment where at least two groups are matched at the start with respect to characteristics such as age, gender or pre-test scores. The crucial point with this type of design is that the groups are not created in a random manner. Therefore, the risk of researcher bias is considerable.

Thus, to be consistent with the post-positivist viewpoint of the researcher, an RCT was the natural decision.

### **2.3 The dilemma of insider research**

Being an insider does have its merits and, as Merton (1972) suggested, sometimes being an insider is the only way to get at the truth, because on the whole, the world or maybe just society is a distrusting place. Merton (1972) elaborated that this distrust is caused by the creation of groups or 'collectives' that, instead of forging relationships, cause greater distrust since the 'collectives' possess little knowledge about each other. Merton suggests initially that using an 'insider' approach can help to further social knowledge. A researcher with insider status does have an immediate advantage in as much as they are bringing

understanding to the research study, which would not be available to an 'outsider'. Conversely the insider could bring preconceived ideas to the situation which could bias the research. Merton continues to say that although there appears to be an insider/outsider dichotomy this is not true, as both positions have their advantages and disadvantages. In fact, Hammersley (1993) stated:

There are no overwhelming advantages to being an insider or an outsider. Each position has advantages and disadvantages though these will take on slightly different weights depending on the particular circumstances and purposes of the research.

Hammersley (1993, p. 433)

This view is supported by Chavez (2008) and Workman (2007), who state that there are advantages and disadvantages to insider status. Advantages include first-hand knowledge of the workplace; ease of contacting key personnel with whom working relationships are already forged. Disadvantages include the researcher being viewed as a 'spy'; increased pressure to ensure that the study is successful, resulting in manipulation of data to give desired results; and the problem of who to tell controversial results to.

Even though there are advantages and disadvantages to having insider status, further complications and considerations arise when the participants are children.

It is vital that the children's consent is obtained prior to their inclusion in any research. According to *United Nations Convention for the Rights of the Child*, (United Nations, 1989) children must be allowed to express themselves:

States Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child.

(United Nations, 1989, p. 15)

The Department of Health (2001) compiled guidance notes to help practitioners/researchers to understand the importance of consent within the jurisdiction of the law. English law states that a child can consent if they are judged to be competent. Competency is not just seen as a child's ability to

understand, but also on the trust the child feels towards the adults who are asking for consent. Although the guidance does state that:

If a child under 16 is competent to consent for himself or herself to a particular intervention, it is still good practice to involve their family in decision-making unless the child specifically asks you not to do so and you cannot persuade them otherwise.

(Department of Health, 2001, p. 5)

The guidance makes it clear that children should be allowed to withdraw their consent at any time and should not be coerced into changing their minds.

In this study, after a thorough description of how the study was going to affect them, each child was asked individually for their consent. Separate letters were sent home to obtain parents' approval. Initially, consent was going to be asked at class level, but the researcher was advised that peer pressure might have affected those who did not want to partake. However, reviewing this approach, it is possible that the children may have felt they had to say yes because an adult was asking. This issue was experienced by Kim (2012), who recruited her own participants. She found that her change in role from Sunday school teacher to researcher seemed to affect the way her participants reacted to her. On the first meeting with a usually relaxed confident child, the child appeared to be nervous and uncomfortable. When Kim asked her questions, the replies were brief and almost monosyllabic. In addition, Kim felt that the child was thinking very carefully about what responses she should give. Kim found this surprising, and then realised that she may not have been a complete insider. Kim felt that the child was responding in a manner which her Sunday school teacher would want her to. That is giving answers that would gain approval. Kim wanted the child to answer her honestly and with respect to how the child felt. Kim felt that her roles as teacher and researcher were different, but the child did not see this.

Fraser (1997) summarises the ethical issues she encountered under the following headings:

- Personal values and potential bias;
- Researcher's role within the organisation;

- Confidentiality and anonymity;
- Role conflict issues;
- Time constraints.

These headings conveniently summarise the points previously made. The only point not discussed is time constraints. Time constraints are an issue that most research is put under and not specific to insider research.

With reference to this study the researcher, although not employed by the school, was a governor, parent helper and parent. And, according to Workman's (2007) definition, she fulfilled the definition of an insider, namely a person who was working within a school environment but was about to carry out a systematic enquiry that was relevant to the 'job'. In addition, the researcher had previously forged relationships with all the teachers whose children would be taking part in the study and had a good working relationship with the headteacher.

The researcher had access to privileged information, as the assessment tools measured the progress of the children. These results have been used but anonymised, and the results only shared with the class teachers.

Although the researcher was not employed by the school, her complex role as parent, volunteer and governor offered the usual complexities of insider researcher. As Faugier & Sargeant (1997) suggest, the strength of the relationship that develops between researcher and participants is the strongest way to ensure genuine cooperation, and subsequently increases the validity of the data obtained. However, we must be cautious, as Moore (2012) suggests the possibility that all research is coercive, since sampling decisions may be influenced by the ease of gaining access to the desired research environment. With this in mind it is possible to use insider status as a strength rather than a cause for contention, once all aspects have been considered and understood.

## **2.4 Conclusion**

In this chapter the positionality of the researcher and its influence on chosen research methods, and the conflicts that surround insider research, have been discussed.



Although not universally accepted as an appropriate research stance for educational research, (post-)positivist positionality does have its place. If quantitative approaches are to be used within educational research, positivism as a positionality should be accepted.

As a consequence of the researcher's positionality it was a natural decision for the inner study to be investigated using a quantitative approach, in this case mainly an RCT.

The researcher was aware of her unique position within the school, and identified the issues that could have affected her as an insider researcher. If researchers are aware of these issues it is hoped that they can remain objective when it comes to doing the research and eventually interpreting the data collected.

### **3 Pilot/feasibility studies literature review**

#### **3.1 Introduction**

This doctoral thesis has several proposed research questions and hypotheses. The first asks whether the quantitative approach used in this pilot/feasibility study is appropriate, but it is equally important that evidence is provided that the use of a pilot/feasibility study approach is suitable. In this chapter evidence is provided that pilot/feasibility studies are tools that should be used increasingly in educational research, as they provide a means of testing a proposed method without incurring the costs of a full-scale study, as well as identifying any methodological issues that may occur during the study.

The first section of this chapter defines the terms 'pilot study', 'feasibility study' and 'pilot/feasibility study,' and highlights the need for consistent understanding of these terms. The following sections describe why a pilot/feasibility study should be undertaken and introduces the idea of using a CONSORT table to ensure that pilot/feasibility studies follow the same guidelines, and therefore provide consistency in their undertaking. The final sections review educational pilot/feasibility studies that have been published and show value in educational research.

#### **3.2 Definitions of the terms 'pilot studies', 'feasibility studies' and 'pilot/feasibility studies'**

An in-depth search into the application of pilot/feasibility studies in research has shown that the understanding of this term is often met with confusion. The reporting of pilot/feasibility studies is still in its infancy and this is reflected in the myriad of definitions that are around. Some instances, such as the National Institute for Health Research (NIHR) guidance (2016), state that pilot studies do not include data analysis, which is contrary to other definitions that are described below. There are definitions within the literature that describe pilot/feasibility studies that are not RCTs or qualitative studies; these will not be discussed here.

In addition, the availability of definitions of pilot and feasibility studies within the social sciences has been very limited. The evidence produced below has been gleaned from medical and nursing journals, where pilot/feasibility studies feature

on a more regular basis. However, even in these disciplines there is a lack of clarity with respect to what pilot and/or feasibility studies are.

Definitions of pilot and feasibility studies have been difficult to pin down. Some researchers use the words interchangeably, therefore assuming no differences; however, others identify explicit differences.

Polit et al. (2001, p. 467) say that feasibility studies are 'small scale version(s), or trial run(s) in preparation for the major study'. However, NIHR (2016) suggest that feasibility studies are undertaken to answer the question 'Should this study be done?' This definition is supported by Morris and Rosenbloom (2017).

The definitions of pilot studies are equally confusing. Baker (1994, pp. 182-3) suggests that a pilot study can be seen as the pre-testing of a particular instrument. This view is partially supported by the NIHR, who state that pilot studies are a smaller version of the main study and check whether the constituent parts of the study can all work together, assuming that the instruments described by Baker (1994) equate to the constituent parts described by NIHR.

Morris and Rosenbloom (2017) split pilot/feasibility studies into three types, confusing the issue further. In reality, they need only have described pilot studies and feasibility studies. They went on to explain that, within these definitions, there are acceptable variations that could be incorporated, such as randomised or non-randomised approaches.

Eldridge et al. (2016a) highlighted the confusion surrounding the terms pilot and feasibility studies and the wide range of definitions, and therefore understanding of what characteristics constituted each. They wished to develop specific definitions for each term that could be used consistently throughout the research community, and thus encourage consistent understanding and approaches when dealing with pilot and/or feasibility studies. To achieve this they undertook a Delphi study. The aim was to formally define feasibility studies and pilot studies in a manner that could be universally adopted. Eldridge et al. (2016a, p. 8) initially defined a feasibility study as 'whether something can be done, should we proceed with it and if so how'. A pilot study was defined as 'a study in which a future study, or part of a future study, is conducted on a smaller scale to ask the question

whether something can be done, should we proceed with it and if so how'. The difference between a feasibility study and a pilot study is therefore on a practical level. The feasibility study is asking theoretically whether a particular intervention should be investigated, whereas the pilot study investigates the appropriateness of the methods being used to investigate the intervention. This outcome of these definitions is that 'all pilot studies are feasibility studies but not all feasibility studies are pilot studies' (Eldridge et al., 2016a, p. 8).

The Delphi study proposed four different definitions based on: the above supposition; the NIHR guidelines; the fact that feasibility and pilot studies are not mutually exclusive; and a fourth definition based on the MRC guidelines (MRC 2000). The proposed definition obtained from the Delphi study, as suggested by Eldridge et al. (2016a, p 18) is that feasibility is 'an overarching concept, with all studies done in preparation for a main study open to being called feasibility studies and with pilot studies as a subset of feasibility studies'. As a result of this, studies of these types should be called pilot and/or feasibility studies. They further suggest that one of the terms 'pilot study', 'feasibility study' or 'pilot/feasibility study' should be in the title or at least in the abstract, as this would help identify this type of study when it comes to electronic searches.

Succinctly put, this means that all pilot studies are feasibility studies which can be used to determine the efficacy of a specific experimental approach, whether the approach is randomised or non-randomised.

A pilot/feasibility study approach is adopted for this study. It is a feasibility study since it is asking the question whether a non-phonics intervention can be used successfully. In addition, it is asking whether valuable information could be gained from conducting this study and describes how the study could be done.

This study is also a pilot study because it is testing an approach and the efficacy of various assessment tools to determine their suitability for assessing specified outcomes.

Therefore throughout this thesis the term pilot/feasibility study will be used, often abbreviated to PFS.

### **3.3 Why do a pilot/feasibility study?**

According to Shader (2015), a PFS is not done to provide undeniable evidence to support or disprove a null hypothesis, since the sample size is always too small to provide statistically significant results. He describes a PFS as being a preliminary/exploratory study used to determine whether it is worth doing a larger-scale study.

Van Teijlingen and Hundley (2001; p. 2) expand this idea by giving specific reasons why a PFS should be done.

- Developing and testing adequacy of research instruments;
- Assessing the feasibility of a full-scale study;
- Designing a research protocol;
- Assessing whether the research protocol is workable and realistic;
- Determining whether the sampling frame and technique are effective;
- Assessing the likely success of recruitment protocols;
- Identifying logistical problems which might occur using the proposed methods;
- Estimating variability in outcomes to help determine sample size;
- Collecting preliminary data;
- Determining what resources (finance, staff) are needed;
- Assessing the proposed data analysis techniques to identify any potential problems;
- Developing a research question and research plan;
- Training a researcher in the techniques and requirements of the study;
- Convince funding bodies that the research team is competent and the study is worth investing in;
- Convincing other stakeholders that the project is worth supporting.

Although there are the above reasons for undertaking PFSs, there are also some limitations. As mentioned previously, PFSs are usually under-powered and therefore any statistical data are usually non-significant. However, as the reason for doing a PFS is to determine the effectiveness of a study protocol and its associated assessment instruments, this should not be seen as a limitation. However, caution should still be used, since it is possible that any analytic issues

may not be seen until a larger-scale study is done. A PFS, as explained above, is undertaken to determine whether something should be done, but because it is statistically underpowered (low number of participants) the analytical tools may fail to detect some issues. These issues will only be detected when larger numbers are used. In addition, Type I errors in a main study would imply that the null hypothesis is rejected when it should have been accepted. In PFSs, Type I errors could mean that a large study is undertaken when it should not have been.

Two further problems can arise. The first is if the data from the PFS are included within the data from the main study. The issue with this is that, if there have been alterations to the research protocol, due to findings from the PFS, the results from that study could be inaccurate. These inaccuracies could affect the overall outcome of the main study. This effect could be positive or negative, affecting the efficacy and reliability of the main study.

The second issue can be if new data are collected from the PFS participants. The issue here is that those participants have already been exposed to the intervention, and it is unknown how they will react to a second exposure of the same intervention. This may have a positive effect in as much as the participants become more adept at the intervention and the results become inflated. Conversely, a negative effect could be seen because the intervention is no longer new and exciting and the participants become apathetic, resulting in a decrease in effect.

It is sometimes necessary to include the PFS data due to the small population size that is available for the study. To help alleviate the issue of including the PFS data, sensitivity analysis can be done, which assesses the degree to which the PFS process has affected the size of the intervention effect.

In addition, Morin (2013) and Kistin (2015) suggest that PFSs are an under-utilised approach which can initially identify issues with respect to logistics, but also more complex issues such as the risk of Type I and Type II errors, or ineffective assessment tools. Jeray & Tanner (2012) go further and explain in detail how a PFS should have well-defined objectives. In addition, they suggest it should pass the 'Who cares?' question, suggesting that some 'worth' should come out of PFSs. They go on to state that a PFS should have integrity, be safe,

and be economically viable. They conclude by asking the question whether a PFS should always be done, and their answer is yes, unless there is sufficient previous research evidence to immediately implement a full-scale study. Once a PFS has been completed, should the results be published? This is a question asked by Becker & Schulte (2008) who state that, although the results obtained may not be conclusive, aspects such as efficacy could be supported, thereby providing evidence that a full study should be undertaken.

Since non-significant results are often a feature of PFSs, they suffer from publication bias, since many journals only wish to publish positive results. In addition, an article that identifies methodological problems is less likely to be published for similar reasons. The fact that the identification of such issues could help other researchers is not acknowledged, and a vital learning tool is not being exploited.

### **3.4 Pilot/feasibility study consistency**

As mentioned in the previous section, there are limitations to the use of pilot/feasibility studies, some of which could be due to the fluid nature of the definitions attributed to the terms 'pilot study' and 'feasibility study'. Eldridge et al. (2016a) embarked on a mission to pin down specific definitions that could then be used to generate specific criteria that could be used for the design and implementation of a pilot and/or feasibility study.

Using the Consolidated Standards of Reporting Trials (CONSORT) statement, which provides guidance to ensure that RCTs are planned and administered in a reproducible manner, Eldridge et al. (2016b) went on to develop a CONSORT statement for pilot and/or feasibility studies.

The importance of consistent and transparent approaches when doing research is paramount, so to ensure the efficacy of this study the CONSORT statement designed by Eldridge et al. (2016b) has been modified for educational research (see Table 3.1). It will be clear that the statement was devised to apply to any form of PFS; also, however, that some of the criteria apply more specifically to RCTs. When the present research is judged against the criteria in chapter 8, they are separated into two groups for this reason.

**Table 3.1: CONSORT statement for pilot and/or feasibility studies in educational research**

| Characteristic  | Reason for inclusion in CONSORT table   |
|---|---|
| 1. Identified as a pilot and/or feasibility study in the title and/or the abstract.   | By including these key words in the title and/or abstract they are more easily identified, especially for the purpose of systematic reviews.  |
| 2. Has there been sufficient discussion and explanation describing the theoretical/scientific background for the proposed larger study and why a pilot/feasibility study would be of value? | It is necessary to provide evidence that the theory and ideas behind the project are sound, and that the proposed methods are suitable for the investigation.   |
| 3. Have definite objectives and hypotheses been determined that are pertinent to both the pilot/feasibility and proposed full studies?  | So that it is clear what the aims and hoped-for outcomes are.   |
| 4. Has ethical approval been obtained?  | If dealing with children or vulnerable adults, this ensures that the project will be done to the highest ethical standards.   |
| 5. Has the method of randomisation been described?  | This will allow for reproducibility.  |
| 6. Have changes to the method that occur during the study been documented and explained?  | If methods had to be altered, this needs to be documented as it could affect the outcome either positively or negatively of the pilot/feasibility study. In addition, these alterations would have to be accounted for in a larger study. |
| 7. Have the methods by which the participants were identified and gave consent been described?  | Transparency of recruitment.  |
| 8. Has the reason for the number of participants taking part been explained?  | Provides background information why that many were involved.  |
| 9. Has the intervention been described and assessment criteria/tools pre-identified?  | Provides information about the approach being used.   |
| 10. Have any alterations to assessment tools that occur during the study been explained?  | Ensures reproducibility.  |
| 11. Was the investigator instrumental in the randomisation or was it blind?   | Provides evidence to determine the risk of selection bias.  |
| 12. Have suitable methods been used to analyse each study objective?  | Ensures that the results documented are a true and accurate representation of the results.  |
| 13. Has intention to treat analysis been adhered to? If not, why not?   | Counteracts attrition bias.   |
| 14. Why did the pilot/feasibility study end?  | Did the pilot/feasibility study end as intended, or were there issues with the design/intervention that caused it to be stopped prematurely?  |



| Characteristic  | Reason for inclusion in CONSORT table   |
|---|---|
| 15. Were all participants included in the final analyses?   | Counteracts the effects of attrition bias.  |
| 16. Has the effect size been calculated?  | Although it may be small, it could provide an indication of the effectiveness of the intervention.                              |
| 17. Have any other analyses been included within the final report which could have a bearing on the implementation of a larger study?                     | Could identify issues which were not foreseen at the planning stage but could impact on the overall effectiveness of the study. |
| 18. Have the limitations of the pilot/feasibility study been described with reference to bias and the overall feasibility of implementing a larger study? | What could be done differently to improve the effectiveness of the trial?   |
| 19. Are the conclusions/discussions representative of the objectives and findings of the pilot/feasibility study?   | Do the conclusions accurately reflect the findings of the study?  |

### **3.5 Review of educational pilot/feasibility studies**

#### **3.5.1 Introduction**

The application of PFSs in research in general is still in its embryonic state, even within medical research. PFSs within medical research cover a wide range of research interests, ranging from assessing the effectiveness of a hand-washing improvement programme (Lee & Lee, 2014), which was a small quasi-experimental study with 20 children, to studies assessing interventions that help stroke victims and their families (Cameron, et al., 2015), a study with three groups receiving either standard care (n=10), self-directed care (n=10) or stroke support (n=11). They concluded that the research design was feasible and that the intervention appeared to give carers an increase in confidence. It was stated that a full-scale study would be undertaken.

The rest of this section will be devoted to educational PFSs and will be divided into two subsections. The first subsection will give examples of PFSs that have been carried out by the Education Endowment Foundation (EEF) and explain their use of the terms 'efficacy' and 'effectiveness' trials, as well as describing the PFSs they have funded. The second subsection will describe further literacy-based PFSs that have been published.

## **3.5.2 Education Endowment Foundation**

### **3.5.2.1 Efficacy and effectiveness trials**

The EEF describe themselves as having a ‘project pipeline’ which allows them to retest interventions and thus ensure that they are effective. The ‘pipeline’ starts with piloting – this is used if an intervention has not been used in English schools. The EEF state that PFSs are much smaller than their efficacy and effectiveness trials, but are generally funded in areas where there is little previous evidence. They have funded 20 PFSs to date (Education Endowment Foundation, 2017).

When an intervention has been shown to be feasible it is tested through either an efficacy or an effectiveness trial. The efficacy stage determines whether the basic idea, if delivered as intended and under perfect conditions, has promise. The effectiveness trial tests the ability of the intervention to become generalisable and used in a larger number of schools. This ability to retest can detect issues in an intervention.

For example, Switch-on Reading is an intervention designed for children who failed to reach National Curriculum Level 4 (NC4) at the end of Key Stage 2, and was implemented when the children entered Year 7. An efficacy trial (Gorard, Switch-on Reading. Evaluation and executive summary, 2014) showed that children made an additional 3 months progress in their reading outcomes compared to children of similar abilities who did not have the intervention. As a result, an effectiveness trial (Patel, et al., 2017) was mounted. However, the results from the effectiveness trial were not as favourable, and in fact showed that the children who received Switch-on Reading made no more progress than the children who did not. The suggested reasons for this discrepancy were:

- The way the intervention was given had changed with respect to content, duration and format of the sessions;
- In the first trial, Switch-on Reading was delivered by its developers, who would have had a better understanding of how it should be delivered;
- There was a suggestion that the class dynamic had changed, and the effectiveness of teaching assistants (TAs) was altered;
- The amount of literacy tuition that was accessible to the students in the control group was higher in the second trial.

The Switch-on Reading programme has not been completely dismissed since it did have a positive result in its first trial. In 2017, EEF were discussing how the original concept might be better implemented to a wider audience.

### 3.5.2.2 Other EEF-funded pilot/feasibility studies

Two other EEF PFSs have been found and are summarised in Table 3.2.

**Table 3.2: Summary of EEF-funded pilot/feasibility studies**

| Authors                | Intervention                     | Number of participants  | Duration                                 | Design                                 |
|------------------------|----------------------------------|---|--|--|
| Dockrell et al. (2015) | Talk for Writing                 | 6 schools – whole school approach – Y3 & Y5 classes followed into Y4 & Y6 | 1 year (data collected at 3 time points) | Quasi-experimental                     |
| See et al. (2015)      | Word and World Reading Programme | 17 schools – year 3 and 4 classes   | 1 year (whole class)                     | Randomised; 9 intervention, 8 control. |

Talk for Writing is a whole-school approach to writing. The PFS was set up with specific aims in mind: first, to conduct a literature review to determine the strength of the pedagogy supporting the Talk for Writing approach; secondly, to determine the feasibility of the approach and provide recommendations for improvements; finally, to provide some quantitative data to suggest the possible impact of the scheme.

The reviewers identified two aspects of pedagogy that pertained to the scheme. The first, the 3I's (imitation, innovation and invention), is regarded as a way of scaffolding children's writing. They highlight that the two initial stages are unique to Talk for Writing and that, by the end of primary school, the third aspect, invention, should be commonplace. Many of the initial stages involve children talking about what they are going to write and repeating it. This repetition reinforces what the children need to remember to write. However, Dockrell et al., (2015) highlight that there is very little evidence to support this.

The second identifiable pedagogy was scaffolding the writer. As mentioned above, the 3I's include scaffolding since the responsibility for writing moves from being teacher-led to pupil-led. Further scaffolding should be provided by effective teacher feedback. This feedback should not be too directive, but allow the

children to consider what they could do to enhance their writing without being told what to do. Another form of scaffolding is the incorporation of collaborative writing, which Talk for Writing does provide.

Six schools agreed to adopt the Talk for Writing approach, and writing data were obtained from Years 3 and 5 at three-time points (January (baseline), June and January). Because the study started midway through the academic year the children had moved up a school year prior to the final assessment point.

The overall findings of the study were that the teachers were enthusiastic about the approach and found it easy to implement. Although the teachers said they saw an improvement in the writing skills of their pupils, the authors felt that there was not sufficient evidence in the data to support this view. Dockrell et al. (2015) further concluded that more research was needed to accurately determine how effective this approach would be. They recommended that further work was needed to refine the approach before it went to a full trial. Finally, they suggested that this approach might be more effective in early years due to its use of oral skills.

Word and World Reading Programme (See et al., 2015) is designed to improve the literacy skills of 7- to 9-year-olds from low-income backgrounds. The aim is to improve 'core knowledge' through the use of 'knowledge-rich' reading materials, vocabulary lists, reading aloud, and other resources such as atlases and globes. The theory behind this approach is that, by building up 'core knowledge', the children will have developed the understanding to comprehend what they are reading.

This approach had not been previously used in England, or in fact elsewhere, though similar programmes had been used in the United States. The aims of the PFS were to:

- Determine the feasibility of running a larger scale study;
- Assess the programme with respect to how it is delivered and find out the views of the teachers and pupils;
- Determine how likely the programme is to improve the literacy skills of the year 3-4 children.

The study was run in 17 schools located in areas of high social deprivation or coastal deprivation, inner-city schools that had a high proportion of children with challenging behaviours, or schools whose catchment areas covered a mostly white working-class background. Within the study, nine schools received the intervention and the remaining eight were the control group. The intervention lasted a year.

Word and World Reading Programme is a whole-class intervention which consists of a 45-minute lesson undertaken twice a week. For this study, the intervention was based on geography and history texts.

The intervention is split into two sections. Part 1 (30-35 minutes) introduces a new text which is read by the teacher, whilst the children follow it in their books. After each paragraph, the teacher asks questions, and the children are asked to answer in full sentences. Once the text has been completed, a discussion is initiated where the children are expected to use prior knowledge to be able to contribute to the discussion. To reinforce learning, pictures and keywords are shown.

Part 2 (10-15 minutes) is the time left to consolidate what has been learnt. This is achieved by the children completing three short mastery tasks in their work books and then doing a keyword exercise. A key component of this intervention is that the teachers should give immediate feedback, so are expected to circulate and feedback as the children are doing their tasks. The children are expected to use full sentences, the emphasis being on the correct use of their new knowledge and the ability to use the keywords correctly.

The findings of the report state that the programme was well received, and the teachers felt that there were improvements in the children's vocabulary and writing skills, as well as overall learning. The children were assessed using Progress in English (PiE); PiE 7 was used for Year 2 children and PiE 8 for Year 3 children. The assessment data showed very little progress, if any ( $d = -0.03$ ). The study was not designed to rigorously test effectiveness in improving scores on PiE, but to test the trial design and suitability of the intervention. It was suggested that the programme was more suitable for older or higher-ability children, as there was very little differentiation in the programme materials.

However, the reviewers suggested that a greater number of schools should be included in any future trial, and the trial should run for a longer period of time. One area of concern detected was that the teachers did not always have sufficient knowledge of the text topics to provide enough information for a full discussion to take place. This is obviously something that would have to be considered, since the primary aim of this programme is knowledge acquisition.

### 3.5.3 Other education pilot/feasibility studies

#### 3.5.3.1 Denton et al. (2010)

This PFS investigated the effectiveness of a summer school reading scheme based on word-level skills (phonemic awareness, phonics) and text-level skills (vocabulary, comprehension). Denton et al. (2010) stated that there is mixed support for the use of a summer reading scheme and, although they suggested that summer schemes would be beneficial, there is little experimental or quasi-experimental research available. Sainsbury et al. (1998a) reported on data obtained from 558 centres across England that offered 50 hours of literacy tuition through the summer holidays. They stated that, although the statistical data obtained from that study were not significant, with children neither improving nor declining significantly, there were notable improvements in the children's attitudes towards reading, confidence and enjoyment. It must be noted that, although statistical analysis showed a small average increase (0.3 standardised score points), the range of scores was large. In fact Sainsbury et al. (1998a) included the following table (Table 3.3).

**Table 3.3: Mean difference in age standardised scores between initial and final assessments and percentages of pupils showing score changes**

| Mean difference (gain) | Standard deviation | Percentage of pupils whose scores increased | Percentage of pupils whose scores remained the same | Percentage of pupils whose scores decreased |
|------------------------|--------------------|---|---|---|
| +0.3                   | 8.2                | 47  | 8   | 45  |

Sainsbury et al. (1998a, p. 12)

This table clearly shows that, although statistical analysis shows an overall gain in reading progress, the reality is that over 50% of the cohort failed to make any progress and in many cases the children's reading ability decreased.

Evidence of a decline in progress was observed on analysis of data from the transition phase (Year 6 to Year 7). Although children who attended the summer school showed some progress during summer school attendance, re-assessment when they entered Year 7 showed their reading had declined. Interestingly a control group that had not attended summer school reading score also showed a decline in their reading ability. The degree of decline in the summer school attenders was similar to the control group's, showing that summer school attendance had no effect on preventing a reduction in reading over the summer holiday.

Denton et al.'s (2010) PFS took place in 4 elementary schools in large urban areas located in the southwest US. The majority of the children (95%) were described as 'economically disadvantaged'. In schools 1 and 2 there were two summer kindergarten classes; one class was taught the experimental reading programme, and the other class was taught the usual summer school curriculum. Schools 3 and 4 had just single classes and were paired, school 3 delivering the experimental reading programme and school 4 delivering the typical school programme. A final sample of 53 (25 treatment, 28 control) children completed the summer school.

The reading scheme was written by the authors specifically for this study and consisted of instruction in:

- Phonemic awareness;
- Phonics;
- Recognition of high-frequency words;
- Sentence reading;
- Listening comprehension (whole group and small group);
- Vocabulary instruction (whole group, and independent practice through journal writing).

Denton et al. (2010, p. 429)

The results showed significant improvement in Letter Word Identification ( $d=0.91$ ), Oral Comprehension ( $d= 0.45$ ), Blending Words ( $d= 0.76$ ) and High Frequency Words ( $d= 0.81$ ) compared to those children who received the typical summer school curriculum.

Although these results are promising it must be noted that there are a number of limitations with this study:

- Small sample size;
- Only three teachers taught the intervention, which made it difficult to determine whether it was the intervention or the teachers that created the difference;
- High attrition rate. Pre-test data were collected for 147 children. Of these, 44 did not go to the summer school; 11 were dropped from the study; 26 stopped going to summer school before the end of the study; 8 had incomplete data; and 5 were discounted because the schools did not obey the rules of randomisation. This left 53 children;
- Randomisation of the children was not adhered to by the schools, thus the results could be biased;
- Paraprofessionals (teaching assistants) were available for the treatment group but not for the control group.

Regardless of these limitations the results supported the idea that further investigation into the implementation of a summer school reading programme was worth the investment. This was also supported by the fact that there is very little experimental research available.

### **3.5.3.2 Sanchez & O'Connor (2015)**

This PFS emerged from the findings of a larger Tier 2 study. That study spanned four years and involved five schools. The intervention was given to kindergarten and first grade children. It became apparent towards the end of the study (three months before cessation) that, although most children were benefiting from the Tier 2 intervention, there were still some children who had not progressed as expected. A single school was chosen to participate in the PFS. In this school, seven third graders and 24 fourth graders were receiving the Tier 2 intervention. Out of this intervention group eight children (three third graders and five fourth graders) were identified as having not made sufficient progress.

The children were given the Tier 2 intervention for three months in the final (fourth) year of the large study. The intervention was given three times a week



for 25 minutes in small groups. During this time children read books (these were texts that the child could read with 80%-85% accuracy) to develop word reading skills, sight word recognition, reading aloud and development of comprehension skills.

For the following three months, the eight children were individually taught a Tier 3 intervention. The children were seen for 40 minutes, three times a week. These sessions were specifically designed for each child. Books were chosen that the child could read with 90%-94% accuracy. The Tier 3 intervention concentrated more on sight word practice, reading aloud and letter patterns.

The study showed that although small-group teaching (Tier 2) does enhance the progress of most children who are falling behind, this is not enough for all. So, the implementation of an individual tuition (Tier 3) intervention is necessary. As with all PFSs, although the results showed promise, there were a number of limitations. Since the study was not designed to provide definitive quantitative data describing student progress, it was not possible to conclude that any progress made was due to the Tier 3 intervention, or that continued use of this approach would allow the children to make further accelerated progress, thus diminishing the gap. The authors also suggested that the intervention they used might not fulfil the criteria of a true Tier 3 intervention, as it might not have been intensive enough.

### **3.5.3.3 Raffaele Mendez et al. (2016)**

This PFS involving 11 children (six first graders, five second graders) investigated the Tier 2 intervention Reading by Design. This focuses on improving children's decoding skills, sight word vocabulary and reading fluency. The intervention was delivered to the first graders as one group, and the second graders as a second group. The intervention was given for 35 minutes, four days a week. A number of different approaches were used:

- The children worked in similar ability pairs and worked through an audio lesson;
- The children worked individually with a reading specialist using SRA Reading Mastery.

In the last 10 minutes motivational games were played to help maintain interest and enthusiasm.

The main limitation of this study was that there was no control group. Therefore, it is not possible to conclusively state that any progress seen was due to the intervention and not due to the standard school curriculum. However, Raffaele Mendez et al. (2016) stated that the rate of progress for these children was greater than for the rest of the cohort, which could be seen as significant, since prior to the intervention their progress was slower. Interestingly, the authors did not mention that their sample size was very small, which as stated previously significantly affects statistical outcomes.

Raffaele Mendez et al. (2016) did suggest that using an intervention containing a number of different approaches, which are designed to engage and stimulate the children, could help struggling early readers make better progress.

#### **3.5.3.4 Puranik et al. (2017)**

Peer Assisted Writing Strategies (PAWS) is an intervention aimed to improve the writing skills of kindergarten children by providing help with alphabet fluency, spelling, and sentence and essay writing. The PFS was designed to assess the feasibility of incorporating this approach into classrooms and the possible effectiveness of PAWS at improving writing skills and see whether PAWS training affects reading ability.

A total of 86 kindergarten children were selected from five classes in two schools. School 1 was described as low socio economic status (SES) but high-performing, and had 2 kindergarten classes with a total of 40 children. Six children from each class were selected for PAWS training, whilst the remaining 28 children had the standard writing tuition. School 2 was described as low SES but low-performing, and had a total of 65 children in three kindergarten classes. Six children were selected from two of the classes, and the remaining 53 children received the standard writing tuition.

The intervention was designed to be given for 30 minutes three times a week, and a total of 35 lessons (20 lessons concentrating on letters and 15 on spelling) were delivered.

Each lesson covered a specific activity. This could be:

- Letter learning – students were shown three letters and told their names and sounds and how to write them;
- What comes after/before – the children were asked to write the letter that comes before or after letters they had learned the previous week;
- Missing letters – children practised spelling decodable words by completing words by filling in the missing letter;
- Cover-copy-check – children practised writing sight words. Two new sight words were taught each lesson.

The limitations of this study were the small sample size and high attrition rates, particularly in School 2. In School 2 attrition was seen to be higher in the control group for all assessments (17% PAWS vs 30% control) except the essay assessment, where the PAWS group attrition was 36%. In addition, there is the issue that the PAWS intervention was taught to small groups of children rather than to full classes. Thus any improvements of the children who did PAWS could be due to group size/intensity of instruction. There was also inconsistency between the schools with respect to dosage. School 1 completed all the 35 lessons, but School 2 completed only 23 lessons because they started three weeks later and finished two weeks earlier.

### **3.5.3.5 Wollscheid et al. (2016)**

In this study three groups of children from two different schools were assessed to see whether there was a difference between the learning of writing by using pen and paper or digital technologies. School 1 had a traditional approach to writing; Group 1 (n=15) therefore had traditional writing instruction and used pen and paper, and were assessed using pen and paper. School 2 used tablets throughout the curriculum, including writing instruction. Two groups were selected from School 2: Group 2 (n=14) had computer-based writing tuition but were assessed using pen and paper, whereas Group 3 (n=18) had computer-based writing tuition and were assessed using a computer.

The aim of this study was twofold. Firstly, they wanted to pilot an investigation into the effects of two different instructional methods of writing (pen and paper

versus tablet) and their suitability to be scaled-up. The second aim was to develop a robust writing test that could be used in a larger study. The test was made up of three parts; the first was a transcription task which was designed to test speed of writing; this was followed by a dictation task which measured writing speed, spelling and memory skills. The final section was free writing, which again measured writing speed, spelling and 'qualitative dimensions of writing' (undefined, but presumably such things as coherence and style).

Comparison of School 1 and School 2 showed no significant difference when both cohorts were assessed using pen and paper. However, comparison of Group 1 and Group 2 from School 2 did show some difference. In the transcription task group 2 wrote an average of 32 words (s.d. = 9), whereas the average of group 3 was 45 words (s.d. = 14). The dictation task did not yield a statistically significant result with respect to writing speed; however, the average number of errors for children in Group 2 was 26.8 (s.d. = 11.8), whereas for Group 3 the average number of errors was 18.7 (s.d. = 7.5). Wollscheid et al. (2016) explained that the majority of errors identified in Group 2 involved failure to use capital letters. In the case of Group 3, who were using the tablets, using the return key automatically capitalised the first letter, therefore this apparently significant result was an artefact of the test format rather than due to the children's writing skills. A final comparison of School 1 with Group 3 did not show statistically significant results, but School 1 made fewer spelling and writing errors than Group 3.

This study has a number of methodological issues and its overall aims are unclear. The schools were selected by the researchers because they had different approaches to writing instruction; therefore, it could be suggested that this was an example of selection bias. The authors do not make it clear if their overall aim is to introduce the use of digital technologies for the teaching of writing as normal practice. Obviously, the ethical issues surrounding a full-scale study with this as its aim are enormous, since to fully test this theory children would have to be taught using one approach or the other. Since the effectiveness of a digital approach to teaching writing is unproven this type of study would be difficult to authorise. The only definite aim and possible success is the development of the writing test, which could be developed further with a larger study.

### 3.5.4 Summary of pilot/feasibility studies

The EEF have a set of three questions that they apply to all pilot/feasibility studies to help determine whether it is worthwhile continuing to an efficacy trial. These questions have been applied to all the pilot/feasibility studies discussed above, both those commissioned by EEF and the others (see Table 3.4).

**Table 3.4: Summary of pilot/feasibility studies**

| Study                         | Questions asked by EEF     |                               |   |
|-------------------------------|----------------------------|-------------------------------|---|
|                               | Was the approach feasible? | Is there evidence of promise? | Is the approach ready for a full trial without further development? |
| Dockrell et al. (2015)        | Yes                        | Mixed                         | No  |
| See et al. (2015)             | Yes                        | Mixed                         | No  |
| Denton et al. (2010)          | Yes                        | Yes                           | Yes   |
| Sanchez & O'Connor (2015)     | Yes                        | Yes                           | No  |
| Raffaele Mendez et al. (2016) | Yes                        | Yes                           | Yes   |
| Puranik et al. (2017)         | Yes                        | Mixed                         | No  |
| Wollscheid et al. (2016)      | Yes                        | Mixed                         | No  |

As can be seen from the Table 3.4, only two of the pilot studies, even according to the authors, were ready to proceed to a full-scale study. The study design for Denton et al., (2010) was sound and the statistical analysis that was undertaken did not show significance, but this was expected with a PFS.

The study conducted by Raffaele Mendez et al. (2016) showed promising results, with effect sizes ranging from 0.66 to 1.21 for the assessment tools. They did state that they did not use randomisation and that, if a large study were undertaken, this would be an approach they would consider. The main issue with this study was that there were four components to the intervention and, due to the design of the study, it was not possible to determine the effectiveness of each component; nor was it possible to determine the effect of the normal classroom environment since a control group was not used. Although the authors suggested they were ready for a full study, a further pilot study would be advisable.

There are number of reasons that Sanchez & O'Connor (2015) study was not ready for a full study. The main reason was that the Tier 3 intervention was not a tried and tested model, having been designed by the authors by merely altering Tier 2 approaches. In addition, a number of factors were also altered, such as

exposure time and size of groups. Before a full study could be done, the Tier 3 intervention would need to be determined and documented by listing prescribed timings, group size, content and method of delivery. Ideally a further PFS would need to be undertaken to determine the efficacy of the Tier 3 intervention.

Puranik et al. (2017) highlighted that one of the main limitations of their design was the small group size, and stated that they wished to do further studies using whole groups. The children in this study were given PAWS as a small-group intervention, but this is a method which the authors would eventually like to be given to whole classes, although they did state that small-group work could be given to children for Tier 2 instruction. Puranik et al. (2017) also stated that the intervention was given by the developers, and they would wish to determine how effective the intervention is when it is taught by teachers.

Of all the PFSs discussed, Wollscheid et al. (2016) faced the most difficult ethical issues. They were proposing a completely different new way of teaching writing, the use of tablets. To be able to fully assess this, a group of children would have to learn to write exclusively on tablets, which could have a detrimental effect on their overall progress. However, those using digital writing could have additional lessons, but any improvement could then be due to the additional practice rather than the technology. In addition to this ethical minefield, the writing test that was developed needed further investigation. There is also the cost element of running a large-scale study due to the need of the use of tablets.

### **3.6 Conclusion**

This chapter discussed the practicalities of using pilot/feasibility studies in educational research. The beginning of the chapter described the difficulties of finding consistent definitions of the terms pilot study, feasibility study and pilot/feasibility study. Without consistent definitions it is difficult to design studies that follow the same guidelines, therefore the use of a CONSORT table would ensure that all studies are done to the same standard and following the same definitions. This would allow studies to be directly compared.

Pilot/feasibility studies are ideal when small sample sizes are being investigated. Small sample sizes rarely provide statistically significant results, but they do provide a means of determining whether a research study has potential, because

they test the suitability of the practical approach and associated assessment tools. As the number of subjects in the current study was limited to 24, a pilot/feasibility study was the most appropriate approach to use.

## **4 Why investigate alternatives to phonics?**

### **4.1 Introduction**

The primary concern of the ‘inner’ study in this research was the most appropriate way of teaching children who are falling behind in reading. It was proposed that, although phonics has its place, alternative non-phonics-based approaches may also have their place. The initial sections of this chapter describe why alternative methods of teaching reading to children falling behind in reading should be investigated; this is then followed by a brief overview of two theories of reading: the simple view of reading and the dual-route model. Jean Chall’s *Learning to Read: the great debate* is briefly discussed, and some of the key points she made are highlighted.

The next section describes how phonics has become the accepted approach to teaching (one aspect of) beginning reading within English schools, so the following section describes its inclusion in the National Curriculum and how the emphasis on phonics teaching has evolved over time.

The emphasis on phonics has been supported by numerous experimental research studies, and in section 4.7 that evidence is reviewed. Since it is proposed that a non-phonics-based approaches should also be considered, section 4.8 reviews the arguments opposing the exclusive use of phonics, whilst section 4.9 discusses the experimental evidence supporting non-phonics-based approaches.

The chapter ends with a brief summary and conclusion.

### **4.2 Why phonics or not phonics?**

The disagreement about the ‘best way’ to help children who are falling behind in reading is as complex as the discussion about which method to teach beginning reading. Support for whole-word or phonics approaches is like a perpetually swinging pendulum, and it often appears that the protagonists who shout the loudest win the popularity vote. Many of the available interventions are phonics-based, not allowing children who may struggle with phonics to learn any different skills that could advance their reading attainment. Providing the most appropriate



means of teaching a child to read should be used and, although phonics does have its place, it is not the only skill that a child needs to learn to read, especially when dealing with a language with an inconsistent orthography.

### **4.3 What is phonics?**

Phonics is a way of teaching beginning reading by correlating speech sounds (phonemes) to their written partners (graphemes).

There are several types of phonics. There is small-unit synthetic phonics, which is recommended by the National Curriculum (DfE, 2013a). This teaches individual phonemes (sounds) and their corresponding graphemes (single letters, digraphs, trigraphs or even four-letter graphemes). When children know a few letter-sound correspondences they are taught how to segment printed words (sound out individual phonemes, using the grapheme-phoneme correspondences) and then blend (synthesise) the sounds together to form words. The children then use these skills to decode unfamiliar words and, by reversing the process, to spell words they wish to write. Synthetic phonics can be taught as a fast and furious approach, with eight sounds being learnt within a two-week period.

Analytic phonics does not teach children to pronounce the individual sounds in a word (at least at first), but to identify (analyse) a specific phoneme, and its corresponding letter, in a group of words, for example /s/ and <s> in 'sun', 'sat', 'sip' and 'set'. Analytic phonics has traditionally been slow, concentrating on one sound a week; some believe this is too slow, but it can be speeded up.

Some authors (e.g. Johnston et al., 2012) use a narrower definition of synthetic phonics than that just given, and then define all approaches not meeting their definition as analytic, thus classifying as analytic some approaches considered by most authors to be synthetic; their definitions are not adopted here.

Confusingly, the approach known as onset-and-rime is sometimes equated with analytic phonics, but since it involves segmentation and blending it is better seen as the large-unit version of synthetic phonics.

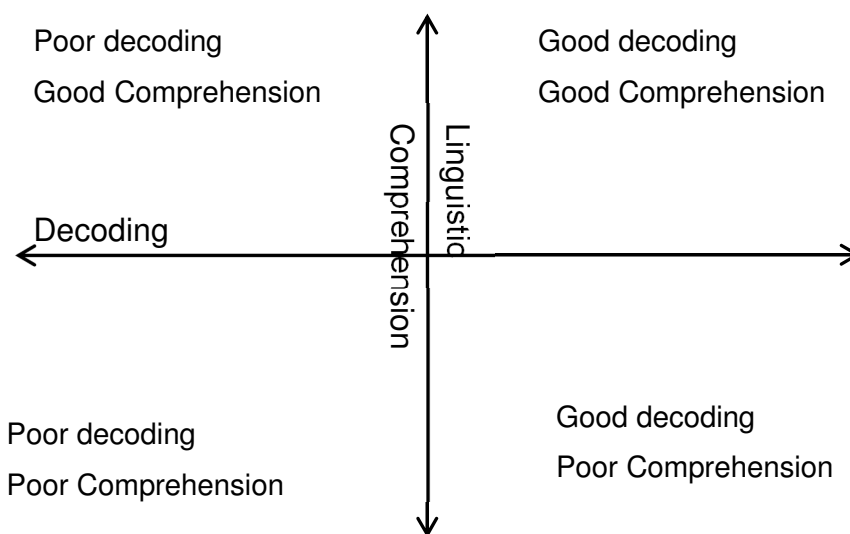
## 4.4 Theories of reading

Reading is a complex process and, as one would imagine, there are a number of theories about how we learn the skills to be able to read. Reading theories can be divided into three groups: bottom-up theories, top-down theories, and metacognitive theories. Top-down theories, such as Schema Theory, posit that background knowledge is crucial to reading, so children are taught whole texts and learn unknown words by using context clues. Those who advocate metacognitive theories believe that the reader needs to be thinking about their reading and this influences the way the reader interprets the text. The final group, bottom-up theories, hypothesise that the process of reading starts with the identification of single letters, progressing to whole words and finally text (a phonics approach). In this section the simple view of reading (SVR) and dual-route models will be described, as they both theorise how bottom-up approaches (such as phonics) can be effective teaching tools.

### 4.4.1 The simple view of reading

The SVR) was proposed by Gough and Tunmer (1986) and shows a natural link between decoding (which is defined as word recognition) and linguistic comprehension (defined as the ability to understand the spoken word). A diagram representing SVR is shown in Figure 4.1.

**Figure 4.1: The simple view of reading**



Source: Adapted from Rose (2006, p. 77)

Gough and Tunmer (1986) believed that, to be a good reader, the balance between decoding and comprehension must be equal. For example, if a child does not have the ability to decode a written word, they cannot find the 'meaning' of that word, affecting the child's ability to understand what they are reading. When a child starts to read they gain 'meaning' from words in several ways, including looking at the word in context, looking at the pictures, or the preferred method of decoding it. The SVR omits the first two cues as it is believed that over-reliance on them can prevent a child from accessing harder texts. Thus, by only including the decoding cue the SVR emphasises the importance of decoding, especially at the beginning reading phase.

Gough & Tunmer (1986) represented this relationship using a simple equation:

$$R = D \times C$$

where R is reading comprehension, D is decoding, and C is speech comprehension. They gave values of between 0 and 1 to D and C (calculated by using various assessments) to obtain values for R. Obviously if D or C is low the impact on R is great and if the value of D or C is zero, then reading comprehension is zero, meaning reading cannot occur.

Tunmer & Chapman (2012) revisited the SVR to see whether D and C were independent. They concluded that the basic two-component structure of the SVR did not need to be modified. But the effects of D and C might need to be re-evaluated as they felt that C had a direct effect on reading, as well as an indirect effect through decoding.

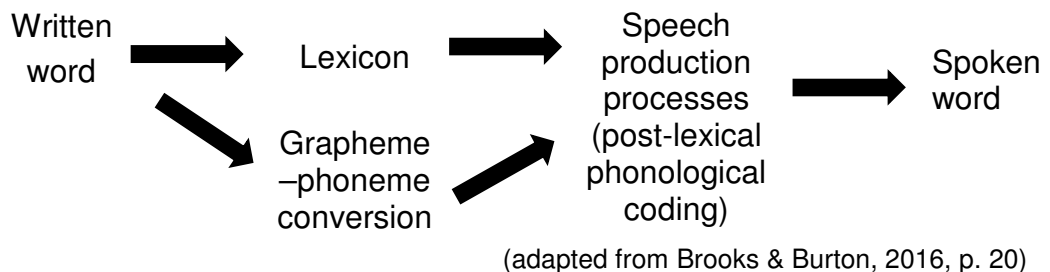
This theory supports that the idea that decoding is important for reading comprehension, thus supporting the teaching of phonics for beginning reading. However, this theory also states the importance of another factor, linguistic comprehension, the need to understand. Decoding of words is not enough.

#### **4.4.2 Dual-route models**

The previous section described how reading skills are developed but it does not describe how the brain identifies individual words and translates them from the written form to the spoken word.

Dual-route models (conveniently summarised in Coltheart, 2006) provide a means of explaining this. They also provide a reason why the teaching of phonics is important (Figure 4.2).

**Figure 4.2: Basic dual-route model of reading aloud**



The model provides an explanation of how words are recognised and articulated, depending on whether they are known/unknown or decodable/irregular.

Two separate pathways have been hypothesised. The first is called the lexical route. It is thought that this route identifies words that are already known by the reader and are stored in the Lexicon (the brain's dictionary). The Lexicon has the ability to immediately identify known words. It is thought that all known words, both decodable/regular and irregular, are stored in the Lexicon.

The second pathway (non-lexical route) is used to decode words that follow regular grapheme-phoneme correspondences, and can also be used to decode (read aloud) non-words which follow those correspondences.

To summarise, reading aloud using the lexical route means that any specific word is 'looked up' in a mental lexicon that contains information about spellings and pronunciations of letter strings that are real words. Reading aloud using the sub-lexical route does not use the mental lexicon, but uses the rules linking graphemes to phonemes (orthography to phonology).

This model again provides substance to the importance of phonics tuition in beginning reading. Prior to words being stored in the Lexicon they must be decoded, if possible, or learnt by sight recognition. Teaching children the ability to decode words helps them build up their own Lexicon. These theories do not state that there is a single method of reading tuition. By identifying decoding as an element of reading they do, however, suggest that phonics is a key factor, but

they do also suggest that there are other factors, such as sight word reading, that are needed for a child to become a skilled reader, and that being able to decode words is not enough.

#### **4.4.3 Summary**

The dual-route theory provides an explanation of how individual words are identified and translated into the spoken word. There are two ways by which the word can be identified: the word is either a) part of the lexicon or b) decoded. It provides evidence for the importance of decoding, but also highlights the importance of the need to be able to identify sight-words. This supports the underlying theory of this study, that non-phonics approaches have their place in the teaching of children to read.

The simple view of reading on the other hand tries to provide a link between decoding and comprehension. The validity of this theory is dependent upon the definition of decoding that is used, which in the case of SVR is defined as word recognition. Initial decoding skills used by beginner readers cause a lack of reading fluency since individual grapheme-phoneme correspondences are worked out, which impacts directly on the ability to understand what has been read. However, if decoding allows the word to be identified instantaneously, without having to verbalise each phoneme, as defined by SVR, the fluency of reading will improve, thereby improving comprehension.

The SVR identifies the need for good word recognition skills which can be developed using phonics if the words are 'decodable' and therefore supports the need of phonics. However, due to the definition of decoding used, the SVR is highlighting that other word recognition skills should be used, such as whole-word skills for reading non-decodable words.

#### **4.5 Jean Chall: The Great Debate**

Jeanne Chall's 1967 work *Learning to Read: the great debate* can be considered as the first thorough systematic (though narrative) review of the phonics/whole-word dispute as it applies to English. Her project was the result of a huge undertaking triggered by the US National Conference on Research in English.

Chall wanted to answer several questions in a way which removed the rhetoric surrounding phonics and whole word approaches;

- How should beginning reading be taught?
- Why is it so difficult to reach agreement on the best approach?
- When is the best time to start beginning reading?

Chall approached the study in several ways. First, she re-assessed previous reading assessment research, concentrating on beginning reading, including investigating whether there was a link between how far children fell behind with their reading and the method by which they were taught.

Secondly, Chall reviewed the different methods/approaches used to teach beginning reading, allowing for direct comparison. She thought that full understanding of how to deliver these approaches/methods was lost in the complexities of the manuals, and any similarities camouflaged. By identifying these similarities, she felt that methods/approaches that were similar and shown to be ineffective could be rejected. The findings could direct new research by identifying approaches that work but also identify methods that were already being used.

Thirdly she interviewed whole-word and phonic proponents, hoping to identify whether their motivation made their approaches successful as they stated. Or as boring as their opponents said.

Finally, Chall reviewed and analysed the readers' workbooks and the teachers' guidebooks of the most used reading schemes.

Chall's approach was well designed to answer the questions she posed (in the opinion of the author). She collected evidence from proponents of phonics and whole-words in equal parts, and critically analysed it to produce conclusions. She was not persuaded by hearsay, which could have adversely influenced her findings. Much of the 'evidence' she received was based on personal preference and opinion. She stated:

I was personally buffeted by persuasive arguments and testimonials for or against a given method.

Chall (1967, p. 7)

Chall's main conclusion was that phonics teaching for beginning reading gives better results. But several caveats were attached to that statement. She emphasised that no specific phonics approach was better than another:

I cannot emphasise too strongly that the evidence *does not endorse any one code-emphasis method over another*. There is no evidence to date that ITA is better than a linguistic approach, that a linguistic approach is better than a systematic-phonics approach, or that a systematic-phonics approach is better than ITA or a linguistic approach.

Chall (1967, p. 307)

She emphasised that phonics should only be taught as a beginning reading method. She stated her discontent at the development of decoding for older children, because it was presumed that, if it is good for beginning readers, it is good for older readers. Chall stated that, once a child has developed the ability to identify the written word, teaching further phonics 'is sheer madness'.

Chall emphasised that good teaching is important. If the teaching is poor, children will not progress, regardless of approach.

In conclusion, as Marilyn Jager Adams (1999) stated,

...as a complement to connected and meaningful reading – systematic phonic instruction is a valuable component of beginning reading instruction.

Adams (1999, p. 39)

Much the same point has been made more recently by Margaret Clark (2015):

In shallow orthographies it may be natural to teach reading by synthetic phonic methods by which letters are decoded to sounds and then combined to form larger units such as syllables. In deep alphabetic orthographies, such as English, a combined method by which children learn basic alphabetic decoding procedures and at the same time master a sight vocabulary of familiar words may be more appropriate.

Clark, M. (2015, p. 10)

The important point here is that phonics should be used as part of a beginning reading approach, but not as the only approach.

Soon after Chall's work, Groff (1977) pondered the question, 'Does phonics help all children and in particular those with reading difficulties?' He did a small survey of elementary school teachers and asked them various questions regarding reading instruction. As expected, he received a wide range of answers which covered both extremes about phonics teaching. What became clear was the thought that it comes to a point where too much phonics is taught and that it has an adverse effect on a child's reading:

It seems equally clear from this study that it was thought that not all remedial readers can profit from additional phonics instruction alone.

Groff (1977, p. 97)

He went on to suggest alternative methods that these children could use:

It can be inferred from the answers given to the questions in this survey that many clinicians felt that other cues to word recognition as gained from the syntax and semantics of unknown words in question should be an essential part of word analysis techniques taught to remedial readers.

Groff (1977, p. 97)

In this respect, Groff can be seen as urging the use of a variety of cues, of which more later.



## 4.6 Phonics in the National Curriculum

The first National Curriculum (NC1) mentions phonics once within the entire 43-page document. It stated, 'Pupils should be able to ...use picture and context cues, words recognised on sight and phonic cues in reading' (DfE, 1989, p. 7). Here the importance of phonics was ranked the same as picture or grammar cues, which suggested that different skills could be used along with phonics to help beginning reading.

The importance of phonics was becoming more apparent in the 1995 National Curriculum (NC2). From a single mention in NC1, phonics had evolved to have an entire section headed 'phonic knowledge'. Under this heading there were six requirements (DfE, 1995, p. 7):

- 'Recognising alliteration, sound patterns and rhyme and relating these to patterns in letters;
- Considering syllables in longer words;
- Identifying initial and final sounds in words, including sounds which rhyme;
- Identifying and using a comprehensive range of letters and sounds (including combinations of letters, blends and digraphs), and paying specific attention to their use in the formation of words;
- Recognising inconsistencies in phonic patterns;
- Recognising that some letters do not always produce a sound themselves but influence the sound of others.'

The phonics approach described in NC2 was not (or not fully) synthetic, which is the preferred method today. The emphasis appeared to be on the use of analytic phonics because the requirements asked for children to be able to identify the initial and final sounds of words, and not all the individual sounds of the letters in a word, which is one of the hallmarks of synthetic phonics. At that point, avoiding full endorsement of synthetic phonics (even though it had strong advocates) may not have been seen as problematic since, as Chall had stated and is still the case, there was not enough research evidence to determine whether one approach was better than another.

The next major emphasis on the use of phonics, more specifically synthetic phonics, was the introduction of the National Literacy Strategy Framework for Teaching (DfEE, 1998). As can be seen below an emphasis on identifying all the sounds in a word was stated:

- Discriminate between the separate sounds in words;
- Learn the letters and letter combinations most commonly used to spell those sounds;
- Read words by sounding out and blending their separate parts;
- Write words by combining the spelling patterns of their sounds.

Source: NLS Framework for Teaching, 3<sup>rd</sup> edn. (DfES, 2001, p. 4)

The NLS divided literacy into three different areas: word-level work, sentence-level work and text-level work. Word-level work included phonics, which became increasingly complex over time and was specifically mentioned up to Year 4 term 3 objectives. This inclusion of phonics up to Year 4 contradicted the findings of Chall who, as previously stated, emphasised that phonics was a strategy to be used for beginning readers only. However, it did support her finding that, when a strategy is deemed to be good, it is implemented throughout the academic curriculum, regardless of its relevance. It cannot be assumed that because a strategy works well for one year group it works well for all year groups. Nor can it be assumed that, if a child fails to progress with phonics, an intervention using phonics will improve a child's reading skills. More of the same approach may not be appropriate. The progress of a child's reading skills is dependent upon the quality of the teaching. Teaching phonics is not achieved by simply reading an instruction manual. Teachers should be trained how to pronounce each phoneme correctly, ensuring as far as possible, for example, that there is no 'schwa' attached to voiceless stop consonant phonemes /p, t, k/. Teachers need to be aware of how phonics helps children to decode words, but also be aware that decoding does not necessarily mean that a child can read with understanding.

This emphasis on phonics was made because there were concerns that children might become too reliant on their context and grammar knowledge when they read, rather than paying attention to the sounds or spellings of the words. The Framework went on to say that suggest that, if children could not decode words by using their sounds or spellings, they would find it difficult to access harder texts

(DfES, 2001, p. 4). The Framework emphasised that, at key stage 1, 'there should be a strong and systematic emphasis on the teaching of phonics and other word-level skills.'

To ensure that children do not become reliant on picture and grammar cues, but identify words directly from the text, it is the responsibility of the person who is listening to a child read to be aware when a picture or grammar cue has been used to decipher an unfamiliar word. At this point the sentence should be re-read, and if necessary or possible, the word decoded using a phonics approach or an acknowledgement that it is a 'tricky' word.

The NLS identified a list of about 200 high-frequency words that should be taught as 'sight recognition' words in the years from Reception to Y2 (DfES, 2001, pp. 60-61). The document did mention (p. 60) that 'Some of these words have irregular or difficult spellings', without stating whether those words should be taught differently from regular words, thus conspicuously evading a key issue, and then further evaded it by saying: '*words will also need to be reinforced through other practice and exploration activities so that they can be easily read out of context*' (DfES, 2001, p. 60). It is impossible to infer from this statement whether 'tricky' words should be taught as sight words and decodable words should be initially taught by using phonics. Thus teachers were left no wiser about which approaches to reading tuition would be used. However, the result of too strong an emphasis on phonics instruction has been identified in chapter 1: children run the risk of trying to decode 'tricky' words by using sounding-out inappropriately. This would result in incorrect use of phonics and ultimately mispronunciation of the unknown word.

*Progression in Phonics* (DfEE, 1999a) was an add-on resource for the National Literacy Strategy and provided resources for whole-class teaching. It explained key terminology, such as segmentation and blending and the alphabetic code. It stated:

Children should become proficient in applying the skills of segmenting and blending to the alphabetic code in order to spell and read. The skills and knowledge can be acquired simultaneously so that as children are learning the earliest steps in segmentation (hearing the phoneme in the initial and final

position in a word) they will be learning some of the letters which represent those phonemes.

DfEE (1999a, p. 7)

A further revision of the National Curriculum (NC3) (DfEE 1999b) changed the terminology and introduced the idea of phonemic awareness and phonic knowledge. The requirements were the same as in 1995, with the addition that children should 'sound and name the letters of the alphabet' (DfEE, 1999b, p. 46).

In 2002 Ofsted investigated the effectiveness of the NLS. The report (Ofsted, 2002) stated that the teaching of phonics in Reception and Key Stage 1 had improved significantly. But this had not followed through to Key Stage 2, contradicting the idea that phonics is a beginning to read strategy.

The NLS was criticised, especially with respect to its ability to effectively teach reading. Solity (2003) suggested that this failure to effectively teach children to read could be due to ineffective training of teachers with respect to phonics. Whether this is a true representation of the facts or conjecture is difficult to determine. Brooks (2003) is an advocate of the use of phonics in the teaching of beginning reading, but emphasised that, although phonics is essential, it is not enough. He discussed the effectiveness of phonics teaching in depth, and explained that, at the grapheme-phoneme level, English orthography is about 75% regular. This statistic cannot be used to calculate what percentage of words are regular or irregular, but it does support the choice, as far as possible, of phonically decodable words for initial instruction, with judicious teaching of essential high-frequency irregular words as sight words.

As Brooks (2003) also stated:

I have long believed that all teachers of language (all primary teachers and all those teaching modern foreign languages, including English to speakers of other languages) should possess an accurate understanding of the phonemes and graphemes of English and of the correspondences between them. Without this there is no common vocabulary for describing English orthography or for analysing children's errors, particularly in spelling.

Brooks (2003, p. 20)

If teachers do not fully understand the principles behind a specific approach (phonics in this case) the quality of teaching will be poor. Brooks (2003) emphasised that teachers need to be confident about their teaching practices. Teachers must understand when it is appropriate to move on to the next stage, or if it is more prudent to allow children to practise their new skills. This flexibility is somewhat hindered with the pressures laid on by the government for children to progress and achieve more than ever before. However, this does not alter the fact that reading progress is reliant upon the efficiency and effectiveness of the teaching.

The government introduced *Playing with sounds: A supplement to progression in phonics* (DfES, 2004) to help improve the teaching of phonics, lowering the age of phonics awareness to 3-year-olds. The literacy hour was the term used to describe how literacy should be taught, according to the NLS. It was very prescriptive and followed the following sequence:

- 15 minutes where the class all worked together using a large print book;
- 15 minutes where they concentrated on specific words (phonics);
- 20 minutes where they read or wrote individually or in small groups;
- 10 minutes where the children recapped the main points in a group.

These additional publications further impressed on schools the importance of phonics, and of ensuring that the teaching of phonics was good. It also provided a template that all schools could follow, meaning that phonics was being taught using a standard approach. The hope was that all children would access excellent teaching of phonics, and all children would become skilled readers.

In 2005 a parliamentary inquiry was launched called '*Teaching Children to Read*' (The Education and Skills Committee, 2005). The focus of this inquiry was to investigate methods used to teach reading and ensure that they were the best available, or whether new approaches should be implemented. The report described why phonics was important and how it should be taught. The report stated that:

In accordance with the available evidence, the DfES now seems to have accepted that phonics is an essential methodology in teaching children to read. The present debate revolves around the status of phonics within early teaching of reading and the type of phonics programme that should be used.

The Education and Skills Committee (2005, p. 16)

The Committee felt that further research should be undertaken to compare the National Literacy Strategy with 'phonics fast and first' approaches. It stated that the research should identify:

- How effective different teaching approaches to reading are, i.e. synthetic phonics, analytic phonics and other methods mentioned in the National Literacy Strategy;
- How effective the mixing of phonics instruction with other methods of teaching reading is, compared to 'phonics fast, first and only';
- The length of time that any gains shown by any method last for;
- What the effect of teaching texts that stretch a child's knowledge of phonics is rather than allowing children only to access instructional texts that fall within their current decoding ability;
- How effective different approaches are with respect to different groups of children, for example, girls vs boys, children with special educational needs, and children from poor socio-economic backgrounds.

The report highlighted the importance of beginning reading and that it was important that any approach that was to be used nationwide should have been thoroughly investigated first. As a result of this Jim Rose undertook his '*Independent review of the teaching of early reading*' (Rose, 2006).

The current increased emphasis on phonics, especially synthetic phonics, in our schools can be dated largely from the publication of the Rose Report (Rose, 2006). The key statement in this report, although often misconstrued, is the following:

The findings of this review argue strongly for the inclusion of a vigorous programme of phonic work to be securely embedded within a broad and language-rich curriculum; that is to say, a curriculum that generates purposeful discussion, interest, application, enjoyment and high achievement across all the areas of learning and experience in the early years and progressively throughout the key stages which follow.

Rose (2006, p. 16)

Rose acknowledged the importance of phonics work, but the key point is that it should only be part of a wider, more encompassing curriculum, as Torgerson et al. (2006a) pointed out as a finding of their systematic review. However, other statements in the report, less keyed to the wider curriculum, have led to the impression that Rose advocated using only phonics in initial teaching, whereas the research evidence at the time only showed that systematic phonics had better results than unsystematic or no phonics. In several places Rose went beyond that evidence and, relying instead on evidence from observed classroom practice and from theory, equated systematic phonics with synthetic phonics. His critics latched on to the (to them) unjustified elision of 'systematic' and 'synthetic', which enabled them to ignore his point about the need to ensure that, whatever form of phonics was used, it was both applied systematically and embedded within a broad curriculum.

Because of the findings of the Rose report (2006) the government developed a synthetic phonics programme entitled *Letters and Sounds* (DfES, 2007) which was sent out to all schools to provide guidance on how to teach phonics. Unlike the NLS, it was expected that most of the phonics teaching would take place in Reception class and key stage 1. Throughout key stage 2 it was expected that children would take the skills they had learnt in key stage 1 and apply them.

*Letters and Sounds* advised that picture and context cues should be avoided and emphasised that regular high-frequency words should be decoded whenever possible, thus clarifying the point avoided by the NLS Framework. It is important

that children use their phonics knowledge, but the use of picture and grammar cues has its place, if only to allow the child to initially identify an unknown word, and then revisit it and decode it, if possible, or acknowledge it as a sight word.

*'Phonics: assessment and tracking guidance'* (DCSF, 2009) provided a way of tracking children's progress through their phonics journey. Again, by issuing this guidance, the government was ensuring that all schools were singing from the same hymn-sheet, and hopefully, ensuring that phonics was being expertly taught.

The introduction of the Year 1 phonics check (Standards and Testing Agency, 2012a) demonstrated the government's commitment to phonics. The phonics check was introduced as a check to ensure that phonics was being taught correctly and children were reaching their age-related expectation (ARE). There were, and still are, concerns about the implementation of such a test.

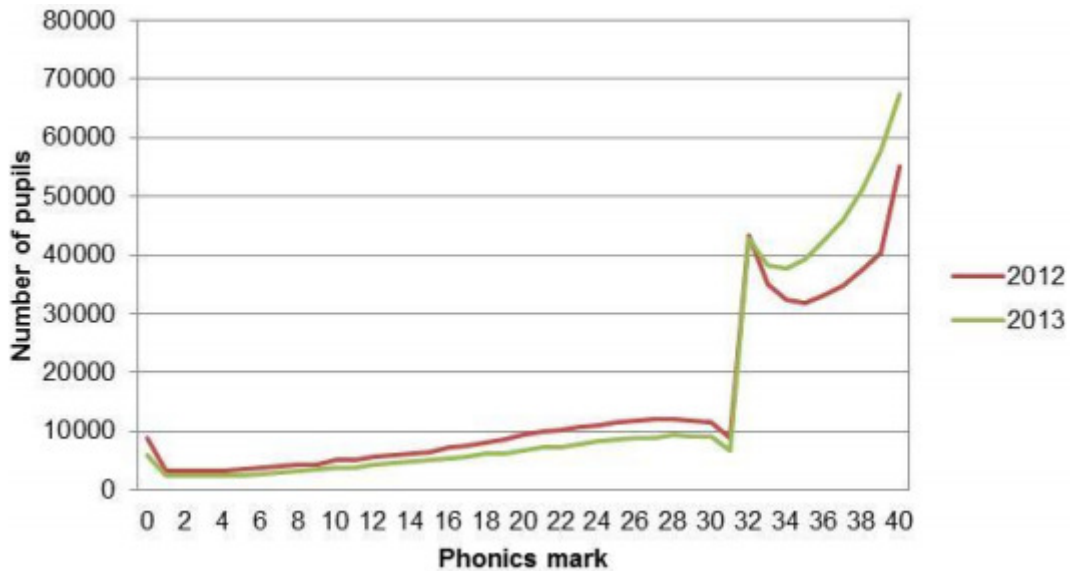
In the five years since the introduction of the phonics test for 5- and 6-year-olds there appears to have been a continuous stream of headlines. There have been concerns about the inclusion of nonsense or pseudo-words (Davis, 2014) with the thought that the ability to blend phonemes together is akin to reading. This was linked to the headline on Davis's article that suggested that more able children were being 'damaged' by being forced to break words down and then blend sounds together. He cited people who maintained this can affect the fluency of reading, and ultimately their children's understanding and comprehension of what they have read. The process of using synthetic phonics slows down the reader's overall reading speed (Richardson, 2014). Due to the inclusion of 'nonsense' words a recent evaluation has shown that there is now an increase of the teaching of nonsense words such as *thob*, *blim* and *flamp* to help the children succeed in the test (Walker, 2014).

In 2013 the government reported that 69% of children passed the Year 1 phonics check, an increase of 11% on the previous year's result. However, concerns around the data have been raised. The phonics test is scored on a 'right or wrong' basis; therefore, when the results are shown graphically, a continuous distribution would be expected (i.e. increasing numbers up to the most common score (the mode) and then tailing off). If a test was particularly easy, you might get a



distribution that created a ceiling effect. However, the results for the phonics check did not show a continuous distribution (see Figure 4.3).

**Figure 4.3: Year 1 phonics screening check mark distribution, 2012 & 2013**



(DfE, 2013b, p. 4)

The distributions shown in Figure 4.3 are very unusual, as noted by Bishop (2013). The fact that there was a general increase in the numbers of children up to the 30-mark area, and then a sudden spike at 32 (almost five times as many children scoring 32 points as 31 points) could be indicative of data manipulation. And since the 'pass mark' of 32 was known prior to the administration of the test, it could be suggested that teachers were weighting their results.

A 2012 technical report (Standards and Testing Agency, 2012b) explained the peak at the pass mark of 32 as follows:

An interpretation of the area around the threshold peak is consistent with teachers accounting for potential misclassification in the check results, and using their teacher judgment to determine if children are indeed working at the expected standard.

Standards and Testing Agency (2012b, pp. 10-11)

The report states that the misclassification accounted for approximately four per cent of the results. If teachers were interpreting the data as described above and deciding whether the children were working at the 'required standard', possibly

regardless of the child's score on the test, should it be assumed that the phonics test is neither needed nor necessary, since teacher assessment is playing a part?

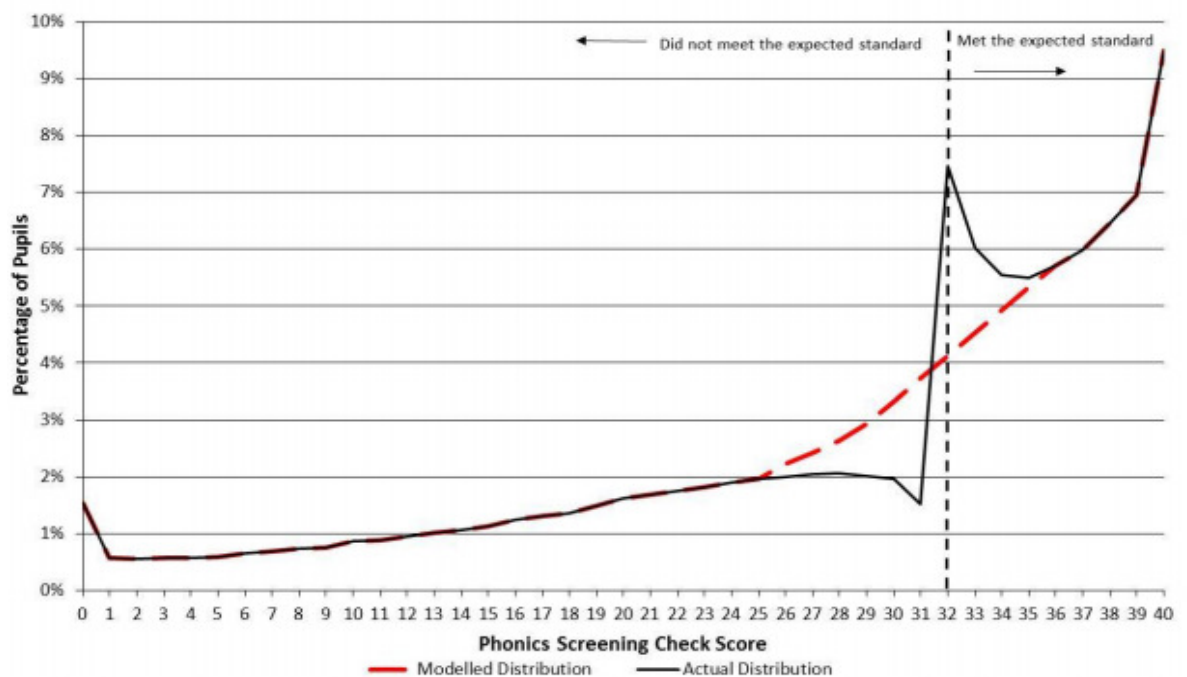
Another report (DfE, 2013c) analysed the spike in more detail and stated:

The change in trend and the spike at 32 suggests [sic] that pupils who are on the borderline of meeting the expected standard may have consciously or unconsciously been given the benefit of the doubt and been marked at 32.

DfE (2013c, p. 28)

The government statisticians then re-examined the 2012 results by removing the spike and interpolating the data (Figure 4.4). Doing this reduced the national 'pass rate' by four percentage points to 54%. A second analysis was done using a logistic regression; by this method the percentage of children who achieved the expected standard was estimated as 46%.

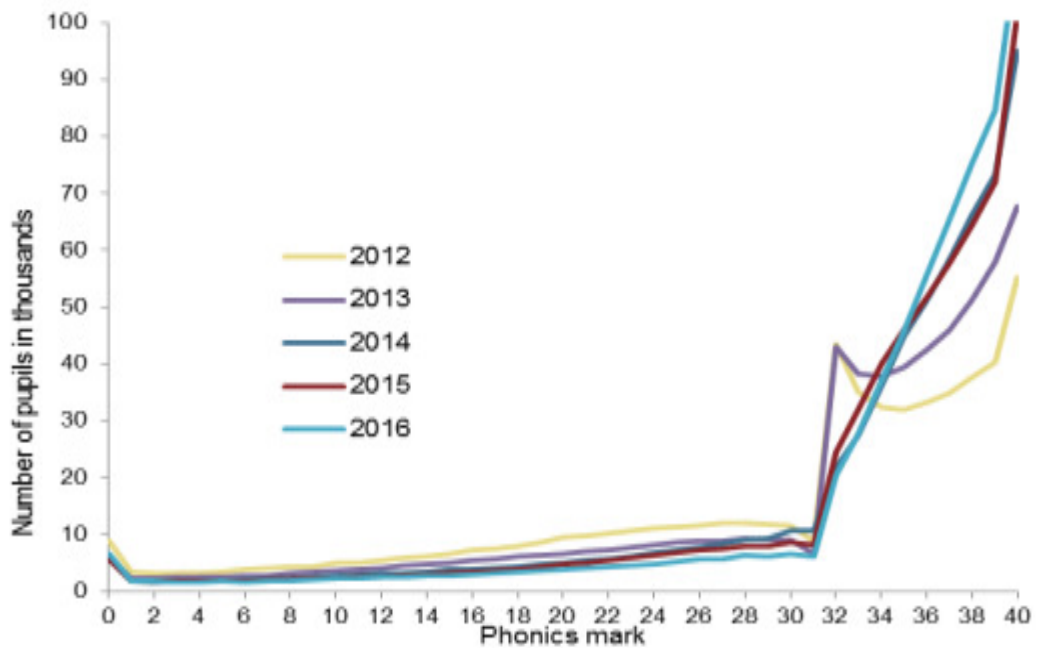
**Figure 4.4: National phonics check results for 2012, actual and remodelled**



(DfE, 2013c, p. 30)

For both years 2012 and 2013 the pass mark was released, but in subsequent years the pass mark was not released. In the years 2014-16 the pass mark was still 32. The distributions for the years 2012-2016 are shown in Figure 4.5.

**Figure 4.5: Phonics check data 2012-2016**



Source: National Pupil Database

DfE (2016, p. 4)

As can be seen, the spike that was observed in 2012 and 2013 was not observed in the following years, which supports the supposition that knowledge of the pass mark may have influenced the marking of the phonics test. Clark, M. (2015) suggested that the results from 2012 and 2013 should not be compared with the succeeding years since the later pass mark was (theoretically) unknown, even though identical.

Research has shown that the phonics check does indeed identify children who have not reached their ARE; however, what happens with this information was a question which was asked by the press (Sellgren, 2013). This was further discussed by Clark, M. (2015) who highlighted that the data were not used to identify whether children might have reading or other learning difficulties. Clark, M. (2015) went on to express her concerns over several other matters. She questioned the use of pseudo-words, why the children must re-take the phonics check if they have failed it previously, and why there are large differences between older and younger children in the year. She calculated that 82% of the oldest children pass, compared to only 65% of the younger children. Rightly she questioned why the government had not picked up on this statistic. She

suggested that, given a further year to develop, these younger children might pass the check without further tuition because they have matured.

Following on from Clark, M. (2015), if these now Year 2 children have re-taken and passed the phonics check, one assumes that their phonics knowledge is that of a Year 1 child. Surely this means that these children are 12 months behind their peers, but as previously stated and supported by Clark, M (2015), what contingencies have been put in place to help these children from falling further behind? A means of assessing children's ability has been designed but its full potential is not being used.

Clark (2017) continues to discuss the efficacy of the phonics check and its inclusion of pseudo-words. Her main concern is the government rhetoric that persuades the population that reading is improving and that the increasing number of children passing the phonics check is evidence of this. But what the government is not divulging is that 26% of children born in July and August are failing this phonics test, whereas just 12% of older children fail. This suggests that younger children find the test more difficult than older children, which could mean the development of phonics skills could be age-related.

Wrigley (2017) is more direct when it comes to the use of phonics and the phonics check. He understands that phonics is a crucial part of beginning reading instruction. What he finds difficult to understand is:

...how a politician, Nick Gibb, now Minister for Schools, has managed to impose his fixed ideas on school and compel teachers to use one particular approach to phonics.

Wrigley (2017, p. 98)

Wrigley continues to say:

...Gibb publicly proclaims ... to favour 'evidence-based teaching' in order to claim legitimacy, [but] it is telling to explore how the limitations of his use of evidence, the generation of moral panics, the political rhetoric and discourse, relates to the de-professionalisation of teachers.

Wrigley (2017, p. 98)

He goes on to say that, although there has been an increase in the numbers of children passing the phonics check, there is no correlating improvement in reading. So, the question must be asked, 'Why are teachers being forced to put Year 1 children under such pressure when there appears to be no benefit to a child's reading progress?'

However, providing contradictory evidence to Wrigley's (2017) claims is the most recent PIRLS study (McGrane et al., 2017), mentioned in chapter 1, which allows direct comparison of the performance of children in the Year 1 phonics check (in 2012) with their attainment in PIRLS (in 2016). The analysis showed that children who gained full marks in the Year 1 phonics check attained higher marks in PIRLS. This was also reflected in the children who had to retake the phonics check in Year 2: those who obtained higher marks in the phonics check also scored higher marks in PIRLS. Nick Gibb and other ministers have claimed that this finding supports the government's emphasis on the importance of phonics in the early stages of reading. However, comparison of KS1 reading results yields similar information. The 2016 PIRLS report states, unsurprisingly, that children who obtained level 1 at the end of KS1 had the lowest PIRLS score of 460, whilst those who gained a level 3 had the highest average PIRLS score of 615. Although the phonics check comparison shows that those children with a higher phonics score did better in PIRLS, the same can be said for those who did well in KS1 reading assessments.

More importantly it must be noted that, although there is a correlation between higher phonics check scores and higher scores in PIRLS, this does not prove causation. To determine causation, it would be necessary to conduct further studies, ideally a randomised control trial. In addition, Northern Ireland (565) and

the Republic of Ireland (567) both scored significantly better than England, and neither country has introduced a phonics check. Therefore, Nick Gibb's claim that the phonics check has improved England's PIRLS score cannot be substantiated from the data gathered.

The most recent National Curriculum, version 4 (DfE 2013a) has a different layout and no longer refers to literacy but to English, which is split into spoken language, reading and writing. Reading is then further split into word reading and comprehension (both listening and reading) – which is reminiscent of the simple view of reading. It states:

Skilled word reading involves both the speedy working out of the pronunciation of unfamiliar printed words (decoding) and the speedy recognition of familiar printed words. Underpinning both is the understanding that the letters on the page represent the sounds in spoken words. This is why phonics should be emphasised in the early teaching of reading to beginners (i.e. unskilled readers) when they start school.

DfE (2013a, p. 13)

Intuitively this seems rather more balanced than earlier statements. The references to unfamiliar and familiar printed words are reminiscent of the Simple view of reading and dual-route models, and the implication that only children who are not yet skilled readers need phonics should hopefully lead to those who are fluent not being put through irrelevant instruction.

Within the main body of NC4, statutory requirements for each year are documented, but phonics is only mentioned directly for Years 1 and 2. After which it states that for the whole of KS2 children should:

... apply their growing knowledge of root words, prefixes and suffixes (morphology and etymology), as listed in English Appendix 1, both to read aloud and to understand the meaning of new words that they meet.

DfE (2013a, p. 35)

Also within NC4, as mentioned in the quotation above, there is an appendix which describes the spelling patterns that should be followed throughout both primary

Key Stages. There is also a greater emphasis on children's need to acquire comprehension skills, the second arm in the simple view of reading.

*Reading: the next steps: supporting higher standards in school* (DfE, 2015c) discusses the importance of reading, and states various statistics showing how important the ability to read is. It states that children who can read well tend to succeed at school. This is not a surprising statement as those that can read can access the curriculum, whilst those that struggle cannot. This then impacts on their adult lives.

However, as shown, the National Curriculum has changed significantly – but maybe that is the issue – the curriculum has changed repeatedly and there has not been sufficient time to allow new approaches to work; nor has there been sufficient training for teachers to ensure that they have the skills to deliver consistently good teaching.

A section of the report discusses the importance of phonics. To encourage the use of phonics the government was developing a phonics partnership grant programme where schools which were seen to teach phonics exceptionally would form a partnership with a struggling school. This would ensure that good practice in phonics teaching was disseminated.

The report succinctly describes the evidence that supports the use of phonics for beginning reading, but, as with all these government documents, it does not provide an explicit means by which to implement these changes. Admittedly the roll-out of phonics teaching is a good approach and should encourage consistency. It mimics the approach that is discussed in Machin et al. (2016), who described how two earlier phonics initiatives were introduced in a staggered manner over the four school years 2005/06, 2006/07, 2008/09 and 2009/10. Although this programme was begun in 2005, no information about it was available until 2016, perhaps because data were being gathered over a long period so that longitudinal comparisons could be made, and conclusions drawn. The two projects, the *Early Reading Development* pilot (ERDp) and *The Communication, Language and Literacy Development Programme* (CLLD), which replaced the ERDp, included the recommendations of the Rose Report, and their staggered introduction enabled a 'natural experiment' to be defined, with

groups of schools recruited later in effect serving as ‘time series’ controls for earlier groups, as shown in Figure 4.6.

**Figure 4.6: Group allocation in Machin et al. (2016, p. 26)**

| Groups            | Phonics Programme | LA                                  | Entry               | Birth Cohort of Students First Exposed to Programme | Year of Age 5 Assessment | Year of Age 7 Assessment | Year of Age 11 Assessment |
|-------------------|-------------------|-------------------------------------|---------------------|---|--------------------------|--------------------------|---------------------------|
| Treatment Group 1 | EDRp              | Schools in 18 LAs                   | 2005/06             | 2001  | 2006                     | 2008                     | 2012                      |
| Treatment Group 2 | CLLD              | Schools in same 18 LAs + 32 new LAs | 2006/07             | 2002  | 2007                     | 2009                     | 2013                      |
| Control group     |                   | Schools in next 50 LAs              | 2008/09 and 2009/10 | 2004  | 2009                     | 2011                     | 2015                      |
|                   |                   | Schools in next 50 LAs              | 2009/10             | 2005  | 2010                     | 2012                     | 2016                      |

Machin et al.’s analysis showed that ‘intensive training’ helped young children with their literacy skills; however, by the age of 11 there was no overall difference between those children who had been part of the intervention and those who had not. Machin et al. explained this outcome by saying that ‘eventually’ most children learn to read, and conceded that there may be unmeasurable benefits to being able to learn to read earlier, but did not expand on this idea. Maybe an added benefit could just be that such children develop a love of books and literature, and thus continue into further education. Despite the non-significant overall result, the study showed that the long-term effects at age 11 were significant for children who entered school at risk of falling behind, for those for whom English is an additional language, and for the economically disadvantaged – precisely the groups one would hope would benefit. However, these findings imply that there were other groups (not reported) who had not benefited, and might actually have regressed between ages 7 and 11.

Despite the partial success of these phonics initiatives, the national statistics on low achievement at key stages 1 and 2 cited in chapter 1 have not yet declined close enough to zero, thus leaving space to investigate, whether non-phonics-based alternatives might work better for some children who have fallen behind. To produce a society of skilled readers it is necessary to have a skilled workforce



who can effectively teach children using approaches that have been empirically shown to be effective. This should result in most children reaching their age-related expectations.

#### **4.7 Experimental research supporting the use of phonics in the teaching of beginning reading**

This section is split into two sub-sections: the experimental evidence on the use of systematic phonics instruction (a) for (mainly) normally-developing children and (b) as an intervention for children falling behind with their reading and spelling. As far as possible, this part of the literature review is based on previous systematic reviews and meta-analyses rather than on single studies.

##### **4.7.1 How effective is systematic phonics teaching for (mainly) normally-developing children of primary age?**

The inclusion of (mainly) in the heading is because the two principal reviews analysed (Ehri et al., 2001; Torgerson et al., 2006a) covered both normally-developing children and those falling behind. It would have been more logical to separate the two categories (and studies focusing specifically on those falling behind are picked up later in the chapter), but the critiques of Ehri et al. summarised below apply across the board, and thus all their findings have to be presented at this point.

As part of the National Reading Panel (National Reading Panel, 2000) review of reading in the US, specific areas of reading education were highlighted as areas of interest. One such area was the use of phonics in the teaching of reading. Ehri et al. (2001) undertook the NRP systematic review and meta-analysis of the experimental evidence on phonics tuition for children in kindergarten and grades 1-6 (years 1-7, ages 5-12). A total of 38 studies yielding comparisons between systematic phonics instruction and unsystematic or no phonics instruction met their inclusion criteria. They produced 66 comparisons from the 38 included studies; 22 of the comparisons were based on participants who were classed as 'normally-developing', and in 40 the participants were 'at risk' (four studies could not be differentiated) – see Table 4.1.

**Table 4.1: Ehri et al.'s mean effect sizes for normally-developing and at-risk children, by age group and overall**

| Age            | Status                     | Number of comparisons | Effect size ( <i>d</i> ) |
|----------------|----------------------------|-----------------------|--------------------------|
| All            | Overall                    | 66                    | 0.41                     |
| KG and Grade 1 | Overall                    | 30                    | 0.55                     |
| Grade 1        | Normally-developing        | 14*                   | 0.48                     |
| KG             | At risk                    | 6                     | 0.58                     |
| Grade 1        | At risk                    | 9                     | 0.74                     |
| Grades 2-6     | Overall                    | 32                    | 0.27                     |
| Grades 2-6     | Normally-developing        | 7                     | 0.27                     |
| Grades 2-6     | At risk – low achievers    | 8                     | 0.15 §                   |
| Grades 2-6     | At risk – reading disabled | 17                    | 0.32                     |

\* There was only one study using normally-developing kindergarteners; this was omitted from the relevant calculation

§ Non-significant; all other results were significant at  $p < 0.05$

Ehri et al. concluded that the evidence showed that systematic phonics teaching helped children to learn to read better than if their teaching included unsystematic or no phonics. They went on to say that phonics instruction is more effective when it is used as a beginning to read strategy and introduced at kindergarten/grade 1 than if it is used at grades 2-6; in particular, their figures suggest that the impact of systematic phonics for at-risk children is much larger in the early than in later grades – again, this supports the investigation of non-phonetic alternatives.

Camilli et al. (2003) tried to recreate the analysis that Ehri et al. had conducted by reviewing all the studies that Ehri et al. had included in their analyses. They rejected one study (because it had provided eight comparisons against control groups, despite having only one intervention group) and added three studies not used by Ehri et al. which they considered did meet Ehri et al.'s inclusion criteria, despite having been excluded. Camilli et al. calculated an overall mean effect size of 0.24, compared to Ehri et al.'s 0.41. Camilli et al. explained that the observed difference between their results and Ehri et al.'s was due to several factors.

As mentioned previously, Ehri et al. (2001) included 38 studies in their meta-analysis, but a total of 66 comparisons were obtained. Camilli et al. raised concerns regarding the validity of producing so many comparisons out of so few

studies. This automatically meant that those studies with more comparisons had more effect on the outcome means than studies that provided only a single comparison. (This point was also made by Torgerson et al., 2006a.) In a further article Camilli et al. (2006) reanalysed Ehri et al.'s results again, and calculated an even smaller effect size of  $d=0.12$ . In fact, they found the greatest effect size was gained between tutoring and small group interventions (Table 4.2).

**Table 4.2: Effect sizes calculated by Camilli et al. (2006)**

| Comparison  | Effect size ( $d$ ) |
|---|---------------------|
| Systematic phonics vs 'some' phonics in treatment group             | 0.123               |
| Systematic phonics vs 'some' literacy activities in treatment group | 0.210               |
| Systematic phonics vs 'some' literacy activities in control group   | -0.403              |
| Tutoring vs small group/whole class instruction                     | 0.489               |

Hammill & Swanson (2006) also re-analysed Ehri et al.'s data, and converted the mean effect sizes ( $d$ ) into mean effect correlations ( $r$ ) and  $r$ -squares. A comparison of the mean effect sizes and the  $r$ -values is shown in Table 4.3.

**Table 4.3: Effect sizes from Ehri et al. compared with  $r$  and  $r^2$  as calculated by Hammill & Swanson**

| Age        | Effect size | $R$  | $r^2$ |
|------------|-------------|------|-------|
| Grade 1    | 0.48        | 0.23 | 0.05  |
| Grades 2-6 | 0.27        | 0.13 | 0.02  |

The  $r^2$  value, which is obtained from the  $r$  value, is a statistic that shows the amount of variance that can be accounted for by the intervention. So, for the above examples Hammill and Swanson were stating that for Grade 1 and Grades 2-6 the percentage of variance that could be attributed to phonics teaching was just 5% and 2% respectively. In fact, Hammill & Swanson stated:

This means that the meagre advantage in reading derived from phonics instruction as compared with other approaches was not sustained over time. The unimpressive  $r$ -type statistics ... lead one to wonder just how solid the support is for systematic phonics.

Hammill & Swanson (2006, p. 22)

Although Ehri et al. had their critics, there were those who came out in support of them. Instead of re-analysing the results obtained by Ehri et al., Stuebing et al. (2008) approached their response to Camilli et al. in a very different manner. Stuebing et al. looked very closely at both studies and arrived at three theories which could account for the differences between Camilli et al.'s and Ehri et al.'s results. They constructed their argument in the manner of proving or disproving three hypotheses. For each of their three hypotheses Stuebing et al. completely unpicked Camilli et al.'s approach. First, Stuebing et al. suggested that the two studies were asking different questions and analysing different sets of factors:

Altogether, the NRP report and Camilli et al. (2003) ask different questions. The NRP question is analogous to asking about the value of receiving the intervention versus not receiving the intervention. The Camilli et al. (2003) report is analogous to asking what is the value of receiving a strong form of the intervention compared to receiving weaker forms of the intervention and relative to factors that moderate the outcomes.

Stuebing et al. (2008, p. 125)

Stuebing et al. imply (correctly) that Ehri et al.'s principal findings were based on main effects without allowing for moderator variables, so it is unsurprising that Camilli et al. reached different conclusions. Hence both sets of figures are valid, for different purposes – but for present purposes, it is the main effect of phonics vs little or no phonics that is of interest.

Stuebing et al.'s second theory was concerned with the additive effects of other literacy approaches on top of phonics or non-phonics approaches; in other words, how much the teachers taught reading skills other than alphabets to encourage reading. These approaches included encouraging independent reading, 'purposeful writing', and using literature to encourage higher-order skills.

Our second hypothesis is that in this stricter comparison, which allows separate empirically derived estimates of the effect of additional literacy activities on top of both some phonics and systematic phonics, effect sizes will be larger when additional literacy activities and tutoring are added to the effects of systematic phonics instruction.

Stuebing et al. (2008, p. 125)

Stuebing et al.'s re-analysis supported the idea that phonics instruction is more effective if it is part of a wider literacy programme.

Stuebing et al.'s final theory was based upon effect sizes and what they actually mean. They quoted Cohen (1988), who explained that the problem with applying terms such as small, medium and large is that they can be viewed as absolute or relative, and therefore not completely understood. It is suggested, for example, that a small effect size may over time become a moderate effect size. Or, for a specific intervention, a small effect size may be recorded but, if it is applied to a large population, thousands of children could benefit. Therefore, their third hypothesis stated:

...that interventions with effect sizes as small as those identified by Hammill and Swanson (2006) for phonics instruction could significantly reduce the number of children with reading problems depending on the base rate used to estimate the incidence of reading difficulties and the effect size associated with different interventions.

Stuebing et al. (2008, p. 126)

They carried out statistical modelling using various estimates of the base rate of reading difficulties and of the effects of interventions, and showed that even quite modest effect sizes can, in theory, substantially reduce the incidence of reading difficulties, thus rebutting Hammill and Swanson's main argument.

In conclusion Stuebing et al. stated:

Camilli et al. (2003, 2006) and Hammill and Swanson (2006) do not contradict the NRP report, concurring in supporting comprehensive approaches to reading instruction.

Stuebing et al. (2008, p. 123)

The most recent review considered here is Suggate (2010). He analysed 85 RCTs or quasi-experimental studies yielding 116 treatment group/control group comparisons, across the ages pre-kindergarten to grade 7 – but these covered a range of interventions, based variously on phonemic awareness, phonics, comprehension, or a mixture. Relevant overall effect sizes were  $d = 0.50$  for phonics (Number of studies = 36) and  $d = 0.58$  for comprehension-based approaches (Number of studies = 37). A regression analysis showed that the benefit of phonics was strongest in kindergarten and slowly diminished across

Grades 1-4 (there were not enough studies at later grades to estimate effect sizes for phonics), while the benefit of comprehension-based approaches was lowest in kindergarten and slowly increased across Grades 1-7, the cross-over occurring in Grade 2.

The systematic reviews and meta-analyses so far mentioned were all conducted in the USA and covered only studies conducted in North America. The first in the UK, conducted by Torgerson et al. (2006a) and confined to phonics, was commissioned alongside the Rose review. Unlike the other reviews just listed, Torgerson et al. (2006a) used only RCTs. Out of the 6114 studies initially identified, only 12 met the criteria decided upon by the panel; these included the first relevant RCT carried out in the UK (Johnston & Watson, 2004, experiment 2). Torgerson et al. did not state an overall mean effect size for phonics, but gave separate ones for reading accuracy (= word recognition) and comprehension. The effect size for reading accuracy was small,  $d= 0.27$ , but highly statistically significant ( $p= 0.007$ ), which suggested that phonics should be included as part of any literacy teaching. The effect size for comprehension was also small, at  $d= 0.24$ , but was statistically non-significant because the number of relevant studies was small ( $N= 4$ ). Therefore, the effect of phonics teaching on improving comprehension skills could not be supported or rejected.

#### **4.7.2 Summary of data supporting the use of systematic phonics for primary-age children**

Table 4.4 shows a comparison between the findings of Ehri et al. (2001) and Torgerson et al. (2006a).

**Table 4.4: Comparison of the findings of Ehri et al. (2001) and Torgerson et al. (2006a)**

| Research question   | Answer given by Ehri et al. (2001) | Answer given by Torgerson et al. (2006a) |
|---|------------------------------------|--|
| Does systematic phonics instruction enable children to make better progress in reading <i>accuracy</i> than unsystematic or no phonics?   | Yes                                | Yes                                      |
| Does systematic phonics instruction enable both normally-developing children and those at risk of failure to make better progress in reading <i>accuracy</i> than unsystematic or no phonics? | Yes                                | Yes                                      |
| Does systematic phonics instruction enable children to make better progress in reading <i>comprehension</i> than unsystematic or no phonics?  | Yes                                | Not clear                                |
| Does systematic phonics instruction enable children to make better progress in spelling than unsystematic or no phonics?  | Yes                                | Not clear                                |
| Does systematic <b>synthetic</b> phonics instruction enable children to make better progress in reading <i>accuracy</i> than systematic <b>analytic</b> phonics?                              | (Not addressed *)                  | Not clear                                |
| Does systematic <b>synthetic</b> phonics instruction enable children to make better progress in reading <i>accuracy</i> than all other forms of phonics?                                      | Yes                                | (Not addressed)                          |

Source: Brooks (2007), with last row added

\* Ehri et al. found only study comparing synthetic and analytic phonics

As can be seen, both Ehri et al. and Torgerson et al. provided evidence that the teaching of systematic phonics to all children is beneficial. However, this finding is only clearly true for reading accuracy. There is no firm evidence that systematic synthetic phonics instruction is better than (for example) systematic analytic phonics, even though the added row in Table 4.4 suggests that systematic synthetic phonics instruction is better than all other forms of phonics lumped together. This partly contradicts the emphasis made by Rose (2006) when he encouraged the use of systematic synthetic phonics. Clark, M. (2015) states that most researchers believe that:

- There is benefit from the inclusion of phonics within the early instruction in learning to read in English, within a broad programme;
- There is *not* evidence to support phonics in isolation as the one best method;
- There is *not* evidence for synthetic phonics as the required approach rather than analytic phonics.

Clark, M. (2015, p. 2)

It is also unclear whether phonics increases comprehension skills, which suggests that maybe a meaning-emphasis approach should be included with phonics, the best of both worlds.

### **4.7.3 Do systematic phonics interventions help falling-behind readers to become more skilled at reading?**

#### **4.7.3.1 Introduction**

This area of research was chosen because the author felt there was insufficient research on the effectiveness of non-phonics interventions with Key Stage 1 children who are falling behind. Even so, the lack of relevant literature was surprising. Some sources were unhelpful. For example, a recent systematic review and meta-analysis (Galuschka et al., 2014) dealt only with reading-disabled teenagers, and some of the interventions classified as 'phonics' were manifestly not. McArthur et al. (2012) based very detailed analyses on just 11 studies, but only one (Hurry & Sylva, 2007, discussed in section 4.7 below) proved relevant.

Identification of applicable studies was necessary; the following criteria were used:

1. Participants in UK years 1 and 2 or North American KG/grade 1;
2. RCT design. This criterion was included to provide direct comparison with the present doctoral research pilot/feasibility study;
3. Participants must be at-risk readers.

#### **4.7.3.2 Studies identified from Ehri et al. (2001) and Torgerson et al. (2006a)**

As shown in Table 4.1 above, Ehri et al. (2001) had identified 15 studies with falling-behind readers in KG/grade 1. These studies were reviewed using the criteria stated above. Five studies were identified and are summarised in Table 4.5.



**Table 4.5: Characteristics of five studies identified from Ehri et al.**

| Author                       | Phonics | Age     | Length of study (dosage) | Group size  |
|------------------------------|---------|---------|--------------------------|-------------|
| Brown & Felton (1990)        | Syn     | Grade 1 | 2 yrs *                  | Small group |
| Mantzicopoulos et al. (1992) | Misc    | Grade 1 | 50 sessions (1hr/wk)     | Tutor       |
| Martinussen & Kirby (1998)   | Syn     | KG      | 8 wks (40-60 min/wk)     | Small group |
| Torgesen et al. (1999)       | Syn     | KG      | 2.5 yr (80 min/wk)       | Tutor       |
| Umbach et al. (1989)         | Syn     | Grade 1 | 1 yr (50 min/day)        | Small group |

\* This study did not state the dosage, only that the intervention took the place of usual reading instruction in the classes

Syn = synthetic phonics; misc = miscellaneous

The first area of concern with respect to the studies identified was the length of time each intervention spanned. Three of the studies (Brown & Felton, Torgesen et al., and Umbach et al.) all had delivery times of at least a year, and in the case of Torgesen et al. it was 2.5 years. The FFT wave 3 intervention used in the current study is just a 10-week intervention, so these studies are not directly comparable. However, they do provide moderate evidence that interventions using a synthetic phonic approach do help at-risk readers to become better readers. However, it cannot be stated whether there is a comparable increase in comprehension skills.

A cross-check with Torgerson et al. (2006a) showed that they had identified and included four of these studies, but had rejected Mantzicopoulos et al. (1992). The 'TEACH' intervention which was investigated in this study was not a direct reading instruction intervention, and there was a very high rate of attrition – 62% of the study group left before the end of the study. On these grounds Torgerson et al. rejected this study, as Ehri et al. should have done.

The mean effect sizes for the four studies, according to Ehri et al. and Torgerson et al., are shown in Table 4.6.

**Table 4.6: Effect sizes of four relevant studies as stated by Ehri et al. and Torgerson et al.**

| Author  | Effect sizes                                     |      |          |                         |                        |                         |
|---|--|------|----------|-------------------------|------------------------|-------------------------|
|   | Ehri et al.                                      |      |          | Torgerson et al. **     |                        |                         |
|   | Reading Accuracy                                 | Comp | Spelling | Reading Accuracy        | Comp                   | Spelling                |
| Brown & Felton (1990)                             | WI = 0.02<br>NW = 0.92                           | -    | 0.51     | 0.24<br>(-0.89 to 1.37) | -                      | -                       |
| Martinussen & Kirby (1998)                        | WI = 0.53<br>Dec = 0.63<br>NW = 0.62             | -    | 0.68     | 0.44<br>(-0.31 to 1.19) | -                      | 0.30<br>(-0.44 to 1.05) |
| Torgesen et al. (1999) *<br>Lindamood<br>Embedded | WI = 0.08<br>NW = 0.58<br>WI = 0.52<br>NW = 0.12 | -    | -        | 0.07<br>(-0.34 to 0.48) | -                      | -                       |
| Umbach et al. (1989)                              | WI = 1.30  | 1.08 | -        | 2.69<br>(1.72 to 3.67)  | 1.08<br>(0.33 to 1.84) |                         |

\* 'Lindamood' and 'Embedded' indicate the two programmes investigated

\*\* Figures in brackets show confidence intervals

Key: WI = word identification; NW = non-words; Dec = decoding; Comp = Comprehension

The main reason for the varying mean effect sizes was differences in methods of calculation. However, what Table 4.6 does show is the shortage of evidence on comprehension and spelling, and wildly varying estimates of effectiveness for reading accuracy, with generally higher effects for non-word reading.

#### 4.7.3.3 Slavin et al. (2011)

Slavin et al. (2011) reviewed a large number of 'Effective programs for struggling readers', as the title of their article has it, and reached the following conclusion (among others):

[A]lmost all successful programs have a strong emphasis on phonics... [O]ne-to-one tutoring programs in which teachers were the tutors had a much more positive weighted mean effect size if they had a strong phonetic [sic – 'phonic' is meant] emphasis (mean ES = +0.62 in 10 studies). One-to-one tutoring programs with less of an emphasis on phonics, specifically Reading Recovery and TEACH, had a weighted mean effect size of +0.23.

Slavin et al. (2011, p. 19)

Reading Recovery is discussed in more detail below. Here it is enough to note that Slavin et al.'s analysis used both RCTs and quasi-experimental studies, and there seems no way, within the article, to disentangle the effect sizes of studies with the two forms of research design from each other.

#### 4.7.3.4 Identifying papers from the What Works Clearinghouse and EEF

Two data repositories were reviewed, the (US) What Works Clearinghouse and the (UK) Education Endowment Foundation. Studies were only included if they fulfilled the criteria mentioned in section 6.3.1 (where the criteria are justified).

The studies are shown in Table 4.7.

**Table 4.7: RCTs identified from EEF and What Works Clearinghouse**

| Authors                 | Country | Intervention                   | Age        | Effect sizes  |               |               |
|-------------------------|---------|--------------------------------|------------|---------------|---------------|---------------|
|                         |         |                                |            | Accuracy      | Comp.         | Spelling      |
| Borman et al. (2008)    | USA     | Open Court Reading             | Grades 1-5 | 0.20          | 0.11          | 0.25          |
| Hatcher et al. (2006) * | England | Reading Intervention Programme | Year 1     | 0.46 to 0.94  |               |               |
| Hurry & Sylva (2007)    | England | Phonological Training          | Year 2     | 0.08          | -0.05         |               |
| McMaster et al. (2005)  | USA     | PALS                           | Grade 1    | -0.09 to 0.44 | -0.17 to 0.49 | -0.16 to 0.40 |
| McNally et al. (2016)   | England | ABRA (Non-ICT)                 | Year 1     | 0.096         | 0.231         |               |
| Vadasy et al. (1997)    | USA     | Sound Partners                 | Grade 1    | 0.19 to 0.37  |               |               |
| Vadasy et al. (2006)    | USA     | Sound Partners                 | KG         | 0.44 to 0.59  |               |               |
| Vadasy & Sanders (2008) | USA     | Sound Partners                 | KG         |               | 0.41 to 0.63  |               |

\* This was not, strictly speaking, an RCT but has been included since, as described by Brooks (2016, p. 122), it was a 'tightly designed and administered quasi-experiment' where the children were matched on pre-test reading age and the duration of the intervention was similar for all groups. In addition, the adults who administered the tests were unaware of which group the children had been allocated to.

Comp. = Comprehension

All the programmes listed have a strong emphasis on phonics, usually accompanied by comprehension instruction (e.g. Open Court Reading), but sometimes not (Phonological Training) – this may help explain the difference in effectiveness between those two interventions. But the overall impression from the data in Table 4.7 is the scatter of results, leading to not very clear conclusions, except that the impact may be stronger on reading accuracy than on comprehension.

#### **4.7.4 Summary of the effectiveness of systematic phonics instruction for children who are at risk of falling behind**

As Torgerson et al. (2006a) concluded, systematic phonics instruction does have a positive effect for children who are falling behind, at least in terms of reading accuracy (word recognition). However, none of the effect sizes are large, and only a few would be classed as medium. The effect of systematic phonics tuition on comprehension is less understood, and it may be that phonics has little impact.

### **4.8 The opposition to phonics**

This section discusses the opposition to systematic phonics which appears to have its foundations firmly planted on personal opinion and ultimately a fight against government wishes. In addition, the teaching of phonics is sometimes poor because teachers understand little about the complexities of reading.

#### **4.8.1 ‘A monstrous regimen of phonics’ (Davis, 2012)**

The title of this paper immediately makes the reader aware of Davis’s stance. He makes it clear that he disagrees with the government’s implementation of blanket phonics tuition in all primary schools in England. He believes that this mandate could persuade many educators that phonics instruction is the only skill required for children to learn to read.

It must be remembered that phonics instruction is a means to an end. It provides children with the preliminary skills needed to decode unfamiliar written words and translate them into spoken words, and understand them. What is unproven is the effect of phonics on comprehension skills, which may affect the fluency of reading.

Davis expressed concern that children were being coached so diligently with their phonic rules that their ability to read some words could be affected. For example, the letter <p> does not always correspond to the phoneme /p/. Look at the words *psychic*, *cupboard*, *corps*, *receipt* and *psalm*. It is unlikely these words would be found in beginner reading books, and Davis was grasping at straws to try to prove his point. Not only do the words given as examples not appear in beginner books; neither are they decodable. The early assumption that the pronunciation of the letter <p> would be /p/ all the time gives way during reading development when it becomes clear that there are exceptions, as there are with other graphemes.

Careful teaching would lead children to realise that (for example) <ph, pt, ps> are special cases. Davis might have had a more convincing argument if he had used high-frequency words such as *do*, *go* and *no*. All end in the letter <o>, but their pronunciations include two different phonemes.

Davis stated that once a word has been segmented and the sounds blended together, there has to be some understanding of the context to understand that the correct word is being read. The example that Davis gave is the word 'mints'. Davis suggests that children could pronounce the word correctly, or pronounce it as 'mince' which would be contextually incorrect. The problem with this is that the normal pronunciation of both words is /mins/ since consonant clusters are normally simplified during running speech, and it is only during decoding that the nuances in the word endings would be identified. Using contextual cues is important, but again this is something that develops over time. For example, the word 'read' can be pronounced in two different ways:

She read the book.

I can read.

Without contextual understanding the word '*read*' could be pronounced incorrectly. This example does support the idea that an understanding of the context is important when reading, a skill that systematic phonics does not of itself provide. But this is a skill that develops over time and cannot be taught as a beginning reading strategy, which is the purpose of phonics.

#### **4.8.2 Carbo (1987)**

Carbo suggests that there may be three possible ways in which children relate to phonics:

1. Children who need phonics, so they become good readers – these children have strong auditory/analytic skills. They can hear the nuances in the spoken word and then link it to the written word.
2. Children who can learn phonics but do not need phonics to be able to read well – these children have the auditory/analytic skills which allow them to interpret phonics. But in addition to this they have a global approach to

reading. This means that they learn reading by writing stories, reading books of their own choice, listening to recorded books.

3. Children who cannot learn how to apply phonics – these children struggle with phonics because they have poor auditory/analytic skills.

Carbo went on to state that often the children found in the third group underachieve in reading due to the fact that they cannot access the curriculum. What she questioned was that not all children benefit from being taught phonics and, that although most children do benefit, other approaches to reading instruction need to be available for those who learn in different ways.

#### **4.8.3 Rhetoric of the phonics vs whole-word debate**

This debate has been raging for centuries and a consensus is unlikely to be met whilst there are divided opinions. Baumann et al. (1998) made the point that much of the disagreement is fuelled by the whims of governments. Politicians notice inflammatory articles in tabloid newspapers decrying the education system and raising concerns about falling standards within schools, and respond by altering policy without consulting those it directly involves.

Baumann et al. (1998) cited the International Reading Association's position statement:

Today, the role of phonics in reading and writing has become as much a political issue as it has an educational one. Teachers and schools have become the focus of unprecedented public scrutiny as the controversy over phonics is played out in the media, state legislatures, school districts and the home.

Baumann et al. (1998, p. 638)

Baumann et al. received 1,207 replies to their survey, and the overwhelming result was that the teachers felt that a balanced approach to reading instruction was needed (89%). The results of the survey stated that the most effective approaches to reading were phonics instruction, contextual analysis, structural analysis, learning sight words and developing a meaning vocabulary. This emphasises that reading, and the teaching of reading are complex activities that need a plethora of approaches to ensure that progress in reading is optimised.

Clem (1990) suggests that there should be compromise instead of fighting a war that has no end:

Those of us who find ourselves supportive of both camps recognize the areas of confusion, but we see no need to choose between the two. Rather, we see an exciting opportunity to work within the philosophy of 'whole learning' while at the same time addressing the important issues grounded in previous instructional research and experience.

Clem (1990, pp. 136-7)

The argument supporting non-phonics-based approaches to the teaching of beginning reading are often clouded because they misconstrue the conclusions of others to strengthen their arguments. Allington (2014), for example, states that the volume of books that are read can improve reading development. His suggestions are plausible and many would say show a degree of logic: the more you do, the better you become. However, in his conclusions he suggests that the findings of the Foorman et al. (2006) study support his suppositions. In fact he states:

Given the findings of Foorman and her colleagues (2006) that the sole aspect of reading lesson design that was related to reading achievement was the volume of reading done during the lessons...

Allington (2014, p. 26)

However, when Foorman et al. (2006) is studied it soon becomes apparent that this is not entirely what was concluded. Foorman et al. concentrated on teacher effectiveness and how the teaching of reading varies from class to class with respect to specific approaches including:

- The amount of time spent on phonics-based activities;
- Time spent on reading comprehension;
- Time spent on reading books compared to the technical aspects of reading.

They suggested that effective teachers implement a variety of approaches to teaching reading. In Grade 1 structured phonic approaches were used, as this

was seen to be the most effective way of teaching beginning reading, and it did have an impact. Foorman et al. (2006) concluded:

[H]ighly rated first-grade teachers positively impacted word attack outcomes by spending more time in phonemic awareness and alphabetic activities compared to non-instructional activities such as disciplining students, interrupting instruction with long transitions or being absent from the classroom. Moreover, effective teachers positively impacted letter–word outcomes by *not* engaging in grammar, mechanics, and spelling.

Foorman et al. (2006, p. 23)

Allington (2014) has obviously over-simplified Foorman et al.'s (2006) findings to strengthen his own conclusions, which places doubt upon his own hypotheses/theories. Foorman et al. (2006) showed that, despite Allington's claim that 'the sole aspect of reading lesson design that was related to reading achievement was the volume of reading done during the lessons', in Grade 1 phonological (= phonic) activities also had an impact.

#### **4.8.4 Stephen Krashen – an advocate for whole-word approaches**

Stephen Krashen feels that there is no need for intensive phonics instruction. He classifies phonics instruction as occurring in one of three ways (Krashen, 2004a):

1. Intensive systematic phonics – this assumes that we learn to read by learning the phonics rules first. This is achieved by 'sounding out or reading out loud' ("decoding to sound"). It is also reliant on the fact that the phonics instruction must be purposefully taught and in a systematic method;
2. Basic phonics – this describes that some rules of phonics are taught, but just the basics. In basic phonics reading is taught by understanding what is read on the page;
3. Zero phonics – this approach states that there is no need to be taught any phonics since everything will be learnt from reading books.

Krashen (2004a) then explained that, contrary to some who the phonics vs whole-word debate, those who support the whole-word stance do not support a zero-phonics approach, but a basic phonics approach.



Krashen (2004b) describes the phonics debate in the terms of differing hypotheses. The first hypothesis, the 'Comprehension Hypothesis', proposes that 'we acquire language and develop literacy when we understand messages by listening or by reading'. Or 'comprehensible input' is received and language gain happens easily, subconsciously and spontaneously. The 'Reading Hypothesis' is similar to the 'comprehension hypothesis,' and proposes that we learn to read by reading. By reading for meaning we acquire our vocabulary, spelling, writing and grammatical skills.

The opposing 'Skill Building Hypothesis' proposes that to learn language and develop literacy skills it is necessary to learn specific rules. These rules are consciously learnt, such as always add an 's' to pluralise written words, and are automatically used when we speak and write. The links between the Comprehension Hypothesis and the whole-word approach of teaching, and between the Skill Building hypothesis and a systematic phonics approach are clearly seen.

Krashen (2004b) went on to state that he felt there is a role for direct teaching of phonics, since it does have its place in learning spelling-sound correspondences. However, he stated that it has its limitations. His main criticism was that often phonic rules can be complex. He stated that some teachers had told him that they had to look up the rules before they teach them. He presumed that the teachers were inexperienced at teaching phonics or had not received sufficient training to be comfortable with teaching certain rules before checking their facts. Krashen viewed this as problematic, suggesting that, if a teacher has to revise the rules, the rules are too difficult for children to understand. But it may be that this reviewing of rules is not a limitation, but just demonstrates that teachers seek clarification so as to teach a specific rule precisely.

Krashen has frequently criticised the NRP's findings on reading and the effectiveness of systematic phonics. In his paper '*More Smoke and Mirrors: A Critique of the National Reading Panel Report on Fluency*' (Krashen, 2001) he criticised the NRP and their failure to acknowledge that sustained silent reading (SSR) has its place in reading instruction. He stated that, in short-term studies (less than one year), SSR students performed as well as or better than 'comparison students' – without stating what the comparison programmes were.

He came to this conclusion by looking at studies that were not accepted by Ehri et al. (2001) and tabulating them as shown in Table 4.8.

**Table 4.8: Krashen’s (2001) results for SSR studies**

| Duration           | Positive | No Difference | Negative |
|--------------------|----------|---------------|----------|
| Less than 7 months | 7        | 13            | 3        |
| 7 months to 1 year | 9        | 11            | 0        |
| More than 1 year   | 8        | 2             | 0        |
| Total              | 24       | 26            | 3        |

Quite why Krashen thought these figures impressive is unclear, since the balance seems to shift in favour of ‘free reading’ only for long-term programmes. He concluded this article by stating:

At worst, the impact of free reading appears to be the same as that of traditional instruction, and it is often better, especially when studies are continued for more than an academic year.

Krashen (2001, p. 122)

Shanahan (2004) and Stahl (2004) countered Krashen’s (2001) point of view regarding the NRP’s finding. First, they suggested that Krashen does not support formal reading instruction, and question how children can learn to read if there is no formal instruction. Secondly, they cast doubt on the effectiveness of SSR and suggest that children do not learn to read through SSR.

With reference to the first objection Krashen (2005) suggested that reading a book is often more educational than the instruction the children may receive. But the idea behind free reading is that children need access to good-quality books and are given the chance and encouraged to read – as though this were universally the case, and there were no deprived households in which these conditions are an impossibility. Krashen then maintained he had never said that SSR was necessarily better than instruction – it can be as good or sometimes better – thus retreating somewhat. But the results from SSR are not immediately seen. It can take over an academic year, as already stated.

Krashen (2005) answered the next objection by explaining that initially there may not be much reading going on because the children are trying to find books they enjoy. The concern about whether books are too easy or too hard is not a

concern. Krashen explained that a hard book can challenge a child, and that repeated reading and help from a teacher will allow the child to progress. If a book is classed as too easy it still provides a basis on which to reinforce skills that will help when it comes to reading other books. This observation is obvious. Re-reading a familiar text can improve a child's confidence and help with word recognition. However, taking the easy option continually will adversely affect reading progress, and may cause children to become disillusioned with reading. However, the flip side of the coin is that children should not read books that are too complex, since they cannot understand what they are reading and thus become frustrated. The key is to find a book that provides challenge without being over-bearing, and to ensure that the child understands what they are reading.

Krashen provided a logical argument that challenged the explicit use of systematic phonics. However, times have changed. In years gone by children learnt to read whilst sitting on a parent's knee listening to stories being read to them. Children would have story time at the end of each day and reading was an implicit part of the school day. But the huge changes in the National Curriculum have meant that the time spent on reading has been severely reduced. As a result, guided reading has been introduced into many schools to increase reading time, but in years gone by silent reading would have been a daily activity. The push to raise standards is the driving force behind the changes in the curriculum, and children are expected to progress quicker and reach milestones earlier, and phonics can help children read/decode quicker. However, with families now leading such busy lives, the skills learned in school are often not practiced at home as often reading together does not occur, and in some cases parents do not possess the necessary skills to help their children to read. Is Krashen correct in his supposition that spending more time reading in schools would be beneficial to children?

Krashen's beliefs, although trying to sound measured and based on empirical results, are insecurely supported by data, since he appears to have cherry-picked the evidence based on his dislike for the phonics approach.

## **4.9 What is the evidence that systematic NON-phonics-based instruction benefits children learning to read and write? How strong is it?**

Finding supporting experimental evidence for non-phonics-based approaches has been difficult. As can be seen from previous sections, much of the discussion based upon the importance of such approaches is based on narrative and opinion. Do systematic NON-phonics-based interventions help falling-behind readers to become more skilled?

This section is based entirely on RCTs, organised in three groups:

- 1) Those evaluating Reading Recovery, because this is the most used and investigated intervention;
- 2) Other teacher-taught schemes;
- 3) Schemes taught by persons other than teachers (mainly volunteers).

### **4.9.1 Reading Recovery**

As Slavin et al. (2011, p. 6) pointed out, '*Reading Recovery* is by far the most widely researched and widely used tutoring program in the world.' Originally devoid of phonics, it has added some attention to this in recent years, especially in the UK version, but most of the research on it was carried out before that. Also, many of the studies used only measures internal to Reading Recovery – those studies are ignored here, as are such internal measures reported in studies which also used independent tests. Even within independent tests only those measuring reading accuracy or comprehension are considered worth reporting here. Sadly, this reduces the amount of reliable evidence considerably – see Table 4.9. All these studies were conducted with Grade 1 pupils in the USA, except Center et al. (1995), which was carried out in Australia.

**Table 4.9: RCTs on Reading Recovery – effect sizes**

| Authors                 | Test  | Effect sizes  |      |
|-------------------------|---|---------------|------|
| Pinnell et al. (1989)   | Comprehensive Test of Basic Skills<br>Comprehension subtest | 0.45          |      |
| Iversen & Tunmer (1993) | Dolch Word Recognition Test                                 | (ns)          |      |
| Pinnell et al., (1994)  | Gates-MacGinitie Reading Test                               | 0.51          |      |
|                         | Woodcock Reading Mastery                                    | 0.49          |      |
| Center et al. (1995)    | (see below)   |               |      |
| Schwartz (2005)         | Degrees of Reading Power Test                               | 0.14 (ns)     |      |
| May et al. (2015)       | Iowa Test of Basic<br>Skills                                | Comprehension | 0.44 |
|                         |   | Word reading  | 0.45 |

All the results not marked 'ns' were significant at  $p < 0.05$  or better. The Iversen & Tunmer (1993) study included not only a standard Reading Recovery group and a control group, but also a modified Reading Recovery group whose programme included phonics; on the Dolch test there were no significant differences at post-test between any of the groups. This is surprising since this study is often touted as showing that adding phonics to Reading Recovery increases its effectiveness, a conclusion which could only be drawn from the internal Reading Recovery measures which are not considered reliable here.

Center et al's (1995) results are summarised in Table 4.10.

**Table 4.10: Effect sizes from Center et al. (1995)**

| Measure                             | Effect size |
|-------------------------------------|-------------|
| Burt word reading test              | 1.55        |
| Neale analysis of reading ability   | 2.03        |
| Passage reading test                | 1.57        |
| Waddington diagnostic spelling test | 1.48        |
| Cloze test                          | 0.42 ns     |
| Word attack skills test             | 1.88        |

Again, all the results not marked 'ns' were significant at  $p < 0.05$  or better, and all the other effect sizes were very large. Taking the results in both Tables together, it seems as though Reading Recovery was effective in most cases.

## 4.9.2 Other teacher-taught schemes

The Pinnell et al. (1994) study investigated not only Reading Recovery but the effects of three other interventions. The four interventions and a brief description of their approaches are shown in Table 4.11 (adapted from Table 1 in Pinnell et al., 1994, p. 13).

**Table 4.11: Overview of the interventions investigated by Pinnell et al. (1994)**

| Key elements                  | Reading Recovery  | Reading Success   | Direct Instruction Skills Plan  | Reading/Writing Group   | Control Group  |
|-------------------------------|---|---|---|---|--|
| n (pre-test)                  | 31  | 41  | 36  | 27  | 190  |
| n (post-test)                 | 31  | 40  | 29  | 24  | 165  |
| Setting for instruction       | 1:1   | 1:1   | 1:1   | Group   | Group  |
| Schedule                      | Daily   | Daily   | Daily   | Daily   | Daily  |
| Allotted time                 | 30 mins   | 30 mins   | 30 mins   | 30-45 mins  | 30-45mins  |
| Activities                    | Reading books<br>Writing messages<br>Word analysis in context<br>Writing fluency practice | Reading books<br>Writing messages<br>Writing fluency practice | Word practice<br>Games<br>Letter-sound practice<br>Reading to students<br>Pre-planned word analysis | Reading books<br>Writing individually<br>Writing group<br>Reading to students | Word practice<br>Worksheets<br>Word study<br>Letter-sound<br>Reading to students<br>Games<br>Computers |
| Stated instructional priority | Strategic processes<br>Fluency<br>Problem solving   | Strategic processes<br>Fluency                                | Mastery of skills<br>Know core of words<br>Decoding skills  | Strategic processes<br>Fluency skills   | Know core of words<br>Decoding skills  |

A total of 403 first grade students from 10 school districts were identified to participate in the study. There were four schools in each district. One school in each of the districts already had Reading Recovery. The other three schools were randomly assigned to one of the other interventions. Each school was asked to identify the 10 lowest-scoring students. Four of these children were randomly assigned to the intervention that was being run in their school. The rest of the students made up the randomised comparison group.

The mean effect sizes are shown in Table 4.12.

**Table 4.12: Post-test mean effect sizes for the four interventions investigated by Pinnell et al.**

|                  | Reading Recovery | Reading Success | Direct Instruction Skills Plan | Reading/Writing Group |
|------------------|------------------|-----------------|--------------------------------|-----------------------|
| Gates-MacGinitie | 0.51 *           | 0.27            | 0.14                           | 0.23                  |
| Woodcock         | 0.49 *           | 0.04            | 0.25                           | 0.23                  |

\* p<0.05; other results ns

The results show that Reading Recovery was the only intervention that provided statistically significant results on both tests, again showing that a non-phonics approach can be successful.

A recent British study which belongs here is was entitled *Literacy and Morphemes* (Worth et al., 2015), one of a significant number of RCTs commissioned and funded since 2012 by the Education Endowment Foundation. However, most of those focusing on literacy were concerned with interventions designed to boost it during the Year 6/Year 7 transition stage. *Literacy and Morphemes* was the only one targeted at key stage 1 children who were falling behind, and was designed to improve spelling and reading comprehension by teaching children about sentence structure and morphemes. A morpheme is defined as ‘The smallest unit of grammar, i.e. the smallest meaningful part of a word, not capable of being divided further’ (Brooks and Burton, 2016, p. 72), and can therefore be a word on its own such as ‘fair’, or an affix such as un- or -ly which cannot be used as a word on its own. The children in the intervention group made slightly less progress than the control group. Thus, there was no evidence that the *Literacy and Morphemes* programme had an impact on literacy ability.

### **4.9.3 Tutored schemes**

The third group of studies were tutor approaches identified from Slavin et al. (2010). The results of the studies, all conducted in the USA, are shown in Table 4.13

**Table 4.13: Results of tutored non-phonics-based schemes**

| Authors                     | Intervention                           | Age        | Mean effect size across tests |
|-----------------------------|--|------------|-------------------------------|
| Baker et al. (2000) *       | SMART<br>(Start making a reader today) | Grades 1-2 | 0.50                          |
| Meier & Invernizzi (2001)   | Book Buddies                           | Grade 1    | 0.89                          |
| Morrow-Howell et al. (2009) | Experience Corps                       | Grades 1-3 | 0.11                          |
| Pullen et al. (2004)        | Help from university students          | Grade 1    | 0.52                          |

\* This study was taught by 'para-educators' (teaching assistants), the rest by community volunteers.

These studies also support the notion that non-phonics-based approaches can benefit young falling-behind children.

Owing to the universal acceptance of phonics, due mainly to the Rose Report (2006) and the systematic reviews by NRP (Ehri, Nunes, Stahl, & Willows, 2001) in the USA and by Torgerson et al. (2006a) in the UK, which have emphasised the importance of phonics in the teaching of beginning reading, coupled with the fact that systematic phonics is now a compulsory part of the curriculum, the likelihood of a significant RCT on alternative approaches being funded is small. However, this does not explain why non-phonics-based interventions are now a rarity and why such approaches are not given a forum in which to show their worth, as the evidence assembled in this section suggests.

#### **4.10 Summary of relevant non-phonics-based experimental data**

The review of relevant research with respect to non-phonics-based approaches is patchy, but shows promise, and thus provides support for conducting a pilot/feasibility study on a non-phonics-based intervention for children who are falling behind in reading.

#### **4.11 Conclusion**

This chapter is split into two broad sections: the first is dedicated to the discussion about phonics approaches, and the second describes non-phonics-based approaches.

The phonics sections began with an explanation about phonics and an explanation of why this aspect of reading tuition was chosen for investigation.



Two bottom-up theories of reading, the simple view of reading and the dual-route theory, and their implications for the use of phonics in the teaching of reading, were then discussed. This was followed by a brief historical overview of phonics and how the use of phonics has been in and out of favour for years.

The following section described how phonics has become an increasingly important part of the National Curriculum, until it appears to be the sole government-approved method of teaching beginning reading. This discussion ended with a discussion about the controversy surrounding the introduction of the phonics check.

The final section of the phonics half of this chapter dealt with the experimental evidence that supports the use of phonics in the teaching of beginning reading.

The second half of this chapter began by discussing the arguments against the universal implementation of phonics. These arguments are largely based on rhetoric and personal opinion. However, the following section does provide experimental evidence supporting the use of non-phonics-based interventions

Evidence has been provided to support the investigation of a non-phonics-based intervention and its effect on struggling readers. Evidence shows that phonics does have a positive effect on beginning reading, but reading is not just decoding, which the phonics check tests and, as Wrigley (2017) states, has no direct effect on improving reading. A reader uses other skills to obtain meaning from text, including grammar cues and knowledge of the subject. This suggests that different approaches to teaching reading should be investigated.

## **5 Review of the effect of reading on spelling, and of research on writing and attitudes to reading**

### **5.1 Introduction**

The secondary and exploratory outcomes investigated in this pilot/feasibility study of whether a non-phonics-based intervention might significantly improve struggling children's literacy or attitudes to reading will be justified in this chapter.

Initially, the theoretical links between reading and spelling will be reviewed by comparing the most accepted spelling theories. A brief discussion of what writing is, is followed by a discussion of how reading development and writing development are inextricably linked.

This is followed by a review of how children's attitudes to reading have been studied and the importance of motivation in a child's learning.

### **5.2 Review of reading and spelling links**

The majority of research looking at the links between reading and spelling took place in the 1980s and 1990s. A number of different theories regarding reading and spelling acquisition have been suggested, and modified as knowledge has been acquired. Henderson (1981) discusses in depth the differing theories about how children learn how to spell. He traces research back to what he calls the cognitive revolution which he dates to the end of World War II. He initially discusses the works of Skinner and Chomsky. He compares the behaviourist views of Skinner with the more mathematical approaches of Chomsky, resulting in the Chomskian view of transformational-generative grammar to describe how spoken language is acquired. Henderson (1981) goes on to discuss spelling theories.

### **5.3 Spelling theories**

The ease of spelling in any language is largely dependent upon the orthography of the language being learnt. In the case of languages such as Finnish which have a regular orthography of one letter (grapheme) per sound (phoneme), spelling is straightforward. However, in the case of English there is not a single sound–single letter relationship. For example, /s/ can be spelt <s> as in 'sit' or <ss> as in 'pass' or <c> as in 'century', and each of the vowel phonemes can be

represented by a number of different graphemes, as shown in the newest version of the National Curriculum (DfE, 2013a, p. 52).

This added level of complexity means that when a child is learning to spell in English they need to be able to identify the correct corresponding grapheme for each phoneme in context, making spelling more difficult to learn. In addition to regular words, there are 'tricky' words that do not follow phonic spelling rules, such as '*said*'. It is paramount that, as children learn to spell, they are made aware that there are two types of word, regular and tricky.

Henderson (1981) gave an overview of how children learn to spell and read. He highlighted specific characteristics/requirements that are needed by children before their literacy journey can begin. These characteristics include language skills, health and their innate intelligence; how important it is within their cultural setting to learn; the importance of written language and how it differs from everyday speech; the understanding of the written form of the word and how this written form evolves from illegible scribbles to understandable prose; whether children have the understanding of what a word is (children speak all the time but do they understand that a specific group of letters makes up a word, that each word is a separate entity); and finally whether children possess the auditory skills to learn to distinguish between phonemes.

Henderson (1980) described his own spelling theory:

- Preliterate prephonetic category – This category contains nursery-age children who do not possess any alphabetic knowledge, but do realise that writing exists. At this level, the children may be able to form some letters and possibly be able to spell their name. If these children are asked to write something down they will scribble, or form symbols that resemble letters or numbers. At this level, the children will write from left to right or right to left, or even both on alternate lines;
- Preliterate phonetic category – The children in this stage are generally found towards the end of kindergarten and the beginning of grade 1. At this level, because the children possess very little alphabetic knowledge, they are aware that letters form part of a word, but only know a few words by sight;

- Letter-name strategy – Children reach this level towards the middle of Grade 1 or at the point where the children understand the concept of the word and have acquired a functional sight vocabulary. At this point, familiar words are spelled correctly. Inventive spellings also occur at this stage, but follow a particular pattern: if a phoneme is matched to a letter name, that letter appears in the phoneme’s place in the word; if there are specific phoneme-grapheme matches then the correct letter will appear in the correct position within a word; if the phoneme does not have a matching letter name, as with the short vowel phonemes, the vowel which is “nearest” in the child’s judgment is substituted. In the case of syllabic consonants, where there is no clear vowel sound (or at most a faint schwa), such as the final / $(\text{ə})$ / in the word *TABLE* / $'\text{teɪb}(\text{ə})$ /, the consonant will be written on its own; finally in the case of preconsonantal nasals such as / $n$ / in *WENT* and / $m$ / in *BUMP* that consonant is omitted;
- Vowel transition – Children reach this level towards the end of Grade 1 or the beginning of Grade 2. At this level children identify the difference between what are conventionally called ‘short’ and ‘long’ vowels (the / $\text{æ}$ ,  $\text{e}$ ,  $\text{ɪ}$ ,  $\text{ɒ}$ ,  $\text{ʌ}$ ;  $\text{eɪ}$ ,  $\text{iː}$ ,  $\text{aɪ}$ ,  $\text{əʊ}$ ,  $\text{juː}$ / correspondences of the five ‘vowel’ letters). Short vowels are spelled correctly, and long vowel sounds start to move towards their correct, often digraphic, spellings.

Henderson (1980) went on to state that after this point it is difficult to identify further distinct stages. He identified progress by looking at the point where errors that children make, such as failing to double consonant letters and using the spelling of the wrong homophone, are no longer apparent. These developments occur throughout Grades 3 and 4 and beyond.

Invernizzi et al. (1994) studied the spelling skills/strategies of 6-, 8- and 14-year-olds, and noticed that the spelling errors they made were not individualistic, but followed trends that supported Henderson (1980). They theorised that the children started with being able to spell simple words where single letters corresponded to single phonemes. The children then progressed to patterns within words and acquired the ability to blend sounds, identify digraphs and the so-called ‘magic <e>’. Finally, children applied their knowledge of words to be

able to spell multisyllabic words. Invernizzi et al. (1994) went on to suggest that spelling instruction should be targeted at the stage that each child has reached in the theoretical spelling model.

Frith (1985) went further and developed a model to describe literacy acquisition, and therefore combined reading and spelling development. Before she developed her model, she needed to take into account certain criteria. A simple developmental model following linear progress and describing a gradual improvement over time does not account for the 'peaks and troughs' of development that occur during literacy acquisition. These 'peaks and troughs' occur for a number of reasons. As children become increasingly adept at decoding they read words with increasing speed until it appears to be instantaneous. This acceleration in reading speed can be accounted for by the increase of entries in the internal lexicon and the speed that words can be looked up. Children may then plateau, or their progress may slow, as they move from one concept to the next, more difficult concept. As a result of these considerations Frith (1985) developed a model consisting of three stages.

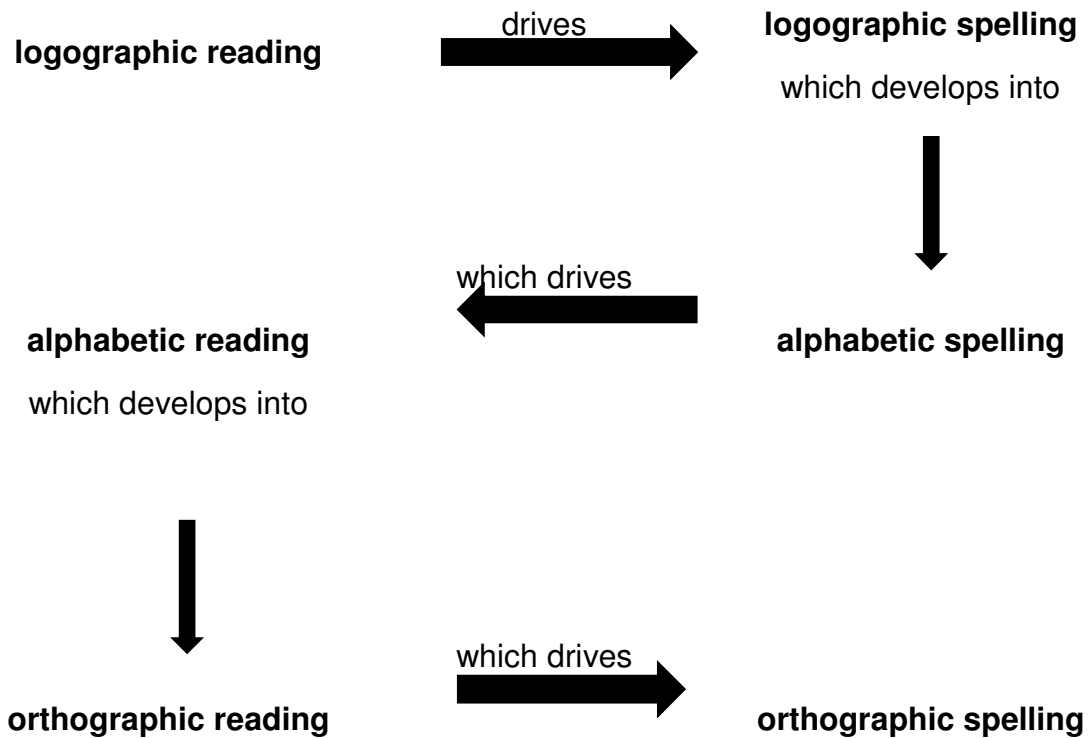
The first stage consists of *logographic skills*. These skills are those used to identify words by: visual familiarity such as the shape of the word (the number of descenders and ascenders), contextual clues, or the fact that it is a familiar logo such as *McDonalds*, *Coke*, *Nike*. At this stage, the order of the letters and phonological factors are not important. With regard to spelling/writing, letters are only just being formed and many may take the form of scribbles.

The second stage consists of *alphabetic skills*. These skills rely on the children's grapheme-phoneme knowledge and allow children to decode words one letter at a time. The order of the letters is vital and allows children to decode unfamiliar words and nonsense words (however, the pronunciation may be incorrect). For spelling, children use a single grapheme for a single phoneme enabling them to write simple CVC and CVCC words.

The third stage consists of *orthographic skills*. By this stage children can read words without having to decode at the single letter-to-phoneme level. Instead they use orthographic units such as morphemes and syllables. Spelling uses the same approach.

Frith (1985) went on to explain that, although the three-stage developmental model applies to both reading and spelling, they do not occur simultaneously (Figure 5.1).

**Figure 5.1: Diagram of Uta Frith's literacy stage theory**



Source: adapted from Brooks and Burton (2017, p. 23)

Frith suggested that, at the logographic stage, literacy learners have to become good logographic readers before they can apply this knowledge to spelling. The theory behind this is based on the number of exposures a child needs to receive before a printed word is committed to memory. For reading, a child identifies certain salient features of the word and needs several exposures to be able to commit the word to memory. For logographic spelling, the child needs to be able to form a mental picture of the word and then replicate that image on paper. This requires more exposures to the word to allow the word to be spelled from memory.

Frith (1985) proposed that the alphabetic skills are first obtained via spelling rather than reading. The main reason for this is that the alphabetic system is

designed for writing, not for reading. To be able to write initial words, only a small number of letters are needed, but to be able to read, all letters of the alphabet must be known, plus (in English) various combinations of them. Children must also be aware that the sounds the letters represent may vary depending upon the surrounding letters. Once these skills have been learnt, children then move on to learning more complex orthographic skills, such as identifying morphemes. Learning these skills speeds up a child's reading progress, but the ability to reproduce these as written words is not as efficient, since the child has not yet had time to commit these orthographic patterns to memory in such a way as to be able to reproduce the spelling automatically.

Frith (1985) further expanded her theory by saying that improving spelling at the simple phoneme-grapheme level will improve the reading of those words. In addition, improving the reading of words with complex correspondences improves the spelling of the words. Frith suggested that the teaching of reading and spelling should be targeted at the level which the child has progressed to, agreeing with Invernizzi et al. (1994). Agreeing with Chall (1967), Frith stated that phonological awareness and phonics tuition should occur in early grades (kindergarten and first grade), but after this more complex orthographic rules should be taught. There is no specific description about what she would classify as more complex orthographic rules.

Ehri (1997) discussed the ideas of Frith (1985) and expanded them. Initially Ehri (1997) identified the complexity of defining the word 'spelling', and suggests this word itself has ambiguity. Spelling can be used to describe the action of spelling a word, but it can also be used to describe how a word is spelt.

Ehri (1997) went on to explain that there are further ambiguities when it comes to what is actually meant by 'spelling'. Spelling could be used to describe two acts: the first is the ability to write a word out correctly, on paper; the second is the ability to identify whether words are spelled correctly or incorrectly as they are read.

Ehri (1997) went on to explain that words have specific spellings, and have two functions for the literacy beginner. The first is the ability to read spellings to work out how to pronounce the words, and subsequently work out their meanings;

within this skill is also the ability to identify if the words are correctly spelled. The second is the ability to write the correct spelling of the word needed. The development of these skills can be accounted for by constructing a conceptual framework, which takes into account the functions of words.

Within this framework are five areas:

- The familiarity of words – Ehri (1994) separated this section into familiar words and unfamiliar words, which she said is an important distinction for beginner readers. When children come across familiar words that have already been added into their mental Lexicon they can read and spell them. Unfamiliar words not in the lexicon are unknown. Over time the number of words in the lexicon increases as a child's alphabetic and world knowledge increases;
- Types of knowledge about spellings – as children's skills become more advanced, they acquire two types of knowledge. The first is knowledge about the alphabetic system. This includes phoneme-grapheme correspondences, blending skills, and the ability to group letters into syllables/morphemes and identify spelling patterns. The second type of knowledge is word-specific knowledge. Word-specific knowledge involves the spelling of words. It is the ability to identify the correct spelling for a word that could be spelled in a number of ways. If we take the word /'telɪfəʊn/, it could be spelled in a number of ways: 'telafone,' 'tellaphown' or 'telafoan'; word-specific knowledge allows the child to produce the correct spelling;
- Types of literacy acts involving words – as described above, reading words to determine their correct pronunciation and then meaning, leading to being able to spell a word and determine whether the word is spelled correctly or not;
- Ways to process words – three ways to 'process words'. The first is by memory; this is how familiar words are read, and these words, due to previous exposure, have been stored in the memory. Invention is a second method of processing words which occurs when children use plausible grapheme-phoneme correspondences to pronounce an acceptable word. The converse is true for spelling. The final method is analogy. Here



children identify features in new words that they have seen in familiar words and use these links to determine an acceptable pronunciation of a word. With respect to spelling, the same principle is used; knowing spelling rules allows children to apply their knowledge to new words;

- Levels of development – these are the stages that form the developmental framework in Frith (1985); however, there are a number of differences. These will be discussed further.

Ehri (1997) agreed with Frith on the first stage of the developmental framework, but calls it the pre-alphabetic stage. She emphasised that children remember words as pictures, which explains why children remember words found in logos. When children come across unfamiliar words they either guess or remain silent. With respect to writing/spelling, the scribbles the children produce can look like cursive writing; however, there is no definite letter formation.

The next stage is where Ehri (1997) differed from Frith (1985). Instead of having a single alphabetic stage, Ehri split it into the *partial alphabetic level* and the *full alphabetic level*.

Ehri (1997) explained that to progress to the *partial alphabetic* or *semiphonetic* level, children must learn the names and phonemes of the letters of the alphabet. Children read words by recall (memorising them) or by context cues. That is, they do not possess sufficient alphabetic knowledge to be able to decode the word. With regard to spelling, children at this level, as with reading, do not possess enough alphabetic knowledge. The children tend to concentrate explicitly on the consonant phonemes. If a person speaks with a rhotic accent, the word *BEAVER* could be spelled *BVR*. (In a non-rhotic accent such as RP, the final schwa vowel would not be 'r-coloured' and unlikely to be spelt with <r>.) Ehri (1997) described these as partial spellings, since they do not contain all the sounds that are present in the words. She explained that the missed phonemes occur because either the children find it difficult to identify the individual phonemes, or they do not know the correct phoneme-grapheme correspondences.

Progression to the *full alphabetic level* is achieved when children know the main phoneme-grapheme correspondences, especially how vowel sounds can be represented in a number of different ways. They also need to know how to

segment words into their individual phonemes. When unfamiliar words are met, children can use their understanding of grapheme-phoneme correspondences to decode them. When spelling words, children use their phoneme-grapheme knowledge to be able to spell regular words. At this stage, some words have been memorised, which allows the children to read unfamiliar words by analogy, that is, by identifying traits in unfamiliar words and linking them to known words, allowing the child to read or spell a word.

The final stage is called the *consolidated alphabetic level*. At this stage children have learnt that strings of letters can form patterns across a series of words. These strings of letters (syllables) can be found at the beginning or ends of words. For example, -IMP, -TION, -ABLE can be found at the ends of words, whilst CON-, EXP- and INT- are found at the beginning of words. By dealing with groups of letters, Bielby (1994) called this ability to group parts of words together 'chunking'. and stated that it makes the decoding of words occurs more quickly, as letter-by-letter decoding does not need to occur. In fact, Frith (1985) and Ehri (1997) suggested that, once syllabic reading occurs, the ability to decode at individual letter level becomes redundant, and those skills are lost over time. These skills are not lost, but stored in one's memory waiting to be reawakened when needed. The same principle occurs with spelling.

Ehri (1997) concluded that reading and spelling are ineradicably linked, so teaching one without teaching the other is not possible. Just teaching reading and hoping that the words collected in a child's memory can be miraculously converted into the ability to spell is misconstrued. Likewise, the teaching of spelling without teaching reading will not result in accomplished readers and spellers.

Analysing the stage theories proposed by Henderson, Frith and Ehri shows there are a number of similarities and differences.

Henderson tends to determine progress by linking it directly with age, which neither Ehri nor Frith does. Ehri and Frith allow children to develop at their own speed, rather than having to reach predetermined milestones that represent 'normal' progress. Henderson's approach tries to equate age with progress and although the 'norm' does exist there are exceptions to these rules. For children

who advance quicker or who take longer to progress, contingencies must be put in place.

Some theories suggest that at the early stages children identify words by their shape. As their alphabetic knowledge increases, this reliance decreases, and the children rely on their phonic knowledge. Frith (1985) refers to this as the logographic stage, and Ehri calls it the pre-alphabetic stage.

Ehri splits the next stage into two (pre-alphabetic stage and full alphabetic stage), where Frith calls it the alphabetic stage. Both authors describe similar skills at this stage. The final stage is Frith's orthographic stage and Ehri's full alphabetic stage. At the end of this stage children should have acquired their full literacy skills and be classed as literate.

The major difference between Ehri and Frith is Frith's idea that progress in spelling and progress in reading do not occur simultaneously. This is not reflected by Henderson or Ehri.

#### **5.4 What is writing?**

Before describing theories of how writing skills develop, it is necessary to explain the importance of writing. First, the links between spelling and writing are inseparable. Even if children can spell words orally, spelling is also, as Ehri (1997) explains, the ability to identify the correct written form. Although this can be present in a text, it is also expected that a child can physically write the word. As time progresses, a child's skill in writing will move from single word writing to whole sentence writing.

Secondly, an explanation is needed of the reasons why writing is necessary. Writing is seen as a symbolic representation of speech. Vygotsky stated that writing was just a manner by which to symbolise speech, where speech is the primary function and writing secondary. The hierarchy of importance arrives since, before writing skills can be developed, it is necessary to be able to communicate orally. Ultimately it is a more permanent way to communicate one's thoughts/feelings/experiences to others. Speech is a much freer way of expressing one's view and is only present momentarily, as the words are spoken and then lost to the air. Writing, on the other hand, is much more rigid with rules

to follow and structures to abide by. Writing can be viewed as a permanent record of an individual's thoughts/feelings.

Currently the ability to write is a lynchpin throughout society. Not only is writing used to assess progress and learning in an education setting, but rules and regulations that should be abided by are in a written format and provide guidance for how society should operate. Writing the rules down reduces the risk of confusion or misinterpretation which may occur if solely oral communication was used.

## **5.5 Emergent literacy**

There is a significant amount of research available that examines beginning reading and spelling. In the case of beginning writing research there is considerably less. Much of the writing research discusses the development of stylistic skills, grammar, or learning a second language, rather than how children take the step from reader to speller to writer. Before beginning to discuss beginning writing, a brief explanation is necessary about emergent literacy, which encompasses beginning writing.

From the 1920s up to the 1970s much of reading research concentrated on 'reading readiness'. Reading readiness is a term that had been used to describe when a child has achieved the requisite skills to be able to learn to read (Teale & Sulzby, 1987). There were two schools of thought. The first group thought that, as children matured, their 'neural ripeness' developed, meaning that they were more able to adapt to the skills needed to learn to read/write. The second group felt that life experiences and exposure to books would accelerate a child's reading readiness.

The maturationists, whose ideas were dominant from the 1930s up to the 1950s, believed that reading would develop with 'neural maturity'. Therefore, with respect to teaching practice if a child was not reading they waited until the child had 'matured'. This idea became gradually replaced by the theory that reading readiness is dependent upon experience and therefore could be taught, which is the basis of reading/writing tuition today.

Teale & Sulzby (1987) went on to state that a child's literacy journey does not start with formal instruction in a school setting, but at home, as they pretend to read and scribble on paper, and that literacy development of children occurs between birth and six years. In addition, they suggest that, rather than trying to separate reading from writing, the term emergent literacy should be used.

## **5.6 Beginning writing and reading**

Following on from the previous section, this section describes the current theories about writing. These all imply that literacy development is dependent on the quality and quantity of relevant experience.

According to Shanahan (2017), there are three theoretical models that are used to describe reading-writing relationships. The first model uses the idea of 'shared knowledge' and the varying cognitive skills that are needed for reading and writing, such as the visual and phonological systems; long and short-term memory; the ability to recognise the increased complexity as a reader/writer moves through individual words, to sentences and finally pieces of text. The 'shared knowledge' approach theorises that there is a single source of knowledge that contains information about both reading and writing. Depending on the task, the appropriate information is accessed.

Fitzgerald & Shanahan (2000) suggest that reading and writing share four 'common knowledges'. The first is described as domain or content knowledge. This knowledge is probably more important with respect to writing, as knowledge of a specific topic/ theme is needed so that writing can occur. Content knowledge is also needed for reading, as it helps with comprehension. Fitzgerald and Shanahan (2000) describe the second knowledge as metaknowledge and describe it as:

‘..several subcategories of knowledge, including knowing about the functions and purposes of reading and writing; knowing that readers and writers interact; monitoring one's meaning-making..’

Fitzgerald & Shanahan (2000, p. 40)

Metaknowledge is therefore thinking about and understanding what, why and how reading and writing are occurring.

The third form of knowledge is understanding features about text, such as the ordering of words; how punctuation is used within text and what to do when punctuation is seen; which direction to read/write the words and how any graphics enhance the text.

Finally, the fourth 'shared knowledge' is the ability to access information from what has been read/written so as to be able to predict, question and summarise what has been read/written.

The second, the socio-cognitive model, describes a relationship between the reader and the author. In this model the author writes in such a manner so as to produce specific feelings in the reader. Shanahan (2017) suggests that the relationship between the reader and writer is like a discussion which suggests interaction between the two. However the writer is the dictating force, as his/her words influence the reader's thoughts and feelings. The reader has no effect on the writer, therefore any relationship is unidirectional.

The third theory states that the skills of writing and reading are separate processes, but can be combined when needed to solve a problem or to achieve a goal. For example, writing and reading can be used as individual skills, but the skills can be combined, providing study skills that allow students to learn more.

Shanahan (2017) goes on to discuss several areas of research that try to link reading and writing development. The first area is that of cognitive relationships. He concludes that the links between reading and writing are bi-directional, in as much as development in reading can have a positive impact on a child's ability to spell and consequently write. However, he hastens to state that, although the link is bi-directional, it is not necessarily an equal sharing of knowledge. Reading is a directed activity where the reader follows what the author has written and applies their phonological analysis to decode the words on the page. When a child is presented with a blank piece of paper and asked to write about a circus, initially they have to draw upon the knowledge they have read or experienced and then express themselves. Although access to the same knowledge base is required the ability to assimilate the information uses different skills, emphasising the importance of teaching both reading and writing.

Socio-cognitive relationships are described as the relationships between the writer and the reader. Shanahan summarises this by saying that if the writer can visualise how to engage the reader then the writing is improved. This is a skill that develops over time, but highlights the idea that writing is communication from writer to reader across time and space. This also suggests that ultimately writers do not write for themselves but to engage the reader.

Shanahan (2017) concludes that research has shown that it is possible to teach the cognitive and linguistic skills required to achieve a good level of reading and writing. He suggests that children should be taught specific writing skills as well as reading skills. This means that children should be asked to think about authorship, and what they want their readers to gain from their writing.

## **5.7 The reader in the writer**

This seminal piece of work *The Reader in the Writer* (Barrs & Cork, 2001), although not concerned with beginning writing, explains in depth the links between reading and writing in Key Stage 2.

Barrs and Cork (2001) based their hypothesis on the understanding that reading and writing are inextricably linked.

The project was designed to investigate the hypothesis that the quality of writing a child produces is dependent upon the quality of their reading and their reading material. The project design ran over a single academic year where the project team would visit and observe:

- How literary texts were used in the classroom;
- How the teaching of writing techniques was linked to the literary texts being used;
- How the children's progress in reading was linked to their progress in writing.

Five schools from Greater London Authorities took part. A single Year 5 class from each of four schools and both Year 5/6 classes from the fifth school were chosen. The teachers were initially asked to identify three children in each class, so that the sample consisted of 8 girls and 10 boys. Of these children, seven were

bilingual. For the final report six children were case-studied (3 girls and 3 boys). All classes were given the same two literary texts to study to offer consistency, *The Green Children* by Kevin Crossley-Holland and *Fire, Bed and Bone* by Henrietta Branford.

The children's writing was analysed at three time-points for specific characteristics which had been identified as showing improvement and increased maturity in writing skills:

- Breaks in the time sequence
- Mental state verbs
- T-units – a T-unit is often used to analyse increasing maturity in writing and was first used by Hunt (1965, p. 20), who defined it as 'one main clause with all subordinate clauses attached to it'.

Children start writing in the first person and recount events as if it is them in the story. Barrs and Cork (2001) conclude that this first-person writing can be enhanced by reading a variety of texts. This allows the children to develop or try out different styles of writing. This experimentation with styles which mimic the style of the text is called an 'echo'. The children can then develop their first-person approach, so they write as if they are a character in a particular situation.

Barrs and Cork (2001) were hoping that they would be able to provide evidence to show whether the same experiences influenced the development of reading and writing equally. They concluded that the evidence supported the idea that the children's reading did develop at the same time as their writing, and that their writing was influenced by their developing reading skills. They based their conclusions on the comments made by the teachers in their completed questionnaires. Statements describing the children's willingness to discuss the characters' thoughts and feelings, and the style of writing in the text, demonstrated the children's awareness of the relationship between reading and writing.



## **5.8 Attitudes towards reading**

### **5.8.1 Introduction**

Attitudes towards reading is an exploratory outcome of this study. It is assumed that, if somebody enjoys or possesses a skill or talent they are adept at, they will have a positive view about it. This is the reason why the attitudes questionnaire was included in this study. It was hoped that, as the children became more proficient at reading, they would have a more positive attitude towards reading. This section also describes the attitudes towards reading surveys undertaken by National Foundation for Educational Research (NFER) and the National Literacy Trust (NLT).

### **5.8.2 Motivation**

Guthrie & Wigfield (2000) discussed the idea that attitude to reading is just part of a child's larger need to learn. They call this motivation to read. Why do children want to read? Guthrie & Wigfield (2000) identify five areas of motivation.

The first area is described as learning orientation, or the motivation to understand what has been read. The next is intrinsic motivation, which comes from within the reader: they read because they enjoy it. As such they will seek out books to read to fuel their curiosity. Thirdly, extrinsic motivation is an external reward provided to the reader for completing a reading task. The fourth concept is self-efficacy, which they describe as having confidence in themselves as a skilled reader. The final motivation is defined as social motivation – the ability to discuss the books that have been read.

A positive attitude towards reading allows children to have confidence in their abilities, which span not just English but all academic subjects, as it allows them to access all areas of the curriculum.

### **5.8.3 Attitudes towards reading surveys**

The attitudes toward reading assessments that will be discussed are all concerned with Key Stage 2 children, but do present an insight into reading attitudes trends.

Brooks et al.'s (1995) survey of eight-year-olds in England and Wales included a reading attitudes questionnaire adapted from that used in the Assessment Of Performance Unit monitoring surveys of 11-year-olds between 1979 and 1988. They found that most children had a positive attitude to reading (approximately 75% of the cohort), and that children who had out-of-school interests were more likely to have a positive attitude towards reading. They highlighted the fact that children who had negative attitudes (25% already at age 8) or found reading difficult needed to be encouraged and taken in hand so that their attainment could improve.

The introduction of the National Literacy Project (NLP) triggered a review by Sainsbury et al. (1998). They conducted a questionnaire survey asking children about their thoughts about reading. The questionnaire consisted of 16 questions and a simple tick response was needed. A mixture of positive and negative statements were given so that the children did not fall into just agreeing with everything. The aspects of reading that were investigated were:

- Enjoyment of reading
- Needing help with reading
- Whether the children liked reading other forms of texts such as magazines and comics.

The questionnaire was given to two groups of Key Stage 2 children; children in Year 3 and then at the end of Year 4; children in Year 5 and then again at Year 6.

The general findings of the report were that three-quarters of children enjoyed reading stories, and this percentage remained constant over the course of the investigation. Slight changes were noticed in certain aspects, especially within the section of the questionnaire examining children needing help with reading. As expected, the percentage of children needing help from adults decreased as the children became more proficient at reading.

The report went on to state that the percentage of children who read regularly remained constant (approximately 75%), with younger children reading every day, and older children reading most days.

More specific findings showed that girls were more likely than boys to enjoy reading, and older children enjoyed reading more than younger children.

Sainsbury and Schagen (2004) revisited the children's attitudes to reading between the years 1998 and 2003, using a slightly modified version of the questionnaire used in 1998. The questionnaires were completed by children in Year 4 and Year 6. They conclude that children in Year 4 and Year 6 find reading less difficult, but unfortunately do not find reading as enjoyable (Table 5.1).

**Table 5.1: Comparison of reading attitudes for the years 1998 and 2003 (percentages)**

|                              | 1998 |          |    | 2003 |          |    |
|------------------------------|------|----------|----|------|----------|----|
| <b>Year 4</b>                | Yes  | Not sure | No | Yes  | Not Sure | No |
| I like reading books         | 77   | 12       | 11 | 71   | 17       | 12 |
| I think reading is difficult | 17   | 27       | 56 | 15   | 19       | 65 |
| <b>Year 6</b>                |      |          |    |      |          |    |
| I like reading books         | 77   | 13       | 10 | 65   | 19       | 16 |
| I think reading is difficult | 14   | 25       | 62 | 10   | 17       | 63 |

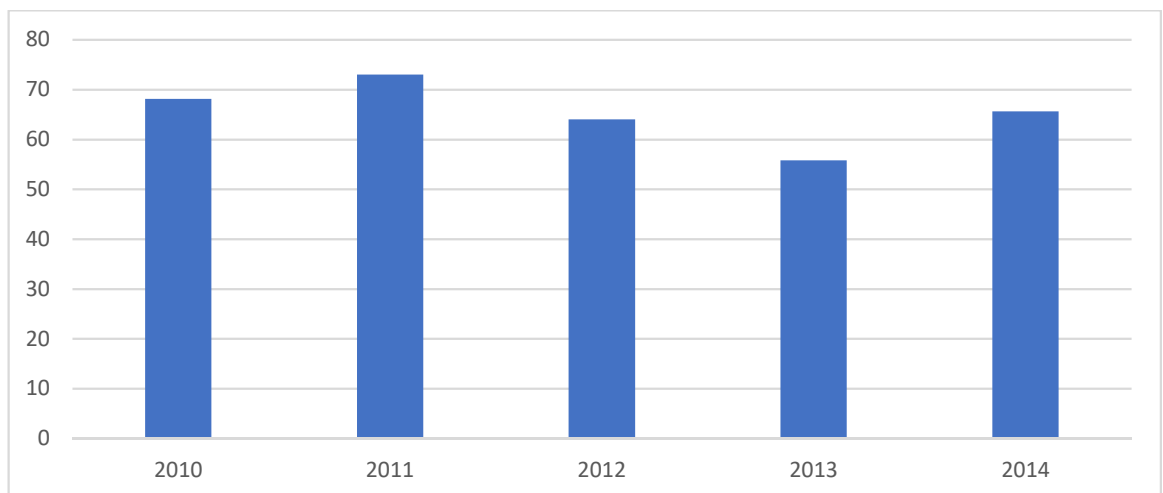
Sainsbury and Schagen (2004) suggested that this decrease in enjoyment could be linked to the way the way reading was taught in the National Literacy Strategy (NLS). They stated that there was no research that definitively supported this supposition. But they explained that the teacher-driven approach of the NLS could have a negative effect, as it does not allow the children to make their own decisions. Using the same questionnaire as they used in 2004, Sainsbury and Schagen (2008) again noted that children were finding reading easier, but were not enjoying reading as much.

Tymms and Merrell (2007) reviewed different studies investigating attitudes to reading, and summarised the results as showing that, generally, primary school children's attitudes were good. But as children get older their enjoyment of reading declined. They did state that with the advent of more television channels and films, there are other distractions that children have to withstand. It could therefore be assumed that, with the rise in the usage of video games, social media and the the internet, the reading of books is lower on a child's priority of things to do.

A more recent review of children’s attitudes to reading was undertaken by the National Literacy Trust (Clark, 2016) in a survey of 32,569 children (aged between eight and eighteen) attending 111 schools throughout the UK. The results obtained from the 2015 survey are compared with previous years’ results going back to 2005, allowing trends in reading to be seen.

Figure 5.2 shows the level of enjoyment gained from reading from 2010 to 2015.

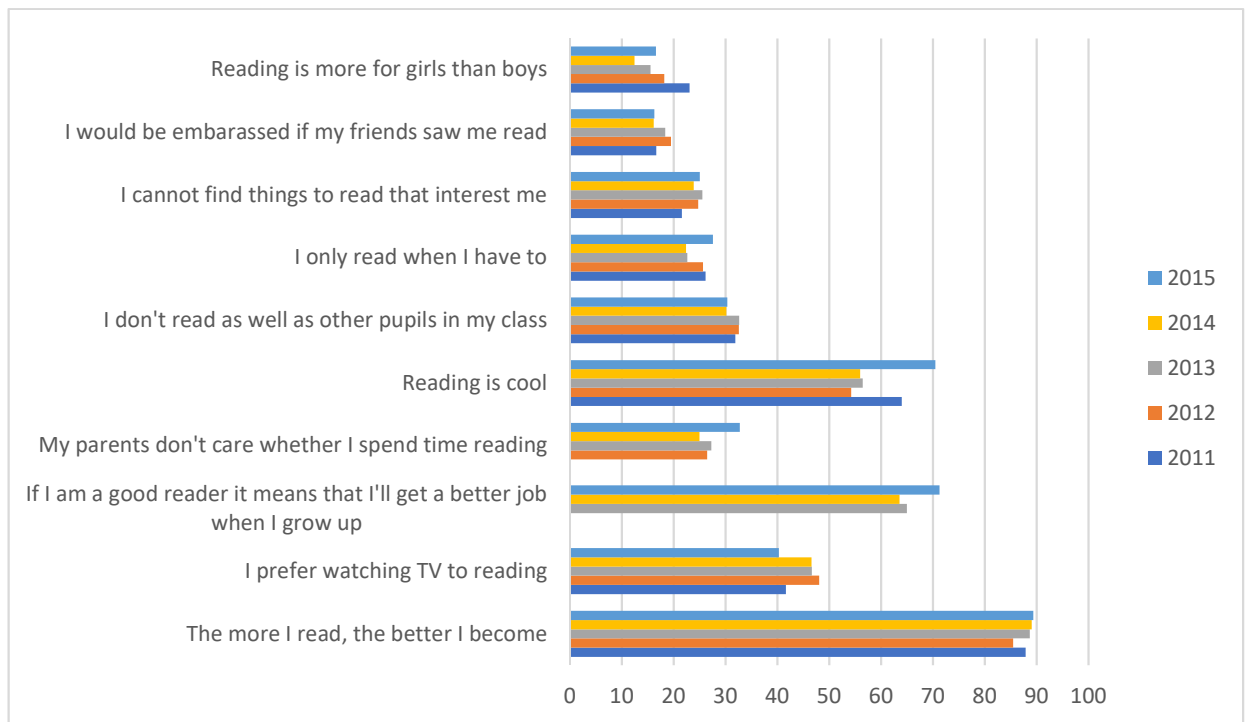
**Figure 5.2: Levels of reading enjoyment, KS2, 2010-2015**



Clark (2016, p. 18)

Clark (2016) went on to identify what motivates children to read. As can be seen from Figure 5.3 the main motivation for reading is the idea that the more efficient a child is at reading the better they become and that this will help them get a better job when they grow up.

**Figure 5.3: Percentages of children with attitudinal statements in 2010 to 2015**



Data obtained from Clark (2012, 2013, 2014, 2015, 2016)

## 5.9 Conclusion

The earlier sections in this chapter have described the relationships between reading and spelling, and reading and writing. With regard to reading and spelling it has been shown that, although these two skills develop at the same time, they do not progress at the same rate. Development in reading skills initially enhances spelling skills, but then, before reading can progress further, spelling skills need to develop. The links between beginning writing and reading have been less well researched, but it is suggested that the more knowledge a child receives, the easier it is for them to write. Increased reading skills allow a child more access to knowledge via books, and therefore their writing skills should be improved. Evidence has been provided to support the inclusion of a secondary outcome investigating spelling and an exploratory outcome investigating writing.

The final section discusses attitudes towards reading and the importance of motivation and consequently a child's self-esteem. From survey results it can be seen that children understand the importance of reading and, if a child is a confident reader, their attitude and interest in reading is more positive. However,

if a child is falling behind, their attitude towards reading is less positive, and their self-esteem could be affected. This supports the inclusion of a reading attitudes questionnaire.

## **6 Method**

### **6.1 Introduction**

As discussed in chapter 2, the methods used in this thesis are primarily quantitative. However, as with all educational research, which inherently deals with human nature/development, there is a qualitative element. The dilemmas of insider research were also reflected upon in chapter 2, as the researcher had several roles within the school, as school governor, parent and parent volunteer. This chapter lays out the process by which the school in which the research was done was chosen, followed by a brief description of the main characteristics of the school. The rest of the chapter concentrates on identification of an appropriate intervention, along with a description of the intervention; a review of experimental evidence on the intervention chosen; the processes used to obtain parental and participant consent; descriptions of the assessment tools and the subsequent statistical analyses planned; and finally, the timetable of key events.

### **6.2 The school**

Before the research proposal was submitted it was necessary to identify a school which would allow a research project to be undertaken on its premises. There were a number of limitations:

- The researcher was not working in a school;
- The researcher was a secondary school science teacher with no formal training in primary education. The researcher had worked as a volunteer in KS1 for four years.

For these reasons the researcher approached a school which was aware of her abilities, namely the school her children attended.

The school takes children from age four to eleven, and is in a small town in a semi-rural area. It is a larger than average primary school (297 students on roll in 2012).

The children come from a wide range of socioeconomic backgrounds, with a larger than expected number from the farming community. Although the proportion of children receiving free school meals is below the national average,

this number is increasing. The percentages of children from ethnic backgrounds or with English as an additional language are well below average; however, there has been a slight increase in the numbers of children with English as a second language, all Polish (there was just one such child in Year 2 at the time of the study, and there were no children from ethnic minorities in Year 1 or 2).

At the time of this study children entered reception class with below the national average attainment, but by the time they left in Year 6 the majority had attained the national average, and in recent years had surpassed the national averages in Maths, English and Science (Ofsted, 2010). This significant increase in attainment was due to a well-organised method of teaching in differentiated groups, with emphasis being put on booster groups, which concentrated on raising children from level 3 to level 4 in the Key Stage 2 (KS2) tests. Those children who were exceeding expected targets were given the opportunity to further excel in these areas.

More recent Ofsted reports (Ofsted, 2014, 2016) conducted after the completion of this research stated that the school's teaching of mathematics and English was not consistent enough throughout the school. The first of these inspections (Ofsted, 2014) highlighted that the teaching of reading was not consistent enough from reception through to Year 6 to ensure that children made expected progress. A recent Ofsted inspection (Ofsted, 2016) stated that there was improvement in reading tuition across the school, but this progress was insufficient.

### **6.2.1 Phonics teaching in the school**

At the time of the study, phonics was predominantly taught in key stage 1, three mornings a week, using the *Letters and Sounds* scheme (DfES, 2007). Depending on the availability of staff, phonics was taught as a whole class or in differentiated ability groups. Reception children started learning phonics 3 to 4 weeks into the autumn term; however, this was dependent upon the children's ability. *Letters and Sounds* was then taught throughout key stage 1. It must be noted that, following the end of the intervention reported here, phonics was being taught in Years 3 and 4 to those children who were falling behind.



## **6.2.2 Intervention teaching in the school**

The researcher was aware that the school policy, at the time this study was undertaken, was only to offer children interventions when they entered Year 3. From listening to children read in Years 1 and 2 it was apparent that, even with regular phonics teaching and reading at home, some children were still struggling to grasp the basic concepts of reading and would benefit from an early intervention. This included children who were falling behind because they were not fortunate enough to be reading regularly at home.

## **6.3 Identification of an appropriate intervention**

### **6.3.1 The criteria used to determine the most suitable intervention**

Identification of the non-phonics intervention was determined by the following criteria:

- I. The intervention needed to be suitable for use with Year 1 and Year 2 pupils (Key Stage 1 (KS1)).
- II. The intervention needed to be non-phonics-based.
- III. The intervention needed to be taught on a 1:1 basis.
- IV. The intervention needed to be designed for 'mainstream' children, not those with additional needs.
- V. The duration of the intervention needed to be no more than 10 weeks.
- VI. Ideally the intervention should help with both reading and spelling.
- VII. The intervention needed to be based on traditional reading books, not computer-based.
- VIII. The intervention needed to be inexpensive to implement.

#### **6.3.1.1 Why focus on Year 1 and Year 2 children?**

This first criterion was very important. It was known by the researcher that some children leave the security of KS1 with poor literacy skills. They enter KS2 to discover that the curriculum is more difficult, and the need for good literacy skills is vital. Assessments of the children show that they are falling behind, and intervention plans are put into place to help the children reach the expected national average. These children are often identified by the classroom teachers

from a young age, but many of the intervention schemes are not recommended for KS1 children. If children were given the opportunity to 'catch up' in KS1 they would be more prepared for KS2 and might not need further intervention.

#### **6.3.1.2 A non-phonics-based intervention**

This criterion was the most important. The whole basis of this research project was to identify whether an alternative method of boosting children's reading would be as effective as the systematic phonics that all children are taught according to the National Curriculum. The main purpose of this study was to determine whether, if a child has not grasped the concept of reading using a phonics approach, is using a phonics-based intervention a sensible way to proceed? Is it possible that providing a child with multi-faceted skills rather than a single skill has greater advantages?

#### **6.3.1.3 Helping children who are at risk of falling behind**

In the project school, children with additional learning needs were identified and Individual Education Plans (IEPs) were drawn up to ensure that they received support where and when needed. Those children who were just failing to achieve their expected targets often did not receive additional help and support. It was decided that these children should be targeted.

#### **6.3.1.4 Time was very important**

The time-frame was very, and therefore it was necessary that the intervention scheme could be completed within 10 weeks.

#### **6.3.1.5 An intervention that focused on reading and spelling**

Reading and spelling development are inextricably linked, as described in chapter 5. Therefore, it was decided that the secondary outcomes of this study would look at the effect of improved reading skills on the corresponding spelling skills. The effect of improved reading skills on writing skills was added after the FFT wave 3 intervention had been chosen and it was apparent that the intervention focused not just on reading but on writing as well.

### **6.3.1.6 Why not use a digital approach?**

The importance of digital literacies is increasing yearly, but at the current time most children are taught to read using books, and therefore an intervention using traditional printed books was chosen.

### **6.3.1.7 Finance**

Finance was a driving factor in the choice of intervention. This research was self-funded and therefore costs had to be carefully investigated.

## **6.3.2 Choosing the intervention**

Although other sources identifying literacy interventions were available, the researcher chose to use Brooks (2007; the 2013 and 2016 editions were not yet available when this research started). Admittedly this was a single source but the researcher was confident in Brooks's (2007) conclusions since only interventions that had been shown to be effective by quantitative methods had been included. Brooks (2007) provided succinct descriptions of each intervention which included main approaches, target age and length of intervention.

Primary interventions were initially reviewed for their target age. The apparent convention of providing interventions for children in Year 3 was highlighted during the review of intervention schemes available. In Brooks (2007) there were a total of 58 primary-age literacy interventions, of which 30 could be used for KS1. This shows that approximately 50% of the interventions available at the time this study started (December 2011) were unsuitable for use with KS1. Twenty interventions, including AcceleRead AcceleWrite, Catch Up Literacy, and Sound Discovery, had only been assessed for use with children in Year 2, and were therefore disregarded. The next group of interventions to be removed from the list of possibilities were computer/digital interventions such as A.R.R.O.W.

The next criterion was the duration of the intervention. For the purpose of this study the intervention needed to be completed within 10 weeks, with daily sessions of approximately 15-20 minutes/day. These timings would allow for 12 children to be part of the intervention group, and subsequently provide sufficient evidence to determine whether a full-scale study was practicable. Many of the interventions had much longer durations, for example Reading Recovery has a

duration of 12-20 weeks with 30 minutes/day, and Toe By Toe has a duration of up to 74 weeks for 60 minutes/day, and were therefore deemed unsuitable.

The only remaining intervention was FFT Wave 3. This has a duration of 10 weeks, and each child is seen for 15-20 minutes/day. It is a non-phonics-based approach, and is taught on a 1:1 basis. The intervention uses traditional books based on book banding, and its aims are to improve a child's reading, spelling and writing skills.

FFT Wave 3 has been the subject of very little research (section 6.4.4) and had not been the subject of an RCT. This provided an opportunity to produce an original piece of research.

#### **6.4 The intervention: Fischer Family Trust Wave 3 (FFT Wave 3)**

FFT Wave 3 was designed to help children who could not access the Early Literacy Support (ELS) intervention. ELS was part of the National Strategies intervention programme (DfES, 2001), and was a Wave 2 intervention programme delivered by teaching assistants to children in Year 1. The small group work was designed to help children meet or indeed exceed their target.

In 2007, ELS was updated (Department for Children Schools and Families, 2007) to take into account the new primary curriculum and the findings of the Rose Report (2006). The intervention was lengthened from 12 to 16 weeks. During this time pre-written lessons were delivered under the assumption that all children were given high-quality, systematic phonics teaching during the Early Years Foundation Stage. The rationale of the ELS was to improve a child's literacy skills by reinforcing phases 1-4 of *Letters and Sounds* and then teaching phase 5.

The six phases of *Letters and Sounds* (DfES, 2007) start in nursery/reception and end in Year 2, and provides children with the skills to be able to decode words by providing Wave 1 guidance and resources which are accessible to most children. Ultimately the aim is for children to be able to recognise words automatically whether they are decodable or 'tricky'. The phrases are described more fully in Appendix 1.

For some children Wave 1 approaches do not provide sufficient support and a Wave 2 intervention (small group work) may be offered. If a child requires more support an individualised Wave 3 intervention can be used. FFT Wave 3 is a Wave 3 intervention which is individualised to each child, who often has significant learning difficulties in a specific area, e.g. literacy or numeracy, or has SEN. There is no statement saying that Wave 3 interventions will accelerate a child's progress to eventually meet age-related expectations.

FFT Wave 3 is intended to help children who have been identified as requiring one-to-one intervention to help with their literacy skills. It comprises a reading day and a writing day, which are described below after the description of how the correct level book is chosen.

#### **6.4.1 Reading level determination**

Prior to the start of the intervention a number of assessments were done which involved observing how the children read: did they read from left to right, did they follow the text onto the next line, could they sound-out specific graphemes? This ensured that the children had good reading behaviours.

The reading scheme used in this study was Rigby Stars, which were graded according to the book bands described in *Book Bands for Guided Reading* (Bickler et al., 2003). At the first FFT Wave 3 meeting the child is asked to read a series of books, starting with the easiest. For each book read, a reading record is filled out (see Appendix 3). The reading record provides a consistent method for analysing how well a child reads a book. It logs each correct/incorrect word, substitutions/omissions, rereading words, and self-corrections, and there is a method of analysing what type of mistake has been made. The difficulty of the book is then determined by working out the percentage of correct words. If the percentage is above 95% the book is too easy, between 90% and 94% the book is classed as instructional, but below 90% the book is too difficult. Thus at the first meeting a number of books can be read before the correct band is determined.

#### **6.4.2 Reading Day**

Once the correct banded book has been determined, the intervention can start. The book is initially read and described by the adult, and difficult words are identified and explained. A discussion of the book takes place, and then the child

reads the book. The principle of FFT Wave 3 is not just to improve a child's reading ability but to encourage good reading behaviours. These include:

- Use of their finger to follow the words to ensure they are reading from left to right and that they are reading the word they are pointing to;
- Looking for cues in the preceding and proceeding text surrounding an unknown word so as to help decipher the word they are struggling with;
- Identifying the first letter of the word (visual prompt), so they know the starting sound of the word;
- Learning to listen to what they have read, and then if necessary self-correct. This self-correction can involve the child re-reading a sentence to ensure the text is understandable, and if necessary altering what they have previously read.

These key features were constantly used to help the child gain key strategies to use throughout their school day. If a child comes across a word they cannot read they are given 10 seconds try and decode it, if this is not achieved the adult pronounces the word and the child is asked to repeat the whole word. The child then reads the complete sentence again, until the sentence is read fluently. The reading section takes approximately 5 minutes.

The next section involves rapid letter work and word-making activities. From the reading activity the adult selects letters and digraphs that the child has struggled with. They are mixed up, and the child sorts them and says them as quickly as possible. This develops into making words, and writing new words down to reinforce the learning.

The final section is the learning of a new word. Foam letters are used to spell a word so the child can look at it. The letters are then mixed up, and the child rearranges the letters to spell the word. This activity is repeated several times.

On subsequent reading days, the book read on the previous reading day is re-read and then a new book is introduced, using the method described above.

### **6.4.3 Writing Day**

Alternating with the reading day is a writing day. The child re-reads the new book from the previous day, and this is checked for accuracy using a reading record. In addition the child's reading strategies are monitored. Various activities are

used to revise any incorrectly read words. Next the child dictates a sentence that the adult records on a lesson plan. The child then writes their sentence into their writing book (A4 plain paper) using a coloured felt tip. Any mistakes are covered with a white sticker and the child continues writing over the error. Plain paper is used rather than lined as it is that hoped the child will concentrate on spelling and composition rather than keeping the words on the line.

The final part of the writing day is the cut-up sentence activity. The child's sentence is written on to a strip of card by the adult and the child reads it out. As the adult cuts the sentence up word by word the child reads the sentence again. At the end of the sentence 'full stop' is said to emphasise the importance of punctuation. The cut-up words and the full stop are jumbled up and the child is asked to reconstruct the sentence, ensuring there is a gap after each word, and the full stop is placed at the end of the sentence without a space between it and the last word. The child then reads out the completed sentence.

The full explanation of FFT Wave 3 can be found in the manual (Fischer Family Trust, 2007). FFT Wave 3 should not be attempted without the full training. The researcher attended training and was given clearance by Jill Canning, author of FFT Wave 3, to deliver the intervention.

#### **6.4.4 Previous research on Fischer Family Trust Wave 3**

There had been little previous research on FFT Wave 3, just enough to show that it had been trialled prior to use in this pilot/feasibility study, and that it was being used in other schools. The only previous data were those of Canning (2004, 2009), which were both single-group pre-test/post-test studies carried out by the developer of the programme.

The pilot study (Canning, FFT wave 3: Report of findings from a 10-week pilot, 2004) involved 81 children initially, although this number decreased to 67, of whom 30 children were from Year 1, 25 from Year 2, and 12 from Year 3. The results are shown in Table 6.1.

**Table 6.1: Summary of results from FFT Wave 3 pilot study (Canning, FFT wave 3: Report of findings from a 10-week pilot, 2004)**

| Measure                   | Maximum | Pre-test |        | Post-test |       |
|---------------------------|---------|----------|--------|-----------|-------|
|                           |         | mean     | (s.d)  | mean      | (s.d) |
| Letter identification     | 52      | 42.1     | (8.8)  | 48.5      | (6.0) |
| Understanding about print | 17      | 12.9     | (3.2)  | 15.8      | (1.4) |
| High frequency words      | 45      | 21.6     | (13.8) | 37.9      | (8.4) |
| Writing vocabulary        | n/a     | 9.7      | (7.8)  | 30.7      | (7.0) |
| Dictation task            | 37      | 20.1     | (9.6)  | 30.7      | (6.9) |
| Book band level           | n/a     | 2.2      | (2.1)  | 7.9       | (5.2) |

The results were calculated as mean scores of all children across Years 1-3. All aspects showed gains in mean scores, some of them substantial in percentage terms, but no overall statistical tests of significance were reported.

The high-frequency words used for assessment were 45 High Frequency Reception Words. Scores in the initial assessment had a range of 0 to 45, but only 2 children scored over 43. In the final assessment 30 children read 43 or more words correctly.

The writing assessment appeared to vary from school to school as the report states:

“They were prompted to write two and three letter High Frequency Words, or words the TA felt they could write.”

Canning (2004, p. 4)

The number of words written at the initial phase ranged between 0 and 41, with only 6 children being able to write over 20 words. Post-intervention, only 6 children wrote fewer than 10 words, whilst the range was between 4 and 72.

As can be seen, there was an increase over the 10-week period, but whether this level of improvement would have exceeded that achieved by other children of similar ability in a classroom environment is unknown.

It is assumed that the dictation task was designed to help the children distinguish one or two different phonemes in CVC words. It was stated that, although there



was an overall gain and 10 children got full marks, 'several' children still found this task difficult.

The final assessment was the children's improvement in reading, which was measured by the number of book bands the children moved up. The average increase was two book bands in 10 weeks, which equates to approximately a seven-month increase in reading age, or a ratio gain of 2.8. The study goes on to state that although there were some large gains, generally for those children who were already reading at Reading Recovery Level 2 or above, a significant number were still on Book Band 3 (Reading Recovery Level 5), which is too low to allow a child access to the curriculum at Year 2.

Analyses of variance at the start of the study showed that the Year 1 children had significantly lower results than the Year 2 and Year 3 children on four out of the six assessments (print understanding, high-frequency words, writing vocabulary, and dictation). This would be as expected, as the children in Year 1 have had less schooling than those in Years 2 and 3. By the end of the intervention Year 1 had lower scores in writing vocabulary, dictation and book level. The report states that the Year 3 group made more progress during the intervention than in the previous two years, but this statement is not substantiated in any way. The Year 3 average gain in book levels was nine (or three book bands), resulting in a mean book level of 11, which is a reading age of approximately 6.0 years, significantly lower than the average chronological age of Year 3 children.

The author concluded that, although children in Year 3 made the most progress, their gains still left them considerably behind their peers, and they would not be able to comfortably access the Year 3 curriculum. In addition to this, FFT Wave 3 does not provide the necessary skills required for the Year 3 curriculum. It was recommended that children should start FFT Wave 3 in Year 1.

In Canning (2009) data on 255 children studied in 2008 were analysed. Whereas the first study was based on a ten-week programme, this study investigated a 15-week period where the children had 58 sessions, an average of 3.8 sessions per week, rather than the advised 5 sessions per week. Canning (2009) stated that one of the most difficult aspects of implementing FFT Wave 3 was the ability of schools to allocate sufficient resources to run the intervention fully.

Since there was no comparison group it was only possible to report mean scores, standard deviations and gains, and some of the data were expressed graphically. Canning (2009) did not report any statistical tests to determine significance. However, she did report the results shown in Table 6.2 and in more detail on book bands in Table 6.3.

**Table 6.2: Assessment results pre- and post-intervention in Canning (2009)**

| Assessments                         | Initial mean score | Initial range | Standard deviation | Final mean score | Final range | Standard deviation |
|-------------------------------------|--------------------|---------------|--------------------|------------------|-------------|--------------------|
| Understanding about print (max. 17) | 13.3               | 3-17          | 3.49               | 16.3             | 10-17       | 1.84               |
| Writing vocabulary                  | 17                 | 0-46          | 9.87               | 30.9             | 11-60       | 5.84               |
| Recording phonemes (max. 55)        | 37.2               | 0-55          | 14.23              | 51.1             | 15-55       | 6.24               |

Source: Canning (2009, p. 8)

She also reported results for letter identification and word reading data, but these have not been included. The letter identification results presented were the product of three different tests, each having a different maximum score (52, 54 or 57). The word reading result was again a product of two different assessments, the original assessment from FFT Wave 3, which gives a mark out of 45, and the *Letters and Sounds* assessment, which gives a score out of 56. The results reproduced in Table 6.2, are those which seemed reliable.

The 'understanding about print' and 'writing vocabulary' assessments appear to have been the same as those used in the earlier study. The gains in Understanding about print were virtually identical in the two studies, but the 2008 group's gain in Writing vocabulary was considerably greater.

**Table 6.3: Book band results from FFT Wave 3 main study (Canning, 2009)**

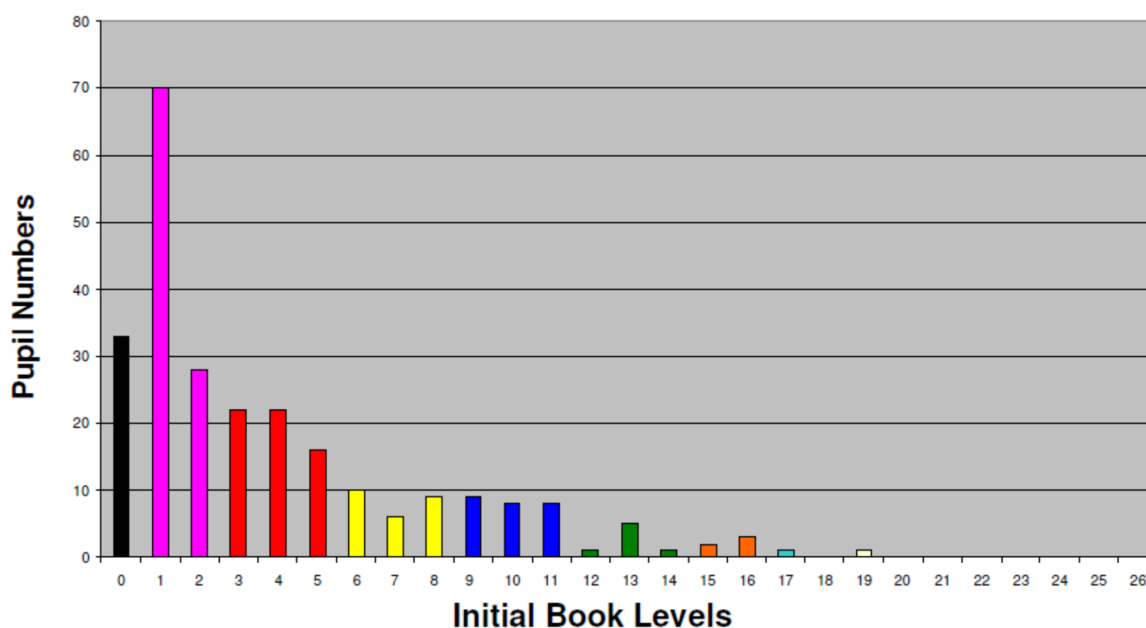
| Year Group | N * | Initial mean book level | (s.d.) | Final mean book level | (s.d.) | Gain in mean book level |
|------------|-----|-------------------------|--------|-----------------------|--------|-------------------------|
| 1          | 113 | 2.2                     | (2.56) | 11.95                 | (4.8)  | 9.7                     |
| 2          | 102 | 4.67                    | (4.14) | 15.22                 | (5.62) | 10.6                    |
| 3          | 12  | 8.9                     | (4.6)  | 19                    | (3.37) | 10.1                    |
| 4          | 18  | 7.25                    | (4.13) | 16.56                 | (5.09) | 9.25                    |
| 5          | 10  | 4.88                    | (6.38) | 14.42                 | (5.47) | 9.6                     |

Source: Canning (2009, p. 6)

\* The relevant Table in Canning (2009) gives only the percentages of the sample in each year group, plus the overall N of 255; the Ns above have been back-calculated from that information.

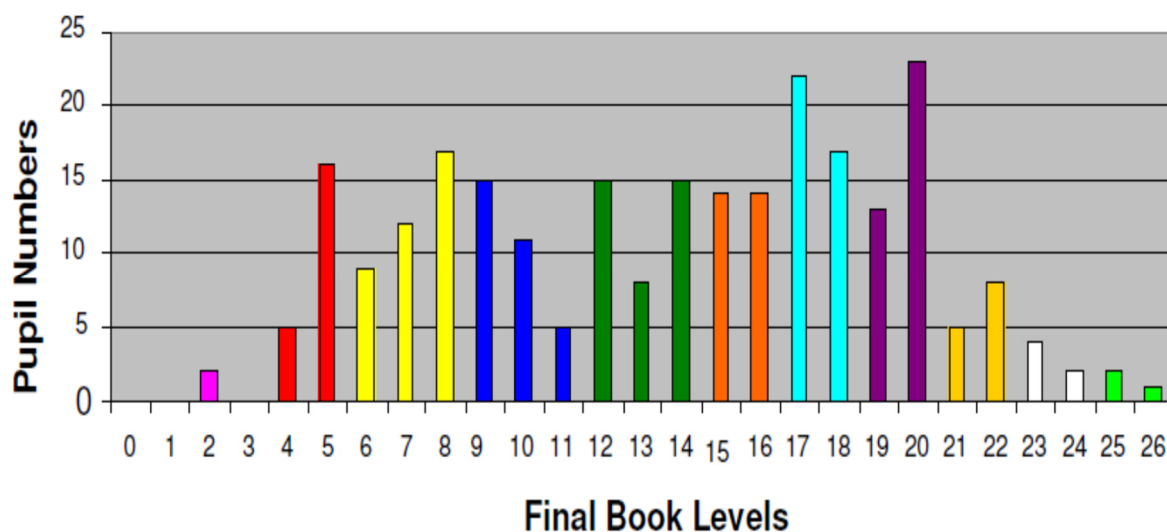
The data on book bands are worth exploring in more detail. Canning (2009) states that the results shown for Y3-Y5 cannot be considered robust since the numbers of participants were small. The graphs below (Figure 6.1 and Figure 6.2) show the pre- and post- book band levels across all year groups.

**Figure 6.1: Book band levels of children at the start of the Canning (2009) intervention**



Source: Canning (2009, p. 5)

**Figure 6.2: Book band levels at the end of the intervention**



Source: Canning (2009, p. 5)

As can be seen from a comparison of Figure 6.1 and Figure 6.2, there was a considerable increase in book band levels. Using Canning's data, it can be calculated that initially 75% of the 255 children were reading at book levels 1-5. After the intervention only 9% were reading at those levels.

Brooks (2013) re-analysed the book band data from both studies and converted them into reading ages, which allow the progress of each cohort to be more easily understood – see Table 6.4. This conversion was possible because Nelson Thornes Publishers have produced a table comparing reading ages with book bands in their PM Benchmark Kit. This table was produced because all children who are entered into Reading Recovery are also assessed on the British Ability Scales Word Reading Test. Nelson Thornes were able to use this information to correlate book bands with reading ages.

**Table 6.4: Pre- and post-test average Reading Recovery book bands and reading ages in years and months, gains in book bands, gains in reading accuracy in months of reading age, and ratio gains derived from Canning (2009)**

| Cohort |            | Pre | Post | Gain | ratio gain |
|--------|------------|-----|------|------|------------|
| 2004   | Book bands | 2.2 | 7.9  | 5.7  |            |
|        | r.a.       | 5:1 | 5:8  | 7    | 2.8        |
| 2008   | Book bands | 3.8 | 13.7 | 9.9  |            |
|        | r.a.       | 5:5 | 6:5  | 12   | 4.8        |

Source: Brooks (2013, p. 161)

The ratio gains shown in Table 6.4 are a way of determining whether a reading intervention has been effective. According to Brooks (2007 and later editions), a ratio gain of 1.0 signifies that a child's reading skills were developing at an expected/normal rate, but they will not be catching up with their peers. Brooks (2007, p. 289) went on to give an outline of the impact of different ratio gains:

- Ratio gains of less than 1.4 are of 'doubtful educational significance';
- Between 1.4 and 2.0 show 'modest impact';
- Between 2.0 and 3.0 show 'useful impact';
- Between 3.0 and 4.0 show 'substantial impact' and,
- Above 4.0 show 'remarkable impact'.

Canning (2004) shows a ratio gain of 2.8 which rates as a 'useful impact' according to Brooks (2007, p. 289), whereas Canning (2009) shows a markedly increased ratio gain of 4.8, which suggests that the intervention had a 'remarkable impact' on the reading skills of the children.

Brooks's (2013) review of FFT Wave 3 continues with the observation that, although progress had been made, the children were not yet functionally literate, especially those above Year 1. He summarises by saying that, although useful progress was made in 2004 and substantial progress in 2008, the children would still have needed support from good quality class teaching.

The results from the 2008 group reinforce the earlier findings of 2004 that the best time to introduce FFT Wave 3 is in Year 1. This should allow the children to access class teaching more effectively and earlier.

#### **6.4.5 Arranging to work with the Fischer Family Trust**

FFT Wave 3 having been decided on as the intervention to be investigated, Fischer Family Trust was contacted. Before it was possible to obtain the documentation to administer the intervention it was necessary to undergo two days' training. Training took place in a primary school in Middlesbrough and was followed by a 1:1 training day with the course director. The training explained what approaches to use; how to speak to the children and explain things; and, more importantly, emphasised the importance of having a consistent approach.

#### **6.5 Obtaining consent**

An appointment was made to visit the headteacher of the chosen school to discuss the proposed research project. The following points were discussed:

- Evidence of ethical approval. This had been obtained through the University of Sheffield before the school was approached (see Appendix 2);
- An outline of the research proposal;
- The reasons why the Fisher Family Trust Wave 3 literacy intervention had been chosen; and
- How many children could take part in the intervention.

Copies of the proposed correspondence to the parents were shown and the headteacher's advice was sought. The Headteacher was informed that full training in FFT Wave 3 had been given and Jill Canning's details were given, if he wished to contact her. He was enthusiastic and stated that the more help these children could get the better. However, he did stipulate that each of the class teachers involved and the Governing Body needed to be asked for their approval as well.

As mentioned previously, one of the main reasons for choosing Wave 3 was that it could benefit children as young as five, and therefore fulfilled part of the criteria stated in section 6.3.1. Within the school there was a Year 1 class, a mixed Year 1/2 class and a Year 2 class; this meant that three class teachers needed to be approached. The proposal was discussed with each of them in turn and the following issues explained:

- The impact that it would have on their teaching;
- How long each child would be out of class for;
- A timetable would be constructed that would ensure that the children did not miss the same lesson on consecutive days;
- The necessity for them to choose the children who would take part in the study (but not which group they would be in, treatment or control in Phase 1), but also that these children should not be on the SEN register, and after that point their input into conducting the research would be over.

The teachers expressed concern about obtaining parental consent. They were informed that the go-ahead of the study was dependent upon obtaining parental and child consent beforehand, which would be undertaken by the researcher. It was made clear that the researcher would work closely with them, and if any issues arose they would be immediately informed. This agreement was reciprocated. All the teachers agreed.

The next step was to obtain permission from the Governing Body (of which the researcher was a member at the time). An outline of the research was discussed. Various questions regarding the project were asked, but the main concern was that the children's educational progress would be affected by them being out of class. It was explained that the children would be out of class for no more than 20 minutes a day, and that it was hoped that the FFT Wave 3 programme would enhance the children's learning skills. This increase in learning skills, when transferred to the classroom, should then allow the children to flourish within the classroom environment. Ultimately the aim of this research was to improve the children's target levels. This prediction was backed up with the fact that many interventions are given on a one-to-one basis outside the main learning environment. For example, Hattie (2012) states an effect size of 0.59 for direct instruction, compared to 0.24 for classroom learning. In addition Hattie (2012) states that an effect size of 0.47 can be observed for early intervention. These figures were calculated using over 900 meta-analyses to determine how students improve at school. In addition, Ehri et al. (2001) calculated mean effect sizes for tutoring and class teaching as 1.09 and 0.37 respectively. Approval was given.

## **6.6 Arranging permission and selecting the participants**

### **6.6.1 Gaining permission from parents**

Letter 1 and an explanation of the research project (Appendix 4) were sent to all parents of Year 1 and Year 2 children (79 children in total). The letter was sent out on several occasions, and eventually handed to specific parents, since it became apparent that the children who might benefit most from the intervention were those whose parents were not returning the permission slips. Eventually, the majority of permission slips were returned. Out of the 79 children, 4 sets of parents did not give permission, 5 did not respond, and the remaining 70 gave permission. Included in Appendix 4 is the letter which was sent to parents whose children were selected to take part in the study.

### **6.6.2 Selection of participants**

After the permission slips had been returned the teachers were asked to identify the 12 Year 1 children and 12 Year 2 children who they felt would benefit from participating in the study. The researcher had calculated that she could give the intervention to 24 children (2 sets of 12, the intervention group in one term, and the control group in the following term).

The only exclusion criterion given to the teachers was that the children should not be on the SEN register. Wave 3 is designed for children who are not reaching the expected target level for their age, but not necessarily for children who have special educational needs. Also, it was part of the experimental design that SEN children were not used, as this would have increased the number of variables, made randomisation more difficult to achieve, and caused the results to be less reliable. If the capacity to recruit a larger participant group had been possible then SEN children could have been included. The teachers chose the children by using a combination of available attainment grades and personal understanding of the needs of each of the children in their classes.

### **6.6.3 Introducing the research to the children**

The children were spoken to initially at a class level, and then on an individual basis. The researcher introduced herself; this was fairly straightforward since she was a regular visitor to the school, which helped combat any initial anxiety the children might have felt. Several points were discussed:



- The researcher was at school just like they were and was learning about how children learn to read and that their help was needed,
- Some children would be seen daily
- All children would help three times during the year,
- The teachers were choosing which children were going to help daily.

At no point were the children coerced into helping and were only shown and asked to sign the consent sheet (Appendix 5) once they had agreed and understood what they were being asked to do. The children were all very receptive and the thought of helping an adult to learn they found quite intriguing.

### **6.7 Measuring the effectiveness of FFT Wave 3**

The outcome measures in this study were split into three groups as described in chapter 1. The primary outcome to be measured was the effectiveness of FFT Wave 3 on reading comprehension. The secondary outcomes investigated the effectiveness of FFT Wave 3 on oral word and sentence reading and spelling. The exploratory outcomes to be measured were book band levels, writing and attitudes to reading. Full statistical analysis would be undertaken only on the primary and secondary outcomes, which were assessed using standardised tests.

### **6.8 Power calculations**

Gorard (2007, p. 14) defined a power calculation as '*an estimate of the ability of the test to separate the effect size from random variation*', in other words, the ability to determine whether the sample size present in a study is large enough to differentiate whether an effect size is due to the intervention, or to chance. Therefore, it is advantageous if the minimum number of participants/subjects can be determined to give a reasonable probability of a statistically significant result. This number is found using a power calculation.

In an ideal world, this number would be calculated prior to the onset of research and determine the size of the sample. To calculate the number of subjects/participants certain criteria need to be decided upon. These would normally be a pre-determined probability (this is usually 80%) of finding an effect size of, for example, 0.5, with a confidence interval of 95%. For these specific values it has

been established that a minimum sample size of 64 subjects/participants per group is required in a 2-arm RCT, such as the main research approach used here, specifically with the primary and secondary outcomes. If the minimum important difference were lowered to 0.3, in line with the typical levels found in educational research, the sample size would be reduced, but still greater than was possible in the current study.

However, in educational research the sample size is often pre-determined by the availability of subjects and for practical reasons, as was the case in this study. It is possible to do a retrospective power calculation to calculate the probability that an effect would be detected.

There are a number of online packages available for power calculation. The one used for this study was developed by Schoenfeld (2012). The power calculation showed there was only an 80% chance of detecting a significant result if a (very large) effect size of 1.2 was observed. Therefore, it was highly unlikely that a statistically significant result would be detected, even if a smaller minimum important difference were postulated.

The main issue preventing this study from being larger was resources. The researcher was doing all the testing and intervention work, and there were not enough hours in the day to be able to have any more children in the study. The time that was made available was also cut considerably, making the process even more difficult. The reasons for this are described in section 6.12. Also, there was no possibility of enlisting help from others or recruiting other subjects.

## **6.9 Assignment to groups using minimisation**

Minimisation is a method of randomising participants by taking into account certain characteristics which might have an adverse effect on the outcome of the result of the study. Treasure and MacRae (1998) described the advantages of minimisation over standard randomisation, and in fact stated that *'If randomisation is the gold standard, minimisation may be the platinum standard.'*

The choice of minimisation as the method of randomisation was also determined by the number of participants that were available and manageable. Torgerson & Torgerson (2007) stated that minimisation is an ideal method for allocating small

numbers of participants. Minimisation identifies important covariates that may predict outcome at the start of the trial, and these covariates are then used to allocate participants. The use of a randomised design allowed the researcher to investigate whether a non-phonics-based approach to literacy improvement was beneficial by allowing her to look at causality rather than just correlation.

In addition, the method of minimisation provides a means of allocating participants into groups which should mitigate the risk of other factors, including unknown ones, influencing the data and of any factor other than the intervention causing the effect, thereby reducing the validity of the results.

The process of minimisation was carried out by the researcher's supervisor, Professor Greg Brooks, according to the method explained in Torgerson & Torgerson (2007), and in complete independence from the researcher, who provided him with the data shown in Appendix 6.

Brooks minimised by using gender, school year (1 vs 2), free school meals or not (x = receiving FSM), and group reading pre-test scores – he dichotomised these by above versus at or below the mean. He decided not to use the other test scores because (a) it would have greatly complicated the process; (b) this comprehension test would provide the gain score of main interest.

Brooks then allocated the first six children by simple randomisation (coin toss), then used a seventh coin toss to decide whether heads or tails should be intervention. He allocated children from number 7 onwards by inspecting the running totals of the four factors in cells W27 and X27 (see Appendix 6) – as indicated in Torgerson & Torgerson (2007), each time he allocated the next child to the group with the lower total. At two points the totals were equal, so again allocation was determined by tossing a coin. This produced 2 groups of 12. The totals came out almost as equal as they could be, and the column totals confirmed that the numbers across the 4 factors were accurately balanced. The allocations were decided, as also shown in Appendix 6.

Greg Brooks then sent the minimisation to Professor David Torgerson at the University of York for confirmation of correct method. Professor Torgerson confirmed that the minimisation had been done correctly.

## **6.10 The assessment instruments**

In total, six instruments were administered to each child in both groups at pre- and post-test: four standardised tests of reading/spelling, an informal writing assessment, and an attitudes questionnaire. The book band levels for each child for the start and the end of the intervention were obtained from the reading records that were taken as part of FFT Wave 3, as described in section 6.4.1.

### **6.10.1 Choice of standardised reading and spelling tests**

As explained above (section 6.7), these were to be used as measures of the primary outcome (reading comprehension) and the secondary outcomes (oral word reading, oral sentence reading, spelling). (For the measures used for the exploratory outcomes see below.)

There are many such tests available to purchase, and identifying the correct tests was difficult. Certain criteria needed to be met.

1. Each assessment pack needed to cover a wide range of attainment, especially lower attainment.
2. Each assessment needed to have more than one form. Each child would be assessed three times during the research, so it was important that the risk of the children recalling answers was minimised. The parallel forms also needed to be statistically equated (which was determined from the test manuals).
3. The reading test needed to assess not just the ability to decode words but also a child's comprehension skills.
4. The tests needed to give reading/spelling ages, standardised scores and National Curriculum Levels (the latter primarily for the class teachers' use).
5. Unfortunately, cost was a driving factor. These tests are expensive to purchase. Denis Vincent, long-term expert on such tests, was able to give advice and to confirm the suitability of the tests for the purpose of measuring progress due to an intervention. He was also able to confirm the statistical robustness of the tests.

There are two main providers of assessment tests, GL Assessment and Hodder Assessment. In general, the standard of GL Assessment is considered more rigorous. However, they were too expensive. The assessment tests provided by Hodder were therefore chosen. Because some of the tests would need to be administered on an individual basis, considering the number of children and the

time constraints, for some aspects it was decided to choose tests that could be administered to groups. Three tests were identified: Hodder Group Reading Tests (Vincent & Crumpler, 2011a), Hodder Oral Reading Tests (Vincent & Crumpler, 2011b), and the Diagnostic Spelling Test (Crumpler & McCarty, 2010).

A reading comprehension assessment to determine improving reading standards was the primary outcome measure in the inner study. This was to determine that children not only had the ability to decode words but understood what they had read. Therefore, it was important that the reading test contained a significant element of 'reading for meaning', which would mean the children would have to understand the text they had read to enable them to choose the correct word out of the four possibilities on offer. The Hodder Group Reading test starts with pictures and a choice of four words, and the children have to circle the correct answer. The test then gets increasingly difficult as short paragraphs are introduced with missing words, and finally a longer passage. Each new passage requires more understanding to be able to select the correct words.

An oral assessment was one of the secondary outcome measures in this research. This was to determine whether children could 'read' random words of increasing difficulty by using their phonics knowledge to decode without necessarily needing to understand what they were saying. The Hodder Oral Reading Test offered three ways of determining 'oral' reading age: reading speed, word reading and sentence reading. The reading speed test was not used as it provides a measure of reading speed which could have an adverse effect on reading accuracy. The oral word and sentence reading tests were included and had parallel forms, which was desirable to reduce the risk of progress due to recall. The oral word and sentence lists comprised of increasingly difficult words, some of which could be pronounced purely by decoding and some not. This allowed the children the chance to decode words that they would not previously have come across, or indeed understand. The Oral Word Reading and Oral Sentence Reading tests are given on an individual basis.

The Diagnostic Spelling Test provides pictures for the children to identify and then write down the correct word. The second section involves the children understanding a written passage. The adult reads the passage, and the children fill in the blanks on their copies with the provided words. This approach means

the children not only have to understand what they are listening to and reading, but also identify the correct spelling of the word if the answer is a homophone.

The Hodder Group Reading test and Diagnostic Spelling Test can both be administered on a whole-class basis, and their child-friendly design suited the purpose of this research. They both came with parallel forms. Each of the tests was suitable for testing children between 5 and 12 years.

All the tests provided conversions of the raw scores into reading/spelling ages as well as into standardised scores. The Hodder Oral Reading tests and Hodder Group Reading test also provided conversions into National Curriculum Levels, which the teachers found very helpful.

### **6.10.2 Statistical robustness of the standardised tests**

The Hodder Group Reading Test was standardised in 1999 and the reliability of HGRT 1 forms A and B was given as 0.96 for the internal reliability coefficient and 0.95 for the inter-form product moment correlation coefficient. These results indicate a high level of reliability between the two forms.

The Hodder Oral Reading Tests A and B were standardised in the 2005/06 school year. The product moment correlation for word reading ability and sentence reading ability was 0.826. This is high, but there is a possibility that pupils could perform better on one of the tests than on the other, which could be educationally relevant. The internal consistency reliability coefficient for both forms of the Word Reading tests is 0.96, and 0.97 for the Sentence Reading tests, which suggests high reliability – but it is stated that caution should be taken since it is assumed that the entire test is attempted.

The standardisation for the Diagnostic Spelling Test was undertaken between September and December 2005. The reliability of the Diagnostic Spelling Test was determined by calculating an internal consistency reliability coefficient which was 0.96. This value is an indication of the high reliability of the test.

### **6.10.3 Regression towards the mean**

Regression towards the mean is a statistical phenomenon that happens when repeated measurements are based on the same individuals over time. Due to

random error, those who achieve very high or very low scores on first testing are likely to have less extreme scores the next time they are measured (their scores become closer to the mean). For example, if children are streamed into ability groups the 'high achievers' could appear to have falling levels over time due to random error. The risk of results exhibiting regression towards the mean is reduced if the tests used have high reliability (those used here did – see previous section). In addition, randomisation reduces the problem, because there is an equal chance that the groups are going to be affected. Since they are both equally likely to show this phenomenon, the effects should cancel each other out.

#### **6.10.4 Administration and marking of the standardised tests (primary and secondary outcomes)**

All the tests were given according to the instructions provided in the manuals (Crumpler & McCarty, 2010; Vincent & Crumpler, 2011a, b). The pre-test group tests were administered by the class teachers; subsequent tests were administered by the researcher whilst the class teachers were present. The tests were marked by the researcher and copies of the results given to the class teachers, who used them as part of their own assessments.

It must be noted that the entire cohorts for Years 1 and 2 were assessed and those results were given to the class teachers.

During the testing phase there was an opportunity for experimenter bias, as the tests were marked by the researcher. The researcher confirms that there was no manipulation of any data. This would have compromised not only the research but also the integrity of the teachers as they were using the data to assess the children in their classes. In hindsight, the assessment papers should have been marked or moderated by a third party, to ensure that the marking was fair.

The word reading and sentence reading tests were administered on an individual basis by the researcher at pre- and post-test. These tests were difficult to administer and were subjective due to the need to decide whether a word had been pronounced correctly or incorrectly. For example, on the word 'candidate' some children would sound out and say each syllable and say 'can-did-ate' or 'can-die-date'. After pronouncing the individual syllables some children would blend the segments together to pronounce the word fluently and correctly, and

the mark was given. However, some children could not blend the segments and pronounced the word only as separate syllables; this was marked as an error.

#### **6.10.5 Scores**

Once the tests had been marked, the raw scores from the standardised tests were converted into reading or spelling ages (these results were passed on to the class teachers). For the final analyses, standard scores were calculated. This was done for all children, though only the results for the 24 children in the RCT were used in the analyses to be reported in the following chapter.

#### **6.10.6 Analyses of standardised test data**

The graphs presented in the following chapter were generated using Microsoft Excel, and the statistical analyses were carried out using IBM SPSS 21. The dependent t-tests were calculated using the paired samples t-test function, the independent t-tests using the independent samples t-test function, and the ANOVAs using the repeated measures function.

### **6.11 Exploratory outcomes**

#### **6.11.1 Book band levels**

The book band levels were determined from the reading records that were completed for each child in the intervention group. Reading records were completed at pre-test and throughout the intervention as the child became more proficient at reading. At the end of the intervention, the number of book band levels that each child had improved by was calculated, and the corresponding national curriculum levels and reading ages were determined using the Nelson Thorne Benchmark Kit. Dependent t-tests and ratio gains were calculated. The data are expressed as raw scores.

#### **6.11.2 Writing assessment**

The children's writing was assessed by comparing their first sentence with the final sentence they wrote as part of the intervention. The total number of words in each sentence was counted and compared. A dependent t-test was used to determine statistical significance. The language content of the sentences was also analysed to see if there was an increase in the complexity of the language used. The data are expressed as raw scores.



### 6.11.3 Attitudes questionnaire

The questionnaire (Appendix 7) was designed for every child to complete, to determine the children's attitude towards reading and books. The researcher hoped to detect trends and possible changes in the children's attitude towards reading if their reading skills improved. Unfortunately, there was no time to trial the questionnaire prior to the study and therefore the ease of use was overestimated. The children were asked to fill out the questionnaire individually, but some found this too difficult, as they found some of the questions ambiguous and did not understand how to complete the questionnaire (even after instruction). As a result, it was necessary to go through the questionnaire in small groups and, where necessary, on an individual basis. The questionnaire was given at pre-test and at the end of the intervention.

The questionnaire answers were collated and allocated a numerical value. Although statistical comparisons of the two groups' data had been planned, in the event none were conducted, for the reasons given in section 7.7.3.

### 6.12 Timetable

Once permission slips had been returned a timetable of events was devised – see Table 6.5.

**Table 6.5: Timetable of research study**

| Date               | Duration | Task  |
|--------------------|----------|---|
| December 2011      | 2 weeks  | Assess all children (Pre-test)              |
| January-March 2012 | 10 weeks | Administer FFT Wave 3 to intervention group |
| April 2012         | 1 week   | Post-test 1                                 |
| May-July 2012      | 10 weeks | Administer FFT Wave 3 to control group      |
| July 2012          | 1 week   | Post-test 2                                 |

There were significant time constraints, so it was imperative that the timings were strictly adhered to.

The proposal stated that each child in the experimental group (and eventually the control group) would be seen for 20 minutes/day. Initially the intervention ran according to this plan. All the teachers were given a two-week rolling timetable in which no child missed the same lesson on any consecutive day, nor the same part of a lesson. However, after the first 5 weeks the teachers expressed

concerns that the children were missing the morning numeracy and literacy lessons. The teachers felt that missing these vital lessons far outweighed any benefit these children may gain from FFT Wave 3. Their concerns were based on the fact that the children who missed the initial plenary returned to class and did not possess the new knowledge to partake in the remaining part of the lesson. The teachers felt that by the time they had explained what they had missed it was the end of the lesson, and the child had not even started the expected task. Thus, in the teacher's opinion FFT Wave 3 was having a detrimental effect on the children's learning.

The teachers believed that altering the timetable would not materially affect the research. In fact, one teacher said that it would not affect the effectiveness of the intervention because they never delivered an intervention scheme as it should be done. Obviously, a comment like that was rather surprising, but since the co-operation of the teachers was vital to the success of the research the timetable was altered. The new timetable allowed contact with the children during PE lessons (each class had one morning lesson/week) and all afternoons. Each session was reduced from 20 minutes to 13 minutes. In addition, it was no longer possible to see each child every day; the best fit was 7 times over a 2-week (10-day) period. The teachers were given a revised timetable, which was accepted.

### **6.13 Summary**

This chapter has described: the process by which the school was chosen, as well as a brief description of the school; the process by which the chosen intervention was selected, as well as a description of FFT Wave 3; and the methods by which the children were selected and randomised. In addition, there has been a full discussion of the assessment instruments used. Finally, there is an explanation of the types of data that were collected and analysed, and the timetable and how it had to be adjusted.

## **7 Results**

### **7.1 Introduction**

This chapter details the statistical results from the pilot/feasibility evaluation of FFT Wave 3, and is divided into seven sections. The first section describes the determination of equivalence of the groups at pre-test for the primary and secondary outcomes, including establishing the suitability of the use of t-tests for that purpose. The second sections describes the statistical analyses undertaken. The third section describes the results of the primary outcome (reading comprehension). The fourth section details the results of the secondary outcomes (oral word reading, oral sentence reading, and spelling). The fifth section summarises the non-parametric analyses of the primary and secondary outcomes. The sixth section details the analyses of the exploratory outcomes (book band analyses, writing analysis and attitudes questionnaire). Finally, a summary is given, with implications for the pilot/feasibility study which enfold the attempted quantitative evaluation of FFT Wave 3.

### **7.2 Primary and secondary outcomes: equivalence of groups at pre-test**

As stated in section 6.10.6, it was envisaged that the main statistical analyses of the primary and secondary outcomes data from the RCT would be carried out using analysis of variance. However, for that approach to be valid (and for analysis of covariance to be unnecessary), the intervention and control groups' pre-test scores should not differ significantly. The researcher was aware that it can be assumed that if randomisation (or in this case minimisation) is done correctly there is no need to do pre-test statistical analysis, since randomisation should have ensured there were no statistical anomalies between the two groups at pre-test. However, it was decided to analyse at pre-test to check that there were no anomalies, and to provide evidence that the minimisation had been successfully undertaken.

It was further envisaged that, even though the minimisation procedure used to allocate the children randomly to groups should have sufficiently ensured the equivalence of their scores at pre-test, that would be checked using independent t-tests. But again, for that approach to be valid (and for manipulation of the pre-

test data to be unnecessary), two assumptions underlying t-tests should be met, namely that each group's scores separately are normally (or close to normally) distributed, and that the two groups' variances do not differ significantly. The appropriate checks of these assumptions are the Shapiro-Wilks test and Levene's test respectively, and the results are shown in Table 7.1. All the results were non-significant, and it could therefore be assumed that both assumptions were met.

**Table 7.1: Results of Shapiro-Wilks and Levene's tests of standardised pre-test data**

| Test                  | Shapiro-Wilks |          | Levene's |
|-----------------------|---------------|----------|----------|
| Group                 | Intervention  | Control  | Between  |
| Instrument            | <i>p</i>      | <i>p</i> | <i>p</i> |
| Group Reading         | 0.81          | 0.33     | 0.50     |
| Oral Word Reading     | 0.45          | 0.54     | 0.40     |
| Oral Sentence Reading | 0.81          | 0.20     | 0.97     |
| Diagnostic Spelling   | 0.74          | 0.27     | 0.77     |

The equivalence of the groups was then checked using two-tailed independent t-tests, two-tailed since it was not possible to pre-determine the direction of any difference. The relevant data are shown in Table 7.2.

**Table 7.2: Pre-test mean scores, by group (standard deviations in brackets)**

| Outcome               |    | Primary       | Secondary    |               |               |
|-----------------------|----|---------------|--------------|---------------|---------------|
| Group                 | N  | Group reading | Oral word    | Oral sentence | Spelling      |
| Intervention          | 12 | 82.17 (7.55)  | 79.50 (6.90) | 82.75 (7.55)  | 92.67 (15.96) |
| Control               | 12 | 81.00 (8.01)  | 75.92 (5.14) | 79.42 (7.06)  | 90.75 (13.12) |
| <i>t</i> (two-tailed) |    | 0.37          | 1.44         | 1.11          | 0.32          |
| <i>p</i>              |    | 0.72          | 0.16         | 0.28          | 0.76          |

The differences shown in Table 7.2 were all non-significant, showing that the minimisation procedure had achieved the desired equating of the groups.

It is also worth remarking that all the mean scores were well below the norm of 100 (though less so on the spelling test than on the others), thus showing that the children recruited for the study were indeed underperforming and suitable for an intervention intended to boost their literacy.

## **7.3 Description of statistical analyses for Phase 1 (RCT) and Phase 2 (follow-up) data**

### **7.3.1 Introduction**

For each standardised assessment tool the same series of statistical analyses was undertaken. More details of the tests used are shown in 7.3.2 and 7.3.3. The results for the primary outcome, reading comprehension, are followed by the results for the secondary outcomes of oral word reading, oral sentence reading and spelling. The (differently-calculated) results for the exploratory outcomes are then shown in the following order: book band analysis, writing analysis and attitudes to reading.

### **7.3.2 Phase 1 analysis**

Phase 1 data for each of the standardised tests are listed here. For each test a repeated measures analysis of variance and two effect sizes are reported.

Two-way mixed analyses of variance were conducted on the pre-test and post-test primary and secondary outcomes, in order to determine whether a significant time by group interaction would show that the intervention group had made significantly more progress than the control group. (The main effects of time and group would be of little interest, time because the intervention group at least would be expected to improve, group because the substantive hypothesis for each test was that the groups' mean scores would differ at post-test.)

Also, for each standardised test, two effect sizes were calculated. Coe (2002, online) defines effect sizes as 'simply a way of quantifying the size of the difference between two groups.'

Effect sizes are expressed in terms of units of standard deviations and can be determined in a number of ways. The two most often used are Cohen's *d* and Hedges' *g*; both involve dividing a difference in mean scores between the intervention and control groups by a standard deviation. Both are reported below, Cohen's *d* because it has been very widely used, and Hedges' *g* because this is now specified by the Education Endowment Foundation (2015).

Statisticians differ sharply over which mean scores and s.d. to use; those used here are the mean gain scores and a pooled value for the post-test s.d.'s. The resulting formula is shown below:

$$d = \frac{\text{mean gain of group 1} - \text{mean gain of group 2}}{s}$$

where  $s$  = the pooled post-test standard deviation, for which the formula used in calculating Cohen's  $d$  is usually:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}$$

(Hartung et al., 2008)

The formula for Hedges'  $g$  differs only in having  $n_1 + n_2 - 2$  as the denominator, which has the effect of reducing the overall values. However, as will become apparent, the differences in this case are trivial.

It should also be noted that confidence intervals on the effect sizes are not reported. This is partly because of the pilot/feasibility nature of the study, but mainly because the effect sizes themselves were either non-significant in any case, or misleading – see discussions below.

### 7.3.3 Phase 2 analysis

Here the two groups' data were analysed separately because comparisons between their scores were no longer relevant. ANOVAs and effect sizes were not conducted on the Phase 2 (post-test and follow-up) data because at post-test the groups were already different, one having received the intervention. Dependent  $t$ -tests were conducted comparing the post-test and follow-up scores for each group separately for each test. This was done to determine whether the post-test/follow-up differences were significant. The tests were performed using SPSS, which automatically does a 2-tailed test. In order to obtain a  $p$ -value for a one-tailed test the 2-tailed value was halved, on the assumption that the data would be normally distributed (IBM, 2008).

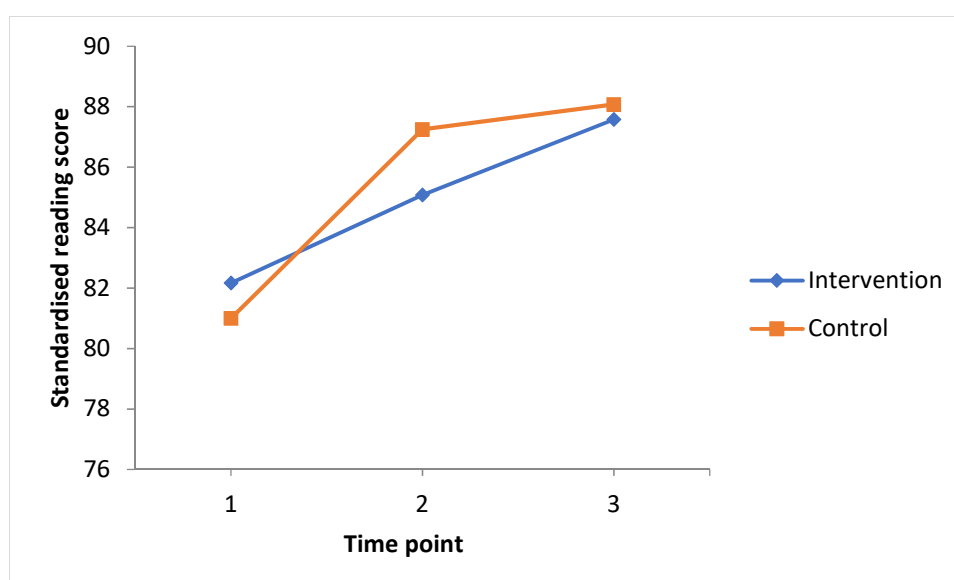
## 7.4 Primary outcome: reading comprehension

The means and standard deviations for the primary outcome at pre-, post- and follow-up were calculated and are shown in Table 7.3 and Figure 7.1.

**Table 7.3: Hodder Group Reading Test mean scores (and standard deviations), by group and assessment points**

| Group        | n  | Pre-test |        | Post-test |         | Follow-up |         |
|--------------|----|----------|--------|-----------|---------|-----------|---------|
|              |    | mean     | (s.d.) | mean      | (s.d.)  | Mean      | (s.d.)  |
| Intervention | 12 | 82.17    | (7.55) | 85.08     | (11.12) | 87.58     | (12.18) |
| Control      | 12 | 81.00    | (8.01) | 87.25     | (9.14)  | 88.08     | (6.86)  |

**Figure 7.1: Hodder Group Reading Test mean scores, by group and assessment points**



The data appear to show that, contrary to the hypothesis, the control group made more progress than the intervention group while the intervention group were receiving the intervention (Phase 1), but that the intervention group caught up during Phase 2 when they were not receiving the intervention. This is a confusing result, but might suggest that the intervention was having a negative effect on reading comprehension.

### 7.4.1 Phase 1 ANOVA

The results of the analysis of variance (Table 7.4) were significant with respect to time (as expected), but non-significant with respect to both group and (crucially) the time\*group interaction, contrary to the hypothesis, even though the groups'

scores ‘crossed over’ between pre- and post-test, such that the control group ‘overtook’ the intervention group. The effect sizes, Cohen’s  $d = -0.34$  and Hedges’  $g = -0.33$ , confirmed that the intervention group made less progress than the control group, and indeed showed a weak effect in favour of the control group.

**Table 7.4: Repeated measures ANOVA, Hodder Group Reading Test, Phase 1**

|            | F     | df | p    |
|------------|-------|----|------|
| Time       | 7.79  | 1  | 0.01 |
| group      | 0.023 | 1  | 0.88 |
| time*group | 1.031 | 1  | 0.32 |
| Error      |       | 22 |      |

#### 7.4.2 Phase 2 *t*-tests

Dependent *t*-tests were then done for each group separately, and the results are shown in Table 7.5.

**Table 7.5: Dependent *t*-test results, Hodder Group Reading test, Phase 2**

| Group              | Time      | Mean  | (s.d.)  | t    | P    |
|--------------------|-----------|-------|---------|------|------|
| Intervention group | Post-test | 85.08 | (11.12) | 1.35 | 0.10 |
|                    | Follow-up | 87.58 | (12.18) |      |      |
| Control Group      | Post-test | 87.25 | (9.14)  | 0.29 | 0.39 |
|                    | Follow-up | 88.08 | (6.86)  |      |      |

The results for both the intervention group and control group were non-significant, suggesting that the intervention had no effect on reading comprehension during the follow-up.

## 7.5 Secondary outcomes

### 7.5.1 Hodder Oral Word Reading

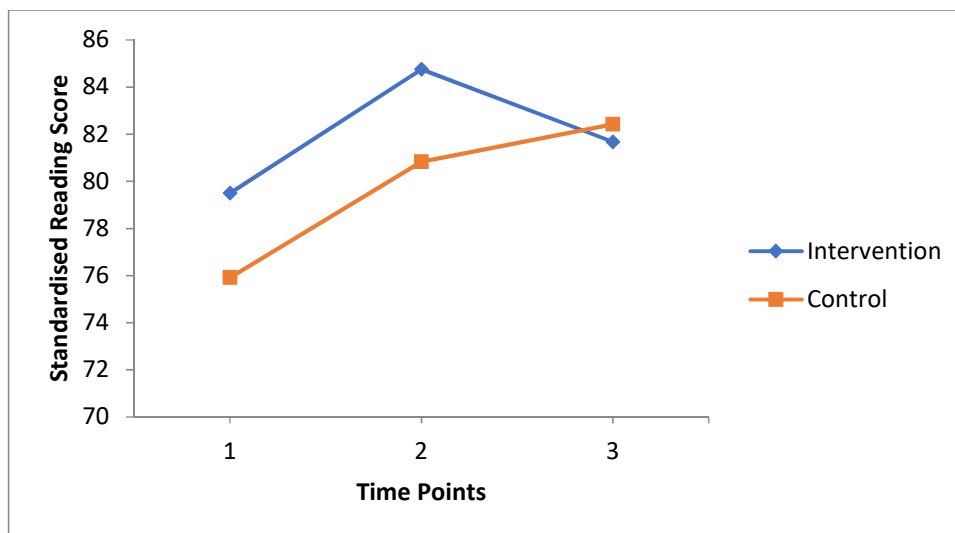
The means and standard deviations of the Hodder Oral Word Reading assessment were calculated for pre-, post- and follow-up time points, and are shown in Table 7.6 and Figure 7.2.



**Table 7.6: Hodder Oral Word Reading Test mean scores (and standard deviations), by group and assessment points**

| Group        | n  | Pre-test |        | Post-test |        | Follow-up |        |
|--------------|----|----------|--------|-----------|--------|-----------|--------|
|              |    | mean     | (s.d.) | mean      | (s.d.) | Mean      | (s.d.) |
| Intervention | 12 | 79.50    | (6.90) | 84.75     | (3.42) | 81.67     | (7.94) |
| Control      | 12 | 75.92    | (5.14) | 80.83     | (5.41) | 82.42     | (7.37) |

**Figure 7.2: Hodder Oral Word Reading Test mean scores, by group and assessment points**



Both the intervention group and the control group made progress during Phase 1. The gradients of the control and intervention groups' gains in Phase 1 suggest that the progress made was similar in both cases. However, during Phase 2 the intervention group actually regressed almost to their starting level. The control group continued to progress, but less rapidly than in Phase 1.

### 7.5.2 Phase 1 ANOVA

The results of the analysis of variance (Table 7.7) were significant with respect to time, but non-significant with respect to both group and the time\*group interaction, and the effect sizes, Cohen's  $d = 0.08$  and Hedges'  $g = 0.075$ , confirmed that the result was not significant.

**Table 7.7: Repeated measures ANOVA, Hodder Oral Word Reading Test, Phase 1**

|            | F     | df | p    |
|------------|-------|----|------|
| time       | 22.25 | 1  | 0.00 |
| group      | 3.86  | 1  | 0.62 |
| time*group | 0.02  | 1  | 0.88 |
| error      |       | 22 |      |

### 7.5.3 Phase 2 t-tests

Dependent *t*-tests were then done for each group separately, and the results are shown in Table 7.8.

**Table 7.8: Dependent *t*-test results, Hodder Oral Word Reading test, Phase 2**

| Group              | Time      | Mean  | (s.d.) | t    | p    |
|--------------------|-----------|-------|--------|------|------|
| Intervention group | Post-test | 84.75 | (3.42) | 1.80 | 0.05 |
|                    | Follow-up | 81.67 | (7.94) |      |      |
| Control Group      | Post-test | 80.83 | (5.41) | 0.90 | 0.19 |
|                    | Follow-up | 82.42 | (7.37) |      |      |

The results showed a significant – but negative (the mean score had declined) – result for the intervention group ( $p=0.05$ ) but not for the control group ( $p=0.19$ ).

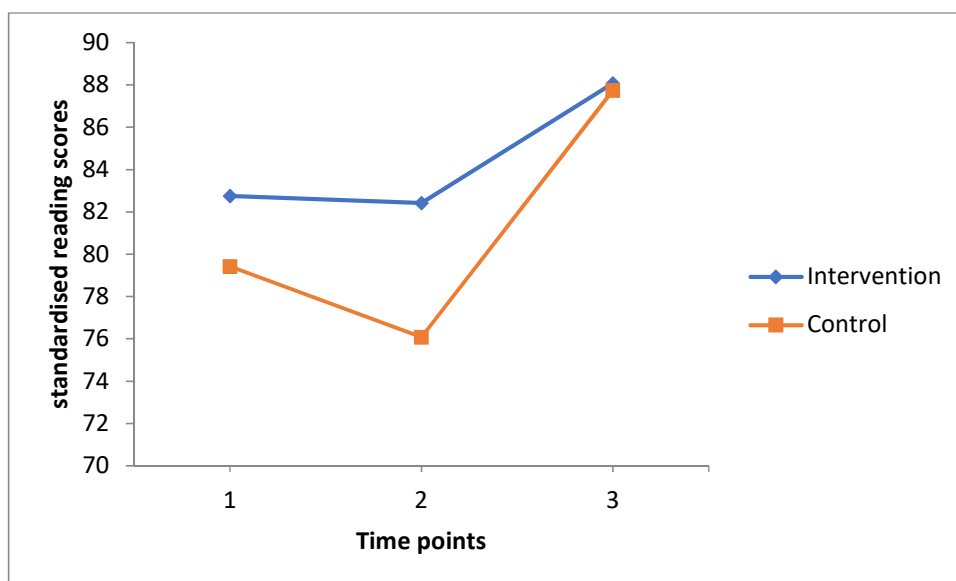
### 7.5.4 Hodder Oral Sentence Reading

The means and standard deviations were calculated for pre- and post-test and follow-up time points and are shown in Table 7.9 and Figure 7.3.

**Table 7.9: Hodder Oral Sentence Reading Test mean scores (and standard deviations), by group and assessment points**

| Group        | n  | Pre-test |        | Post-test |        | Follow-up |        |
|--------------|----|----------|--------|-----------|--------|-----------|--------|
|              |    | mean     | (s.d.) | mean      | (s.d.) | Mean      | (s.d.) |
| Intervention | 12 | 82.83    | (7.55) | 82.42     | (8.13) | 88.08     | (6.78) |
| Control      | 12 | 79.42    | (7.06) | 76.08     | (5.79) | 87.75     | (6.18) |

**Figure 7.3: Hodder Oral Sentence Reading Test mean scores, by group and assessment points**



The intervention group did not make any progress in Phase 1, whereas the control group’s performance decreased. During Phase 2 the intervention group made progress, but the control group’s progress was greater, and by the follow-up they had attained the same level as the intervention group. The control group had received the intervention by this point.

### 7.5.5 Phase 1 ANOVA

The results of the analysis of variance (Table 7.10) were significant with respect to time, but non-significant with respect to both group and the time\*group interaction. The effect sizes, Cohen’s  $d = 0.44$  and Hedges’  $g = 0.42$ , although larger than the others reported here, are misleading: the intervention group made no improvement, and its advantage over the control group was due solely to that group having fallen back considerably during the RCT.

**Table 7.10: Repeated measures ANOVA, Hodder Oral Sentence Reading Test, Phase 1**

|            | F    | df | p    |
|------------|------|----|------|
| time       | 2.58 | 1  | 0.12 |
| group      | 3.17 | 1  | 0.09 |
| time*group | 1.73 | 1  | 0.20 |
| error      |      | 22 |      |

### 7.5.6 Phase 2 *t*-tests

Dependent *t*-tests were then done for each group separately, and the results are shown in Table 7.11.

**Table 7.11: Dependent *t*-test results, Hodder Oral Sentence Reading test, Phase 2**

| Group              | Time      | Mean  | (s.d.) | t    | p    |
|--------------------|-----------|-------|--------|------|------|
| Intervention group | Post-test | 82.42 | (8.13) | 5.20 | 0.00 |
|                    | Follow-up | 88.08 | (6.78) |      |      |
| Control Group      | Post-test | 76.08 | (5.79) | 9.11 | 0.00 |
|                    | Follow-up | 87.75 | (6.18) |      |      |

The results for the intervention group and the control group were both significant. At this point the control group and intervention group had both received the intervention and the results suggests that the intervention continued to aid in oral sentence reading after the intervention has finished. In addition, the control group made significant improvement after the intervention.

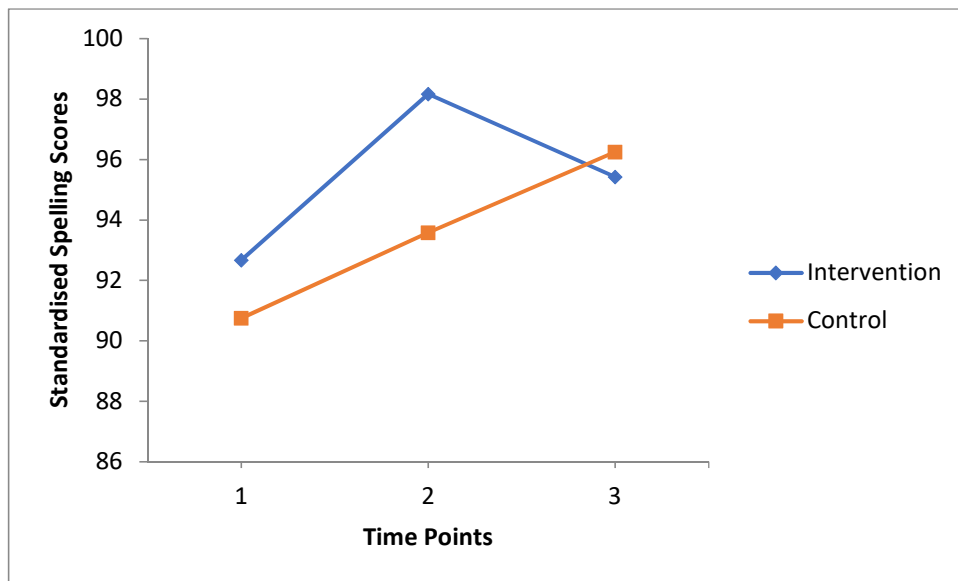
### 7.5.7 Diagnostic Spelling Test

The means and standard deviations at pre-, post- and follow-up were calculated and are shown in Table 7.12 and Figure 7.4.

**Table 7.12: Diagnostic Spelling Test mean scores (and standard deviations), by group and assessment points**

| Group        | n  | Pre-test |         | Post-test |         | Follow-up |         |
|--------------|----|----------|---------|-----------|---------|-----------|---------|
|              |    | mean     | (s.d.)  | mean      | (s.d.)  | Mean      | (s.d.)  |
| Intervention | 12 | 92.67    | (15.96) | 98.17     | (15.49) | 95.42     | (19.01) |
| Control      | 12 | 90.75    | (13.20) | 93.58     | (17.61) | 96.25     | (17.88) |

**Figure 7.4: Diagnostic Spelling Test mean scores, by group and assessment points**



The data appear to support the hypothesis: the intervention group made more progress than the control group whilst the intervention group were receiving the intervention (Phase 1), but the intervention group’s gain was not maintained, and their mean score dropped (just) below that of the control group once that group had received the intervention.

### 7.5.8 Phase 1 ANOVA

The results of the analysis of variance (Table 7.13) were, unexpectedly, non-significant with respect to time, and also non-significant with respect to both group and the time\*group interaction, contrary to the hypothesis. The effect sizes, Cohen’s  $d = 0.17$  and Hedges’  $g = 0.16$ , confirmed that the difference in progress was too small to be significant.

**Table 7.13: Repeated measures ANOVA, Diagnostic Spelling Test, Phase 1**

|            | F    | Df | Sig  |
|------------|------|----|------|
| Time       | 2.16 | 1  | 0.16 |
| Group      | 0.32 | 1  | 0.58 |
| time*group | 0.22 | 1  | 0.64 |
| Error      |      | 22 |      |

### 7.5.9 Phase 2 *t*-tests

Dependent *t*-tests were then done for each group separately, and the results are shown in Table 7.14.

**Table 7.14: Dependent *t*-test results, Diagnostic Spelling test, Phase 2**

| Group              | Time      | Mean  | (s.d.)  | t    | p    |
|--------------------|-----------|-------|---------|------|------|
| Intervention group | Post-test | 98.17 | (15.49) | 1.41 | 0.09 |
|                    | Follow-up | 95.42 | (19.02) |      |      |
| Control Group      | Post-test | 93.58 | (17.61) | 1.86 | 0.04 |
|                    | Follow-up | 96.25 | (17.88) |      |      |

The dependent *t*-test for the intervention group is not statistically significant, but that for the control group is significant. The results appear to show that over time the intervention's effect on spelling decreases. However, the control group who had just finished the intervention showed a significant improvement, suggesting that the intervention may have a short-term positive effect on spelling.

## 7.6 Primary and secondary outcomes: Non-parametric analysis

Mann-Whitney U tests were run on differences between groups on each test at each of the three occasions of testing, 12 such tests in all – see Appendix 9. All but one were non-significant, the exception being the Oral Sentence Reading Test at time 2 (post-test). But this merely confirmed the misleading result shown in Table 7.9 and Figure 7.3 the control group's performance was significantly poorer than the intervention group's at that point, because the control group had gone backwards. These tests contributed only confirmation of the parametric test results.

### 7.6.1 Summary of primary and secondary outcome results

The statistical result for the primary outcome, the Hodder reading comprehension test, was negative because the control group made better progress than the intervention group. However, even though the effect size confirmed that, the effect was weak and non-significant. Phase 2 analysis supported this result, since *t*-test results for the intervention and control groups were both non-significant.

For the secondary outcome of oral word reading the results were non-significant. Further analysis at Phase 2 further supported this result. A significant but negative result was obtained for the intervention group (they fell back during this phase), and a non-significant positive trend was seen for the control group.

For the secondary outcome of oral sentence reading the results (both t-tests and effect sizes) appeared to show a significant result in favour of the intervention group, but this was misleading. It only came about due to the control group having lost considerable ground. At Phase 2 both groups made statistically significant progress.

Finally, the Phase 1 Diagnostic Spelling Test results were non-significant. In Phase 2 the result for the intervention group was not statistically significant. However, that for the control group was significant.

Across both Phases for all four primary and secondary outcomes the results were very confusing. There was no consistent pattern of progress for either group, or to the backslides. The intervention group did improve on three measures during Phase 1 – but not enough to outperform the control group significantly. The control improved on only two measures in Phase 2 when they were receiving the intervention. There is no obvious explanation for any of this.

## **7.7 Exploratory outcomes**

### **7.7.1 Book Band analysis**

The book band levels pre- and post-intervention were recorded for the intervention group (only) and converted into national curriculum levels and reading ages as previously described (section 6.4.4) – see Table 7.15. The raw data used to calculate the gains shown in the table are shown in full in Appendix 8.

**Table 7.15: Book band, national curriculum level and reading age results**

| Child        | Change in book band levels | Change in NC level (sub-levels) | Change in r.a. (months) |
|--------------|----------------------------|---------------------------------|-------------------------|
| 1            | 4                          | 4                               | 24                      |
| 4            | 2                          | 1                               |                         |
| 6            | 3                          | 1                               |                         |
| 7            | 2                          | 1                               |                         |
| 10           | 4                          | 3                               | 24                      |
| 12           | 2                          | 2                               | 12                      |
| 13           | 3                          | 1                               |                         |
| 14           | 2                          | 2                               | 12                      |
| 15           | 2                          | 1                               |                         |
| 17           | 2                          | 2                               | 12                      |
| 19           | 4                          | 2                               | 18                      |
| 22           | 2                          | 2                               | 12                      |
| <b>total</b> | <b>32</b>                  | <b>22</b>                       | <b>114</b>              |
| Mean         | 2.67                       | 1.83                            | 9.5 / 16.3              |
| (s.d.)       | (0.89)                     | (0.94)                          | (9.39 / 5.71)           |
| Ratio gain   | n/a                        | n/a                             | 3.8 / 6.5               |

Assuming that 9 book bands should be covered in 7 or 8 terms, children should progress through the bands at about 1-2 bands/term. Considering the intervention was 10 weeks, which is just short of a standard term of 12 weeks, all these children exceeded that expectation, and the mean number of levels the children improved by was 2.67 (s.d. = 0.89). The results obtained here are consistent with the findings of Canning (2004), who found that the average increase in book bands over the 10-week period of her first study was 2 bands.

With respect to NC levels it is expected that children should progress by 2 sub-levels per year. As can be seen all the children progressed by at least 1 sub-level in 10 weeks, which, if sustained, would lead to a gain of approximately 4 sub-levels in a school year. The mean increase in NC levels was 1.83 sub-levels (s.d. = 0.94).

Reading age data for five children (4, 6, 7, 13 and 15) are not shown. This is because the Nelson PM Benchmark kit does not give specific reading ages for book bands pink to green (levels 1-14); the authors state only that the reading ages of children in levels 1-14 are between 5.0 years and 6.5 years. They explain



that they cannot give specific reading ages because the ‘fine grading of the PM levelling makes it inappropriate to give a specific reading age.’ Even if these five children are attributed a reading age gain of zero, the mean gain of the 12 children was 9.5 months (s.d. = 9.39; RG = 3.8), or 16.3 months (s.d. = 5.71; RG = 6.5) if the children for whom reading ages could not be calculated are excluded.

So as to be able to perform a *t*-test on the NC level data, a numerical value was allocated to each level. NC level 1C became 1, 1B became 2 and so on. For those children who were working towards NC level 1C a starting value of zero was attributed. The results were then analysed using a dependent *t*-test; parallel analyses were done on the NC levels and reading ages. The results are shown in Table 7.16.

**Table 7.16: Dependent *t*-test results for book band, NC and reading age analyses**

| Test   | Time | Mean  | (s.d.)  | t    | p    |
|--|------|-------|---------|------|------|
| Change in book band levels   | Pre  | 3.75  | (2.8)   | 7.75 | 0.00 |
|  | Post | 6.58  | (2.75)  |      |      |
| Change in NC levels  | Pre  | 1.17  | (1.59)  | 6.13 | 0.00 |
|  | Post | 3.08  | (2.15)  |      |      |
| Change in reading age, entire group (months)                       | Pre  | 69.50 | (12.39) | 3.35 | 0.00 |
|  | Post | 79.50 | (18.47) |      |      |
| Change in reading age for those whose reading age changed (months) | Pre  | 76.29 | (12.35) | 6.22 | 0.00 |
|  | Post | 93.43 | (9.07)  |      |      |

The results in Table 7.16 show that all results, even the reading age gain of the entire group, were significant and appeared to support the substantive hypothesis. However, these outcomes were based only on the intervention group, and were contradicted by the more rigorous analyses of the standardised test data from both groups, to which more credence must be given.

### 7.7.2 Writing analysis

In addition to the reading day there was a writing day in the FFT Wave 3 intervention. The writing days concentrated on the formation of a sentence that was generally linked to the book that had been read, but on occasion could relate to something the child had done that day/weekend.

Table 7.17 below shows the sentences that the children in the intervention group wrote (none were collected from the control group) on the first writing day, and then approximately 10 weeks later on their final writing day.

**Table 7.17: The first and final sentences written by the intervention group**

| <b>Child no.</b> | <b>First Sentence</b>                         | <b>Final Sentence</b>  |
|------------------|---|--|
| 1                | It went to the dog.                           | Dad forgot his shoes.  |
| 4                | I saw a dinosaur.                             | Are you too tired? asked Ravi.                               |
| 6                | The dog came running up.                      | I think I am back home.                                      |
| 7                | It is nice.                                   | I forgot my shoes.   |
| 10               | It went to dad.                               | I want my ball back.   |
| 12               | It is a dinosaur.                             | Chloe couldn't catch the fly.                                |
| 13               | Mrs Clark gave Tom some ice cubes.            | The prince was scared but the princess wasn't scared.        |
| 14               | The girl really liked the purple hat.         | Six nice elves came to make a hundred pairs of shoes.        |
| 15               | Owl is going to surprise rabbit.              | You can see Monty in a plastic display box.                  |
| 17               | I am telling you off because you are naughty. | The shoes are shiny.   |
| 19               | It is funny.                                  | The elves are making red shoes.                              |
| 22               | Owl flew down to see rabbit.                  | I like this page because the prince is pushing the servants. |

Table 7.18 shows the number of words the children wrote in their first and final sentences and the difference between them.

**Table 7.18: Number of words written by each child in their first and final sentences**

| Child no. | Number of words |        | Change |
|-----------|-----------------|--------|--------|
|           | Start           | End    |        |
| 1         | 5               | 4      | -1     |
| 4         | 4               | 6      | 2      |
| 6         | 5               | 6      | 1      |
| 7         | 3               | 4      | 1      |
| 10        | 4               | 5      | 1      |
| 12        | 4               | 5      | 1      |
| 13        | 7               | 9      | 2      |
| 14        | 6               | 11     | 5      |
| 15        | 6               | 9      | 3      |
| 17        | 9               | 4      | -5     |
| 19        | 3               | 6      | 3      |
| 22        | 6               | 11     | 5      |
| Mean      | 5.17            | 6.67   | 1.5    |
| (s.d)     | (1.75)          | (2.64) | (2.68) |

Although the means show that there was an increase in the average sentence length between the beginning and the end of the intervention, a *t*-test was conducted to check this statistically, and showed that the average increase was significant ( $t=1.91$ ,  $p=0.04$ ). The 12 initial sentences contain 9 simple sentences (one main clause, no subordinate clauses). Sentences 17 and 22 are complex, containing one main clause and one subordinate clause; in the case of sentence 22 'to see rabbit' is the (non-finite) subordinate clause. The remaining sentence, sentence 6, could fall into the complex category if 'running up' is counted as a subordinate clause.

The final set contains 7 simple sentences; sentence 13 is a compound sentence (having 2 main clauses) and 4 sentences (4, 6, 14 and 22) are complex (counting 'to make a hundred pairs of shoes' as a subordinate clause).

In the first series of sentences there are no negatives, but in the second series sentences 12 and 13 both have negatives. This feature is particularly important since correct use of negatives is a more advanced linguistic skill as it presupposes a 'context of plausible denial'.

In the first set the main verbs of 10 of the sentences are simple, while sentences 15 and 17 show continuous/progressive verbs (*going* and *telling*). In the final group of sentences there is only one continuous/progressive main verb, but a continuous/progressive verb is present in the subordinate clause of sentence 22.

Looking at adjectives, in the first set of sentences there are 4 adjectives (*nice*, *purple*, *naughty* and *funny*). This number almost doubles to 7 in the final sentences (*tired*, *scared* (x2), *nice*, *plastic*, *shiny* and *red*). This indicates an increase in understanding the use of descriptive words.

Although few in number and short in length (as this is the requirement of the intervention) the two sets of sentences show a definite increase in complexity with respect to length and use of more complex vocabulary. Some of this improvement was no doubt due to maturation and intervening teaching and learning, but it seems unlikely that this was the complete explanation.

### **7.7.3 Summary of writing analysis**

It has been statistically shown, even with a small sample, that there was a significant increase in the amount of writing the children who did the FFT Wave 3 programme produced; and qualitative comparisons of their first and final sentences suggested an increase in complexity.

### **7.7.4 Attitudes towards reading**

The questionnaire data are shown in Table 7.19 and Table 7.20. The results are shown separately for each question for each group and for pre- and post-test. For reference purposes the questions have been numbered, in the tables below.

**Table 7.19: Intervention group questionnaire results**

| Question  | Pre-test       |                |                | Post-test |           |           |
|---|----------------|----------------|----------------|-----------|-----------|-----------|
|   | Happy          | Okay           | Sad            | Happy     | Okay      | Sad       |
| When I have stories read to me, I feel                        | 10             | 1              | 1              | 10        | 1         | 1         |
| When I look at books, I feel                                  | 6              | 5              | 1              | 9         | 2         | 1         |
| When I read my reading books, I feel                          | 8              | 2              | 2              | 5         | 3         | 4         |
| When I have to talk about what I have read, I feel            | 6              | 4              | 2              | 4         | 4         | 4         |
| When I help my friends with their reading, I feel             | 6 <sup>a</sup> | 1 <sup>a</sup> | 3 <sup>a</sup> | 9         | 2         | 1         |
| When I have to write, I feel                                  | 7              | 4              | 1              | 7         | 3         | 2         |
| When I do my spellings, I feel                                | 7              | 3              | 2              | 6         | 2         | 4         |
| I like books.   | 9              | 1              | 2              | 10        | 2         | 0         |
| <b>Total questions 1-10</b>                                   | <b>59</b>      | <b>21</b>      | <b>14</b>      | <b>60</b> | <b>19</b> | <b>17</b> |
| I like playing on the computer.                               | 12             | 0              | 0              | 10        | 2         | 0         |
| I easily follow the instructions on the computer screen.      | 4              | 3              | 5              | 8         | 3         | 1         |
| I find it easy to learn how to do new things on the computer. | 9              | 3              | 0              | 8         | 2         | 2         |
| I prefer reading things on a screen than in a book.           | 8              | 1              | 3              | 7         | 2         | 3         |
| <b>Total digital literacies (questions 11-14)</b>             | <b>33</b>      | <b>7</b>       | <b>8</b>       | <b>33</b> | <b>9</b>  | <b>6</b>  |

<sup>a</sup> one child said that they did not help their friends with reading

**Table 7.20: Control group questionnaire results**

| Question  | Pre-test       |                |                | Post-test |           |           |
|---|----------------|----------------|----------------|-----------|-----------|-----------|
|   | Happy          | Okay           | Sad            | Happy     | Okay      | Sad       |
| When I have stories read to me, I feel                        | 10             | 2              | 0              | 8         | 3         | 1         |
| When I look at books, I feel                                  | 5              | 5              | 2              | 8         | 1         | 3         |
| When I read my reading books, I feel                          | 8              | 2              | 2              | 2         | 8         | 2         |
| When I have to talk about what I have read, I feel            | 6              | 5              | 1              | 3         | 2         | 7         |
| When I help my friends with their reading, I feel             | 9 <sup>b</sup> | 0 <sup>b</sup> | 1 <sup>b</sup> | 11        | 0         | 1         |
| When I have to write, I feel                                  | 7              | 3              | 2              | 4         | 3         | 5         |
| When I do my spellings, I feel                                | 8              | 2              | 2              | 6         | 3         | 3         |
| I like books.   | 10             | 1              | 1              | 7         | 4         | 1         |
| <b>Total questions 1-10</b>                                   | <b>63</b>      | <b>20</b>      | <b>11</b>      | <b>49</b> | <b>24</b> | <b>23</b> |
| I like playing on the computer.                               | 10             | 2              | 0              | 11        | 1         | 0         |
| I easily follow the instructions on the computer screen.      | 7 <sup>c</sup> | 1 <sup>c</sup> | 3 <sup>c</sup> | 5         | 5         | 2         |
| I find it easy to learn how to do new things on the computer. | 6              | 4              | 2              | 6         | 4         | 2         |
| I prefer reading things on a screen than in a book.           | 9              | 1              | 2              | 7         | 2         | 3         |
| <b>Total digital literacies (questions 11-14)</b>             | <b>33</b>      | <b>8</b>       | <b>6</b>       | <b>30</b> | <b>13</b> | <b>5</b>  |

<sup>b</sup> two children would not give answers; <sup>c</sup> One child just replied 'I only play'

The aggregated data have been shown separately for questions 1-10 and 11-14 since those groups of questions investigated different aspects of reading. Questions 1-10 asked the children about their attitudes towards traditional books, whereas questions 11-14 asked about digital technologies. In this digital era, the importance of the use of digital literacy technologies cannot be overestimated, as it is now not unusual to see young children 'using' smartphones and tablets. However, when children start school and start to learn how to read, the main mode of instruction is the use of books. The children's attitudes towards books, especially those children who are falling behind in reading, could provide valuable information. Could the children come to appreciate books more if they possessed more proficient skills to access them?

During the course of obtaining ethics approval it was advised that questions about digital literacies should be included, even though these were tangential to the

research being undertaken. However, the responses to the final questions do show that most of these children preferred to read on computers rather than read books.

Further statistical analyses were not done on the questionnaire data. Inspection of the aggregated data showed very little difference between the pre- and post-test data for the intervention group. Statistical tests would merely have confirmed non-significance. There did appear to be a possibly significant pre/post difference for the control group, but in the 'wrong' direction: the control group's attitudes appeared to have deteriorated somewhat. Applying between-group comparisons might therefore have produced a misleading impression, as occurred in the Hodder Oral Sentence Reading analysis (section 7.5.4)

## **7.8 Summary**

In this chapter, the standardised test data for phases 1 (RCT) and 2 (follow-up), the statistical data for book band levels and associated analyses and the children's writing, and the questionnaire results have been analysed. Discussion of the results follows in chapter 8.

## **8 Discussion**

### **8.1 Introduction**

The aims of this chapter are to draw conclusions, state limitations, and discuss and provide evidence that this is an original piece of research which contributes additional knowledge to the realm of educational research, and is a successful piece of doctoral research. The next three sections discuss the quality of the RCT, then of the rest of the quantitative evaluation (i.e. both aspects of the inner study), and of the PFS (outer study). This order, deliberately the reverse of that at the beginning of this thesis, wraps up the inner study first, then feeds its conclusions into the wrapping-up of the outer study. Then a substantial section is devoted to lessons learnt from conducting a pilot/feasibility study.

The next three sections discuss the limitations of the research with respect to its design; the possibility of Type I and Type II errors; and, briefly, the lessons learned by the researcher.

The last section provides conclusions for each of the hypotheses stated in chapter 1, and a final summary stating the conclusions of the research.

### **8.2 Testing the quality of the randomised control trial**

The rigour of the RCT reported here would be vital to the success of any future study of Fischer Family Trust Wave 3. It was therefore important to review its quality.

The results of an RCT are not immediately trustworthy solely by virtue of the method by which they are obtained. Torgerson et al. (2006a) identified a number of key aspects of an RCT that they felt were critical to ensure that any RCT could be classified as '*gold standard*' (Torgerson & Torgerson, 2007, p. 333). Regardless of the size of the RCT, if the key components are not included the trial can contain a fatal flaw. These flaws can be avoided by systematically auditing the internal validity of the study.

A review by Torgerson et al. (2006b) looked at adult literacy and numeracy and used an audit checklist to determine the validity of their RCT. By replacing the



words adult(s) with child(ren) that table has been adapted to test the validity of the RCT in this study (Table 8.1).

**Table 8.1                      Audit of RCT against an adaptation of Torgerson et al.'s (2006b) checklist of characteristics of valid RCTs**

| Characteristic   | Present in this study?   |
|--|--|
| 1. Was the study population adequately described? (i.e. were the important characteristics of the randomised children described, e.g. age, gender?)  | Yes  |
| 2. Was the minimum important difference described? (i.e. was the smallest educationally important effect size described?)  | Yes – in section 6.8   |
| 3. Was the target sample size adequately determined?   | No. The sample size ( $n = 24$ ) was fixed externally  |
| 4. Was intention to treat analysis used? (i.e. were all children who were randomised included in the follow-up and analysis?)  | Yes  |
| 5. Was the unit of randomisation described? (i.e. individual children or groups of children)   | Yes  |
| 6. Were the participants allocated using random number tables, coin flip, computer generation?   | No – minimisation was used, as this provides a more robust randomisation for small numbers   |
| 7. Was the randomisation process concealed from the investigator? (i.e. was the researcher recruiting children to the trial blinded to the children's allocation until after the children had been included in the trial?) | Yes – although the researcher knew the children she did not know which children were allocated to which group until after the minimisation had been done |
| 8. Was the intervention described and were assessment criteria/tools pre-identified?   | Yes - FFT Wave 3 is discussed in detail in section 6.4 and the choice of assessment tools in section 6.10  |
| 9. Were follow-up measures administered blind? (i.e. were the researchers who administered the outcome measures blind to treatment allocation?)  | No – as the researcher was the investigator she knew which child was allocated to which group  |
| 10. Was estimated effect on primary and secondary outcome measures stated?   | No, because of the small sample and the dearth of prior research to use for guidelines   |
| 11. Was precision of effect size estimated (confidence intervals)?   | No – see section 7.3.2   |
| 12. Were summary data presented in sufficient detail to permit alternative analyses or replication?  | Yes  |
| 13. Was the discussion of the study findings consistent with the data?   | Yes (however, since this is a doctoral thesis, it is the decision of others to judge the consistency)  |

Five of Torgerson et al.'s (2006b) characteristics resulted in a negative answer. Characteristic 3 asks whether the target sample size was calculated prior to the RCT, using a power calculation. This was not possible, since it was only possible to give the intervention to a maximum of 24 children. A retrospective power calculation was done, and these results are shown in section 6.8.

Characteristic 6 asks how randomisation was achieved. As described in the method chapter (section 6.9), randomisation was done by minimisation, which has been described as the 'platinum standard'. Minimisation was used because of the small sample.

Characteristic 9 asks whether the follow-up measures (tests) were administered blind. The researcher was fully and unavoidably aware of which children were in which group since she was the sole investigator. All four tests were administered as fairly as possible. The Oral Word Reading test and Oral Sentence Reading test were given on an individual basis, and therefore could have been open to manipulation by the researcher but, as described in section 6.10.3, strict guidelines were adhered to. The Hodder Group Reading Test and the Diagnostic Spelling Test were given in a classroom environment, under strict test conditions, and the tests were marked in strict accordance with the mark schemes.

Characteristic 10 asks whether the estimated effects on primary and secondary outcome measures were stated. The answer to this was No because the sample size was small and there was too little prior research to base estimates on.

Confidence intervals on the effect sizes (characteristic 11) were not reported for the reasons given in section 7.3.2.

A question which is implied in Torgerson et al.'s characteristic 4 is whether the original allocation of the children to groups was kept throughout. It was – the allocation schedule was not altered.

The findings of this analysis can be summed up by saying that the criteria for a robust RCT were adhered to as far as was possible for a small-scale, one-researcher study. The remaining possibilities of bias are discussed below.

### **8.3 Judging the quality of the exploratory measures**

The measures based on book bands clearly illustrated the perils of one-group designs. The findings were positive, but were like the last gift to emerge from Pandora's Box, false hope. Such measures should not be re-used.

The writing measure also produced a positive, but limited, result. Longer samples would be needed to facilitate more detailed analysis, and should be gathered from the control group too, in order to strengthen confidence in any findings.

The attitudes questionnaire would have benefited from a trial run. It became apparent that the children found some of the questions difficult to understand, as well as not understanding the meaning of the 'smiley' faces to express how they felt. However, in many such evaluations it is important to get some idea of whether the intervention suits the children involved. As with the writing samples, data from the control group should, ideally, also be gathered.

#### **8.4 Testing the quality of the pilot/feasibility study**

In chapter 3 the importance of conducting a rigorous PFS was discussed and a proposed CONSORT table was constructed. The completed CONSORT table is shown below (Table 8.2) but, as stated in chapter 3, items relevant specifically to the RCT and discussed above are omitted.

All criteria relevant for the rigorous implementation of a PFS (and separate from those for RCTs), according to the PFS CONSORT table, were fulfilled. Where fuller explanations are required, references to specific parts of this research are cited. The use of the CONSORT table has ensured that all aspects of the PFS have been addressed in the reporting. The CONSORT table provides a 'crib sheet' to ensure that any alterations to the original research design have been documented. This is a safeguard to ensure that, if this design were to be used again, the alterations in approach would be immediately apparent and be introduced into the larger study design.

The use of a CONSORT table in the assessment of a PFS is highly recommended. Ideally, the same CONSORT table should be used throughout all educational research, but before this can occur it would be necessary that a Delphi survey is undertaken to ensure that definitions of key terms and PFS criteria are agreed.

**Table 8.2: Completed CONSORT table to determine the validity of the PFS**

| Characteristic  | Present in this study?   |
|---|--|
| 1. Identified as a pilot/feasibility study in the title and/or the abstract.  | Yes – in both  |
| 2. Has there been sufficient discussion and explanation describing the theoretical/scientific background for the proposed larger study and why a pilot/feasibility study would be of value? | Yes – in chapter 2   |
| 3. Have definite objectives and hypotheses been determined that are pertinent to both the pilot/feasibility and the proposed full studies?  | Yes – in section 1.10.1  |
| 4. Has ethical approval been obtained?  | Yes – see Appendix 1   |
| 5. Have changes to the design that occurred during the study been documented and explained?   | Yes – in section 6.12  |
| 6. Have the methods by which the participants were identified and gave consent been described?  | Yes – in section 6.6   |
| 7. Have any alterations to assessment tools that occurred during the study been explained?  | There were no such alterations.  |
| 8. Why did the study end?   | It had concluded at its expected point.                                      |
| 9. Have any other analyses been included within the final report which could have a bearing on the implementation of a larger study?  | No – unnecessary   |
| 10. Have the limitations of the pilot/ feasibility study been described with reference to bias and the overall feasibility of implementing a larger study?                                  | Yes – see section 8.2.1  |
| 11. Are the conclusions/discussions representative of the objectives and findings of the pilot/feasibility study?   | Yes – see this chapter, but, as with the RCT, it is for others to judge this |

#### 8.4.1 Suitability and limitations of the PFS

As discussed fully in Chapter 3, PFSs are used to determine whether a proposed study design is ‘fit for purpose’, and primarily to identify any major flaws in design which could have a negative impact on a larger study. In this research a PFS approach was used due to the small number of participants that could take part, and the acknowledgement that the results obtained could be statistically non-significant – a fact that is more readily accepted when a PFS is undertaken.

The ‘outer’ study which comprised the PFS performed as expected. The small-scale approach did not highlight any major methodological design issues that would affect a larger-scale study, and only two minor ones. The first issue that could have an impact on the research was the alteration in the timings, when the teachers asked that the children not be taken out of morning lessons (except PE). This aspect could not have been accounted for and was unexpected. It is possible

that the researcher's insider status had an impact here; if she had been an external researcher, the teachers might have been less likely to ask for alterations in the timings. If a similar study were to be undertaken a more formal agreement between the school and the researcher should be drawn up to ensure that timings, for example, are agreed beforehand and cannot be altered.

The other problem is the risk of bias, especially with respect to the administration and marking of the assessment tests. These were administered and marked by the researcher and therefore could be subject to reporting bias. There is no way of proving that this did not occur despite the researcher's best efforts. With respect to other aspects of bias that can be found in RCTs, and consequently in PFSs, all possible attempts to prevent bias were used. The researcher did not select the participants; therefore, selection bias was avoided. All participants were present at the end of the research; therefore, no attrition bias was observed.

All possible precautions were implemented to reduce the risk of unconscious bias and adverse effects of any additional variables. Since this was a PFS the financial implications of ensuring that the study could be achieved without any risk of bias would have been prohibitive. However, in a full-scale study it would be essential to put stricter precautions against bias in place.

In summary the methodological approaches used in this PFS could be scaled up to be used in a larger-scale study, provided some adjustments were made.

## **8.5 Research designs and rigour**

The pilot/feasibility design of this project has shown that the design was mainly fit for purpose and could be used for a larger study. The main assessment tools performed well and, given a larger sample size and therefore increased statistical power, might have detected any differences in gains between the intervention and control groups. The reading attitudes questionnaire would have benefitted from being trialled prior to the onset of the research. It performed adequately but needs further development.

The choice of an RCT approach was based on the observation that RCTs can produce more robust results (as discussed in section 2.2.6) than those obtained from single-group pre- and post-test studies. However, as just implied, this does

rely on there being sufficient power. In its absence, the confusing and inconclusive pattern of results on the primary and secondary outcomes is less surprising. Although results for the exploratory outcomes of writing and book bands were more promising, they were still inconclusive because no data on these aspects were gathered from the control group.

## **8.6 Limitations**

There are limitations with all research, and this research was no different. Two of the limitations described in this section were known at the start, the size of the RCT, due to the study being a PFS, and the researcher's insider position. A third limitation was the reduction in time with each child imposed by the school.

Since this was a PFS the RCT was small ( $n=24$ ). This was because there was a single investigator, who could only provide the intervention to a maximum of 12 children in each group in the time available. This would have allowed the children to be seen on a daily basis and receive the intervention as its designer intended. This immediately implied that the results obtained were unlikely to be statistically significant. But the reason for undertaking a PFS is to ensure that the correct design and method(s) have been chosen to investigate a specific research question, and their suitability for a much larger study.

Although it was expected that non-significant results would be obtained, the implementation of the RCT portion of this study needed to be done to a high standard, and the requirements for a valid RCT followed, as far as possible. It was a well-planned piece of research and is a contribution to knowledge since, as demonstrated by the literature review, there has been very little research looking at non-phonics-based interventions and their effectiveness (or not) for children who are falling behind.

Due to the low numbers in the trial, each child was a significant percentage of the total, approximately 8% of the sample per group. Thus, when a child or two refused to complete a test or were not focused, it had a significant effect on the data. If the sample had been larger, these missing data would not have had such a noticeable effect.

The researcher's insider status is discussed fully in chapter 2. But, to summarise, as well as being the investigator, she was a parent, a parent governor and volunteer helper at the school – which allowed full understanding of how the school operated – and already knew the class teachers and many of the children with whom she would be working; because of this familiarity the children were quick to settle, and mainly worked well.

Although the researcher's position made fitting into the school easier, it may also have hindered her. If she had been an external researcher who had asked the school staff to sign up for the intervention, and they had agreed to it being implemented as intended, they might not have asked for the timetable to be altered so significantly.

The reduction in the time allowed for the intervention was severe. Initially each child in the intervention group was seen every day for 20 minutes (total time in each two-week period: 200 minutes); this was reduced to seven times over a two-week period for 13 minutes a time, equivalent to 91 minutes in each two-week period. This was a reduction of 109 minutes/fortnight, or 54.5%, in contact time. These were not the conditions under which Canning (2004, 2009) had trialled FFT Wave 3 and intended it to be used, but it was a matter of fitting in with the teachers' requirements. They expressed concerns that the children were missing numeracy and literacy, especially the start of the lesson where the discussion about that day's work was happening, and they felt the children were being disadvantaged.

The constant pressure on teachers to raise standards in numeracy and literacy prevents schools fully appreciating the importance of delivering interventions according to the instruction manuals. FFT Wave 3 requirements are 20 minutes a day per child, which could be seen as intrusive, but if the intervention works and children reach their age-related expectations that should be sufficient evidence to prove that a little inconvenience is worthwhile.

The study would have benefited if the researcher had been 'blind' to the children's group allocation and had the assessments administered and marked by a third party. This would have required extra funding, which was not available. However, this approach would definitely have removed the risk of bias in the assessments.

Although the researcher did not consciously treat the children or mark their assessments in a biased way, she may have been done so unconsciously.

A further limitation was the choice of non-phonics-based interventions available for the age group being assessed. In the intervening 10 years there have been a further 10 intervention schemes designed for KS1 (Brooks, 2016). This suggests an increase in the understanding of the importance of early interventions and identifying children with literacy problems. By addressing these issues earlier, children should gain the necessary literacy skills needed to access the curriculum and avoid the risk of becoming despondent, which would have a negative effect.

### **8.7 Type I and Type II errors**

A total of 41 individual statistical analyses were done on the data obtained. According to the definition of Type I errors, the fact that so many analyses were done means it is probable that at least one result, possibly two, could be inaccurate. This suggests that it is possible that null hypotheses may have been rejected when they should not have been.

Type II errors occur when studies are under-powered due to a lack of participants. As discussed in section 6.8, this study was under-powered and therefore may have suffered from Type II errors, meaning that null hypotheses may not have been rejected when they should have been.

### **8.8 What experience has been gained?**

During this rigorous process of research, valuable lessons have been learnt, and new analytical skills and the ability to conduct a substantial piece of educational research have been gained, starting with development of a hypothesis, identification of a suitable method by which to investigate the hypothesis and, above all, conducting a pilot/feasibility study to determine the advisability of doing a large-scale study. The ability to use statistical analyses and the skill of analytical writing have been difficult to learn, and are skills that will be forever developing for any researcher.

It has become apparent that any educational researcher is dependent upon the good graces of the teachers who are a necessary aspect of such research,



resulting in the need to be flexible and accommodating, which may adversely affect the outcome of the research project.

The most rewarding experience was working with the children who needed more support with their reading, spelling and writing skills.

Although the null results were expected it is disappointing that the Archimedes 'Eureka' moment did not occur. However, as Einstein so eloquently said:

No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

Quoted by Calaprice (1996, p. 224)

## **8.9 Principal conclusions**

The use of CONSORT tables to ensure consistency of approaches when PFSs and RCTs are being used should become the accepted norm. They provide a structure that all researchers can use, and allow studies to be directly compared in meta-analyses or systematic reviews.

The PFS (fully) and RCT (mainly) were well designed, as supported by the completed CONSORT tables, and the overall methodology of wrapping a PFS around a small-scale quantitative study could be scaled up to be used in a larger-scale study, with some adjustments. However, in future research a signed agreement explaining the school's responsibilities and researcher's responsibilities should be drawn up, thus preventing any issues with timetabling.

The assessment tests in future studies would be best marked and administered by a third party who has no knowledge of the allocation of the participants; this would remove the chance of reporting bias.

The results obtained from the 'inner' study are at best inconclusive and currently do not provide sufficient evidence to support a larger study of FFT Wave 3.

## **8.10 Principal and subsidiary research questions and hypotheses**

As described in chapter 1 there are a number of different hypotheses to be answered.

### **8.10.1 Quantitative study research questions and hypotheses**

#### **8.10.1.1 Primary outcome**

##### **Research question:**

Can a non-phonics-based intervention significantly improve struggling children's reading comprehension?

##### **Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's reading comprehension.

##### **Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's reading comprehension.

##### **Conclusion:**

The statistical evidence does not justify rejecting the null hypothesis.

#### **8.10.1.2 Secondary outcomes**

##### **Research question:**

Can a non-phonics-based intervention significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling?

##### **Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling.

**Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's oral word reading and/or oral sentence reading and/or spelling.

**Conclusion:**

The statistical evidence does not justify rejecting the null hypothesis.

**8.10.1.3 Exploratory outcomes****Research question:**

Can a non-phonics-based intervention improve struggling children's progression through book band levels and/or writing ability and/or attitudes to reading and books?

**Null hypothesis:**

A non-phonics-based intervention does not significantly improve struggling children's progression through book band levels or writing ability or attitude to reading and books.

**Substantive hypothesis:**

A non-phonics-based intervention does significantly improve struggling children's progression through book band levels and/or writing ability and/or attitude to reading and books.

**Conclusion:**

The book band analysis and writing analysis support the substantive hypothesis. However, the reading attitudes questionnaire analysis does not justify rejecting the null hypothesis.

**8.10.2 Pilot/feasibility study research question and hypotheses**

Is the quantitative evaluation within this pilot/feasibility study suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study?

**Null hypothesis:**

The quantitative evaluation within this pilot/feasibility study is not suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study.

**Substantive hypothesis:**

The quantitative evaluation within this pilot/feasibility study is suitable for assessing the effectiveness of a non-phonics-based intervention scheme in a larger study, given suitable adjustments.

**8.10.3 Conclusion**

Although the statistical data are inconclusive, it has been determined that the substantive hypothesis of the PFS largely holds true.

Finally, in addition to the use of the CONSORT table, the Education Endowment Foundation has formulated three specific questions that need to be answered to determine whether a pilot/feasibility study has provided sufficient evidence to move the study onto the next phase: these questions were applied to this pilot/feasibility study (see Table 8.3).

**Table 8.3. EEF PFS questions answered with respect to this research**

| Questions asked by EEF     |                               |   |
|----------------------------|-------------------------------|---|
| Was the approach feasible? | Is there evidence of promise? | Is the approach ready for a full trial without further development? |
| Yes                        | Uncertain                     | No  |

The answers show that, although the approach is feasible, the evidence supporting FFT Wave 3 as a non-phonics-based intervention is not yet strong enough to expand research on it to a full-scale study. Thus the inner, quantitative study failed to show conclusively whether the FFT Wave 3 intervention had real impact. On the other hand, the PFS successfully showed that, with adjustments, a quantitative (RCT) approach could be a suitable method for assessing a non-phonics-based intervention.

## References

- Adams, M. J. (1999). *Beginning to Read: Thinking and Learning about Print*. Massachusetts: MIT Press.
- Allington, R. L. (2014). How reading volume affects both reading fluency and reading achievement. *International Electronic Journal of Elementary Education*, 7(1), 13-26.
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, 35(4), 494-519.
- Baker, T. L. (1994). *Doing social research*. New York: McGraw-Hill Inc.
- Barrs, M., & Cork, V. (2001). *The reader in the writer* (First ed.). London: Centre for Language in Primary Education.
- Baumann, J. F., Hoffman, J. V., Moon, J., & Duffy-Hester, A. M. (1998). Where are the teachers voices in the phonics/whole language debate? Results from a survey of U.S. elementary classroom teachers. *The Reading Teacher*, 51(8), 636-650.
- Becker, P. T., & Schulte, H. D. (2008). Publishing pilot interevention studies. *Research in Nursing and Health*, 31, 1-3.
- Bickler, S., Baker, S., & Hobsbaum, A. (2003). *Book Bands for Guided Reading* (Third ed.). Cambridge: The Wolsey Press.
- Bielby, N. (1994). *Making sense of reading the new phonics and its practical implications* (1st ed.). Glasgow: Scholastics Publications Ltd.
- BIS. (2011). *2011 Skills for Life Survey:Headline findings*. London: Department for Business, Innovation and Skills.
- Bishop, D. (2013). *Good and bad news on the phonics screen*. Retrieved May 31, 2014, from <http://deevybee.blogspot.co.uk/2013/10/good-and-bad-news-on-phonics-screen.html>
- Bonell, C., Maisey, R., Speight, S., Purdon, S., Keogh, P., Wollny, I., Wellings, K. (2013). Randomized controlled trial of 'teens and toddlers':A teen pregnancy prevention intervention combining youth development and voluntary service in a nursery. *Journal of Adolescence*, 36(5), 859-870.
- Borman, G. D., Dowling, N. M., & Schneck, C. (2008). A multisite cluster randomised field trial of Open Court Reading. *Educational Evaluation and Policy Analysis*, 30(4), 389-407.
- Bredo, E. (2009). Getting over the methodological wars. *Educational Researcher*, 38(6), 441-448.
- Brooks, G. (2002). *What works for children with Literacy Difficulties? The effectiveness of intervention schemes*. Retrieved June 14, 2014, from <http://dera.ioe.ac.uk/4662/1/RR380.pdf>

- Brooks, G. (2003). *Sound sense: the phonics element of the National Literacy Strategy*. Retrieved 10 16, 16, from [http://dera.ioe.ac.uk/4938/5/nls\\_phonics0303gbrooks.pdf](http://dera.ioe.ac.uk/4938/5/nls_phonics0303gbrooks.pdf)
- Brooks, G. (2007). *What works for pupils with literacy difficulties? The effectiveness of intervention schemes*. Nottingham: Department for Children, Schools and Families.
- Brooks, G. (2013). *What works for children and young people with literacy difficulties? The effectiveness of intervention schemes*. Bracknell: Dyslexia SPLD Trust.
- Brooks, G. (2016). *What works for children and young people with literacy difficulties*. Retrieved June 12, 2016, from <http://www.dyslexiaaction.org.uk/files/dyslexiaaction/what-works-5th-edition.pdf>
- Brooks, G., & Burton, M. (2016, March). *Elinet Analytic Glossary of the teaching and Learning of Literacy*. Retrieved September 14, 2016, from [http://www.elinet.eu/fileadmin/ELINET/Redaktion/user\\_upload/1.\\_Elinet\\_Analytic\\_Glossary\\_of\\_the\\_Initial\\_Teaching\\_and\\_Learning\\_of\\_Literacy1.pdf](http://www.elinet.eu/fileadmin/ELINET/Redaktion/user_upload/1._Elinet_Analytic_Glossary_of_the_Initial_Teaching_and_Learning_of_Literacy1.pdf)
- Brooks, G., & Burton, M. (2017, March (revised edition)). *ELINET analytic glossary of the initial teaching and learning of literacy*. Retrieved July 19, 2017, from [http://www.elinet.eu/fileadmin/ELINET/Redaktion/user\\_upload/Elinet\\_analytic\\_glossary.pdf](http://www.elinet.eu/fileadmin/ELINET/Redaktion/user_upload/Elinet_analytic_glossary.pdf)
- Brooks, G., Schagen, I., & Nastat, P. (1995). *Trends in reading at eight*. Slough: nfer.
- Brown, I. S., & Felton, R. H. (1990). Effects of instruction on beginning reading skills in children at risk of reading disability. *Reading and Writing: An Interdisciplinary Journal*, 2, 223-241.
- Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, Autumn*, 1-21.
- Calaprice, A. (1996). *The Quotable Einstein*. (A. Caraprice, Ed.) Princeton: Princeton University Press.
- Cameron, J. I., Naglie, G., Green, T. L., Gignac, M. A., Bayley, M., Huijbregts, M., . . . Czerwonka, A. (2015). A feasibility and pilot randomised controlled trial of the "Timing it Right Stroke Family Support Program". *Clinical Rehabilitation*, 29(11), 1129-1140.
- Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching children to read: The fragile link between science and federal education policy. *Education Policy Analysis Archive*, 11(15), 1-51.
- Camilli, G., Wolfe, P. M., & Smith, M. L. (2006). Meta-analysis and reading policy: Perspectives on teaching children to read. *The Elementary School Journal*, 107(1), 27-36.
- Canning, J. (2004). *FFT wave 3: Report of findings from a 10-week pilot*. Letchworth: Fischer Family Trust.
- Canning, J. (2009). *FFT Wave 3*. Middlesbrough: Fischer Family Trust.
- Carbo, M. (1987). What works isn't always phonics? *The Phi Delta Kappan*, 68(6), 431-435.

- Center, Y., Wheldall, K., Freeman, Y., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30(2), 240-263.
- Chall, J. (1967). *Learning to Read: The Great Debate*. New York: McGraw-Hill.
- Chavez, C. (2008). Conceptualizing from the inside: Advantages, complications and demands on insider positionality. *The Qualitative Report*, 13(3), 474-494.
- Clark, C. (2012). *Children's and young peoples reading today: Findings from the 2011 National Literacy Trust's annual survey*. London: National Literacy Trust.
- Clark, C. (2013). *Children's and young people's reading in 2012: Findings from the 2012 National Literacy Trust's annual survey*. London: National Literacy Trust.
- Clark, C. (2014). *Children's and young people's reading in 2013*. London: National Literacy Trust.
- Clark, C. (2015). *Children's and young people's reading in 2014: Findings from the 2014 National Literacy Trusts annual survey*. London: National Literacy Trust.
- Clark, C. (2016). *Children's and young person's reading in 2015*. London: National Literacy Trust.
- Clark, M. M. (2015). *An evidence-based critique of synthetic phonics in literacy learning*. Retrieved October 3, 2016, from <http://nape.org.uk/wp-content/uploads/2015/03/Synthetic-Phonics.pdf>
- Clark, M. M. (2017). The phonics screening check: its development consequences, intended and unintended. In M. M. Clark (Ed.), *Reading the evidence* (pp. 81-97). Birmingham: Witley Press.
- Clem, K. S. (1990). Moving beyond conflict and confusion in literacy instruction. *Early Child Development and Care*, 61(1), 131-138.
- Coe, R. (2002). *It's the Effect Size, Stupid. What effect size is and why its important*. Retrieved May 18, 2014, from <http://www.leeds.ac.uk/educol/documents/00002182.htm>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale NJ: Erlbaum.
- Coltheart, M. (2006). Dual-route and connectionist models of reading: an overview. *London Review of Education*, 4(1), 5-17.
- Cree, A., Kay, A., & Steward, J. (2012). *The economic and social cost of illiteracy: A snapshot of illiteracy in a global context*. UK: World Literacy Foundation.
- Crumpler, M., & McCarty, C. (2010). *Diagnostic Spelling Tests Manual* (First ed.). London: Hodder and Stoughton Ltd.
- Davis, A. (2012). A Monstrous Regimen of Synthetic Phonics: Fantasies of Research-based Teaching' Methods' Versus Real Teaching. *Journal of Philosophy of Education*, 46(4), 560-573.
- Davis, A. (2014). *The Guardian*. Retrieved May 8, 2014, from <http://www.theguardian.com/teacher-network/teacher-blog/2014/mar/04/reading-lessons-phonics-world-book-day>

- DCSF. (2007). *National Archives*. Retrieved September 14, 2016, from <http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/00767-2007BKT-EN.pdf>
- DCSF. (2009). *Phonics: assessment and tracking guidance*. London: Crown.
- Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010). A pilot study of a kindergarten school reading program in high-poverty urban schools. *The Elementary School Journal*, 110(4), 423-439.
- Department for Children Schools and Families. (2007). *Early Literacy Support*. Nottingham: DCSF Publications.
- Department of Health. (2001). *Seeking consent: working with children*. London: Department of Health.
- DfE. (1989). *National Curriculum for Keystages 1 and 2*. London: HMSO.
- DfE. (1995). *Key Stages 1 and 2 of the National Curriculum*. London: The Stationary Office.
- DfE. (2013a). *The National Curriculum in England Key stages 1 and 2 framework document*. London: Crown.
- DfE. (2013b). *Phonics screening check and national curriculum assessments at Key Stage 1 in England 2012/2013*. Retrieved May 17, 2014, from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/245813/SFR37-2013\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/245813/SFR37-2013_Text.pdf)
- DfE. (2013c, May 20). *Topic Note: 2012 Phonics Screening*. Retrieved May 31, 2014, from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/200000/DFE-RR287.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200000/DFE-RR287.pdf)
- DfE. (2015a). *Phonics screening check and national curriculum assessments at Key Stage 1 in England, 2015*. Retrieved 10 19, 2016, from Department of Education: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/463002/SFR32\\_KS1\\_2015\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/463002/SFR32_KS1_2015_Text.pdf)
- DfE. (2015b). *National curriculum assessments: Key Stage 2, 2015 (revised)*. Retrieved 10 19, 2016, from Department of Education: <https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-2015-revised>
- DfE. (2015c). *Reading: the next steps: supporting higher standards in schools*. London: Crown.
- DfE. (2016, 29 September). *Phonics screening check and Key stage 1 assessments in England, 2016*. Retrieved November 7, 2016, from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/556769/SFR42\\_Phonics\\_KS1\\_2016.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/556769/SFR42_Phonics_KS1_2016.pdf)
- DfEE. (1998). *The National Literacy Strategy: a framework for teaching*. London: DfEE.
- DfEE. (1999a). *Phonics Progression in Phonics: Materials for whole-class teaching*. Retrieved October 13, 2016, from <http://www.schoolslinks.co.uk/ProgressionInPhonics.pdf>



- DfEE. (1999b). *The National Curriculum*. London: DfEE/QCA.
- DfES. (2001, November). *The National Literacy and Numeracy Strategy*. Retrieved September 13, 2016, from <https://core.ac.uk/download/pdf/4157139.pdf>
- DfES. (2001). *The National Literacy Strategy: framework for teaching 3rd editon*. Retrieved 10 14, 2016, from <http://dera.ioe.ac.uk/4699/>
- DfES. (2004). *Playing with sounds: a supplement to Progression in Phonics*. London: Crown.
- DfES. (2007). *Letters and Sounds*. Nottingham: DfES Publications.
- Dockrell, J., Marshall, C., & Wyse, D. (2015). *Talk for Writing Evaluation and executive summary*. London: Education Endowment Foundation.
- Dudovskiy, J. (2016). *Positivism Research Philosophy*. Retrieved September 10, 2016, from <http://research-methodology.net/research-philosophy/positivism/>
- Education Endowment Foundation. (2015). *Policy on analysis for EEF evaluations*. London: Education Endowment Foundation. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Research\\_Report/2015\\_Analysis\\_for\\_EEF\\_evaluations.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf)
- Education Endowment Foundation. (2017). *Why does the EEF fund pilot projects?* Retrieved September 4, 2017, from <https://educationendowmentfoundation.org.uk/news/why-does-the-eef-fund-pilot-projects/>
- Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to Spell* (pp. 237-269). New Jersey: Lawrence Erlbaum Associates Inc.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the Reading Panel's Meta Analysis. *Review of Educational Research*, 71(3), 393-447.
- Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016b). CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *British Medical Journal*, 355(5239), 1-29.
- Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Sally, H., Coleman, C. L., & Bond, C. M. (2016a). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS ONE*, 1-22.
- Every Child a Chance Trust. (2009). *The long term cost of literacy difficulties. 2nd Edition*. London: Every Child a Chance Trust .
- Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing*, 26(4), 790-797.
- Fischer Family Trust. (2007). FFT Wave 3 Programme. Letchworth Garden City: Fischer Education Project Ltd.
- Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35, 39-50.

- Fletcher, R., & Barnes, H. E. (2016). *Auguste Comte: French Philosopher*. Retrieved September 16, 2016, from <https://www.britannica.com/biography/Auguste-Comte>
- Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in Grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*, 1-29.
- Fraser, D. (1997). Ethical dilemmas and practical problems for the practitioner researcher. *Educational Action Research, 5*(1), 161-171.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface Dyslexia, Neuropsychological and Cognitive Studies of Phonological Reading*. (pp. 301-330). London: Erlbaum.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment for children and adolescents with reading disabilities: A meta-analysis of randomised controlled trials. *PLoS ONE, 9*(2), e89900. doi:10.1371/journal.pone.0089900.
- Giddens, A. (1974). *New rules of sociological method*. New York: Basic Books.
- Gorard, S. (2007). *Quantitative methods in educational research* (1st ed.). Kings Lynn: Continuum.
- Gorard, S. (2014). *Switch-on Reading. Evaluation and executive summary*. Retrieved from Education Endowment Foundation: [https://educationendowmentfoundation.org.uk/public/files/Publications/Campaigns/Evaluation\\_Reports/EEF\\_Project\\_Report\\_SwitchOnReading.pdf](https://educationendowmentfoundation.org.uk/public/files/Publications/Campaigns/Evaluation_Reports/EEF_Project_Report_SwitchOnReading.pdf)
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6-10.
- Groff, P. (1977). Critic's Corner: Views on phonics of reading clinicians. *Reading World, 17*(2), 93-98.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III*. Mahwah NJ: Lawrence Erlbaum.
- Hammersley, M. (1993). On the teacher as researcher. *Educational Action Research, 1*(3), 425-445.
- Hammill, D. D., & Swanson, H. L. (2006). The National Reading Panel's meta-analysis of phonics: Another point of view. *The Elementary School Journal, 107*(1), 12-26.
- Hartung, J., Knapp, J., & Sinha, B. K. (2008). *Statistical Meta-Analysis with Application*. Hoboken New Jersey: Wiley.
- Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., Snowling, M. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomised controlled trial. *Journal of Child Psychology and Psychiatry, 47*(8), 820-827.
- Hattie, J. (2012). *Visible learning for teachers: maximising impact on learning*. Oxford: Routledge.

- Henderson, E. H. (1980). Word knowledge and reading disability. In E. H. Henderson, & J. W. Beers (Eds.), *Developmental and cognitive aspects of learning to spell: A reflection of word knowledge* (pp. 138-148). Newark: International Reading association.
- Henderson, E. H. (1981). *Learning to read and spell: The child's knowledge of words*. Illinois: Northern Illinois University Press.
- Himmelfarb Health Sciences Library. (2011). *Randomized controlled trial*. Retrieved October 9, 2016, from <https://himmelfarb.gwu.edu/tutorials/studydesign101/rcts.html>
- Howe, K. R. (2009). Positivist dogmas, rhetoric and the Education science question. *Educational Researcher*, 38(6), 428-440.
- Hunt, K. (1965). *Grammatical structures written at three grade levels. NCTE Research report No. 3*. Champaign, IL USA: NCTE.
- Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Reserach in Reading*, 30(3), 227-248.
- Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers'*. Slough: NFER.
- IBM. (2008). *One-tail test for t-test independent samples*. Retrieved July 5, 2015, from <http://www-01.ibm.com/support/docview.wss?uid=swg21476176>
- Invernizzi, M., Abouzeid, M., & Gill, j. T. (1994). Using students' invented spellings as a guide for spelling instruction that emphasizes word study. . *Elementary School Journal*, 95, 155-167.
- Iversen, S., & Tunmer, W. E. (1993). Phonological processing skills and the reading recovery program. *Journal of Educational Psychology*, 85(1), 112-126.
- Jakobsen, T. G. (2013). *Theory of science-what is positivism*. Retrieved September 12, 2016, from <http://www.popularsocialscience.com/2013/02/15/theory-of-science-what-is-positivism/>
- Jeray, K. J., & Tanner, S. L. (2012). Pilot randomised trials: Is there a need? *The Journal of Bone and Joint Surgery*, 1, 15-18.
- Johnson , B., & Christensen, L. (2012). *Educational research: Quantitative, Qualitative and Mixed Approaches* (2nd ed.). New York: Sage.
- Johnston, R. S., McGeown, S., & Watson, J. E. (2012). Long-term effects of synthetic versus analytic phonics teaching on the reading and spelling ability of 10 year old boys and girls. *Reading and Writing*, 25, 1365.
- Kim, H. W. (2012). Research with children: challenges and dilemmas as an insider researcher. *Early Child Development and Care*, 182(2), 263-276.
- Kistin, C. (2015). Pilot Studies: A critical but potentially misused component of interventional research. *JAMA*, 314(15), 1561-1562.
- Krashen , S. (2004b). The Phonics Debate. *Language Magazine*.

- Krashen, S. (2001). More smoke and mirrors: A critique of the National Reading Panel report on fluency. *The Phi Delta Kappan*, 83(2), 119-123.
- Krashen, S. (2004a). Basic phonics. *TexTESOL Newsletter*, pp. 2-4.
- Krashen, S. (2005). Is in-school reading good for children? Why the National Reading Panel report is (still) wrong. *The Phi Delta Kappan*, 86(6), 444-447.
- Lee, R. L., & Lee, P. H. (2014). To evaluate the effects of a simplified hand washing improvement program in schoolchildren with mild intellectual disability: A pilot study. *Research in Developmental Disabilities*, 35, 3014-3025.
- Leitch, S. (2006). *Prosperity for all in the global economy-world class skills*. Norwich: TSO.
- Machin, S., McNally, S., & Viarengo, M. (2016). *"Teaching to Teach" Literacy*. London: The Centre for Economic Performance Publications Unit.
- Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the SEARCH/TEACH tutoring approach with middle-class students at risk of reading failure. *The Elementary Journal*, 92(5), 573-586.
- Martinussen, R. L., & Kirby, J. R. (1998). Instruction in successive phonological processing to improve reading acquisition skills. *Developmental Disabilities Bulletin*, 26(2), 19-39.
- May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., Tognatta, N. (2015). Year one results from the multisite randomised evaluation of the i3 scale-up of Reading Recovery. *American Educational Research Journal*, 52(3), 547-581.
- McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohonen, S., Anandakumar, T., . . . Castles, A. (2012). Phonics raining for English-speaking poor readers. *Cochrane Database of Systematic Reviews*(12), 1- 103.
- McGrane, J., Stiff, J., Baird, J.-A., Lenkeit, J., & Hopfenbeck, T. (2017). *Progress in International Reading Literacy Study (PIRLS): National Report for England*. Oxford: OUCEA.
- McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, 71(4), 445-463.
- McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). *ABRA: Online reading support evaluation report and executive summary*. London: Education Endowment Foundation.
- Medical Research Council. (1948). Streptomycin treatment of pulmonary tuberculosis: a report of the streptomycin in tuberculosis trials committee . *British Medical Journal*, 2, 769-82.
- Meier, J. D., & Invernizzi, M. (2001). Book Buddies in the Bronx: Testing a model for America Reads. *Journal of Education for Students Placed at Risk (JESPAR)*, 6(4), 319-333.
- Merton, R. K. (1972). Insiders and outsiders: A chapter in the sociology of knowledge. *American Journal of Sociology*, 78(1), 9-47.

- Moore, J. (2012). A personal insight into researcher positionality. *Nurse Researcher*, 19(4), 11-14.
- Morin, K. H. (2013). Value of a pilot study. *Journal of Nursing Education*, 52(10), 547-549.
- Morris, N. S., & Rosenbloom, D. A. (2017). Defining and understanding pilot and other feasibility studies. *American Journal of Nursing*, 117(3), 38-45.
- Morrow-Howell, N., Jonson-Reid, M., McCrary, S., Lee, Y., & Spitznagel, E. (2009). *Evaluation of experience corps: student reading outcomes (CSD Research Report 09-01)*. St. Louis MO: Washington University, Centre for Social Development.
- MRC. (2000). *A framework for development and evaluation of RCTs for complex interventions to improve health*. Retrieved October 12, 2016, from <https://www.mrc.ac.uk/documents/pdf/rcts-for-complex-interventions-to-improve-health/>
- National Literacy Trust. (2011a). *Raising UK literacy levels*. Retrieved October 10, 2016, from [http://www.literacytrust.org.uk/news/4160\\_adult\\_skills\\_survey\\_results\\_released\\_today](http://www.literacytrust.org.uk/news/4160_adult_skills_survey_results_released_today)
- National Literacy Trust. (2011b). *Illiterate Adults in England*. Retrieved May 14, 2014, from [http://www.literacytrust.org.uk/adult\\_literacy/illiterate\\_adults\\_in\\_england](http://www.literacytrust.org.uk/adult_literacy/illiterate_adults_in_england)
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence based assessment of the scientific research on reading and its implications for reading instruction*. Rockville MD: NICHD Clearinghouse.
- NIHR. (2016). *Guide to feasibility and pilot studies: a guide for NIHR research design service advisors*. Retrieved from <https://www.rds-eastmidlands.nihr.ac.uk/resources/Guide%20to%20Supporting%20Feasibility%20and%20Pilot%20Studies%20MAY%202016%20FINAL.pdf>
- Oakley, A. (1998, October 31). Experimentation and social interventions: a forgotten but important history. *British Medical Journal*, 317, 1239-42.
- Ofsted. (2002). *The National Literacy Strategy: the first four years 1998-2002*. Northampton: Arkle Print.
- Ofsted. (2010). *Harpur Hill Primary School Inspection Report*. London: Ofsted.
- Ofsted. (2014). *Harpur Hill Primary School Ofsted Report*. London: Ofsted.
- Ofsted. (2016). *Harpur Hill Primary School Inspection Report*. London: Ofsted.
- Oliver, S., Bagnall, A., Thomas, J., Shepherd, J., Sowden, A., White, I., Garrett, Z. (2010). 'Randomised controlled trials for policy interventions: a review of reviews and meta regression'. *Health Technology Assessment*, 14(16), 1-189.
- Patel, R., Jabin, N., Bussard, L., Cartagena, J., Haywood, S., & Lumpkin, M. (2017). *Switch-on effectiveness trial*. London: Education Endowment Foundation.

- Pinnell, G. S. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal*, 90(2), 160-183.
- Pinnell, G. S., Lyons, C. A., Deford, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high risk first graders. *Reading Research Quarterly*, 29(1), 8-39.
- Polit, D. F., Beck, C. T., & Hungler, B. P. (2001). *Essentials of Nursing Research: Methods, Appraisal and Utilization*. Philadelphia: Lippincott Williams & Wilkins.
- Pullen, P. C., Lane, H. B., & Monaghan, M. C. (2004). Effects of a volunteer tutoring model on the early literacy development of struggling first grade students. *Reading Research and Instruction*, 43(4), 21-40.
- Puranik, C. S., Patchan, M. M., Lemons, C. J., & Al Otaiba, S. (2017). Using peer assisted strategies to teach early writing: results of a pilot study to examine and promise. *Reading and Writing*, 30(1), 25-50.
- Raffaele Mendez, L. M., Pelzman, C. A., & Frank, M. J. (2016). Engaging struggling early readers to promote reading success: A pilot study of reading by design. *Reading and Writing Quarterly*, 32(3), 273-297.
- Richardson, H. (2014, January 28). 'Able readers damaged by phonics' academic says. Retrieved May 8, 2014, from <http://www.bbc.co.uk/news/education-25917646>
- Rose, J. (2006). *Independent review of the teaching of early reading*. Nottingham: DfES.
- Ryan, A. B. (2006). *Post-positivist approaches to research*. Retrieved September 30, 2016, from [http://eprints.maynoothuniversity.ie/874/1/post-positivist\\_approaches\\_to\\_research.pdf](http://eprints.maynoothuniversity.ie/874/1/post-positivist_approaches_to_research.pdf)
- Sainsbury, M., Caspall, L., McDonald, A., Ravenscroft, L., & Schagen, I. (1998a). *Evaluation of the 1998 summer schools program 1998, full report*. Slough: nfer.
- Sainsbury, M., & Clarkson, R. (2008). *Attitudes to reading at ages nine and eleven: full report*. Slough UK: nfer.
- Sainsbury, M., & Schagen, I. (2004). Attitudes to reading at age nine and eleven. *Journal of Research in Reading*, 27(4), 373-386.
- Sainsbury, M., Schagen, I., Whetton, C., Hagues, N., & Minnis, M. (1998b). *Evaluation of the National Literacy Project. Cohort 1 1996-1998*. Slough: nfer.
- Sanchez, V. M., & O'Connor, R. E. (2015). Building tier 3 intervention for long-term slow growers in grades 3-4: A pilot study. *Learning Disabilities Research and Practice*, 30(4), 171-181.
- Sanson-Fisher, R., Bonevski, B., Green, L., & D'Este, C. (2007). Limitations of the randomised controlled trial in evaluating population-based health interventions. *American Journal of Preventative Medicine*, 33(2), 155-161.
- Schoenfeld, D. (2012). *Statistical Considerations for a parallel trial where the outcome is a measurement*. Retrieved June 30, 2014, from [http://hedwig.mgh.harvard.edu/sample\\_size/js/js\\_parallel\\_quant.html](http://hedwig.mgh.harvard.edu/sample_size/js/js_parallel_quant.html)
- Schwartz, R. M. (2005). Literacy learning of at-risk first grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97(2), 257-267.

- See, B. H., Gorard, S., & Siddiqui, N. (2015). *Word and World Reading*. London: Education Endowment Foundation.
- Sellgren, K. (2013, June 5). *Phonics test 'accurate but unnecessary'*. Retrieved October 25, 2016, from <http://www.bbc.co.uk/news/education-22783399>
- Shader, R. I. (2015). Proof of feasibility: What a pilot study is and is not. *Clinical Therapeutics*, 37(7), 1379-1380.
- Shanahan, T. (2004). Critiques of the National Reading Panel Report: Their implications for research, policy and practice. In P. McCardle, & V. Chhabra (Eds.), *The Voice of Evidence in Reading Research* (pp. 235-265). Baltimore: Paul.H.Brookes.
- Shanahan, T. (2017). Relationships between reading and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 194-207). New York: The Guildford Press.
- Sibbald, B. (1998). Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, 316, 201.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best evidence synthesis. *Educational Research Review*, 1-26.
- Solity, J. (2003). *Teaching phonics in context: A critique of the National Literacy Strategy*. Retrieved 10 13, 2016, from [http://dera.ioe.ac.uk/4938/7/nls\\_phonics0303jsolity.pdf](http://dera.ioe.ac.uk/4938/7/nls_phonics0303jsolity.pdf)
- Solity, J., & Vousden, J. (2009). Real books vs reading scheme: a new perspective from instructional psychology. *Educational Psychology*, 29(4), 469-511.
- Stahl, S. (2004). What do we know about fluency? Findings of the National Reading Panel. In P. McCardle, & V. Chhabra (Eds.), *The Voice of Evidence in Reading Research* (pp. 187-211). Baltimore: Paul.H.Brookes.
- Standards and Testing Agency. (2012a). *Assessment framework for the development of the Year 1 phonics screening check*. 2012: Crown.
- Standards and Testing Agency. (2012b). *Phonics Screening Check; Technical report*. Retrieved May 31, 2014, from <https://www.gov.uk/government/publications/phonics-screening-check-2012-technical-report>
- Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. *British Journal of Educational Psychology*, 73, 585 - 598.
- Stuebing, K. K., Barth, A. E., Cirino, P. T., Francis, D. J., & Fletcher, J. M. (2008). A response to recent reanalyses of the National Reading Panel report: Effects of systematic phonics instruction are practically significant. *Journal of Educational Psychology*, 100(1), 123-134.
- Suggate, S. P. (2010). What we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, 46(6), 1556-1579.

- Taylor, P. C. (2013). *Educational research paradigms: From positivism to multiparadigmatic*. Retrieved October 10, 2016, from <http://www.meaningcentered.org/educational-research-paradigms-from-positivism-to-multiparadigmatic/>
- Teale, W. H., & Sulzby, E. (1987). Introduction: Emergent literacy as a perspective for examining how young children become readers and writers. In W. H. Teale, & E. Sulzby (Eds.), *Emergent Literacy* (pp. vii -xxv). New Jersey: Ablex Publishing Corporation.
- The Education and Skills Committee. (2005). *Teaching Children to Read*. London: The Stationary Office Limited.
- Torgerson, C. J., & Torgerson, D. J. (2007). The use of minimization to form comparison groups in educational research. *Educational Studies, 33*(3), 333-337.
- Torgerson, C. J., Brooks, G., & Hall, J. (2006a). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. University of Sheffield. Nottingham: DfES Publications.
- Torgerson, C., Porthouse, J., & Brooks, G. (2006b). A systematic review of controlled trials evaluating interventions in adult literacy and numeracy. *Journal of Research in Reading, 28*(2), 87-107.
- Torgerson, C., Torgerson, D., Birks, Y., & Porthouse, J. (2005). A comparison of randomised controlled trials in health and education. *British Educational Research Journal, 31*(6), 761-785.
- Torgesen, J., Wagner, R., Rashotte, C., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing Reading Failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579-593.
- Treasure, T., & MacRae, K. D. (1998). Minimisation: the platinum standard for trials. *British Medical Journal, 317*, 362-368.
- Trochim, W. M. (2006). *Deduction and Induction*. Retrieved June 12, 2014, from <http://www.socialresearchmethods.net/kb/dedind.php>
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), 453-466.
- Tymms, P., & Merrell, C. (2007). *Standards and quality in English primary schools over time: The national evidence*. Cambridge: Primary Review, Faculty of Education, University of Cambridge.
- Umbach, B., Darch, C., & Halpin, G. (1989). Teaching reading to low performing first graders in rural schools. *Journal of Instructional Psychology, 16*(3), 112-121.
- United Nations. (1989). *The United Nations Convention on the Rights of the Child*. London: UNICEF.
- Vadasy, P. F., & Sanders, E. A. (2008). Code-orientated instruction for kindergarten students at risk for reading difficulties: a replication and comparison of instructional groupings. *Read Write, 21*, 929-963.



- Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997). The effectiveness of one-to-one tutoring for at-risk beginning readers. *Learning Disability Quarterly, 20*(2), 126-139.
- Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-orientated instruction for kindergarten students at risk for reading difficulties : A randomised field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508-528.
- van Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*(35), 1-4.
- Vincent, D., & Crumpler, M. (2011a). *Hodder Group Reading Tests Manual* (Second ed.). London: Hodder Education.
- Vincent, D., & Crumpler, M. (2011b). *Hodder Oral Reading Tests Manual* (First ed.). London: Hodder Education.
- Walker, M., Bartlett, S., Betts, H., Sainsbury, M., & Worth, J. (2014). *Phonics screening check evaluation. Research report*. Slough UK: nfer.
- Walsh, K. (2012). Quantitative vs qualitative research: A false dichotomy. *Journal of Research Nursing, 17*(9), 9-11.
- West, A., & Spring, B. (n.d.). *Randomised controlled trials*. Retrieved 10 10, 2016, from [http://www.ebbp.org/course\\_outlines/rcts.pdf](http://www.ebbp.org/course_outlines/rcts.pdf)
- Wollscheid, S., Sjaastad, J., Tomte, C., & Lover, N. (2016). The effect of pen and paper or tablet computer on early writing -A pilot study. *Computers and Education, 98*, 70-80.
- Workman, B. (2007). "Casing the joint" Explorations by the insider-researcher preparing for worked-based projects. *Journal of Workplace Learning, 19*(3), 146-160.
- World Literacy Foundation. (2015). *The economic and social cost of illiteracy: A snapshot of illiteracy in a global context* . London: World Literacy Foundation.
- Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015). *Improving numeracy and literacy: Evaluation report and executive summary*. Retrieved October 27, 2016, from [https://educationendowmentfoundation.org.uk/public/files/Projects/EEF\\_Project\\_Report\\_ImprovingNumeracyAndLiteracyInKeyStage1.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/EEF_Project_Report_ImprovingNumeracyAndLiteracyInKeyStage1.pdf)
- Wrigley, T. (2017). Synthetic phonics and the phonics check: the hidden politics of early literacy. In M. M. Clark (Ed.), *Reading the evidence* (pp. 98-114). Birmingham: Witley Press Ltd.

## Appendix 1: Description of Letters and Sounds phases

*Phase one* (Nursery/reception) is divided into seven different aspects: environmental sounds, instrumental sounds, body sounds, rhythm and rhyme, alliteration, voice sounds, and oral segmenting and blending. This phase was designed to help develop the children's listening skills to be able to link sounds with letters in the order that they occur in words.

*Phase two* starts in Reception and lasts up to six weeks. Children learn 19 letters of the alphabet and one of the corresponding phonemes for each letter. They learn to blend these sounds together to make simple VC and CVC words and to segment words into their separate phonemes to allow them to read simple phrases. They are also taught some high frequency 'tricky' words such as *the*, *to*, *go* and *no*.

In *Phase three* (Reception; up to 12 weeks) 25 graphemes, most of them digraphs (for example <ch, oo, oa>), and the remaining 7 letters of the alphabet are taught, again with one of the corresponding phonemes for each letter. The children continue to practise segmenting and blending.

By *Phase four* (Reception; 4 to 6 weeks) the children should know 42 phonemes and their corresponding graphemes and be able to segment and blend CVC words. This phase is designed to consolidate what the children have already learned and apply it to reading and spelling more effectively.

In *Phase five* (throughout Year 1) new graphemes and alternative pronunciations of already known graphemes are taught. Children become increasingly quick at recognising graphemes of more than one letter, and at blending. They are also able to identify the correct grapheme to be used in writing specific words. In the final phase, *Phase six* (throughout Year 2), the aim is to ensure that the children are fluent readers and accurate spellers by the end of KS1.

## Appendix 2: Ethical Approval



The  
University  
Of  
Sheffield.

The  
School  
Of Education.

Louise Davis

Head of School  
Professor Jackie Marsh

Department of Educational Studies  
The Education Building  
388 Glossop Road  
Sheffield S10 2JA

29 November 2011

Telephone: +44 (0)114 222 9096  
Fax: +44 (0)114 2228105  
Email: [jacquie.gillott@sheffield.ac.uk](mailto:jacquie.gillott@sheffield.ac.uk)

Dear Louise

**Re: Can a non-phonics based intervention scheme accelerate the progress in reading and spelling of Year 1 and Year 2 children who are falling behind?**

Thank you for your application for ethical review for the above project. The reviewers have now considered this and have agreed that your application be approved with the following optional amendments.

(Please see attached reviewers' comments)

Yours sincerely

Mrs Jacquie Gillott  
Programme Secretary

## ETHICS REVIEWER'S COMMENTS FORM

This form is for use by members of academic staff in the School of Education when reviewing a research ethics application.

### Note to reviewers and applicants:

The ethical review process in the School of Education is designed to provide critical response on ethical issues identified in research proposals. For this reason, reviewers' comments are not anonymous\*. The comments given here are intended to help applicants (and where appropriate their academic supervisors) to revise their research plans where necessary to ensure that their research is conducted to high ethical standards.

The contents of this form remain internal to the University, and should not be used for wider dissemination without written permission from the Ethics Reviewer named here and the Chair of the Ethics Review Panel.

|  |   |
|--|---|
| <b>1. Name of Ethics Reviewer*:</b><br><small>Reviewers who wish to make anonymous responses should contact the Chair of the Ethics Review Panel before completing the review.</small>   | Toni Ibarz  |
| <b>2. Research Project Title:</b>  | Can a non-phonics based intervention scheme accelerate the progress in reading and spelling of Year 1 and Year 2 children who are falling behind? |
| <b>3. Principal Investigator (and name of Tutor/Supervisor in the case of student applications):</b>   | Louise Davies/Greg Brooks   |
| <b>4. Academic Department / School:</b>  | Education   |
| <p><b>5. I confirm that I do not have a conflict of interest with the project application</b></p> <p style="text-align: center;">Or</p> <p><b>The following details may be considered as a conflict of interest.</b> (If a possible conflict of interest is declared, the Chair of the Ethical Review Panel will take this into account)</p> |   |

| 6. I confirm that, in my judgment, the application should: |   |  |  |
|--|---|--|--|
| Be approved:   | Be approved with suggested amendments in '7' below: | Be approved providing requirements specified in '8' below are met:<br>and/or | <u>NOT</u> be approved for the reason(s) given in '9' below: |
|  | x   |  |  |

|   |
|---|
| <p>7. Approved with the following suggested, optional amendments (i.e. it is left to the discretion of the applicant whether or not to accept the amendments and, if accepted, the ethics reviewers do not need to see the amendments):</p> <p>I have made a few suggestions in the parental and information letters. I attach the two forms with my suggestions.</p> |
| <p>8. Approved providing the following, compulsory requirements are met (i.e. the ethics reviewers need to see the required changes):</p>   |
| <p>9. Not approved for the following reason(s):</p>   |
| <p>10. Date of Ethics Review: 10<sup>th</sup> November 2011</p>   |

## ETHICS REVIEWER'S COMMENTS FORM

This form is for use by members of academic staff in the School of Education when reviewing a research ethics application.

### Note to reviewers and applicants:

The ethical review process in the School of Education is designed to provide critical response on ethical issues identified in research proposals. For this reason, reviewers' comments are not anonymous\*. The comments given here are intended to help applicants (and where appropriate their academic supervisors) to revise their research plans where necessary to ensure that their research is conducted to high ethical standards.

The contents of this form remain internal to the University, and should not be used for wider dissemination without written permission from the Ethics Reviewer named here and the Chair of the Ethics Review Panel.

|  |   |
|--|---|
| <b>1. Name of Ethics Reviewer*:</b><br><small>Reviewers who wish to make anonymous responses should contact the Chair of the Ethics Review Panel before completing the review.</small> | Kate Pahl   |
| <b>2. Research Project Title:</b>  | Can a non-phonics based intervention scheme accelerate the progress in reading and spelling of Year 1 and Year 2 children who are falling behind? |
| <b>3. Principal Investigator (and name of Tutor/Supervisor in the case of student applications):</b>   | Louise Davies/Greg Brooks   |
| <b>4. Academic Department / School:</b>  | Education   |

|   |
|---|
| <p><b>5. I confirm that I do not have a conflict of interest with the project application</b></p> <p style="text-align: center;"><b>Or</b></p> <p><b>The following details may be considered as a conflict of interest. (If a possible conflict of interest is declared, the Chair of the Ethical Review Panel will take this into account)</b></p> |
|---|

| <b>6. I confirm that, in my judgment, the application should:</b> |   |   |   |
|---|---|---|---|
| Be approved:  | Be approved with suggested amendments in '7' below: | Be approved providing requirements specified in '8' below are met:<br><small>and/or</small> | NOT be approved for the reason(s) given in '9' below: |
|   | x   |   |   |

**7. Approved with the following suggested, optional amendments (i.e. It is left to the discretion of the applicant whether or not to accept the amendments and, if accepted, the ethics reviewers do not need to see the amendments):**

I can see you have now got a strategy to get the children's consent. However, I can also see there might be an element of coercion involved in having as your plan involves one form the children all have to sign.

I also wondered what would happen if a child said no. This needs to be a viable option, not 'the elephant in the room' which you hope will not happen.

I also thought that you could have explained more clearly the process by which you will gain informed consent. To me, this still feels like a project that hasn't quite got children's ethical procedures at the heart of it. You might like to read some articles in *Children and Society*, particularly the last two or three years, about how social science views the issue of children's consent in research.

**8. Approved providing the following, compulsory requirements are met (i.e. the ethics reviewers need to see the required changes):**

**9. Not approved for the following reason(s):**

**10. Date of Ethics Review: 10<sup>th</sup> November 2011**

## Appendix 3: Reading Record

Running Record (RCM) - 4 - 1

| Name <input type="text"/>  |  | Class <input type="text"/>  |  | Date <input type="text"/>   |   |  |
|--|--|---|--|---|---|--|
| Number of words read + Number of errors made <input type="text"/>  |  | Error ratio (1: ) <input type="text"/>  |  | Accuracy % <input type="text"/>   |   |  |
| Comment on children's reading, strategies used, phrasing, expression, pace: <input type="text"/><br><input type="text"/><br><input type="text"/>   |  |   |  |   |   |  |
| Pages  | Title ..... Level <input type="text"/> | Ea  | SCa  | Information   |   |  |
|  |  |   |  | Used  | Errors  |  |
| <input type="checkbox"/>   | <input type="checkbox"/>               | <input type="checkbox"/>  | <input type="checkbox"/>   | <input type="checkbox"/>  | <input type="checkbox"/>  |  |
| <p>*1 Establish error ratio (NOTE: SC is NOT an error)</p> <p>Number of words read divided by the number of errors</p> <p>---100 gives error ratio 1:10 (1 error in 10 words)</p> <p>---15 ..... 1</p> <p>---5 gives error ratio 1:5 (1 error in 5 words)</p> <p>---20 ..... 2</p> |  | <p>Easy<sup>a</sup></p> <p>1:200 ---99.5%</p> <p>1:100 ---99%</p> <p>1:50 ---98%</p> <p>1:33 ---97%</p> <p>1:25 ---96%</p> <p>1:20 ---95%</p> | <p>Instructional<sup>a</sup></p> <p>1:17 ---94%</p> <p>1:14 ---93%</p> <p>1:12.5 ---92%</p> <p>1:11.75 ---91%</p> <p>1:10 ---90%</p> | <p>Hard<sup>a</sup></p> <p>1:9 ---89%</p> <p>1:8 ---87.5%</p> <p>1:7 ---85.5%</p> <p>1:6 ---83%</p> <p>1:5 ---80%</p> <p>1:4 ---75%</p> <p>1:3 ---66%</p> <p>1:2 ---50%</p> | <p>*2 Use error ratio to determine accuracy level from the tables</p> <p>e.g. 1:10 = 90% ..... 1:14 = 93% ..... 2</p> |  |



## Appendix 4: Letter for Parents

Dear Parent/Carer

I am currently a student at the University of Sheffield and am about to undertake some educational research looking at how children learn to read. Mr O'Connor and the Governors of the school have kindly agreed to let me do my research at Harpur Hill Primary School. The research involves using a proven method of teaching of reading and writing developed by the Fischer Family Trust, and will involve the children having a 10 minute/day session with me on an individual basis for 10 weeks. The research I am proposing has been approved by the University of Sheffield and has gained full ethics approval. I am a qualified teacher and have been CRB checked.

Although I will be working with a small number of children more closely during the Spring and Summer terms of 2012, I would appreciate your consent to assess your child's reading and spelling, during December 2011, April 2012 and July 2012. Any information I gain would be passed onto your child's teacher as well as being used in my research. I must emphasise that I will not disrupt your child's education and will work closely with your child's teacher to ensure that this does not occur.

Your child's name will not be published. They would appear as a reference number in any written work. In addition I would also like to have your permission to use your child's date of birth and gender, (these criteria are used in educational research).

Please complete the slip below and return it to school as soon as possible.

Thank you for taking the time to read this letter.

If you have any questions please do not hesitate to contact on 0788 6181173

Louise Davis

-----  
-----

I do/do not give permission for \_\_\_\_\_ to take part in Louise Davis' research project.

Signed \_\_\_\_\_

Date \_\_\_\_\_

## **Parental Information Sheet**

### Why is this research taking place?

The aim of this research is to improve the literacy skills of 5- 7-year-olds using a non-phonics approach. After the Rose report recommended that all children be taught to read using a synthetic phonics approach (breaking the words down into their individual sounds (phonemes) and then blending the sounds together to make the word) all schools and new reading schemes adopted this approach. My concern is that, although this approach works for many children and for many words, there are some children who find this approach difficult, and unfortunately most intervention schemes are phonics based. Wave 3 is a proven non-phonics based intervention scheme which is divided into two sections, a reading day and a writing day. During the reading day children read a new book, learn to spell and write a new word and do letter work activities. During the writing day children pick a sentence out of their book to write, learn at least one new word and do sentence construction activities. By using this complementary method, it is hoped that children will progress more rapidly and gain in confidence and self-esteem as they become more confident readers.

### Will my child's education be affected?

Your child's education will not be adversely affected. Your child will be receiving additional individual help with their reading and spelling, which will help your child by providing additional strategies for them to use and subsequently improve their skills. I will be in constant communication with your child's class teacher and if any problems arise, they will be dealt with promptly.

### Why has my child been chosen?

Your child has been chosen because their class teacher feels they will benefit from additional one-to-one teaching.

### How many children will be taking part?

Twelve children from Year 1 and twelve children from Year 2 will be chosen, and then split into two groups.

### Which group will my child be in?

Your child will be placed randomly into a group. Group 1 will follow the Wave 3 programme in the Spring term, and Group 2 will complete it in the Summer term.

### What will my child have to do?

The reading and spelling ages of your child will be determined at three points through the year (December, April and July). If your child is in Group 1, he or she will see me for 15 minutes a day between January and April, and we will work through the programme, alternating between reading and spelling. If your child is in Group 2, he or she will see me for 15 minutes a day between April and July, and we will work through the programme in the same way.

If you have any questions please contact me on 0788 6181173.

Louise Davis

## Letter to parents whose child had been chosen for the trial

Dear .....

I am currently a student at the University of Sheffield and am about to undertake some educational research looking at how children learn to read. Mr O'Connor and the Governors of the school have kindly agreed to let me do my research at Harpur Hill Primary School. The research involves using a proven method of teaching of reading and writing developed by the Fischer Family Trust, and will involve the children having a 15 minutes/day session with me on an individual basis for 10 weeks. The research I am proposing has been approved by the University of Sheffield and has gained full ethics approval. I am a qualified teacher and have been CRB checked.

Your child's teacher has suggested that your child would benefit from this alternative approach and I am writing to you to ask for your permission for your child to take part. Your child's education will not be affected adversely (on the contrary, it should have a positive effect) and I will be in full communication with your child's teacher at all times.

Your child's name will not be published. They would appear as a reference number in any written work. However, I would also like to have your permission to use your child's date of birth, gender, teacher assessment levels and whether they have free school meals (these criteria are used in educational research).

I enclose an information sheet that provides more details.

Please complete the slip below and return it to school as soon as possible.

Thank you for taking the time to read this letter.

Louise Davis

-----  
I do/do not give permission for \_\_\_\_\_ in \_\_\_\_\_ class  
to take part in Louise Davis' research project.

Signed \_\_\_\_\_ Date \_\_\_\_\_

# Appendix 5: Children's consent sheet

## Class Pupil Consent Form

Mrs. Davis has spoken to me on my own explained about the project that she is working on.

I understand that I do not have to help if I do not want to.

I understand that my name will not be used.



| Child's Name | Signature/Smiley Face |                          |
|--------------|-----------------------|--------------------------|
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |
| =            | =                     | <input type="checkbox"/> |



## Appendix 6: Minimisation data and results

Greg Randomisation - Microsoft Excel

|    | A       | B          | C      | D    | E                  | F                        | G                 | H                     | I   | J |
|----|---------|------------|--------|------|--------------------|--------------------------|-------------------|-----------------------|-----|---|
|    | Ref num | Date of Bi | Gender | Year | Group Reading Test | Diagnostic Spelling Test | Oral Word Reading | Oral Sentence reading | FSM |   |
| 1  | 1       | 20.09.2005 | Boy    | 1    | 77                 | 70                       | 77                | 76                    | X   |   |
| 2  | 2       | 01.03.2006 | Boy    | 1    | 75                 | 94                       | 74                | 74                    |     |   |
| 3  | 3       | 17.08.2006 | Girl   | 1    | 92                 | 109                      | 82                | 85                    |     |   |
| 4  | 4       | 11.05.2006 | Boy    | 1    | 90                 | 95                       | 70                | 70                    |     |   |
| 5  | 5       | 12.06.2006 | Girl   | 1    | 87                 | 103                      | 81                | 75                    |     |   |
| 6  | 6       | 09.05.2006 | Girl   | 1    | 83                 | 96                       | 78                | 84                    |     |   |
| 7  | 7       | 21.01.2006 | Girl   | 1    | 70-                | 92                       | 77                | 74                    |     |   |
| 8  | 8       | 07.06.2006 | Boy    | 1    | 90                 | 103                      | 70-               | 75                    |     |   |
| 9  | 9       | 08.08.2006 | Girl   | 1    | 89                 | 105                      | 70-               | 76                    |     |   |
| 10 | 10      | 11.05.2006 | Girl   | 1    | 85                 | 101                      | 81                | 81                    |     |   |
| 11 | 11      | 15.06.2006 | Boy    | 1    | 87                 | 95                       | 70-               | 70-                   |     |   |
| 12 | 12      | 21.09.2005 | Girl   | 1    | 77                 | 82                       | 70-               | 71                    |     |   |
| 13 | 13      | 12.04.2005 | Girl   | 2    | 78                 | 90                       | 85                | 89                    |     |   |
| 14 | 14      | 22.11.04   | Girl   | 2    | 77                 | 70-                      | 77                | 83                    |     |   |
| 15 | 15      | 22.07.2005 | Boy    | 2    | 98                 | 127                      | 92                | 95                    |     |   |
| 16 | 16      | 20.10.2004 | Boy    | 2    | 78                 | 92                       | 80                | 85                    | X   |   |
| 17 | 17      | 07.06.2005 | Boy    | 2    | 81                 | 85                       | 87                | 88                    |     |   |
| 18 | 18      | 01.11.2004 | Boy    | 2    | 70-                | 70-                      | 73                | 75                    |     |   |
| 19 | 19      | 26.08.2005 | Boy    | 2    | 88                 | 91                       | 77                | 87                    |     |   |
| 20 | 20      | 04.08.2005 | Boy    | 2    | 76                 | 96                       | 77                | 87                    |     |   |
| 21 | 21      | 18.03.2005 | Boy    | 2    | 70-                | 70-                      | 75                | 81                    |     |   |
| 22 | 22      | 02.01.2005 | Boy    | 2    | 83                 | 106                      | 85                | 91                    | X   |   |
| 23 | 23      | 23.07.2005 | Girl   | 2    | 82                 | 84                       | 77                | 87                    |     |   |
| 24 | 24      | 23.12.2004 | Boy    | 2    | 78                 | 78                       | 84                | 89                    |     |   |

minimisation data LZU100 - MICROSOFT EXCEL

File Home Insert Page Layout Formulas Data Review View

Clipboard: Paste, Cut, Copy, Format Painter

Font: Calibri, 11, Bold, Italic, Underline, Color, Background Color, Merge & Center, Wrap Text

Alignment: Left, Center, Right, Top, Bottom, Merge & Center, Wrap Text

Number: General, Percentage, Number, Text, Scientific, Fraction, Date, Time, Custom

Styles: Conditional Formatting, Format as Table, Cell Styles

Cells: Insert, Delete, Format





|    | A       | B      | C    | D     | E         | F   | G     | H    | I     | J    | K     | L    | M    | N    | O    | P      | Q      | R      | S      | T      | U      | V      | W      | X     | Y     | Z     | AA    |        |  |
|----|---------|--------|------|-------|-----------|-----|-------|------|-------|------|-------|------|------|------|------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|-------|-------|--------|--|
| 1  | Ref num | Gender | Year | Group | Reading T | FSM | Group | Iboy | Igirl | Cboy | Cgirl | IYr1 | IYr2 | CYr1 | CYr2 | Iread+ | Iread- | Cread+ | Cread- | Iread+ | Iread- | Cread+ | Cread- | Iftot | Cftot | Group | Int N | Cont N |  |
| 2  | 1       | Boy    | 1    | 77    | X         | I   | 1     |      |       |      | 1     |      |      |      |      |        |        | 1      |        |        |        |        |        |       |       | I     | 1     | 0      |  |
| 3  | 2       | Boy    | 1    | 75    |           | C   |       | 1    |       |      |       |      |      |      |      |        |        | 1      |        |        |        |        |        |       |       | C     | 1     | 1      |  |
| 4  | 3       | Girl   | 1    | 92    |           | C   |       |      | 1     |      |       |      |      |      |      |        |        | 1      |        |        |        |        |        |       |       | C     | 1     | 2      |  |
| 5  | 4       | Boy    | 1    | 90    |           | C   |       |      |       | 1    |       |      |      |      |      |        |        | 1      |        |        |        |        |        |       |       | C     | 1     | 3      |  |
| 6  | 5       | Girl   | 1    | 87    |           | C   |       |      |       |      | 1     |      |      |      |      |        |        | 1      |        |        |        |        |        |       |       | C     | 1     | 4      |  |
| 7  | 6       | Girl   | 1    | 83    |           | I   |       | 1    |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 2     | 4      |  |
| 8  | 7       | Girl   | 1    | 70-   |           | I   |       | 1    |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 3     | 4      |  |
| 9  | 8       | Boy    | 1    | 90    |           | I   |       |      | 1     |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 4     | 4      |  |
| 10 | 9       | Girl   | 1    | 89    |           | I   |       |      |       | 1    |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 4     | 5      |  |
| 11 | 10      | Girl   | 1    | 85    |           | I   |       |      |       |      | 1     |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 5     | 5      |  |
| 12 | 11      | Boy    | 1    | 87    |           | I   |       |      |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 5     | 6      |  |
| 13 | 12      | Girl   | 1    | 77    |           | I   |       |      |       |      | 1     |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 6     | 6      |  |
| 14 | 13      | Girl   | 2    | 78    |           | I   |       |      |       |      |       | 1    |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 7     | 6      |  |
| 15 | 14      | Girl   | 2    | 77    |           | I   |       |      |       |      |       | 1    |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 8     | 6      |  |
| 16 | 15      | Boy    | 2    | 98    |           | I   |       |      |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 9     | 6      |  |
| 17 | 16      | Boy    | 2    | 78    | X         |     |       |      |       | 1    |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 9     | 7      |  |
| 18 | 17      | Boy    | 2    | 81    |           |     |       | 1    |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 10    | 7      |  |
| 19 | 18      | Boy    | 2    | 70-   |           |     |       |      | 1     |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 10    | 8      |  |
| 20 | 19      | Boy    | 2    | 88    |           |     |       | 1    |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 11    | 8      |  |
| 21 | 20      | Boy    | 2    | 76    |           |     |       |      |       | 1    |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 11    | 9      |  |
| 22 | 21      | Boy    | 2    | 70-   |           |     |       |      |       |      | 1     |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 11    | 10     |  |
| 23 | 22      | Boy    | 2    | 83    | X         |     |       | 1    |       |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 12    | 10     |  |
| 24 | 23      | Girl   | 2    | 82    |           |     |       |      | 1     |      |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 12    | 11     |  |
| 25 | 24      | Boy    | 2    | 78    |           |     |       |      |       | 1    |       |      |      |      |      |        |        |        |        | 1      |        |        |        |       |       | I     | 12    | 12     |  |
| 26 | mean    |        |      |       |           |     |       |      |       |      |       |      |      |      |      |        |        |        |        |        |        |        |        |       |       |       |       |        |  |
| 27 | Totals  |        |      |       |           |     |       | 6    | 6     | 7    | 5     | 6    | 6    | 6    | 6    | 7      | 5      | 6      | 6      | 2      | 10     | 1      | 11     | 21    | 20    |       |       |        |  |
| 28 |         |        |      |       |           |     |       |      |       |      |       |      |      |      |      |        |        |        |        |        |        |        |        |       |       |       |       |        |  |













## Appendix 7: Attitudes Questionnaire

Ref Number: \_\_\_\_\_ Date completed: \_\_\_\_\_  
 Date of Birth: \_\_\_\_\_ School Year: \_\_\_\_\_



### My thoughts about reading

|   |   |  |   |
|---|---|--|---|
| When I have stories read to me, I feel...             |    |    |    |
| When I look at books, I feel...                       |    |    |    |
| When I read my reading books, I feel...               |    |    |    |
| When I have to talk about what I have read, I feel... |    |    |    |
| When I help my friends with their reading, I feel...  |    |    |    |
| When I have to write, I feel...                       |   |   |   |
| When I do my spellings, I feel...                     |  |  |  |
| I like books.   |  |  |  |

|   |   |  |   |
|---|---|--|---|
| I like playing on the computer.                             |  |  |  |
| I easily follow the instructions on the computer screen.    |  |  |  |
| I find it easy to learn how to do new things on a computer. |  |  |  |
| I prefer reading things on a screen than in a book.         |  |  |  |

## Appendix 8: Book level raw data

| α                  | Pre-intervention <sup>α</sup> |                         |                      | Post-intervention <sup>α</sup> |                         |                      |
|--------------------|-------------------------------|-------------------------|----------------------|--------------------------------|-------------------------|----------------------|
|                    | N.C. level <sup>α</sup>       | Book-level <sup>α</sup> | R.A <sup>α</sup>     | N.C. level <sup>α</sup>        | Book-level <sup>α</sup> | R.A <sup>α</sup>     |
| Child <sup>α</sup> |                               |                         |                      |                                |                         |                      |
| 1 <sup>α</sup>     | Towards 1C <sup>α</sup>       | 1 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1C <sup>α</sup>                | 3 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> |
| 6 <sup>α</sup>     | Towards 1C <sup>α</sup>       | 1 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1C <sup>α</sup>                | 4 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> |
| 7 <sup>α</sup>     | Towards 1C <sup>α</sup>       | 1 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1C <sup>α</sup>                | 3 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> |
| 8 <sup>α</sup>     | Towards 1C <sup>α</sup>       | 2 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1C <sup>α</sup>                | 5 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> |
| 10 <sup>α</sup>    | Towards 1C <sup>α</sup>       | 2 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1C <sup>α</sup>                | 4 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> |
| 12 <sup>α</sup>    | Towards 1C <sup>α</sup>       | 2 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 1B <sup>α</sup>                | 6 <sup>α</sup>          | 6.5-7.0 <sup>α</sup> |
| 13 <sup>α</sup>    | 1B <sup>α</sup>               | 6 <sup>α</sup>          | 6.5-7.0 <sup>α</sup> | 2A <sup>α</sup>                | 10 <sup>α</sup>         | 8.5-9.0 <sup>α</sup> |
| 14 <sup>α</sup>    | Towards 1C <sup>α</sup>       | 2 <sup>α</sup>          | 5.0-6.5 <sup>α</sup> | 2C <sup>α</sup>                | 8 <sup>α</sup>          | 7.5-8.0 <sup>α</sup> |
| 15 <sup>α</sup>    | 2C <sup>α</sup>               | 8 <sup>α</sup>          | 7.0-7.5 <sup>α</sup> | 2A <sup>α</sup>                | 10 <sup>α</sup>         | 8.0-8.5 <sup>α</sup> |
| 17 <sup>α</sup>    | 1B <sup>α</sup>               | 6 <sup>α</sup>          | 6.5-7.0 <sup>α</sup> | 2C <sup>α</sup>                | 8 <sup>α</sup>          | 7.5-8.0 <sup>α</sup> |
| 19 <sup>α</sup>    | 1B <sup>α</sup>               | 6 <sup>α</sup>          | 6.5-7.0 <sup>α</sup> | 2C <sup>α</sup>                | 8 <sup>α</sup>          | 7.5-8.0 <sup>α</sup> |
| 22 <sup>α</sup>    | 2C <sup>α</sup>               | 8 <sup>α</sup>          | 7.5-8.0 <sup>α</sup> | 2A <sup>α</sup>                | 10 <sup>α</sup>         | 8.5-9.0 <sup>α</sup> |

¶



## Appendix 9: Mann-Whitney results

Hypothesis Test Summary

|    | Null Hypothesis   | Test                                    | Sig.               | Decision                    |
|----|---|---|--------------------|-----------------------------|
| 1  | The distribution of Group reading test 1 standard score is the same across categories of Group.         | Independent-Samples Mann-Whitney U Test | .799 <sup>1</sup>  | Retain the null hypothesis. |
| 2  | The distribution of Group reading test 2 standard score is the same across categories of Group.         | Independent-Samples Mann-Whitney U Test | .514 <sup>1</sup>  | Retain the null hypothesis. |
| 3  | The distribution of Reading 3 is the same across categories of Group.                                   | Independent-Samples Mann-Whitney U Test | .799 <sup>1</sup>  | Retain the null hypothesis. |
| 4  | The distribution of Diagnostic spelling test standardised score is the same across categories of Group. | Independent-Samples Mann-Whitney U Test | .977 <sup>1</sup>  | Retain the null hypothesis. |
| 5  | The distribution of Diagnostic spelling test 2 standard score is the same across categories of Group.   | Independent-Samples Mann-Whitney U Test | .671 <sup>1</sup>  | Retain the null hypothesis. |
| 6  | The distribution of Diagnostic spelling 3 is the same across categories of Group.                       | Independent-Samples Mann-Whitney U Test | .799 <sup>1</sup>  | Retain the null hypothesis. |
| 7  | The distribution of Oral sentence 1 standardised score is the same across categories of Group.          | Independent-Samples Mann-Whitney U Test | .291 <sup>1</sup>  | Retain the null hypothesis. |
| 8  | The distribution of Oral sentence 2 standardised score is the same across categories of Group.          | Independent-Samples Mann-Whitney U Test | .039 <sup>1</sup>  | Reject the null hypothesis. |
| 9  | The distribution of Oral Sentence reading 3 is the same across categories of Group.                     | Independent-Samples Mann-Whitney U Test | 1.000 <sup>1</sup> | Retain the null hypothesis. |
| 10 | The distribution of Oral test word standardised score is the same across categories of Group.           | Independent-Samples Mann-Whitney U Test | .178 <sup>1</sup>  | Retain the null hypothesis. |

Asymptotic significances are displayed. The significance level is .05.

<sup>1</sup>Exact significance is displayed for this test.

### Hypothesis Test Summary

|    | Null Hypothesis   | Test                                    | Sig.              | Decision                    |
|----|---|---|-------------------|-----------------------------|
| 11 | The distribution of Oral test word 2 standardised score is the same across categories of Group. | Independent-Samples Mann-Whitney U Test | .078 <sup>1</sup> | Retain the null hypothesis. |
| 12 | The distribution of Oral word reading 3 is the same across categories of Group.                 | Independent-Samples Mann-Whitney U Test | .799 <sup>1</sup> | Retain the null hypothesis. |

Asymptotic significances are displayed. The significance level is .05.

<sup>1</sup>Exact significance is displayed for this test.