Responses to a School-Based Literacy Intervention for Poor Readers in Junior School. Evaluating Academic and Attitudinal Impact, using a RCT Methodology.

Patricia Jane Lucas
University College London
Doctor of Philosophy
Abstract

The acquisition of reading and writing is a key outcome of education; poor literacy is associated with difficulties throughout life. Many remedial reading programmes have been developed to address poor reading in school, but there has been a paucity of methodologically rigorous evaluations of such interventions. Where they have taken place, evaluations have been limited to either progress in reading and reading related skill, or to affective aspects of literacy. A small group intervention was introduced for children with very poor reading skills entering Year 3 (age 7/8 years) in a London school. A rigorous methodology (Randomised Controlled Trial, using waiting list controls) was employed to examine the effectiveness of the intervention on both achievement and affective dimensions. Little change was observed. Significant change was seen only in reading accuracy, and here the effect of the intervention was small (accounting for only 6% of the variance). There was some evidence that the children in the intervention group became less positive about books over the period of the intervention. A comparison between immediate and delayed (i.e. waiting list control) intervention groups showed that timing of the intervention did not affect progress. However, a significant proportion of children who were assigned to waiting did not go on to receive the intervention, either because of movement to other schools or because teachers' assessment of need changed. Follow-up data (up to 30 months post baseline) demonstrated that reading problems persisted in this group as a whole. An additional pilot study of functional literacy assessment was employed. Although the group identified as poor readers at age 7 years remained poor readers at age 9 years they were no less able, nor less likely to accurately follow written instructions to complete a cooking task.
Contents

CHAPTER 1 Literacy, Illiteracy and Educational Policy in the UK

1.1 Introduction ................................................................. 11
1.2 Poor readers in English schools ........................................... 11
1.2.2 Characteristics of the School ............................................ 13
1.3 Plan for Thesis ................................................................. 16

CHAPTER 2 The problem of poor readers; what do we know about intervening?

2.1 Introduction ................................................................. 19
2.2 The epidemiology of illiteracy: Prevalence and consequences of poor reading skills ................................................................. 21
2.2.2 Association between Poor Achievement and Poor Attitudes ............................................. 24
2.3 Reading acquisition and reading difficulties ....................... 28
2.4 What we know about how to intervene - Features of intervention programmes ................................................................. 37
2.4.2 The case for early intervention ............................................. 38
2.4.3 Student characteristics - General reading difficulties, specific reading difficulties and Dyslexia ................................................................. 43
2.4.4 The case for global training programmes versus specific skill training programmes ................................................................. 45
2.4.5 The case for small-group intervention ................................. 51
2.5 How do we assess effectiveness of intervention? ................ 52
2.6 Conclusions - the gaps in the literature .................................. 60

Chapter 3 Evaluation Methodology ........................................ 62
3.1 Introduction ................................................................. 62
3.2 Experimental Design ........................................................... 62
3.3 Selection of Subjects .......................................................... 64
3.4 Selection of Intervention Children ...................................... 67
3.5 The teaching programme ..................................................... 69
3.6 Standardised Assessment Tools .......................................... 71
3.6.1 Young Group Reading Test 3rd Edition (Young 1980) ....................... 72
3.6.2 Young Group Maths Test 3rd Edition (Young 1996) ............................ 73
3.6.3 British Picture Vocabulary Scale, Short Form (Dunn, Dunn, & Whetton 1982) .................. 74
3.6.4 Raven’s Progressive Matrices (Raven & Raven 1996) ......................... 75
3.6.5 Reading Attitude Scale (McKenna & Kear 1990) ........................... 76
3.6.6 Neale Analysis of Reading Ability (Neale 1989) ............................ 77
3.6.7 Phonological Assessment Battery (Fredrickson, Frith, & Reason 1997) ........... 80
3.6.8 Battle Self Esteem Inventory (Battle 1981) ..................................... 81
3.7 Non-standardised Measures ............................................... 83
3.7.1 Writing Sample ................................................................. 83
3.7.2 Teacher Assessment ........................................................... 87
3.7.3 School Attendance ............................................................ 88
3.7.4 Semi-Structured Pupil Interview ......................................... 89
3.7.5 Reading Self-Concept ........................................................... 91
3.7.6 Rating of Enjoyment for Reading & Writing ............................ 92
3.7.7 Reading Preferences ............................................................... 92
Chapter 4 Results and discussion of baseline data

4.1 Introduction ................................................................. 98
4.2 Sample ........................................................................ 100
4.3 Results & Discussion .................................................. 101
4.3.1 Response Rates ....................................................... 102
4.3.2 Standards of reading and other academic skills .......... 103
4.3.3 Study sample standards of reading and reading related skills .............................................. 108
4.3.4 Analysis by Gender .................................................. 122
4.3.5 Home language effects on performance ................. 124
4.3.6 Attitudes expressed toward books and reading, Self-esteem and general mood score (whole sample and study sample) ......................................................... 114
4.3.7 Relationships between variables ............................ 117
4.3.8 Characteristics of Intervention Group ...................... 119
4.3.9 Population trends as reflected in intervention groups ......................................................... 128
4.3.10 Post intervention differences ................................. 209
4.3.11 Baseline differences .............................................. 207

Chapter 5 Results of Randomised Controlled Trial

5.1 Introduction ................................................................. 133
5.2 The use of age adjusted scores .................................... 135
5.3 Post Intervention Results and Discussion .................... 138
5.3.1 Neale Analysis of Reading Ability ............................ 138
5.3.2 Phonological Assessment Battery ............................ 144
5.3.3 Writing skills .......................................................... 152
5.3.4 Reading Strategy Recall ........................................... 160
5.3.5 Pupil Attitudes ......................................................... 161
5.3.6 Attendance .............................................................. 174
5.3.7 Teacher Assessment ................................................ 176
5.3.8 Qualitative Feedback on interventions from pupils and from reading teachers ......................... 180
5.4 Conclusions ................................................................. 194

Chapter 6 Long-term literacy achievement in intervention groups

6.1 Introduction ................................................................. 198
6.2 Method ......................................................................... 201
6.2.2 Sample ..................................................................... 201
6.2.3 Data collection ........................................................ 202
6.2.4 Selection and use of measures ................................. 203
6.2.5 Functional Literacy assessment .............................. 204
6.3 Results and Discussion ................................................ 207
6.3.1 Baseline differences .............................................. 207
6.3.2 Post intervention differences ................................. 209
6.3.3 Functional reading task .......................................... 215
6.4 Conclusions ................................................................. 216
Index of Figures

Figure 2:1 Dual-Route Model of Word Recognition, Ellis & Young 1996.................30
Figure 3:1 Overall Plan for Evaluation.................................................................64
Figure 3:2 Eligibility of Pupils in Target Year Group...........................................66
Figure 4:1 Eligibility of Pupils in Target Year.....................................................101
Figure 4:2 Young’s Group Reading Test; Whole Sample Distribution...............104
Figure 4:3 Young’s Group Maths Test; Whole Sample Distribution......................105
Figure 4:4 British Picture Vocabulary Scale; Whole Sample Distribution..............106
Figure 4:5 Non Verbal IQ; Whole Sample Distribution........................................107
Figure 4:6 Neale Analysis of Reading Ability; Study Sample Distribution..............110
Figure 4:7 Phonological Assessment Battery; Study Sample Distribution...............111
Figure 4:8 Reading Attitude Survey; Whole Sample Distribution..........................115
Figure 4:9 Selection for Intervention; Distribution by Gender..............................128
Figure 4:10 Selection for Intervention; Distribution by Home Language.................130
Figure 5:1 NARA Accuracy; Scattergram showing change over time....................139
Figure 5:2 NARA Comprehension; Scattergram showing change over time ..........141
Figure 5:3 NARA Rate; Scattergram showing change over time............................143
Figure 5:4 PhAB Spoonerisms; Scattergram Showing Change Over Time...............145
Figure 5:5 PhAB Semantic Fluency; Scattergram Showing Change Over Time........147
Figure 5:6 PhAB Alliteration Fluency; Scattergram Showing Change Over Time......148
Figure 5:7 PhAB Rhyme Fluency; Scattergram Showing Change Over Time............150
Figure 5:8 PhAB Nonword Reading; Scattergram Showing Change Over Time........151
Figure 5:9 Writing Sample Story Length; Scattergram Showing Change Over Time...154
Figure 5:10 Writing Sample Vocabulary; Scattergram Showing Change Over Time....155
Figure 5:11 Writing Sample Spelling; Scattergram Showing Change Over Time........157
Figure 5:12 Writing Sample Content; Scattergram Showing Change Over Time.......158
Figure 5:13 Reading Strategy Recall; Change Over Time.....................................161
Figure 5:14 Reading Attitude Survey; Change Over Time....................................163
Figure 5:15 Battle SEI; Change Over Time.........................................................165
Figure 5:16 General Mood Score; Change Over Time..........................................166
Figure 5:17 Print Awareness; Change Over Time................................................167
Figure 5:18 Current Reading Self-Concept; Change Over Time...........................168
Figure 5:19 Future Reading Self-Concept; Change Over Time.............................169
Figure 5:20 Current Other Self-Concept; Change Over Time...............................170
Figure 5:21 Future Self-Concept; Change Over Time........................................171
Figure 5:22 Average Attendance Per Half Term................................................176
Figure 5:23 Teacher Assessment of Reading Fluency..........................................178
Figure 5:24 Teacher Assessment of School Adjustment........................................179
Figure 5:25 Pupil Feedback of Lesson Content..................................................182
Figure 5:26 Pupil Feedback of Best Things in Lessons.........................................186
Figure 5:27 Pupil Feedback of Worst Things in Lessons.......................................187
Figure 5:28 Pupil Feedback of Desire for More Lessons......................................188
Figure 6:1 Overall Plan for Evaluation...............................................................202
Figure 6:2 Instructions for Making Chocolate Crispy Cakes..................................205
Figure 6:3 18-Month Follow-Up; Mean Group Changes......................................212
Figure 7:1 Number of Entries in Reading Diaries................................................225
Figure 7:2 Relationship Between Comprehension Change and Baseline Vocabulary.243
Figure 7:3 Relationship Between Nonword Reading Change and Baseline Vocabulary .................................................................244
Figure 7:4 Relationship Between Accuracy Change and Baseline Vocabulary ...........246
Figure 8:1 Comparison of Selection Methods.................................................................263
Index of Tables

Table 2:1 Summary of Evaluation Methodologies ...............................................................58
Table 3:1 Battle SEI scores – US standardisation data .........................................................82
Table 4:1 Whole Sample Response Rates .........................................................................103
Table 4:2 Study Sample Response Rates .........................................................................103
Table 4:3 Phonological Assessment Battery; Correlation Matrix ......................................113
Table 4:4 Writing Sample; Whole Sample Data .................................................................113
Table 4:5 Self Esteem and General Mood; Study Sample Data .........................................115
Table 4:6 All Baseline Measures; Correlation Matrix .......................................................117
Table 4:7 Baseline Measures; Comparison of Intervention Groups to Class Peers ..........120
Table 4:8 Baseline Measures; Comparison of Intervention Group to Class Peers ..........121
Table 4:9 Baseline Measures; Comparison by Gender ......................................................122
Table 4:10 Baseline Measures; Comparison by Gender ...................................................123
Table 4:11 Baseline Measures; Comparison by Home Language .....................................125
Table 4:12 Baseline Measures; Comparison by Home Language .....................................126
Table 4:13 Distribution of Poor Comprehenders in Study Groups ..................................131
Table 5:1 Overall Plan for Evaluation ...............................................................................135
Table 5:2 NARA Accuracy; Pre/Post Intervention Data ....................................................139
Table 5:3 NARA Comprehension; Pre/Post Intervention Data ........................................140
Table 5:4 NARA Rate; Pre/Post Intervention Data ...........................................................143
Table 5:5 PhAB Spoonerisms; Pre/Post Intervention Data ...............................................145
Table 5:6 PhAB Semantic Fluency; Pre/Post Intervention Data .......................................146
Table 5:7 PhAB Alliteration Fluency; Pre/Post Intervention Data ....................................148
Table 5:8 PhAB Rhyme Fluency; Pre/Post Intervention Data ..........................................149
Table 5:9 PhAB Nonword Reading; Pre/Post Intervention Data ......................................151
Table 5:10 Writing Sample Story Length; Pre/Post Intervention Data ...............................153
Table 5:11 Writing Sample Vocabulary; Pre/Post Intervention Data ...............................155
Table 5:12 Writing Sample Spelling; Pre/Post Intervention Data .....................................157
Table 5:13 Writing Sample Content; Pre/Post Intervention Data .....................................158
Table 5:14 Modal Rating of Reading and Writing Enjoyment ..........................................171
Table 5:15 Reported Home Reading Frequency ................................................................172
Table 5:16 Teacher Assessment of Reading Fluency; Pre/Post Intervention Data .............177
Table 5:17 Teacher Assessment of School Adjustment; Pre/Post Intervention Data .......178
Table 6:1 Scoring System for Functional Literacy Task ....................................................206
Table 6:2 Baseline Differences - All Intervention and Comparison Groups .......................208
Table 6:3 18-Month Assessments ......................................................................................210
Table 6:4 30-Month Assessments ......................................................................................213
Table 6:5 Functional Reading Task Scores ........................................................................215
Table 7:1 Selection and Inclusion in Intervention Groups ..................................................227
Table 7:2 NARA; Comparison of Immediate and Delayed Intervention Groups .............230
Table 7:3 PhAB; Comparison of Immediate and Delayed Intervention Groups ..............232
Table 7:4 Writing Sample; Comparison of Immediate and Delayed Intervention Groups ..........................................................233
Table 7:5 Academic Achievement; Comparison of Immediate and Delayed Intervention Groups ..................................................................................................................235
Table 7:6 Attitudinal and Affective Measures; Comparison of Immediate and Delayed Intervention Groups ..................................................................................................................236
Table 7:7 Initial Analysis of Factors Affecting Change ........................................... 240
Table 7:8 Linear Regressions using identified likely factors ........................................ 242
Table 8:1 Comparison of Selection Methods ............................................................... 262
Table 8:2 All Attitudinal Measures .............................................................................. 269
Index of Appendices

Appendix 1 – Teaching Methods.................................................................306
Appendix 2 – Pupil Interview Schedule and constituent parts..................314
Appendix 3 – Self Concept Assessment Tool...........................................317
CHAPTER 1 Literacy, Illiteracy and Educational Policy in the UK

1.1 Introduction

It has been argued that the most important role which we ascribe to children in the developed countries today is as learners or, perhaps, consumers of education (Wagg, 1996). In the UK it has been argued that this interpretation of childhood followed from the 1870 Education Act and was a fundamental change from earlier understandings of childhood (Hendrick, 1997). Thus, education has been key to our understanding of childhood as a period of learning and as distinct from adulthood. The right to benefit from good education has become universal and the compulsory school enrolment is no longer contentious. The hope that underpins state financed education is that it will ensure that everyone growing up in Britain should be at least literate and numerate when they leave school. “Success” or “failure” in education is a matter of personal, community and national concern and recently there has been a growing realisation that the provision of schooling does not always ensure this outcome. Our societal concern is no longer with the number of children who are enrolled at school, but in the number of school leavers who achieve academic goals such as literacy. It has been shown that levels of literacy have been stable over the period 1948–1996 (Barber, 1997; Brooks et al 1999), and that around 15% of adult have limited literacy. If we are interested in decreasing this level we need to focus our attention to those who are not achieving adequate levels of literacy and establish the reasons for this and methods of remediation. Research has the potential to contribute directly to policy in two ways; to establish why, given similar environments, some children fail to learn to read and to establish what these children need in order to achieve adequate levels of reading skill. This purpose of this thesis is to contribute to the second of these questions by adding to the body of literature that provides solid evidence for the treatment of reading disabilities.
1.2 Poor readers in English schools

The context for this project was the increasing realisation of the importance of providing additional support to poor readers that was occurring nationally at the time the project began in 1997. This was the year that the National Literacy Strategy (NLS) was introduced by the incoming government. The Literacy Task Force, (Barber, 1997) set targets that by 2002 80% of all 11 year olds should reach level 4 on Key Stage 2 SATs. The introduction of the National Curriculum and Standard Assessment Tests by the previous government had made it possible to examine both national and local performance on tests of literacy against these markers. These changes were made in response to increasing evidence that in international comparisons children in British schools were performing poorly (Elley 1992). In 1997 only 57% of primary leavers were achieving the hoped for level 4. Thus, non-achievement of literacy (defined by failure to reach NC Level 4 in English) was found to be 43% in Britain compared to less than 5% in Australia and New Zealand (Barber 1997). In this report Barber notes that “a distinctive feature of British performance is the existence of a long ‘tail’ of underachievement” suggesting that it was not that some children were performing well, but rather that disproportionately many were doing badly. These performance findings reflected the results of earlier studies that suggested that there were problems with the standards of teaching of reading in many schools, where not all teachers or schools were following recommended practice (Cato & Whetton, 1991).

The NLS was idealised as the having the potential to reform along the lines of other studies e.g. “Success for All” (Madden et al., 1993; Slavin, 1997) as a three-phased approach. Firstly, improvements in initial teaching to address the needs of the 80% of children who would reach literacy without problems were to be made. This was to be achieved using the literacy hour, which aimed to increase the quantity of direct teaching of literacy and to standardise the content of teaching. A framework of teaching outcomes was recommended for all schools. Secondly, early one-to-one intervention would be provided for the 18% who experience early, but resolvable problems. Thirdly, the final 2% of children who have intractable and ongoing problems with reading that would then receive intensive interventions at KS2 (Beard, 1999; Beard, 2000). This provision should ensure that every child has more than one opportunity to benefit from
additional input. It took some time for this final stage to filter through to practice, and at the time this project began no centrally organised provision for poor readers at KS2 was made. During the final year of this project a programme of small group work within class (ALS or Additional Literacy Support) was brought in at Year 3 (aged 7/8 years), to be delivered by teaching assistants. It should be noted that ALS provides for small group teaching, in class, provided by learning assistants rather than qualified teachers. Third wave support providing 1:1 support from a qualified teacher is not yet fully operational. Beyond early years education (i.e. during primary and junior or KS1 & 2) there is little or no provision for poor readers.

1.2.2 Characteristics of the School

The school in which this project was based is a large local authority primary school in South London. At the outset of the project it had 543 pupils on the school roll, with a three-form entry, and a further 25 in its nursery section. This places it among the largest 21% of English primary schools. It is a school noted for its committed staff, innovative teaching and good practice results (Marshall, 2000; McAlpine, 1996) but it is located in an area of high social deprivation and despite its efforts its pupils have historically had poor achievement relative to national and local averages. The table below (Table 1:1) shows a breakdown of statistics for the school compared to local borough and nation-wide levels for eligibility for free school meals (which is used as a proxy measure of poverty), children with English as an additional language, children with statements of special educational need and KS2 SAT results. The years from which data is drawn are shown in brackets. Whenever possible, data is shown for the year 1997, when the project began. The figures demonstrate that the school has higher levels of social deprivation than local or national norms; higher than average numbers of children receiving free school meals and children with statements of special educational needs. The school, like many schools with similar profiles, has high levels of pupil turnover. At the time the project began 63% of children nationally achieved Level 4 in Key Stage 2 SATs. Level 4 is the benchmark set by government as the level that children are expected to achieve to demonstrate proficiency in this area. In this school the figure is lower than the national average, but higher than the borough average. Given the evidence of deprivation in the intake, this suggests that the school was
helping more children than other local schools to achieve this. This is confirmed in later PANDA reports, which indicate that SAT results for children from this school were in line, above or well above the scores of children in similar schools. Nevertheless, more than 40% of the children who left this school were not considered proficient in reading and writing.

<table>
<thead>
<tr>
<th></th>
<th>School</th>
<th>Borough</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Eligible for Free School Meals</td>
<td>64.9% (1996)</td>
<td>46.3% (1995-1998)</td>
<td>18.9% (National, 1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.1% (Inner London 1996)</td>
</tr>
<tr>
<td>% Children with English an additional language (1996)</td>
<td>35.4% 1</td>
<td>18.2%</td>
<td>7.8% (1997)</td>
</tr>
<tr>
<td>% Children with statement of special need (1996)</td>
<td>2.5%</td>
<td>1.7%</td>
<td>1.4%</td>
</tr>
<tr>
<td>% Children Reaching Level 4 or above in KS2 SATs for English (1997)</td>
<td>59%</td>
<td>53%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Table 1:1 Characteristics of Study School (1996/1996)(Source DFEE and LEA statistics)

The year previous to this inception of this project Her Majesty’s Inspectors had conducted a review of reading teaching in this and two other London Boroughs (HMI, 1996). Their conclusions were that several problems existed across these boroughs and good teaching was observed in only about a quarter of classes in each age group observed. Specific problems identified included lack of clarity about the “place and purpose of phonic work”, teaching to age groups for which teachers did not have adequate training and many teachers themselves felt they were not well equipped to teach reading. Using the Neale Analysis of Reading Ability (Neale, 1989) only 21% of Year 2 pupils were found to be reading at or above their age level, with the reading age of nearly 2/3 more than one year behind their chronological age.

---

1 35.4% EAL, the school records show where children speak a language other than English at home, rather than those for whom English is their second or additional language. This figure is therefore slightly inflated with reference to borough and national statistics.
The school approached the research team to place a joint bid for funding in 1996. The school had an existing relationship with both the funding body and researchers from an earlier study across a number of local schools examining the impact of a counselling intervention (Sherr & Sterne, 1999). The school was in the process of amalgamating the infant and junior school. The head (who was the existing head of the junior school) felt that a greater focus on success in literacy was important for the progress of many children in the school. He observed that those children who went on to drop out of secondary schooling and those who began to commit crimes in their teenage years were more likely to have left his school with reading and writing problems. A decision was made that the greatest changes needed to be made at KS2. Funding for provision of Reading Recovery to a small number of children in KS1 had already been secured the previous year, but intervention outside of classroom provision was relatively new to the school.

Researchers were asked, then, to advise on the design of the scheme to allow a thorough evaluation of its effects. The decision to place the intervention at entry to Year 3 was, therefore, decided in advance of the research programme using much of the evidence described in Chapter 2 (2.4.1). The advantages and disadvantages of this decision are considered there.

The task for research was to place the programme in the wider research agenda, such that lessons from it would apply to other schools and the policy community. In order to make recommendation to policy makers and practitioners there is a need to ensure that we are providing advice drawn from the best available evidence. The question of what the best evidence might constitute is considered in Chapter 2 (2.5) For this study to contribute to the body of literature on reading remediation a careful consideration of what it might achieve was made. Acknowledge the limitations of size, the recommendation to the school was that considerable effort would need to be placed on achieving a methodologically rigorous evaluation with a wide scope for examining change. The existing, co-operative relationship between researchers and school enabled this intervention to be incorporated into a research model.
1.3 Plan for Thesis

The aim of this project was to examine an intervention with poor reader in Year 3 for effects across a spectrum of skill based and attitudinal outcomes. The evaluation methodology selected for this task was a randomised controlled trial, using a waiting list control and follow-up planned for the length of the project. The thesis forms a description of the rationale, methodology, findings and conclusions from this research. The following chapters will be set out:

Chapter 2 – The problem of poor readers: what do we know about intervening?

There is a substantial literature on the nature and treatment of reading difficulties. In order to consider a contribution to this body of knowledge it is necessary to establish the background established by the existing work and identify question remaining for research. A review of the literature is necessary to establish these facts, and this forms Chapter two of this thesis. The quality of the evidence base to date will be examined with particular reference to the interventions employed, the outcomes included and the study groups selected. The need for work which employs a rigorous research methodology in an applied setting, which incorporates both academic and attitudinal outcomes and which looks at intervention with older children will be argued. The age group chosen for intervention in this case is at entry to Junior school (age 7/8 years) and the intervention is a small group intervention delivered out of class by a reading recovery trained teacher. The literature concerning evaluation methodology is also reviewed and conclusions about the methodology necessary are made.

Chapter 3 - Methodology

Leading from the conclusions of Chapter 2, the methodology for the evaluation of the intervention is described. Careful consideration was made of the most appropriate assessment tools was made, and a discussion of this process is also included.
Chapter 4 - Results and discussion of baseline data

The characteristics of the study group are described in some depth to provide the background and baseline for the study. Comparisons will be made between intervention groups and school peers and with the national age-group peers. It will be concluded that these children constitute a group with profound reading problems. These difficulties do not seem to have had a negative impact on their attitudes toward school or reading, or in their self-esteem and self-concepts. These findings are both with respect to their within school peers and with respect to external samples of children of the same age. This is taken to be an encouraging sign for the long-term adjustment of the children.

Chapter 5: Results of Randomised Controlled Trial

The core of this study is a randomised controlled trial over the period of the intervention. The impact of the intervention is analysed across the skill-based and attitudinal outcomes. Conclusions are drawn about the benefits of the intervention in this case, and the potential for generalisability.

Chapter 6: Long term literacy achievement in intervention groups

The progress of those who received the intervention is important in the longer term, as well as over the period of the intervention. The strategy employed to ascertain such possible changes is set out in Chapter six. The evidence for changes across domains of achievement already are assessed. In addition a pilot study for assessing functional literacy in this group is discussed.

Chapter 7: Determinants of progress in intervention groups

Examining change for the whole group it is possible to consider whether the intervention had differential effects on individual children. The aim of this analysis is to determine whether any variables could be identified which are predictive of change.

Chapter 8: Critical analysis of evaluation and findings

Having arrived at a considered methodology, and reviewed the findings of this study, it is necessary to determine whether the findings of this study would cause us to question some of the assumption made in
planning the evaluation. A critical analysis of the study methods and results will be undertaken in Chapter Eight. This will enable conclusion and recommendation for research and practice to be drawn from the work.
CHAPTER 2 The problem of poor readers; what do we know about intervening?

2.1 Introduction

In the context of poor academic achievement in school there are two questions that are key to developing policy and practice; is there a need for additional intervention, and what intervention would be appropriate? Compulsory education in the UK and in other countries is an intervention perceived to be for the good of the individual, the whole population and the economy. It is important that the content and structure of education should be informed by knowledge of the most helpful approaches for young people, both in terms of their achievement and their quality of life. For those young people for whom school is not proving to be effective, those with poor academic achievement, the received wisdom is that additional intervention is always advisable, but there remains considerable debate as to the form such intervention should take. This debate on the need and provision for intervention quickly becomes circular, since only effective interventions are needed, while intervention is only appropriate where a need has been identified. This thesis was devised principally to contribute to the question of effectiveness, recognising that effectiveness would have an impact on questions of appropriateness. The question posed within this thesis is; what evidence is there that intervention with poor readers and writers is more effective than non-intervention? In order to answer this question it is necessary to establish what evidence we have of the need for intervention and of the effectiveness of interventions for children with poor reading and writing skills to date. The background to this task is then threefold:
1. What evidence is there of the need for intervention?

A consideration of levels of illiteracy and the consequences of illiteracy will inform a debate on the need for intervention. It is important to establish why literacy in English is considered to be a key aim of the education system, and what happens to those who leave the UK education system not proficient in reading and writing. A consideration of groups who may be at increased risk of either poor literacy, or of poor outcomes associated with poor literacy, is also included.

2. What do we know about intervention for children with poor reading?

There exists an extensive body of literature regarding the acquisition of pre-reading and reading skills, the cognitive pathways of reading, writing and reading development. What bearing does this evidence have on the decisions associated with intervention (when and how to intervene)? What evidence do we have of the relative success of remedial treatment for reading difficulties to date? This information will be used to identify an intervention strategy that is likely to be successful, and to identify questions remaining regarding the effectiveness of intervention.

3. How can we demonstrate effectiveness and appropriateness of intervention?

Most of the literature on literacy intervention to date assumes that intervention is beneficial. This assumption has had important impacts on the nature of the research undertaken, and the conclusions drawn. It is important to consider the validity of this assumption. The available literature will be examined to determine the extent to which this question has been addressed to date, and to establish the most appropriate research methods for evaluating effectiveness.
2.2 The epidemiology of illiteracy: Prevalence and consequences of poor reading skills

The importance of achieving literacy has received much prominence in recent years amongst policy makers and academics (Barber 1997). This has been as the result of a growing awareness that illiteracy in adulthood is associated very strongly with long term unemployment, poverty, ill health and crime (Rice, Howes & Connell 1998). It has been estimated that men in the UK with poor levels of literacy are four times more likely to be unemployed than their counterparts with university level education (Literacy Trust 2001). Beard (2000; 1999) quotes a study conducted by Ernst and Young (1983), which found evidence that 60% of all UK jobs now require reasonable literacy. In comparison to this, research from the Basic Skills Agency found that 15% of 21 year olds in this country have limited literacy (Barber 1997). The cost of illiteracy has been put at over £10 billion per year (Beard 2000; Beard 1999). Although poor levels of literacy are not exclusive to Britain, it has been suggested that there is a distinctive “tail of underachievement” in literacy here, compared to other western countries (Elley 1992).

Moreover the “Matthew Effect” (Stanovich 1986) predicts that individual differences in standards of reading will grow wider with age in response to bi-directional responses to cognitive, reading related and attitudinal factors. This theory suggests that those who read well will, through greater exposure, acquire better skills, wider knowledge and more confidence in reading which will, in turn, facilitate better learning. In contrast those who struggle will not have access to these advantages and will fall progressively further behind (Nara-Bulter 1987). It has also been shown that poor school behaviour is associated with poor achievement at school (Adams 1990). Poor behaviour is likely to impair future school progress because attention on
educational tasks is likely to be reduced so a cycle of poor achievement resulting in poor
behaviour in turn harming achievement can be set up.

The problem of poor reading is not evenly spread across the population; poorer areas and
poorer families show poorer literacy (Muijs 1997). The relationship between poor outcomes
and low SES (socio-economic status) is complex however, and a review of literacy levels
compared to uptake of free school meals (used as a proxy measure for social disadvantage)
showed that some of the schools with highest uptake had score better in KS2 SATs than those
with very low uptake. However, these were the exceptions rather than the rule (Barber 1997).
In particular there are factors that may confound evidence of poor levels of achievement
associated with low SES. For example behaviour problems are more frequent in inner city
schools, and these in turn are also associated with poor reading acquisition (Jorm et al. 1986).

Inner city schools may be less likely to identify those who are performing below their ability
levels (Nara-Bulter 1987); even if children are still performing reasonably well, this
underperformance will depress average achievement scores in such areas. The mechanisms for
the relationships between better resources and higher school achievement may not be
immediately transparent. Children from non-literate environments are at a distinct
disadvantage. In her work with a non-literate family in the United States Victoria Purcell-Gates
describes this encounter:

"I suggested that be [the son of the family] make an instruction booklet . . . so that other children could make
kites if they wished. He looked at me dumbfounded . . . "Why?" he demanded . . . "Why, I'll show 'em if'n
they want to know!" he exclaimed. "Well, you may not be here when they are here," I proceeded . . . "Well, then
you can show 'em." (Purcell-Gates 1994) p.45-46
This young boy perceives that there is no purpose to written text, and may be unlikely to place much importance on learning to read or write. The lack of literacy in a family has profound consequences for the next generation, and closes off one of the main forms of communication we use to them. Innovative work has been undertaken to understand how the relationship between economic resources and educational achievement might come about. For example, in a small scale study in one UK inner-city school Stuart and her team (Stuart et al, 1998) examined access to books, home environment and reading achievement in children from "middle class" and "working class" homes. They found no significant differences in book ownership or access, or time spent listening to children read. The middle class parents did report spending more time reading to their children and spending more time playing "I Spy" games. The authors propose these experiences may have contributed to the acquisition of literacy in the children. Games such as "I spy" may have helped children to learn letter-sound correspondences and perhaps enhanced their phonological awareness and phoneme segmentation skills. It seems likely that many factors will be involved, but whilst the mechanism may still not be clear, the relationship seems well proven.

There has been a convincing body of evidence suggesting that boys have particular problems with reading and spelling (QCA 2000) and this evidence has not been restricted to the UK (e.g. Phillips et al, 2002). Recent research has begun to question the evidence which has placed boys at a disadvantage in reading. Davies and Brember found that boys are located at the extremes of performance, therefore whilst more scored very poorly, their average scores were higher than girls (Davies & Brember 1996). The evidence that this difference in boys is innate is scant and it is has been argued that it may be due to teachers' negative stereotyping of boys as non-readers (NFER 1999). Moreover there is a suggestion that it is the higher propensity of boys to
2.2.2 Association between Poor Achievement and Poor Attitudes

As well as academic and employment limitations resulting from poor literacy, there appear to be affective consequences of poor literacy. The relationship between failing to achieve in school, attitudes toward school, and self-perception are complex. Cross sectional studies have established that there is a link between poor performance and poor self-concept e.g. (Hay, Ashman, & van Kraayenoord 1997). Such cross sectional studies do not, however, establish causality. Do children who have poor achievement develop poor affective responses to school or does poor affective responses result in poor achievement? This relationship is likely to be complex and bi-directional. Affective dimensions of education such as emotional responses to school and perceived ability are thought to contribute to both personal and academic growth (Chapman & Tunmer 1997; Fontana 1997; Hansford & Hattie 1982; Muijs 1997). In particular it has been asserted that while poor reading self-perception is a result of early reading difficulties, this lack of self-belief inhibits later reading development (Chapman & Tunmer 1997). Low attainers lose confidence in their reading abilities and develop negative reading attitudes (Bridge 1990). These same studies however also, inadvertently perhaps, have shown that self-concept is a surprisingly stable attribute, Chapman & Tunmer report a highly significant path coefficient between self-concept between the ages of 5 and 6 (p<0.001) and between the ages of 6 and 7(p<0.001). Similarly, in a large (n=1001, across 51 schools) study
of Flemish primary aged children Muijs reports a regression coefficient for self-concept of 0.45 (p<0.001) between the ages of 9.5 years and 10.5 years. Early academic self-concept was the largest predictor of later academic self-concept. School grades were also significantly, although of lesser (regression coefficient=0.33) degree, predictive of self-concept at time 2. Importantly, both of these studies are prospective cohort studies, which means that they have demonstrated the pattern of change for individual children over time. The stability shown between time points using different methods of assessment suggests that while this factor may be associated with academic achievement the relationship is one which is difficult to break. An explicit aim of teaching is often to establish a method for addressing self-concept as well as academic knowledge and this aim is one to which global training regimes for reading intervention subscribe.

A meta-analysis of 128 studies looking at the relationship between self-concept of performance and actual performance showed an overall significant relationship between them. However, correlations ranged from 0.2 to 0.5 with an average of 0.22 that, while significant, is not a particularly strong correlation. There was evidence that age affected the relationship, with the strength of relationship to grow stronger with age, peaking at secondary schooling (Hansford & Hattie 1982). There are considerable problems with the sampling strategy used in the meta-analysis and its specific conclusions need to be treated with caution. It appears a decision was made, and then abandoned, to use a random sample of the 702 studies identified in the published literature in this area. Instead a compromise between random selection and convenience of retrieval was adopted. Methodologies for systematic reviewing would reject this approach (Egger et al., 2000; Sterne et al., 2000). However, in summarising 128 studies, accepting that the sampling may have been somewhat biased, the relationship demonstrated
was very largely positive. Of the 1136 effects sizes produced 944 were positive, 22 zero and
170 negative effects, with a range of -0.77 to 0.96. Thus, although the conclusions regarding
sub-group analysis might be questioned since these will have smaller samples, the overall trend
for relatedness seems to be acceptable. The pooled conclusions of 128 studies, even when not
systematically collected, is likely to be representative. It would be more informative if a smaller
sample of the 702 identified were examined, by looking for all those referring to a particular age
group, or by identifying acceptable methodologies in advance (e.g. prospective cohort only).

Work attempting to determine whether the 'Matthew effect', or the increasing gap between
good readers and poor readers over time, takes place and to examine its causal mechanisms. In
a prospective cohort of children in the Netherlands Bast & Reitsma (1998) demonstrated
relationship between later reading skill and early factors such as word recognition and
vocabulary, which suggest that those with good early skills make significantly faster progress
than those without. They proposed the importance of indirect effects on literacy achieved, for
example home literacy did not appear to have a relationship with reading achievement, but was
significantly related to vocabulary which in turn was related to reading achievement. Similarly,
word recognition skills predicted both frequency of leisure reading (indicating attitudes to
reading) and later progress, thus ease of reading may mediate attitude, which in turn contributes
to exposure and progress. The model emerging from this study is one of complex
relationships between many aspects of reading, and affective variables. There is some difficulty
in drawing conclusions from this study to the general population, since the subjects selected for
inclusion in the cohort were those with the best and worst early skills. A stronger case would
be built by examining results across an entire age cohort. It is conceivable that mechanisms
that operate for those with particularly good or particularly poor skills are exceptional in some
unknown way. The reasons for this focus on good and poor pre-readers are clear.

Whilst we know about something of the relationship between poor reading and poor reading attitudes we know less about the impact of intervention on this relationship. There is a shortage of studies which attempt to assess both the academic or reading achievement and the attitudinal or affective impact of interventions. Intervention studies predominantly focus on reading or reading skill changes in the intervention groups. In section 2.5 below a list of 12 recent intervention studies is made, of these only four studies attempted to assess any outcomes that could be viewed as representing affective changes in children, e.g. attendance, or teacher rated behaviour (Chapman, Tunmer, & Prochnow 2001; Madden et al. 1993; O' Shaughnessy & Swanson 2000; Somerville & Leach 1988; Wade & Moore 1998). In the light of the body of literature concerned with the association between attitude and performance it would seem important to establish whether changes in academic performance can be seen in response to changes made in attitude or vice versa. If poor attitude is caused by poor performance, one might hypothesise that a scheme to help struggling readers is likely to improve their attitude toward reading, conversely if poor attitudes results in avoidance improving attitude and increasing exposure might result in improved performance. Intervention studies need to address both these variables in order to look for such relationships.

It seems clear that there is a case for intervening in poor reading performance if successful intervention strategies can be identified. Poor population literacy has been shown to be a result of restricted opportunities for groups within that population, and to contribute to the continuance of groups with restricted opportunities. Consequences of reading difficulties include immediate consequences in school of poor access to education, poor self-esteem and
poor attitudes to school and continue into long-term impacts in poverty and unemployment in adulthood. It seems reasonable, then, to assume that a need for intervention has been demonstrated within the UK.

2.3 Reading acquisition and reading difficulties

When instigating an intervention it is necessary to consider the evidence available for strategies that are likely to be most successful. Over the past two decades our understanding of the cognitive processes involved in reading has increased greatly. This process has begun to inform practice in teaching reading and writing both in initial teaching and in remedial teaching. The task of reading has been described in its constituent parts and each step has been considered in order to arrive at a unified theory that can explain both the processes by which this skill is acquired, and the processes that skilled readers use when reading. Comprehension of text and decoding of text are clearly equally important tasks to describe, but decoding necessarily underpins comprehension and it is here that an understanding of the cognition of reading must begin. In order to consider the impact of research and theory on teaching practice a brief overview of the issues that arise within this field is given here and the implications of this evidence for selecting programmes of teaching will be made explicit.

Considering this task of “decoding” words in isolation from comprehension for the present, a skilled reader of the English language can be observed to do two tasks when reading a new piece of text. Firstly they read words they know well and have read many times, but it is likely that they will also be asked to read words that they read seldom or have never read before. Such unknown words can be read without being recognised for meaning. A model of reading
cognition must therefore be able to explain how it is we can read words that have no meaning, as well as words that do have an immediate meaning for us. Expert readers can read words that do not exist (non-words) if they are written within the regular rules of English. This might be thought of as a “code-breaking” skill, written words clearly contain information on pronunciation that does not correspond to meaning. We can also, however, read a large number of words that comprise alphabetic patterns which do not follow these rules, names are often particularly good examples of this e.g. Gloucester, the correct reading of this word obviously does not follow the written pattern and we are using the meaning of the word alone to inform pronunciation. Cognitive models of reading have attempted to explain how these two reading skills may equate to processing skills.

In response to these two reading skills, cognitive models were developed which rely on a dual route model of reading; that two routes exist for reading written language - a fast track visual route for familiar words, and a decoding system relying on letter (or grapheme) to sound (or phoneme) conversion. In this way some, regular unknown words can be read via grapheme-phoneme conversion and known words can be read via the visual input system. Both these systems feed into a comprehension or semantic system that will provide meaning for known words. Research evidence leading to this conclusion was reviewed and synthesised to create the following summary figure (Figure 2:1) relating to a cognitive model of word understanding (Ellis & Young 1996) of which the right hand side corresponds to reading. We can see the suggestion that two separate systems apply, where words recognised on sight are processed through a “visual input” lexicon whilst unknown words take a route through conversion at the level of sounds to establish whether it is a known or unknown spoken word. Evidence for this distribution of routes is provided by studies which have shown that well-known words are read

...
more quickly than novel words, even when the known word is irregular (Ehri 1995). Many models have been developed of which this is cited as an example, but while these models vary in their organisation they agree on their underlying theory of “dual processing”.

**Dual-Route Model of Word Recognition, Ellis & Young 1996**

![Dual-Route Model of Word Recognition](image)

Evidence for such a model has been provided by impairment studies, in which it has been shown that impairment to individual pathways has effects on decoding, for example studies of brain damaged patients who can spell words in phonetically plausible ways, but lose the ability to produce correct irregular spellings (Beauvois, Derouesne, & Saillant 1980) or of patients who
can spell real words correctly, but lose the ability to produce phonetically plausible versions of nonwords (Shallice, Warrington, & McCarthy 1983). Evidence from priming studies have suggested that these routes are not entirely independent. For example, (Barry & Seymour 1988) showed that when the spellings of non-words was contingent on previously presented real words, common spelling patterns were more likely to be used but this could be influenced by presenting real words with less common spelling patterns. That is, that the lexical route can be shown to affect the sub-lexical route. In general the division between phonetically “regular” and “irregular” words has been criticised as over-simplistic and it is thought that the unit of analysis is likely not to be at the level of phonemes but at the level of phonetic patterns or, as Nunes, Bryant & Bindman have suggested more recently, morphemes (Nunes, Bryant, & Bindman 1997a; Nunes, Bryant, & Bindman 1997b; Nunes, Bryant, & Bindman 1997c). Notwithstanding these refinements of the model, the power of the underlying theoretical model of a combination of lexical and sub-lexical routes for reading and spelling is still relevant.

Cognitive psychologists argue about whether this model represents two entirely separate and independent processing routes (Coltheart et al. 1993; Coltheart & Rastle 1994; Stanovich 1991) or whether it represents a single processing route where these processes are in some way unified (Norris 1984; Seidenberg & McClelland 1989; Seidenberg & McClelland 1990; Wagner et al. 1997). These cognitive theories have been developed into “stages” of reading development, wherein the argument between discrete or unified models continues (Snowling 1998; Stuart & Coltheart 1988). These differing views on processing become important when considering the teaching of reading. Notterman & Drewry (1993) looked at the ways in which theoretical approaches to the cognition of reading may influence reading instruction. The question which emerges is
whether in early teaching we need to emphasise processes which rely on exposure to words and meaning (corresponding to the visual processing route) or processes which make explicit grapheme/phoneme correspondence, blending and production. Do we have reason to suppose that either of these processes is more difficult to acquire, is more amenable to additional support, or whether teaching either approach alone is sufficient? We need to establish how these models correspond to teaching practice, and how findings from reading acquisition can, in turn, inform such models.

The process of sight-recognition has been described in terms of phases of development (Ehri 1995; Frith 1985) which make explicit the stages in which sight recognition of words becomes increasingly mature. It is postulated that children learning to read use a variety of strategies to recognise words, beginning with features of the word other than the alphabetic cues (such as word shape) moving through to increasing use of alphabetic cues and moving to a stage when whole words and letter patterns occurring across several words are easily recognised and processed. A consequence of this sequence of learning is that exposure to written words is crucial in the acquisition of this skill. As exposure increases both the number of words recognised and the speed of reading will increase, impacting on reading fluency and comprehension. It seems likely that this staged reading acquisition will rely on self-learning facilitated by suggestions of cues to word recognition. Teaching in this realm is likely to focus on a “top down” approach to reading where the repeated exposure is crucial, but there are no explicit rules or descriptions of strategies that can be taught. A group of teaching techniques such as “whole language” or “real books” approaches (Goodman 1982; Smith 1978) rely exclusively on such top down approaches. Such teaching theory was particularly popular in the 1970s and 1980s, but still remains in some contexts such as “whole word” or “look and say”
approaches. The value of these teaching strategies is the emphasis on the nature of written words as entire units and the function of written language as a whole. However, as teaching strategies in the absence of more direct instruction they have been largely discredited as ineffective (Oakhill & Beard 1994; Turner 1990a; Turner 1990b). Turner pointed out that reading and writing are culturally transmitted phenomena, not products of any innate skill and therefore are likely to need direct instruction, making explicit the strategies required in reading. Later work, particularly that by Oakhill & Beard (1994) showed that the top down, or “apprenticeship” approaches themselves were creating difficulties for children in learning to read. These approaches leave children to construe the initial steps for themselves and researchers have demonstrated that top down approaches alone are insufficient to equip all children with the skills needed to become expert readers.

We must then consider the evidence in favour of bottom up approaches applicable to the model exemplified in Figure 2:1 above. A clear candidate for exploration in this context is the process by which the written letters are converted into recognisable sound units. It would appear that the key ability underlying this pathway is to be able to segment words into their constituent sounds (phonemes) and to be able to manipulate these. The huge importance of such phonological skills in learning to read has been robustly demonstrated (Bryant, Bradley, Maclean M., & Crossland 1989; Bradley & Bryant 1978; Bradley & Bryant 1983; Bryant et al. 1990; Goswami & Bryant 1990; Maclean, Bryant, & Bradley 1987; Sterne 1996). It would seem that phonological skills are the “code-breaking” skills necessary to classify and process written language. In response to this body of evidence both early reading and remedial reading programmes have incorporated training in phonological awareness and focus on making explicit grapheme-phoneme relationships (for example see Stuart 1999). Evidence is clear that
children who struggle with these skills are at a disadvantage when learning to read (Bradley & Bryant 1983; Bryant et al. 1989; Maclean, Bryant, & Bradley 1987), and that explicit teaching of skills promotes reading development (Bus, van IJzendoorn, & Pellegrini 1995; Byrne & Fielding-Barnsley 1995; Hatcher, Hulme, & Ellis 1994). Recent evidence has explored the role of morphology (units of meaning, for example ‘-ed’ meaning past tense) on written language development, arguing that morphology is another system that children use when learning to write (Nunes, Bryant, & Bindman 1997; Bryant & Nunes 1998; Treiman & Cassar 1996). As such evidence builds it may be appropriate to incorporate this too into teaching strategies. If the importance of these underlying skills to reading acquisition is accepted, then one implication is that early teaching and early intervention are crucial. If what holds children back from achieving reading fluency is underlying, pre-reading skills then clearly these need to be addressed before progress in reading can be expected.

Narrow, bottom up approaches have been criticised as not providing the exposure to books considered necessary to promote literacy. As mentioned earlier, “decoding” of words may be necessary to the acquisition of comprehension, but without comprehension it does not enable a child to read. Moreover, exposure to books increases reading practice and vocabulary both of which are likely to have a significant impact on skilled reading acquisition. Finally, as discussed in the section 2.1.1, the impact of reading experience on attitudes to reading is important as this will also effect reading practice.

Skills-based bottom-up programmes may avoid this problem in that early, successful reading books will promote literate behaviour and lead in turn to an enriched book exposure. Since the value of book exposure has not been disproved, it begs the question whether narrow training programmes remain appropriate at older ages. The process of learning to read becomes self-
limiting at the point when a child fails to learn to read. Exposure to written language will be
reduced relative to those who are reading fluently, and remedial programmes may need to
explicitly address exposure to print. For example there is evidence that it is lack of flexibility in
approaching the task that marks poor readers (Sprenger-Charolles 1991) implying a role for
practice. Favoured programmes for children who have failed to learn to read have therefore
included elements of promoting affective responses and increased exposure to books.

Despite this body of evidence, there remains debate on the approach that teachers themselves
take in the classroom in England. It has been found that teachers do not rely exclusively on a
single approach and prefer for teaching reading in an indirect way, rather than by direct
instruction (Blatchford, Ireson, & Joscelyne 1994), and that teachers of young children in
English schools in particular still hold views which could be identified as “apprenticeship” or
“real books” approaches. These methods are felt to contribute toward the affective response
of children to books and contribute to the “literate environment” that is a pre-requisite for
reading and writing development (Sulzby & Teale 1991). Exposure to reading practices does
appear to play an important role in both acquiring reading skills (Bus, van IJzendoorn, &
Pellgrini 1995) and comprehending the purpose of written language (Purcell-Gates 1994). The
impact of the literacy hour introduced in 1998 and its more prescriptive approach to teaching
methods is not yet clear. Although success for the National Literacy Strategy as a whole has
been claimed (OFSTED, 2002) changes cannot be attributed to any particular aspect of the
NLS, nor can the possibility of improvements in results stemming from improvement in
training for the tests rather than improved literacy be discounted. It may take some time for
teachers to form an opinion of the usefulness of the methods recommended within the
national literacy strategy and to accept some of its recommendations as standard practice. It
has long been recognised that empirical evidence takes some time to change attitudes to practice (Rieben & Perfetti 1991).

As well as considering the cognitive tasks involved in reading and spelling, cognitive psychologists have contributed significantly to an understanding of the progression of skill acquisition (Snowling 1998). By understanding the way that reading development changes over time, and how skills may depend upon or support others, we may gain insight into the most efficacious teaching methods. Once again, two alternative theories have been proposed. One camp of psychologists have concentrated on a stage model of reading and spelling development exemplified by the model developed by Frith (1985). These models suppose a sequential acquisition of skills, whereby skills in reading spelling alternately rely on and support each other, moving through increasingly more cognitively sophisticated processing. In Frith's model, initial acquisition is in the visual domain supported by partial cues, moving through alphabetic spelling, which supports their reading, which in turn supports a move to the use of orthographic and morphologic strategies. Whilst stage models provided a useful description of the individual stages, the concept of linear movement through these stages by individual children has been challenged since not all children were seen to follow this pattern. Moreover, researchers pointed out that this pattern of acquisition seemed to be influenced by teaching practice and according the language they were learning (Johnston & Thompson 1989; Wimmer & Hummer 1990). However, the theories discussed above concerning the importance of phonological skills as a pre cursor of the acquisition of reading and writing in alphabetic languages does seem to be universal. Later models of reading and spelling acquisition have not favoured explicit descriptions of discrete stages, but modelled a connectionist approach, mirroring developments in other areas of cognitive psychology, whereby initial skills are
required in order to set up a structure which is built upon using reinforcement at decision nodes (Seidenberg & McClelland 1989). In other words a parallel processing framework is suggested, where training and practice reinforce particular processing routes. If we consider this as the model for reading acquisition it has important consequences for the teaching of reading. This model would suggest that explicit teaching of initial skills supported by subsequent intensive practice and supervised learning would be the most successful model for teaching. In this context reading disorders can be modelled by constraining the processing framework (Snowling 1998) and identification of these constraints for each child would be seen as crucial to treatment. However, the concentration on the "decoding" features of reading acquisition needs to be balanced with an awareness of the function of reading, whereby acquisition of vocabulary and comprehension are equally as important.

2.4 What we know about how to intervene - Features of intervention programmes

As acknowledged by Tunmer, Chapman and colleagues (Tunmer et al. 2002), the wealth of information we have amassed on how children learn to read can inform practice, but doesn’t address the question of how we should teach children to read. Awareness of the key stages and skills that are necessary to advance, and sensitivity to indicators of problems will aid educationalists to plan strategies for teaching. However, other research resources must be called upon to pinpoint which teaching practices are likely to help children to achieve reading goals. In the context of this thesis, of particular interest is the research concerning how to teach children who are having difficulty learning to read and how to remediate established reading difficulties? There are a number of features of intervention programmes that we can
consider in this context to identify what strategies have been shown to be successful or appropriate to this context. Those discussed here are timing of intervention, student mix, global vs specific training programmes and teacher/student ratios. The crux of the question we must pose as researchers is not only to identify what existing research has concluded, but to identify pertinent questions which deserve research attention.

2.4.2 The case for early intervention

The balance of research evidence supports the view that early intervention is more effective than later intervention (Hiebert & Taylor 2002; Hurry 2000). The definition of “early” intervention is dependent on the school system, but can be thought of as intervention within the first two years of school entry, thus in the UK an intervention is deemed to be early if it takes place within Key Stage 1. As an example of the differential impact of earlier intervention we can consider the US programme Success for All, which provides support across a number of primary school years (Madden, Slavin, Karwiet, Dolan, & Wasik 1993). This research concluded that intervention was most successful when provided early and for those children in the lowest quarter of their cohorts. This team also pointed to the importance of intensity of intervention to gain success. The pattern of early intervention producing better results is not surprising in the light of the research evidence presented above that reading and spelling acquisition are built upon earlier skill acquisition. However it is also true that early intervention is not sufficient to secure success, other factors must be taken into consideration. Further the research evidence does not necessarily point to later intervention being unsuccessful, but results are mixed.

When intervention is delayed, the evidence to support its implementation is lacking. Interventions in Key Stage 1 (up to the age of 7 years) have been exhaustively researched and
evaluated (for review see Hurry 2000), but far less work has concentrated on even slightly older children. In recent years this lack has been acknowledged and more work has been undertaken at Key Stage 2 (Clipson-Boyles 2000; Fawcett et al. 2001). The National Literacy Strategy (NLS) assumes that intervention would occur early (in Key Stage 1) for most children, leaving few children in need of continuing and ongoing support into Key Stage 2 (DfES, 2003). This strategy is built on the premise that early intervention is the more successful option. The argument over whether to intervene with poor readers early or late has attracted much interest and research. The potential to prevent generalisation to other areas of the curriculum by resolving reading problems early is seen as a key benefit to early intervention. In addition the teaching process may be easier while the child's behaviours are less entrenched. These arguments could be summarised by saying that early intervention will prevent a later escalation of difficulties. In addition there is the consideration voiced earlier in this chapter of 'Matthew effects', that is that the gap between good and poor performance will become larger as children progress. However, the success of early intervention relies on our ability to reliably identify those with reading difficulties while still early in their progress in reading. Inaccurate identification will either tend to over-include children whose difficulties may resolve without additional support or to under-include children whose difficulties are not yet apparent. These difficulties are inherent in the assumption that there is continuity between early and later skills, that a child who has an early reading problem will necessarily have a problem later in reading and vice versa. This assumption may not be true. For example, in a study of 187 pupils in Canada over 6 years Phillips and colleagues (Phillips et al., 2002) found that categorisation as a "below average" reader in Grade 1 did not predict categorisation as a "below average" reader in Grade 6 on standard assessment tests. Although none of those in the below average group became better than average readers, and no above average readers became poor readers, of
nineteen pupils categorised as below average readers in Grade 1, 9 were still classified as below average readers and 10 as average readers 5 years later. In fact the pattern of scores over time showed a regression toward the mean, with progressively less variation between scores at each grade assessment. Whilst the authors of this study comment that their sample size is larger than other studies, it remains a very small number for a longitudinal study demonstrating causality. The study is also constrained by its focus on only a homogenous “socio-economic environment”, in whole language classrooms, including no teaching of phonics. It is possible that greater numbers of children were classified as “below average” readers early in their school career because of the teaching style adopted, and in other contexts fewer would have experienced early problems which they later overcame. Only a small proportion of children moved from an “average” to a “below average” categorisation (18 of 139; 12.9%), but this is greater than the number of “below average” who remained so. Children in these schools would have been ill-served by a single early intervention, as 10 of those who didn’t go on to need help would have received it against 18 who developed problems only later.

There are a number of pragmatic and theoretical reasons why, despite its desirability, early intervention may not be appropriate for some children. Pragmatically, schools may not offer early intervention because of lack of resources, or children may not be identified at assessment for early intervention. There may be particular problems for children with high mobility between schools. In itself frequent school movement has a negative impact on reading achievement, with frequent movement accounting for about 50% of later differences between mobile and non-mobile children (Temple & Reynolds 1999). These mobile children are also more likely to miss out on early intervention where offered, simply because they are not at a
school when it is offered, or because they be less likely to be identified as poor readers because of lack of continuity of schooling.

Allowing for circumstances in which early intervention is available given all resources available and a non-mobile population, it may be important to offer late intervention in addition to early intervention since effective early intervention hinges on the ability of teachers to recognise those at risk of reading failure early in their schooling. In her review Hurry concludes that the best opportunity for intervention is when it is introduced early, using transparent assessment tools to identify those at risk (Hurry, 2000). What the most appropriate assessment techniques might be is not clear, and Hurry acknowledges that there is a lack of literature examining the “importance of assessment per se” (p.24). An example of a study examining the difficulties of early identification of reading problems is given in Mantzicopoulos & Morrison (1994). In exploring the predictive ability of one early identification tool, they found a high proportion of both false negatives (50%) suggesting that this tool under-identified those with reading problems. This single study is limited in its scope assessing only one assessment tool among 270 children in California. Other studies making such a direct comparison between prediction and outcome were not found, although such data may have been buried in other studies not located. Although it would be useful to assess a cohort of children across a number of early identification tools and track progress, the ethics of identifying need and not addressing it make such a study unlikely. The difficulty of recognising need in school samples is exacerbated by the relationship between behaviour problems and reading problems. Poor behaviour is likely to mask underlying problems leading to late recognition of these problems, but conversely poor behaviour is more common in children who are struggling at school.
Whilst we may accept the evidence for the better outcomes associated with early intervention, we can see that this doesn’t negate the need for evidence as to the effects of later intervention. It would be instructive to know the effects of later intervention on those who have already received early intervention, and the effects of later intervention for those who have not received early intervention. Either of these strategies for research implies the selection of an intervention timing clearly later than that usually recommended within early intervention strategies, selected for its validity within a school setting. In this case the timing of the intervention selected was the transition from Key Stage 1 to Key Stage 2. The appropriateness of this point is confirmed by the subsequent introduction of Additional Literacy Support with the NLS in the same age group. The transition from Year 2 to Year 3 (aged 7 rising 8 years) may be a particularly key time for an intervention. This is the age at which most children are assumed to have reached functional literacy (Meek 1982). It is also thought that “any remedial approach introduced once a child is eight or more is likely to be both much less effective and much more costly” (Barber 1997). There is a substantial change in curriculum and social demands made on children at this stage (Clipson-Boyles 1997; Clipson-Boyles 2000) and those unable to tackle these new demands are likely to fall ever further behind. The research project considered in this thesis was constrained to deliver the intervention at entry to Year 3. This decision was made by the school involved, informed by much of the debate above, but also in response to the specific needs of the school. Thus, early intervention with its greater likelihood of change was unavailable as a strategy. This forced choice represents an interesting context for research, however. It is an accepted limitation of many research programmes that they are delivered as a demonstration programme, and the implementation of these interventions in
real-world conditions is likely to be problematic. This point is discussed further when considering the implications of the study in Chapter 8 (8.2.2).

2.4.3 Student characteristics - General reading difficulties, specific reading difficulties and Dyslexia

When faced with children who are having difficulty acquiring fluent reading skills there is a question as to whether to attempt to diagnose and address specific learning disabilities or to put in place a general teaching programme with the aim of teaching universal skills. We can ask whether identifying particular groups of children, such as those with phonological dyslexia, would be of help when planning an intervention, and whether they are likely to benefit from a general intervention programme, rather than a specific programme for phonological dyslexia?

Early assumptions that discrepancy between reading abilities and IQ distinguish children with specific learning disability (SLD) are not universally accepted (Fletcher 1992). Rutter & Yule (1975) established this principle that disparity between IQ and reading suggests SLD. This relies on the assumption that IQ predicts reading achievement (Naglieri 2001), and therefore a discrepancy between progress in reading as predicted by IQ and actual reading progress implies that an additional explanation of the reading difficulties is required. This additional explanation was interpreted as a specific difficulty with reading acquisition. The original work of Rutter and Yule has been criticised for not excluding children with neurological disorders from their sample. Later work showed that the bimodality demonstrated by Rutter and Yule was not apparent when more stringent inclusion criteria were in place (Shaywitz et al. 1992; Share et al. 1997). The view that IQ predicts reading achievement has also been shown to be unstable (Vellutino 2001; Vellutino, Scanlon, & Lyon 2000). Fletcher (1992) summarised the evidence
for the relationship and concluded that "differences between children . . . are either non-existent or small, and of questionable significance". Certainly a simplistic definition of difficulties implied by the IQ discrepancy argument is unlikely to hold true. Recent attempts to use verbal IQ to determine likelihood of response to intervention, or to identify different patterns of need in intervention have also shown that IQ discrepancy is not a reliable marker for reading difficulties (Stage et al. 2003). If we are attempting to use IQ as a diagnostic tool, we need evidence that it distinguishes between different "types" of learning disabilities evidenced by patterns of difficulty or remediation common to those with such "types". The evidence to date does not support the use of IQ in this fashion.

In addition to the difficulty of using simplistic diagnostic heuristics in any one school there are likely to be few children with any one diagnosable specific learning disability. Those with severe difficulties and clearly diagnosable problems are likely to have progressed within the special needs system to a point with additional support is available. This point was referred to as receiving a statement of special needs, or to have been designated as "school action plus" in the new SEN code of practice. Inner city schools in deprived areas have been shown to have greater proportions of poor readers, and greater numbers of children with special educational needs. In this context the efficacy of attempting to distinguish between children with reading difficulties according to possible cause and offering tailored reading programmes would be slight. It is unlikely that a strategy relying on extensive diagnostic testing and differential teaching methods would be practicable in any single school. If this strategy was preferred it implies that provision would be made across a local authority area, pooling resources and providing teaching to children across schools. In this intervention study it was decided that a general reading programme would be administered since this would seem to meet the need in
the population best. Non-verbal IQ was included as a baseline measure in order to be able to address some of the issues raised by this discussion.

2.4.4 The case for global training programmes versus specific skill training programmes

There has been some success recorded for specific skills training programmes (e.g. Bryant & Bradley 1995; Greaney, Tunmer, & Chapman 1997). However, when specific skill training is combined with general reading training it has been found to be most beneficial. Hatcher, Hulme & Ellis (1994) found that when comparing progress of 7 year old poor readers, a phonology training plus reading training programme was more effective than phonology or reading alone in terms of progress in reading. The reading plus phonology group showed a significant improvement relative to controls on Neale Analysis accuracy and comprehension. The differences between reading only and phonology only groups and the control group were found to be non-significant. More recently reviews of evidence have also concluded that phonology training is an essential aspect of training programmes, but is not sufficient to ensure progress (Bus & van IJzendoorn 1999; Hurry 2000). Indeed, Hurry (2000) concludes that spelling practice may be an important aspect of reading instruction.

A case has been advanced for separate interventions for children with comprehension difficulties (Cornoldi & Oakhill 1996; Oakhill & Yuill 1996; Rupley & Willson 1997; Yuill 1996). For some years Oakhill and Yuill have argued that children with poor comprehension skills are often under-recognised and under-served by current intervention research. They have produced evidence that some children do not have particular difficulties decoding text
accurately, but are less able to understand the text as a whole. These poor comprehenders are more likely to benefit from specific skill training in this area. Poor comprehenders are thought to differ from good comprehenders in their word recognition skills (Haller, Child, & Walberg 1988), inference making skills, understanding of story structure and comprehension monitoring (Cornoldi & Oakhill 1996), although word recognition skills alone are unlikely to identify poor comprehenders. A meta-analysis of meta-cognitive or metacomprehension training programmes has been shown to have a significant effect on comprehension skills across a number of studies (Haller, Child, & Walberg 1988). There is little doubt that included within a group of poor readers will be a group of poor comprehenders, and it may be that for these children training which focuses on word recognition will not be sufficient to address comprehension difficulties, but that additional meta-cognitive training will be needed. On the one hand this evidence implies that specific training is needed for those with comprehension difficulties, but pragmatically it would also support the use of generalised reading training programmes over those with a specific skill focus. It is unlikely that a large enough group of children with specific comprehension difficulties will exist in any one school to justify a training programme focussed specifically on their needs. Whilst research might argue that this would produce best results, the reality of in-school provision is likely to be that a general programme which includes meta-cognitive features and which can be sensitive to the needs of children is likely to be the most efficient provision available to children in this category. The efficacy of such provision will rely crucially on the sensitivity of the teacher to specific problems, and their ability to adapt general programmes to the needs of the individual.

Evidence demonstrating a link between the quantity of reading and reading consequences would also support the argument for global teaching approaches for poor readers. It has been
shown that better readers tend to read more than poor readers (Allen et al, 1992; Anderson, Wilson & Fielding, 1988) and that exposure to books makes a difference to reading progress (Neuman 1999). Evidence is also strong that reading to pre-literate children makes a difference. Bus, van IJzendoorn, & Pellgrini (1995) report a meta-analysis of 15 studies examining the effects of book reading on reading acquisition and found a strong association between more joint book reading at pre-school age and rate of reading acquisition at school. However once children became independent book readers this effect reduces. This suggests that an early advantage can be overcome with increased exposure. It also provides support for the view that non-reading is a self-limiting condition, since those who cannot read will not benefit from extended exposure to reading.

It would appear that there is some evidence in support of mixed reading programmes, on the clear condition that they include direct instruction on bottom-up features of reading acquisition. One intervention programme that has been widely used in this context is Reading Recovery (Clay 1987; Clay 1989; Clay 1993; Clay 1995). Reading Recovery integrates the teaching of phonological awareness with the use of contextual information to assist reading (Adams 1990). It takes a Vygotskyian approach to reading instruction by placing the child at the centre of a resource for their structuring of reading skills. One advantage of Reading Recovery is that it attempts to integrate specific skill training with increased appreciation and enjoyment of books and increase children’s education in literate language rather than a narrow training field (DES 1990). It is designed for use with young children (aged 6 years) in order to benefit from the advantages of early intervention. Evaluations of Reading Recovery have been conducted in the UK, Australia and New Zealand and have assessed Reading Recovery in comparison to other interventions, in the longer term and on attitudinal as well as skill based

The results of Hurry and Sylva's study make the most comparison to the study reported here, since they use children in this country and included a within-school control group for Reading Recovery. The control group reported by Sylva & Hurry is a little problematic, since they identified the six children in the year group with the lowest reading scores, assigned the lowest three or four to RR and the other two or three constituted the within-school control group. Hence at the outset they were not true equivalent controls, because their scores on the reading measures used were significantly higher at the outset than their RR contemporaries. Despite this initial disparity in favour of the control children, RR children were found to have significantly higher reading scores post intervention than their within-school controls. However, they also found that by three years post intervention the differences between RR children and control children were no longer as clear. In addition they found that when the amount of specialist help given to children was estimated, it was found that RR children received more help than the control children during the intervention period. Whilst there was some evidence that the group had moved forward (as assessed by the number still classified as poor readers relative to their peers) this change was apparent for both RR children and within school controls. This study did not attempt to assess school adjustment, or attitudes toward reading. This study creates a precedent of initial gain being shown in a RR programme in the UK. The second important finding from this work was that those children from homes with high social disadvantage and with especially low reading scores were those for whom most long term impact was observed.
Wade and Moore used the same methodology as Sylva and Hurry (comparing those who had been through Reading Recovery with the next best readers in the same school i.e. those who just missed getting Reading Recovery) on a group of children in New Zealand and Australia. They report that four years later the RR children had a significantly higher reading age than the comparison group (Wade & Moore 1998). However, the results are not adjusted for scores at baseline. Wade and Moore, however, did ask children about their attitudes toward reading (1998a, 1998b), and they concluded that ex-RR children had a generally more positive attitude toward reading than comparison children. Once again, no baseline data was reported.

Pinnell and colleagues (Pinnell, Lyons, Deford, Bryck, & Selzter 1994) compared the progress of children across five programmes — Reading Recovery, Reading Success, Direct Instruction Skills Plan and Reading/Writing group plus control group. This group followed 403 children (238 boys and 165 girls) from 10 different school districts in the US, ensuring a good proportion of urban schools. Data regarding eligibility for aid to dependent children (the equivalent of free school meals in the UK) was not complete because not all school districts would release data, but in those where reported the proportion ranged from 8.5% to 41.93%. Most of the children participating where classified as “white” (76.4%), the remainder were classified as “black”, although we don’t know if this is representative of the local population. They compared the progress of children to within-school controls, and to children from neighbouring schools taught using different intervention techniques. The techniques compared were Reading Recovery, Reading Success, Direct Instruction Skills Plan and a Reading/Writing Group (a small group intervention taught by a Reading Recovery trained
teacher). The last two of these were taught in small groups, while the others were delivered individually. There were daily lessons for at least 30 minutes for a random sample of the poorest readers in Grade 1 in each school. Reading Recovery was the most successful of these both in terms of initial post intervention success and follow-up success (at entry to following school year). Reading Recovery produced significant effects on text reading level, dictation, comprehension and vocabulary (Gates-MacGinitie) and Woodcock Reading tests at the end of the intervention period, and at the beginning of the following school year. They further argue that they proved the need for individual tuition; saying that individual teaching proved necessary (group instruction was less successful) but not sufficient (individual instruction did not guarantee success) to achieve progress. Significant change was seen in the dictation test for the reading/writing group following intervention. Note, however, that the comparisons were made only with within-school controls, rather than direct comparison of one method against another. Reading Recovery was not randomly allocated to schools, but existing RR schools were used, whereas the other schools were randomly assigned one of the other intervention programmes, which was then introduced. It is conceivable that the differences, then, were associated with the progress of children in schools where RR was already in place. These schools may have been more supportive of progress in literacy across the board, enabling those with poorest performance to make better progress regardless of intervention. Or these schools may have avoided problems associated with the introduction of new systems. While this study suggests individual RR was most beneficial, it does not conclusively prove it to be necessary. It raises particular questions with regard to group versus individual training which are discussed further below (Section 2.4.5).
Due to the evidence from comparable samples that RR is an effective programme for poor readers in disadvantaged areas, it was selected as an appropriate model for planning this reading intervention. However, RR is designed to be used individually with children aged 6 years. The decision had already been made that this programme was to be introduced to children aged 7 rising 8 years so the content of the lessons would have to vary from that prescribed by an RR scheme. Moreover it was also felt that a small-group training programme would suit the needs of the school better and this too broke away from the RR model. It was decided that an intervention would be put in place that should be based around the principles of RR, that is with a combination of discrete skills training, increased exposure to books, text reading, focus on attitudes to books and combining reading and writing, but that the programme should be adapted for this older age group and made appropriate to small groups of children. For example, in this context it was no longer possible to complete running records every day so these were decreased to once a week, the choice of books was expanded and extensive use was made of shared reading experiences within the groups.

2.4.5 The case for small-group intervention

Although it has been argued that the use of one-to-one tuition is crucial for progress in RR (Pinnell et al. 1994), small group interventions for poor readers are not without precedent in this country (Taylor, Hanson, & Watts 1997) and indeed the adaptation of Reading Recovery to this format has been attempted (Begoray 2001; Osuna 2000). There is some evidence that small group training has been successful in improving reading skills in poor readers (Nicolson
et al. 1999). Reading is a social process and encouraging the social aspect of reading through shared reading may advance skill. Indeed, shared reading may in itself be a useful intervention in some circumstances (Overett & Donald 1998). Elbaum et al. (1994) in a systematic review found no evidence for advantage of an RR-tutored small group above one-to-one tutoring in two studies. The debate on the relative efficacy of group interventions versus one-to-one interventions has not been finally resolved.

There is a second driver for using group interventions, that of cost. If equivalent, or nearly equivalent, results can be achieved when teaching groups of four children as when teaching individually, group teaching is clearly more cost-effective. Although cost is not often raised as a research issue, it is one which is clearly important for policy makers both at the local (school management, local education authority) and national (governmental) levels and the question of cost has been particularly raised with reference to RR (Dyer 1992; Pinnell et al. 1994; Hurry & Sylva 1998). Since RR uses highly trained teachers for individual instruction it is the most expensive form of remediation to provide. Since evidence would suggest that teachers deliver more effective generalised interventions than non-teachers (Hurry 2000), the solution to this is to determine whether comparable effects can be achieved using groups of children in a RR setting. Since long-term benefits of literacy are so great, any intervention that can be shown to make an appreciable improvement to literacy will almost certainly confer cost savings in the long term.

2.5 How do we assess effectiveness of intervention?

Interventions which hope to reduce inequalities for children in health, education and social care
attract considerable funding and research interest. The targeting of funds to those groups which are most in need of support is part of the success of a programme. However, if it is to be effective, a programme needs to be informed by the evidence available in the area. Recommendations from research for policy can only be as good as the evaluations that inform them. When interventions are put in place, particularly when they substitute for standard provision, it must be clear whether they do harm, have a negligible effect, or whether they help those involved. Some interventions introduced in order to benefit children have been demonstrated to do harm. For example, interventions using peers to facilitate or model behaviour change have been shown to do short and long term harm to young people at risk of criminal behaviour (Dishion, McCord, & Poulin 1999), increasing criminal behaviour, mortality and unemployment. Particularly concerning in this case is evidence that simply identifying children as struggling resulted in poorer examination results (Defty & Fitz-Gibbon 2001).

For any evaluation then, the question arises as to which is the best evaluation methodology. The “gold standard” of intervention evaluation is the randomised controlled trial, in which a group which is eligible for the intervention is randomly divided into a group which receives it and a group which does not. Often the needs of service delivery are seen to be directly at odds with such a methodology, for instance in “denying” the provision of treatment to a group where need has been identified (Karoly et al. 1998). In order to overcome this problem, non-experimental or quasi-experimental research designs are often employed. In the case of non-experimental methodologies, there will simply be no evidence that any changes observed could be attributed to the intervention that has been put in place. Quasi-experimental studies usually rely on either comparison between interventions, or the creation of a matched sample group. If results are compared between two or more interventions (e.g. O’Shaughnessy & Swanson,
2000; Torgesen et al., 2001), the key question of whether the change observed in any group would have occurred without intervention remains unanswered. When matched samples are used (e.g. Fawcett et al., 2001; Madden et al., 1993; Nicolson et al., 1999), the criteria by which samples are matched become key. In each of the studies cited, comparison children were matched for chronological and reading age and, where possible, gender but often drawn from other schools. It would be difficult to argue that children in different schools had equivalent classroom experiences, and that therefore the input of classroom provision on these groups could be assumed to be equal. There remains, of course, the possibility that an un-anticipated factor (and therefore unmatched) may be contributing disproportionately to change. Whilst these short-comings in research design predominate, interventions of questionable or unproven efficacy can continue to be recommended, since evidence that would demonstrate their lack of effect is not being produced.

The shortage of evaluations with properly described comparison groups in this field has been recognised by reviewers (Brooks et al. 1999; Brooks, 2002; Bus & van IJzendoorn 2000; Elbaum et al. 2000; Turner 1997), who mention the large number of studies which they are unable to include or whose results they cannot condone because of this lack. That the use of this methodology has been unpopular in applied settings in general (Oakley 1998) and educational settings in particular has been recognised by researchers in psychology. It has been noted that, “This concept [randomisation], and with it the rejection of the concept of achieving equation through matching . . . has been difficult for educational researchers to accept” p.2 (Maclean, Bryant, & Bradley 1987). The difficulty of using controlled methodologies can be clearly seen if we consider one study from Canada as an exemplar case (Begoray 2001). Although the intention of this study had been to use random assignment to control groups,
resource teachers were "vehemently in favour" of a non-randomised, non-control group approach and indeed altered the content of the intervention to favour this approach. The result of this departure from research methodology is that the study can tell us nothing about whether or not change that was observed had any relation to the use of additional support. Change would be expected to occur, but research aims to interpret this change and attempts to demonstrate causality. Interestingly, historically the use of control groups in social research was accepted (Oakley 1998) and the use of meta-analysis to provide systematic synthesis of controlled trials has early roots in education (Glass 1976) that have not been continued until recently (Elbourne, Oakley, & Gough 2001, Coe & Constable, 2000, Petrosino et al 2000).

This problem is not specific to the complex interventions employed in reading remediation, but exists also in the area of phonological training where experimentally sound regimes are easier to employ. In a review critiquing the methodology employed in phonological awareness interventions several features of significance come to light (Troia 1999). He identified 39 phonological awareness training studies that included results from a comparison group. Only 54% used random assignment to treatment and control/comparison condition; this lack of randomisation undermines both the equivalence of groups and the frequent use of analysis which assumes group independence. Moreover, only 12 of the studies (30.1%) used classroom-based interventions. Although conducting studies in classes produce many problems in controlling for external features of environment, for instance teacher-bias, without it we cannot be sure of the application of treatment programmes in this environment.

The intervention literature can be reviewed to examine the extent to which this lack of studies
with experimental design holds true. Elbaum et al., (2000) have reviewed the data from one-to-one reading interventions in the USA from 1975-1998, restricting this to the review of studies with appropriate comparison or control groups. In order to extend this search, and to contribute to the wider world literature, a systematic search for recent studies published in English of reading interventions for non-US primary school aged children. The searches were conducted using Psychinfo¹ and British Educational Abstracts², for studies published covering the 10 year period between 1992 and 2002, limited to the appropriate age group.

Before the search terms for control or comparison groups were added, the search on Psychinfo produced 240 studies. With the addition of the terms for controlled studies, 46 studies were identified using Psychinfo of which 32 were American studies. This means that up to 81% of the studies in this field did not mention in their title, abstract or key words any terms referring to the use of a control group. The remaining 14 studies were examined for relevance to the topic. Four studies (Christer, 1999; Whedall, 2000; Center 1999; Whyte, 1993) were excluded because they did not evaluate an intervention for poor readers, one study was excluded because it was a single-case study (Frazier, Norbury & Chiat, 2000), one study because it used severely dyslexic children in an out-patient clinic using neuro-psychological techniques; leaving nine relevant studies for discussion. Before inclusion of the search terms for control or comparison groups were added, the search on BEI produced 18 non-US studies. With the addition of the

¹ Search strategy for psychinfo; [(remedial reading/ or reading ability/ or reading achievement/ or reading development/ or reading disabilities/ or reading education/ or reading skills/) and (EARLY INTERVENTION or Educational Program Evaluation or {intervention.mp})] and (Experimental design/ or Experimental Controls/ or control group.mp/ or comparison group.mp) limit to school age and childhood and English and yr=1992-2002
² Search strategy for BEI; [reading ability/ or reading achievement/ or reading difficulties/ or reading failure/ or reading improvement/ or reading research/ or reading skills/ or reading strategies/ or remedial reading/] and [intervention/ or intervention.mp/ or experimental programmes/ or improvement programmes/ or remedial programmes/ or summer programmes] limit to English and primary education and yr=1992-2002 and [control groups/ or control group.mp or comparison group.mp or experimental groups/
terms for controlled studies, no studies were identified. These 18 studies were then handsearched to identify any suitable studies missed by the search. This produced two studies not already identified by the Psychinfo search (Elliott et al 2000; Gustafson et al, 2000). The characteristics of included studies are summarised in Table 2.1.

This table gives the country in which the training took place, the sample size, the intervention used, characteristics of the control or comparison group selection and treatment and outcomes assessed. Studies are listed in date, then alphabetical order. The present study was designed in 1997, by which time only three of the studies quoted below were published and these are highlighted in italics. The table can be used to examine size and type of study, and importantly to examine the selection methods most often used for determining control or comparison groups.

### Summary of Evaluation Methodologies

<table>
<thead>
<tr>
<th>Study (Chronological order)</th>
<th>Intervention</th>
<th>Research Design</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>wcett et al. 2001 (K)</td>
<td>Group based in school support</td>
<td>40 expt 62 comparison</td>
<td>Lowest scores on screening measures. Matched on Chronological and reading age, from comparable schools</td>
</tr>
<tr>
<td>pson-Boyles 2000 (K)</td>
<td>Catch up (Pilot findings)</td>
<td>17 expt 14 matched time 14 control</td>
<td>Those at Level 1 on KS1 SATs, i.e. those somewhat, but not substantially behind norms from 9 schools, 6 pupils from each school. Schools designated as catch up, control, or matched time (that is same time spent but unguided).</td>
</tr>
<tr>
<td>ott et al, 2000 (K)</td>
<td>Volunteer support in class</td>
<td>68 expt 72 control</td>
<td>Three reception classes in deprived area, classes randomised to support or control</td>
</tr>
<tr>
<td>Staffson, Suellsson et al., 80</td>
<td>Phonological Training</td>
<td>33 expt 16 age matched controls</td>
<td>49 children across 14 schools selected by schools for special instruction in reading. Schools</td>
</tr>
</tbody>
</table>

or matched groups/ or pretests posttests/ or quasiexperiemental design/ or scientific methodology or research design/
<table>
<thead>
<tr>
<th>Study (Chronological order)</th>
<th>Intervention</th>
<th>Research Design</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ácool &amp; Topping, 2000</td>
<td>Paired reading with parents</td>
<td>24 expt 24 control</td>
<td>Text reading and comprehension (in Portuguese)</td>
</tr>
<tr>
<td>Sackay, 1999</td>
<td>Small group instruction plus home reading</td>
<td>8 expt 1 8 expt 2 8 controls</td>
<td>NARA Qualitative assessment of attitudes</td>
</tr>
<tr>
<td>Colson, Fawcett, Oss, Nicolson, &amp; Asen, 1999</td>
<td>Flexible group intervention</td>
<td>62 expt 40 comparison</td>
<td>Reading Spelling</td>
</tr>
<tr>
<td>Pore &amp; Wade, 1998</td>
<td>Reading Recovery</td>
<td>121 expt 121 comparison</td>
<td>Reading accuracy and comprehension (NARA) Generated questions about attitudes to reading Reported amount of reading Eliciting of reading strategies Writing skills</td>
</tr>
<tr>
<td>Inter, 1995</td>
<td>Reading Recovery</td>
<td>28 expt 34 true controls 36 comparison</td>
<td>Clay diagnostic survey Text reading Spelling Phonemic awareness Syntactic awareness Word attack skills</td>
</tr>
<tr>
<td>Woos &amp; Willatts, 1994</td>
<td>Musical skills training (pilot findings)</td>
<td>6 expt 6 controls</td>
<td>Schonell reading test</td>
</tr>
<tr>
<td>Utter, Halme &amp; Be 1994</td>
<td>Phonological and/or reading training</td>
<td>32 expt 1 (reading with phonology) 31 expt 2 (reading alone) 30 expt 3 (phonology alone) 31 matched control</td>
<td>19 cognitive tests used, 12 reported including word and text reading, arithmetic, phonological skills and spelling.</td>
</tr>
</tbody>
</table>

Table 2.1 Summary of Evaluation Methodologies
Of these 11 studies identified in the brief review conducted here, one (Moore & Wade 1998) used a group who were not comparable, but the next worst readers; three used matched between school controls (Fawcett et al. 2001; Hatcher, Hulme, & Ellis 1994; Nicolson et al. 1999); one matched within school controls (Douglas & Willatts 1994) and four groups randomised to treatment or control at the school or class level. One study randomised at the level of the individual (MacKay 1999), but this study had a very small sample size (8 pupils per group). As discussed previously there is a clear methodological advantage to randomisation of control groups, in that it ensures systematic differences between control and comparison groups don’t arise. Randomising at the level of the group (for example randomising schools or areas) appears to be a common solution to the difficulties found in the use of randomised methodologies in this area. Such randomisation means that change should not be analysed at the level of the individual, a fact that is often neglected in such studies. In addition differences between schools and classes are not controlled for. Moreover, when randomising at the level of the group conclusions cannot then be made at the level of the individual, we can only say whether the school or class benefited rather than use subgroups to identify particular children who may have benefited. The need for new research to take into account the methodological importance of control groups at the inception of this project is made clear by the number of studies post-dating this study which choose to attempt to use controlled methodologies. In this search, in the years 1992-1997 only three studies were identified, whereas the five years following produced 9 studies. This convergence in the literature highlights the increasing awareness at the time of this study of the need for studies for this type, but the lack of randomised study designs providing true control equivalence is clear.

A number of key factors contribute to good research standards including control groups,
sample size, subject selection from a range of schools and outcome measures. A study that includes a number of schools with children from each school selected by externally verifiable criteria and then randomised to treatment or control at the level of the individual might be thought preferable. The difficulties presented by such an approach are recognised in the literature (Gustafson 2000). In practice, studies such as that described by Gustafson, where groups are randomised at school level, and within school matched comparisons are drawn from younger children, are formulated as a method to persuade practitioners to consent to studies in the form that it is required. The study by Mackay overcomes this problem by working closely with one school to ensure a rigorous methodology is applied, and while this restricts sample size the methodological advantages are considerable. The need for rigorous evaluations of effectiveness within the field of education research is increasingly recognised and groups conducting secondary (or review level) research such as the EPPI Centre in London (www.ioe.ac.uk) and the international Campbell Collaboration (www.campbellcollaboration.org) have highlighted this need in primary research.

2.6 Conclusions – the gaps in the literature

The considerable problems faced by children who have difficulty reading and writing are difficult to change once established. In general, it has been found that intervention delivered early, with a broad scope of skill training and including phonological training, is the most effective. Although early intervention is desirable, it may not always be appropriate or sufficient, and resources need to be developed for remediation later in school careers.

There is a particular problem in a bi-directional relationship between reading and reading
attitudes, in that those who find it difficult to read practice reading less, and practicing reading less may limit progress in reading. Struggling in any area of school achievement can result in poor attitudes to school, and in the long term this may result in disaffection, truancy and dropout. Changing attitudinal outcomes may then influence academic outcomes and the impact of reading interventions on attitudes should be assessed alongside their impact on reading and writing.

In response to these issues it was decided that an intervention would be put in place at the beginning of Key Stage 2 to support those children entering junior school with problems in their reading and writing. It was hoped that progress could be made if an intensive programme was put in place. The intervention was to be delivered to children with general reading difficulties, predominantly in small groups and through a general skills programme, drawing on the principles of Reading Recovery.

In order to develop resources, rigorous evaluations of trial interventions need to be put in place such that policy decisions can be made on the basis of the best evidence available. The 'gold standard' of trial evaluations is the randomised controlled trial. Although there are difficulties with adapting this methodology to a social science domain, associated with ethics and no-treatment equivalency, it is important that this should be one of a battery of evaluation methodologies that are employed. There is a paucity of RCTs in the literature to date. The evaluation was therefore planned as a RCT of a school based intervention for poor readers, assessing both literacy and attitudinal outcomes. The details of the evaluation plan and the constraints placed on the research methodology of this applied setting will be discussed in Chapter 3.
Chapter 3 Evaluation Methodology

3.1 Introduction

This chapter will describe in detail the methods used to evaluate the intervention, beginning with a description of the experimental design employed. In Chapter 2 the literature that informed the methodology was reviewed, and conclusions were drawn as to the best study design, the relevant outcomes, and the intervention of interest. The application of these recommendations to this study is the concern of this chapter. The characteristics of the school in which the intervention took place, and the process of selection and randomisation of children are included in sections 3.3 and 3.4. These are followed by a discussion of individual outcome measures employed, including the reasons for their inclusion, a description of their use and results, piloting protocols where appropriate and difficulties encountered in data collection.

3.2 Experimental Design

Studies aimed at comparing, in the short term, two or more instructional methods for specific literacy skills (for example, Hatcher, Hulme, & Ellis 1994; Sylva & Hurry 1995; Pinnell et al. 1994) have provided a basis for the structure and content of this project. The main body of this study was created to employ a Randomised Control Trial methodology in a school-based intervention evaluation. Teachers selected a group of children who they felt were in need of additional support with their reading and writing; these children were allocated to receive help in either the first or second half of the school year (see section 3.4 Selection of Intervention Children). The majority of
children were allocated randomly to one or other teaching block, but a small number of children (n=6) were prioritised to the first teaching block because of exceptional circumstances. The children allocated to the second teaching block formed a group of "waiting list controls", who met the same criteria of need as the first intervention group, but who did not receive additional teaching during the first half of the year. This control group "disappeared" once they received their teaching in the second half of the year. A group of peers were selected to act as a comparison group. The purpose of this group was to provide data to compare with data from the intervention children in the long term (i.e. once the control group had been lost). The controlled trial evaluation only lasts for the first half of the year for each successive cohort; after this point the design reverted to a longitudinal trial evaluation. This design enabled the implementation of a methodologically desirable RCT whilst meeting our ethical obligation to teach all the children whose need was identified.

The study used three succeeding cohorts of children entering year 3 in the target school. Each cohort was monitored in the following years to examine longer term change. Figure 3:1 shows a plan of the research design. The first half term of each year was given over to assessing the incoming group of children, identifying children for study groups, and baseline assessments. The intervention period ran from mid-term of the Autumn term to the Easter break. In the closing weeks of the Spring term the post-intervention assessments were carried out with the control group prior to their beginning their intervention period at the beginning of the Summer term. The post-intervention assessments for the intervention group were conducted immediately after the Easter break. The comparison between groups up to this point constituted the RCT section of the evaluation. The second intervention period lasted another term and half, until mid term of the following Autumn term, when post-intervention assessments
were conducted for this group only. A comparison group was assessed at the same
times of year. Follow-up assessments were conducted in the second half of the Spring
Term for Cohorts 1 and 2 for years following the intervention year.

**Overall Plan for Evaluation**

<table>
<thead>
<tr>
<th>Year</th>
<th>Cohort</th>
<th>Autumn Term</th>
<th>Spring Term</th>
<th>Summer Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 1st intervention</td>
<td>End 1st intervention Post-measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONTROL PERIOD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 3rd intervention</td>
<td>End 3rd intervention Post-measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONTROL PERIOD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 5th intervention</td>
<td>End 5th intervention Post-measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONTROL PERIOD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Overall Plan for Evaluation

Two groups of assessments were used, those using the whole year group as the sample
(whole class test) and those using only intervention or comparison groups as a sample
(in-depth tests). Follow up measures were taken immediately post intervention, then
6/12 months post intervention for cohorts one and two, and 12/18 months post
intervention for cohort one. This was the maximum follow-up possible in the three
year structure allowed for.

Assessment measures were selected to cover a broad range of skills and achievements,
which were appraised according to their use in the literature, efficacy in practice and
face validity and are discussed in detail in sections 3.6 and 3.7. Assessments were
administered according to a standard protocol and order over at least two individual
sessions with children.

### 3.3 Selection of Subjects

The school involved in this study was chosen because of the needs of its pupils. It is
located in a deprived area and many of the children struggle to succeed at school. It is a
school that prides itself on its inclusive policy and aims to give every child the best opportunity for education. The commitment on the part of the school management to improve the education they provide, particularly for those who are at risk of school failure, meant that they were amenable to the introduction of a new service and willing for an evaluation to take place alongside it. For a full description of the characteristics of the school see Section 4.1.1.

A total of 219 children entered Year 3 at the school over the period of three years. This is the number of children who actually began the school term, as opposed to the number on the school roll, as some registered children never arrived. The flow chart below shows (Figure 2) the eligibility and consent rates for all children entering Year 3. All boxes in this flow chart are numbered to aid identification. Four children began the term but left before the study began (box 2.3). This illustrates the mobility of this school population. All children with SEN (special educational need) statements were excluded from the research population because they were already receiving additional support as part of the statementing system (box 2.2). Parental consent was sought and over the three years of the study only two refusals were received (98.9% response rate, box 3.2). The total number of children available to be included in the study was 204.
Class teachers were asked to identify all children who they felt were in need of literacy support (based on their assessment of class work). A total of 73 children were referred over the three years, from which the reading teacher selected the children to be included in the study (see section 3.3; Selection of Subjects). A target of 22 children per year was set for inclusion in the intervention groups (this being the maximum number that could be included). In fact a total of 59 children were selected over the 3 years; 22 in cohort 1, 21 in cohort 2 and 16 in cohort 3.
At the same time as children were being selected for inclusion in the intervention groups a number of assessments were conducted with the whole year group. The results from these were examined and the group of children who fell into the bottom quartile of the population on any of these tests\(^1\) were identified. Children who fell into this group, but were not included in the intervention groups comprised the "comparison group". Since these children fell into the bottom quartile on the whole class tests they were felt to provide a reasonable comparison to those children identified as having literacy problems. This comparison group was not chosen to compare progress over the period of the intervention, but in order to determine whether the intervention groups “caught up” with their class peers in their literacy skills in the short or long term.

Only children in the study groups (intervention, control and comparison groups) then completed further assessments.

3.4 Selection of Intervention Children

Year two teachers were asked to identify those children from the previous year they thought to be at risk of reading failure. These referrals, together with KS1 SATs results and a writing sample, and Year 3 teacher assessments were used to identify a group of children whose achievement was poor. Class teachers did not refer children who had a statement of educational needs and who were therefore already receiving additional support and, thus, children with Statements of Special Needs were excluded from the

\(^1\) Whole year tests were; Young Group Maths Test, Young Group Reading Test, British Picture Vocabulary Scale and Raven’s Progressive Matrices
study. The specialist reading teacher saw each of these children (n=73) individually and completed the following assessments:

- letter/sound identification (Clay 1993)
- consonant clusters, consonant digraph check*
- words recognised on sight (Canberra Word Test, Clay 1993)
- most commonly used words* (Collins Cobuild 1995)
- written vocabulary (Clay 1993)
- hearing sounds in words (Clay 1993)
- mid-vowel dictation* (Brand 1987)
- running record (Clay 1993)

(* = used where appropriate)

The reading teacher used this information, supplemented by her own expertise, to identify a group of children for literacy support. The Head teacher approved these selections. At this point, the Head teacher raised cases for special consideration for a few children who were prioritised to the first intervention group (n=6). These cases were not prioritised on the basis of need, but for reasons internal to the school and were restricted to those instances where the school management felt they would be in breach of their duty of care not to provide additional support immediately. This departure from the randomisation process was accounted for in the analysis (see Section 5.2). After the names of these children were removed, the researcher randomly assigned all others to first or second intervention period.
Within each intervention period there was facility for children to be distributed between two tuition groups of four and three children taught individually. These groupings were allocated by the reading teacher according to her assessment of compatibility between children. Children were selected for individual tuition if the teacher believed they would have problems interacting with the group or because their reading level differed from others due to receive intervention at that time (more or less proficient).

3.5 The teaching programme

A full description of the teaching methods employed in the lessons is supplied as Appendix 1 to this document. In summary, the children were withdrawn from class every a day for half an hour to join the reading lessons. This out of class time was timetabled so that it did not overlap with literacy hour, numeracy hour or the daily act of worship. Children were allocated to attend individually or in a group of four children. This allocation did not change during the teaching period. Groups were drawn from different classes, and were matched according to level of reading. The content of each lesson varied according to the needs of the children at that time, but the underlying structure was taken from Reading Recovery (Clay, 1987; Clay, 1993; Clay 1995) and included:

- practising high frequency words;
- re-reading familiar texts;
- taking a running record of a new text at least once a week for each child;
- focusing on word work (onset and rime, etc);
- writing a sentence;
- sequencing the cut up sentence;
- introducing a new book.
A number of techniques were used to facilitate the acquisition of reading skills. These included practising of high frequency words, re-reading familiar books and poems, discussion of techniques and cues for reading (e.g. reading on for meaning), onset and rime and alphabet work. In years two and three, weekly lists of words were given to class teachers for use in spelling tests, so that the children were focussing on the same word patterns in class and in their reading lessons.

Thus the programme of work included an approach to books that encouraged children to develop a knowledge of author and style and a development of comprehension skills combined with a focus on the constituent parts of words. Thus for those at a lower reading level, this may have included segmenting simple words into their phonemes for decoding, and for others the use and recognition of spelling patterns. In all cases the explicit naming of techniques used to decode unfamiliar or unknown words was an important part of group and individual activities.

A specialist reading teacher was employed to deliver the intervention classes. This teacher was a reading recovery trained teacher, with some years experience of reading recovery tutoring in South London. She had been employed in the school part time for one year prior to the beginning of this project to deliver reading recovery to a small number of Year 1 pupils, and provide support for 11 children in Key Stage 2. The funding from this project enabled her to work full time in the school and to work exclusively with Key Stage 2 children. During the course of the project she was also completing additional training for teaching children diagnosed with dyslexia. The development of the intervention programme was largely in her hands. She was asked to plan the content and structure in advance of the project beginning and to document changes as much as possible. She described the programme she was delivering as built on the principles and practice of reading recovery, adapted to use in an older age group. She was observed by researchers teaching reading recovery classes, and in these intervention classes.

As discussed previously (2.4.4) it was decided to use small group teaching in order to be able to teach more children, and also to enable children to use group learning processes. There is precedence for using adapted RR techniques e.g. use of RR techniques with US
grade 2 (equal to UK Year 3) in small groups (Begoray 2001).

3.6 Standardised Assessment Tools

Measures were selected to assess both the global and specific effects of receiving literacy support. Measures were initially identified from their use in the literature. Subsequently these tests were streamlined according to their efficacy and their utility in practice as shown by their implementation at first baseline (cohort one). Measures needed to be considered in the light of their ease of use, the amount of classroom disruption and time they required and their appeal to the children. This evidence was used to consolidate the final selection of measures.

The standardised measures used were chosen carefully from those available since the characteristics of the measures used determine the usefulness of outcome to potential conclusions. Measures needed to meet a number of validity criteria; that evidence existed that they measured that which they purported to measure, that they were appropriate for the ability levels of the children in the sample over the period of the intervention, that they were acceptable to the school and teachers and that they were acceptable to the children. The body of assessments available for use was lacking in several respects, particularly with reference to children who were struggling with reading and in some areas no acceptable validated measures could be located and new assessments were developed for this study.
All measures are discussed in full below, beginning with those measures that were used for whole year group assessments, then progressing to those from which data was collected from study groups only.

3.6.1 Young Group Reading Test 3rd Edition (Young 1980)

Particular attention was given to the selection of reading measures, and particularly of the need for an assessment that could be used to gain whole year data balanced with the need for a detailed reading test. It has been established that in order to assess reading abilities, particularly comprehension abilities, an individually administered text reading test is necessary (Nation & Snowling 1997). However, there was also a need for a test that could be easily administered to the entire year group. A decision was made that the solution was to use two reading tests, a group administered reading test for the whole year group and an individually administered reading test for those in study groups (see 3.6.6 below). The group reading test selected was the Young Group Reading Test (YGRT). The YGRT is designed to identify children's reading ability using a cloze procedure. Children are presented with a list of words and a picture or incomplete sentences, and are asked to identify which word matches the picture or completes the sentence. The picture task assesses word reading and includes only simple words, the 30 sentences provided are of increasing difficulty and assesses text reading and comprehension. There is a time constraint, although it is generous (allowing 14 minutes in total). The measure was completed in a whole class situation at baseline.

This test was chosen because it was designed for children in British schools entering Year 3 (although it is applicable to older less able children), and is administered in a
whole class situation. This was important for whole class tests because it enabled efficient collection of data from the entire year group.

Scores are usually transformed into quotients - i.e. adjusted for age such that scores form a normal distribution with an average score of 100.

It provides a composite score for reading accuracy and comprehension.

3.6.2 Young Group Maths Test 3rd Edition (Young 1996)

As well as assessing children for literacy skills, measures that covered other areas of academic achievement were included. The aims of this were twofold, first to establish whether the children in the intervention differed from their peers on measures of literacy only or across the board, and secondly to establish whether change was seen in these areas over the period of the intervention. Once again this measure was selected for ease of use across the entire year group. The Young Group Maths Test assesses children's mathematical ability, half of the test items are presented aurally and the remainder in the form of written sums. There is a time constraint, although it is generous (allowing 25 minutes). The measure was completed in groups of six children at baseline and at first follow-up.

This test was chosen because it was designed for children in British schools entering Year 3 (although it is applicable to older less able children), and is administered in groups.
Scores are usually transformed into quotients - i.e. adjusted for age such that scores form a normal distribution with an average score of 100.

It provides a score for mathematical competence.

3.6.3 British Picture Vocabulary Scale, Short Form (Dunn, Dunn, & Whetton 1982)

Since a large number of children at the school have a home language other than English (54.5% in Yrs 1-6), it was felt to be important to have some measure of children's fluency in English. Moreover, there is evidence that children with specific language impairments also experience literacy difficulties (Bishop & Leonard 2000). In particular children with early, mild, undiagnosed language disorder may appear later to have problems of literacy that include, and reach beyond, problems associated with phonology (Snowling and Stackhouse in Bishop & Leonard 2000). The groups within the project were selected on the basis of their reading and writing ability, so it was necessary that this English language measure be independent of other aspects of literacy. The British Picture Vocabulary Scale (BPVS) provides a good test for this purpose as it assesses vocabulary comprehension without using reading or writing. Children are seen individually and asked to select a picture, from a group of four, which illustrates a word spoken in a passive voice by an administrator (e.g. “furniture”). It provides a score that represents the extent of a child's vocabulary compared to other children of the same age.
Scores are in the form of a quotient score - i.e. adjusted for age such that scores form a normal distribution with an average score of 100.

3.6.4 Raven’s Progressive Matrices (Raven & Raven 1996)

IQ was assessed for similar reasons to those for using a maths test, that is to establish differences between intervention children and their peers. This assessment was not included in order to distinguish between those with specific reading difficulties and “garden variety” poor readers (for discussion see Chapter 2 section 2.4.2). It was included to examine the differences and changes observed in study groups over time on broader measures of achievement than literacy skills alone. As with the BPVS, it was important to identify an IQ test that did not rely on literacy ability.

Raven’s Standard Progressive Matrices is a pattern recognition test that was selected as an appropriate test of non-verbal IQ. The Standard Progressive Matrices had the advantage of being suitable for administration in groups and being suitable for children from the age of baseline testing (age 7-8 years) through to adulthood. The test asks children to select from a group (of 6 or 8) the pattern that completes a given sequence. There are 5 sections of increasing difficulty within the test, each containing 12 items, and making 60 items in total. The test was completed in a whole class situation and children were instructed to “have a go” at as many items as they could, making guesses when they weren’t sure and stopping when they couldn’t do any more. Scores are out of a total of 60 and a minimum of 2 and can be compared to published normative data from a sample of British children.
Scores can be given as a total score compared to peers, or interpreted using the population distribution of scores.

### 3.6.5 Reading Attitude Scale (McKenna & Kear 1990)

It was felt to be very important that some measure of attitudes toward reading was included in the baseline tests because motivation is very important if a child is to become an experienced reader. A child who has the skills to read but chooses not to read is unlikely to gain the depth of knowledge necessary to become a fluent reader. However, it proved to be very difficult to find an appropriate measure of reading attitude. Reading self-concept was better dealt with for British subjects (e.g. Norwich 1982; Blatchford 1997) and this was included in the pupil interview, see 3.7.5. An exhaustive search for an assessment of reading attitude across psychological and educational databases, excluding those measures designed to assess self-concept, produced just four studies, all using U.S. samples. Two used 85 and 88 item questionnaires (Alexander & Elgin 1986; Wallbrown, Brown, & Engin 1987 respectively) and were discounted as being over lengthy and unwieldy. Of the remaining two (Estes 1971 and McKenna & Kear 1990)) the Estes tool was thought to be inappropriate because the language was rather complex (e.g. "books should not be read except for class requirements"). Moreover, McKenna and Kear addressed the fact that it was intended for children by using a familiar cartoon character (Garfield) to express mood. Since it was designed for use with American children some of the questions had to be slightly adapted and children still found some of the questions a little "odd sounding".
The Reading Attitude Scale is a series of 20 questions about children's feelings regarding books and reading in different situations. Children are asked to choose their response from a 4 point Likert scale represented by pictures of a cartoon cat in different moods. This measure was administered to all children at baseline. Questions were read aloud to children in a whole class situation and they were asked to "choose the Garfield that best fitted how they would feel".

Scores are out of a maximum of 80 (very positive response) and a minimum of 20 (very negative response).

3.6.6 Neale Analysis of Reading Ability (Neale 1989)

As discussed above (3.6.1), selection of an appropriate individually administered reading test was thought to be key to the rigour of the evaluation. The Neale Analysis of Reading Ability (NARA) was selected as the individually administered reading test. The NARA is a widely used and well-accepted measure of reading for children with reading ages of up to 13 years. It is extensively used in schools by teachers (Turner 1990b) and educational psychologists and it is very commonly used for research purposes (for example; Hatcher, Hulme, & Ellis 1994; Bradley & Bryant 1983; Moore & Wade 1998; Reason & Morfidi 2001; MacKay 1999; Center 1995). Indeed it was used by OFSTED in the locality of this school for inspection of reading levels in the previous year (HMI 1996). It has been given as an example of a reading test that distinguishes those who are experiencing comprehension difficulties as well as decoding difficulties (Nation & Snowling 1997). Single-word tests (even when embedded in sentences such as in the YGRT) do not differentiate comprehension and decoding difficulties and so may not detect children with these problems.
This test was administered to all children in study groups, that is those who were to attend the intervention classes and to comparison children. Children were tested pre and post intervention and at follow-up times.

Raw scores are usually transformed into reading ages in years and months. Separate reading ages are obtained for each of:

- Reading Accuracy
- Reading Speed
- Reading Comprehension

There has been some controversy surrounding the NARA questioning its wide-spread use. Criticisms have been raised about the second set of passages (form 2) (Stothard & Hulme 1991). Stothard and Hulme found that boys scored less well than girls for both accuracy and comprehension on Form 2, although the difference was only significant for accuracy scores. A sample of children was re-tested with Form 1 and boys were found to have significantly higher scores at re-test with Form 1. Girls also showed higher retest scores although this difference was not significant. Moreover the comprehension questions for passages 5 and 6 of Form 2 were particularly criticised since it was found that the scores for comprehension questions on passage 6 were significantly higher than those on passage 5 (i.e. passage 5 questions were harder than passage 6 questions). It was concluded that Forms 1 & 2 were not parallel forms and that there were problems with Form 2 that make its use inappropriate. Moreover it was felt that the revised Neale was biased against boys. No information is given concerning
the selection of children for testing, and the sample size was small (girls n=25, boy n=25).

A later trial (Gregory & Gregory 1994) failed to find the same problems. They did find that the Neale produced much lower reading age scores than did the BAS reading test administered to the same children, but they found no significant differences between the scores of girls and boys, or between the two forms and no significant interaction. They did, however, find a high correlation between scores on the two tests suggesting that they do test the same abilities. In all 93 children were tested, 30 from a clinical setting (referred from educational psychologists) and 63 from years 1-3 in a local school. This trial is more robust than Stothard and Hulme’s trial in terms of numbers, methodology and interpretation.

Finally McKay (McKay 1996) repeated the standardisation using a stratified random sampling procedure based on SES in Australia (n=250) and found no differences between Forms 1 & 2 or between girls and boys in a much larger sample.

The one question unanswered by these studies is that of the inappropriate comprehension questions for passage 5 Form 2. This passage was taken in the British version from the Diagnostic Tutor version of the Australian version because it was felt to be more appropriate for the environment and this may have upset the gradation of difficulty. McKay criticises Stothard & Hulme on the basis that these passages were "written for older children (10-12 year of age) who were not adequately represented in the sample". This may or may not be true, but again since the current study involves poor readers who would be unlikely to reach these more difficult passages this was not thought to present a problem.
In sum, the questions raised about the validity of the NARA seem to have been comprehensively answered by subsequent studies and there is no evidence to suggest that it is not an appropriate assessment tool in this instance.

3.6.7 Phonological Assessment Battery (Fredrickson, Frith, & Reason 1997)

A phonological test was included in the light of the weight of evidence that suggests the importance of phonological awareness and skill in the process of learning to read (for review of evidence see (Goswami & Bryant 1990; Wagner et al. 1997). Phonological skills have therefore been used as outcomes measures for intervention evaluations (e.g. Center et al. 1995).

The Phonological Assessment Battery was chosen because it was appropriate for the study sample, in that it was designed for use with British children within the specified age range. It has a range of eight tests within the battery to assess different phonological skills. The entire battery was administered to 32 children from the study groups at the beginning of the study. Using this pilot data, three tests were excluded on the grounds that they either failed to discriminate between children (the naming speed test), or because children were largely performing at ceiling levels (the alliteration and rhyme tests).
This shortened version of the battery was administered to all children who were to attend the intervention classes and to comparison children. Children were tested individually pre-intervention, post-intervention and at follow up.

Scores for each test are usually transformed using chronological age into quotients - i.e. adjusted for age such that scores form a normal distribution with an average score of 100. Separate scores for each of the following tests were obtained:

- Spoonerisms Test
- Semantic Fluency
- Alliteration Fluency
- Rhyme Fluency
- Non-Word Reading

3.6.8 Battle Self Esteem Inventory (Battle 1981)

There is evidence that children with poor school achievement are likely to have poor self-concept and low self-esteem (Fontana 1997;Muijs 1997). It was therefore important to include a measure of self-esteem to explore differences between intervention children and their peers and to establish whether the intervention had an effect on self-esteem. The Battle Self Esteem Inventory is a 30-item questionnaire for children that assesses self-esteem. Questions are designed to cover four areas; general, social/peer related, academic and parent/home related self-esteem. The inventory in its entirety was thought to be too long for inclusion and the school management team felt
that questions regarding home and children's relationship with their parents were intrusive, inappropriate for the children in this school (since many had complex and difficult home lives) and not relevant to a literacy intervention. In consideration of these points, questions pertaining to parent/home related self-esteem were not used. The remainder of the inventory consisted of items on three subscales plus two “lie items”. A representative sample of 9 questions were selected at random, plus one "lie" item. Four out of 10 general self esteem, three out of five social and two out of five academic self-esteem items were selected plus one out of the five lie items. Two additional dummy questions were introduced to make the overall tone of the questioning more positive as it was felt this would be more acceptable for children. These questions were presented to all intervention and comparison children. Sub-scale averages were recorded, so an adjusted total, excluding parental items and using the appropriate proportion of each sub-scale, could be calculated to verify that the shortened measure approximates scores from the original. Mean scores from the published manual are shown in Error! Reference source not found. below along with the whole sample mean from this study, lie items excluded in both cases. Comparability of means was high, and the validity of the shortened version of the Battle was accepted.

<table>
<thead>
<tr>
<th></th>
<th>U.S. Grade 3</th>
<th>U.S. Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means from Manual, out of 25 - lie items excluded</td>
<td>16.96</td>
<td>16.13</td>
</tr>
<tr>
<td>Manual scores adjusted</td>
<td>5.32</td>
<td>5.43</td>
</tr>
<tr>
<td>Sample score (n=110)</td>
<td>5.86</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Battle SEI scores — US standardisation data

Scores are out of a maximum of 10 (high self-esteem) and a minimum of 0 (low self-esteem).
3.7 Non-standardised Measures

3.7.1 Writing Sample

As well as reading skills, writing skills were assessed. Children were asked to complete an unsupported writing sample in class. Teachers were asked to set children the task of re-telling a well-known story in their own words in 45 minutes. A copy of these samples was collected for analysis. This task was chosen because it suited both the needs of the school and the research project. Children were routinely asked to complete such a sample at entry to Year Three already, so no additional work for class teachers was required. Teachers were able to adapt the task to the literature they were currently using in class. All children were writing the same story and were not constrained by having to invent a new story. This means that the task does not assess children’s imagination, or story production skills. It also limits the variables between children, children would have a standard vocabulary to draw from and the same story concepts would be included. The task also had ecological validity, being a familiar task in an everyday setting.

The disadvantage of this measure was that the information is more difficult to interpret because a wide range of skills are involved in one task, and no standardised data exists.

At baseline this strategy of combining the needs of school and research worked very well, and all year groups completed the same story (Cinderella). However, there were some problems at post intervention. Teachers were told to repeat the exercise with another story to make the sample as similar as possible, but recognising the demands of
completing their current literacy teaching programme. This meant that each year group completed different stories, and one of these was inappropriate and could not be used. One year group repeated Cinderella, which has the advantage of making comparison with the baseline sample easier, but the disadvantage of allowing for practise to influence outcome. Another year group re-told the story of the Frog Prince, which has the converse advantage and disadvantages. The third year group changed a well known story so that it had a different ending. This clearly introduces new skills into the task and the data from this group was not entirely comparable. A decision was made that this constituted a different task; story length would be affected by how quickly or how well an alternative ending could be imagined by the child, no pool of available story concepts could be derived (see scoring protocol). These writing samples were therefore not used in analysis. Data for this time point is therefore not complete, being drawn from two year groups rather than three at baseline.

A similar approach was taken to that used by More and Wade (1998), attempting to quantify several aspects of the quality of writing produced by the children. Moore and Wade included measures of errors of grammar and punctuation, but these were not included in this sample. Several children in this sample had difficulty producing more than a few words of text so that such complex analysis was inappropriate. When assessing vocabulary, however, grammatical manipulations of words with a common root (e.g. flower, flowery; walked, walking) were each counted, giving a proxy measure of grammatical sophistication. Unlike Moore and Wade, because each class retold the same story a measure of story content could be collected and was valid for comparison between children. Initially a count was made of misformed letters (reversals, use of capitals only etc.) but in practice it was found that in a small number of children had very poor letter formation, while most made very gross mistakes making comparison
difficult. Moreover, the process of classifying errors, or the extent of errors was complex needing a clear definition of acceptable production for each letter, in other words what constitutes an error was difficult to define. Given that this skill is related more to fine motor co-ordination than literacy it was abandoned. A scoring protocol was developed as follows:

- Total number of words used; a count was made of all complete words recorded
- Percentage of words spelt correctly. A set of rules was described for deciding correct spelling counts:
  - One mark for each reported word (e.g. on apostrophe = 4 incorrect words)
  - If word is repeated with attempts to correct spelling only count final version
  - Grammatical correction was not recorded — (e.g. “they’re”, “their”, “there” correctly spelt regardless of grammatical context).
  - Invented words (e.g. “ain’t”) were recorded as incorrect spellings.
  - Alternative spellings were allowed (e.g. “Godmother”, “God mother”)

This was presented as a proportion per hundred words. This measure alone tends to over-rate children who wrote few words but spelt correctly over those who wrote longer stories or used a wider vocabulary.

- Vocabulary as a percentage of different words; as discussed above each word used, including grammatical manipulations of the same word, but excluding plurals, was counted and compared to the total number of words written to produce a percentage proportion.
- Story content: a count of the number of story concepts that each child included in their story was recorded. A list of all concepts appropriate to the story was derived from the story samples themselves and their repetition or omission in other writing
sample was recorded. For example: a reference to the fairy godmother would be counted as the inclusion of one story concept, and the changing of a pumpkin into a coach as a second. Since children were explicitly asked to repeat the story, concepts not appropriate to the story were not included. This decision was made in order to ease comparison between stories, and since it was felt that producing an inappropriate piece of work should be penalised in some way. Thus one child produced a piece of writing which scored highly for vocabulary and spelling, but scored zero for story content as none of the elements of the Cinderella story were included (it was simply a different story).

Since this was an unstandardised test it was necessary to conduct a test of inter-observer reliability on the scoring system. This was conducted on a randomly selected sample of 30 papers from cohorts 1 and 2 baseline sample. In these instances, agreement is often assessed by examining the correlation between the observers, but in the case of continuous data sets this method is not reliable (Bland & Altman 1986). A strong correlation implies a strong agreement in the direction and strength of two measures, but will not distinguish between two markers one of whom records consistently fewer or more instances than the other. For this reason, a measure of agreement using the mean difference between observers and the standard deviation of the differences.

Writing samples were marked by two researchers blind to the assessments of the other marker (each of whom contributed equally to the marking of the whole sample). The mean difference can be compared to score ranges to ascertain whether the difference is acceptable. Small differences between markers demonstrates good agreement. Agreement was found to be good for total number of words (mean difference between markers=1.28, s.d.=2.53, score range 4-300); correct spellings (mean difference=1.93, s.d.=2.43, score range=4-121); vocabulary (mean difference=0.67, s.d.=0.90. score range=4-94) and content (mean difference =0.22, s.d.=0.42, score range 0-21).
3.7.2 Teacher Assessment

Teacher assessments were included in an attempt to get a broad assessment of classroom performance and pupils’ school adjustment. A teacher assessment may provide a better reflection of how a child is coping in school than how they perform under scrutiny (as in a test situation). Teachers were asked to rate each child on the Primary Reading Scale (Barrs et al. 1988; 1996) a five point scale describing increased fluency in reading. This measure has been shown to be a reliable reflection of teacher assessments both over time and between teachers (Barrs et al. 1988; Barrs et al. 1996; Hallam 1997).

Teachers were asked to judge where children fell on this scale:

1. For current reading fluency
2. For the level teachers felt each child had the potential to reach by the end of Key Stage 2
3. For the level teachers felt each child would actually reach by the end of Key Stage 2

Teachers were also asked to rate children on the following aspects of school adjustment:

1. How confident is (name) about their own reading, on a 5 point scale where 5 is very confident and 1 is not at all confident?
2. How disruptive is (name) in class, on a 5 point scale where 5 is very often disruptive and 1 is never disruptive?
3. How much does (name) participate in whole class work (carpet work), on a 5 point scale where 5 is very often and 1 is never?

Teacher assessments were completed in a brief interview or questionnaire. Interviews were preferred by teachers, but this involved finding times during the school day to interview teachers and this proved to be more difficult. Questionnaires were provided for those who could not find time for interviews. Despite the best efforts of the researcher, this data is not complete because of reluctance on the part of the teachers combined with the use of temporary staff making follow-up impossible. All permanent members of staff completed the interview, suggesting that this problem reflected teacher shortage in the area rather than the lack of willingness on their part. A few brief feedback questions were included in the interviews with the aim of asking teachers about how they felt children had progressed, its impact on themselves and their experience of the running of the project.

3.7.3 School Attendance

Poor school attendance is likely to be a cause of poor school achievement, and poor relationships with teachers and peers. It was hoped that the introduction of an intervention, and particularly the involvement of carers and families, would encourage families to send children to school. Attendance records were collected for children in the study groups from the school. Attendance is marked for presence or absence every morning and afternoon by class teachers, so children have a maximum of two attendance periods per day. Records also record reasons for absence and lateness. Unauthorised absences, that is absences for which no reason is recorded, are taken to be a measure of
truancy from school and are used as an indication of the success of schools. In this case all absences, authorised and unauthorised, were recorded since unauthorised absences are rare in this age group, and because the effects of persistent school absence is not restricted to those who truant. Neither was lateness recorded as an absence, since lateness was predominantly an arrival within half an hour of the school day beginning. Very late children were recorded as being absent for that session. Attendance was recorded as a percentage of total sessions available (to allow for bank holidays, inset days etc) every two weeks over the intervention year for all children in study groups.

3.7.4 Semi-Structured Pupil Interview

Pupils were interviewed by the researcher, who was familiar to the children, using a variety of techniques to elicit information about the children’s attitudes toward school and reading. The interviews were conducted alone, out of class in a quiet place where the researcher and pupil could talk undisturbed. The interviews took approximately 20 minutes and followed on from the standardised test discussed above. Children were encouraged to concentrate on the topic at hand, but the tone of the interview was informal and allowed for general conversation in order to put pupils at their ease. Children were assured of their confidentiality regarding their opinions about school, but were also told that if the researcher were concerned for their welfare, outside help would be sought. In response to problems emerging in interviews, concerns were raised with teachers, and followed up by the school staff where appropriate.

Children were initially asked background questions designed to get an overall impression of attitudes toward school, and to put them at ease with the interview.
These were followed with the following structured elements and the Battle self-esteem instrument described above. For full text of interview, including all items below see Appendix II.

In addition to these questions, all children who took part in intervention classes were asked about their experience at the end of the intervention period. In the first intervention group no themes for these conversations were set in advance and the conversations were recorded using a small tape recorder. These interviews were transcribed and examined by the researcher. A number of recurring themes emerged which were then used as topics for conversation in later interviews. These were:

- Learning – what they had learnt, what were the aims of the lessons
- Attitudes toward the reading teacher
- Enjoyment of the lessons
- Conceptions of reasons for attending lessons
- Feelings regarding ending of the lessons

These later conversations were not recorded, and instead notes were taken by the researcher during the conversation and later coded by response. The decision to use note taking rather than tape recordings was taken when it became clear in the initial interviews that the use of the tape recorder was a distraction for the children and tended to become a topic of conversation in itself.
3.7.5  *Reading Self-Concept*

Self-concept has been a frequently used measure of school-adjustment and self-esteem concerning school performance (Chapman, Tunmer, & Prochnow 2001; Rumbaugh & Brown 2000). The impact of intervention on self-concept has been examined in two meta-analyses (Hattie 1992; Elbaum & Vaughn 2001) both of which concluded that academic interventions have been successful in changing self-concepts for children with learning difficulties. Using sub-group analyses Elbaum & Vaughn suggest that this effect is stronger in primary aged children. Self-concept is an important measure to include as an outcome of intervention, but many different aspects of self-concept have been assessed across different domains e.g. subject specific, physical appearance, popularity and as a single global assessment. Measures of self-concept identified from the literature have often involved placing themselves in descriptive groups of children (i.e. “I read as well as some children, but not as well as others”, Blatchford 1997; “are you a good reader” Chapman & Tunmer 1997). This method was felt to be inappropriate for this project because it was targeted to those who were poor readers, and may be sensitive to such a comparison. Alternatively longer sets of tests items have been used to gain a measure of reading self-concept (Chapman & Tunmer 1997). This method was again rejected for this group because asking children to read a long list was not feasible in a group of poor readers. In this context we were particularly interested in academic self-concept and therefore the consideration of whether this forms part of a single attribute of self-concept (Harter 1986) or whether different aspects of self-concept should be seen as different measures (Marsh & Hattie 1996).

Reading self-concept was examined using a visual analogue scale of reading competency created for this purpose. Children were given a “learning to read line” and told that the
further a child was along the line the more they knew about reading (an example, plus
the text used to introduce it are given in Appendix III). They were asked to place a
cross where they saw themselves on this line. They were then asked to place a cross (on
a second line) where they thought they would get to by the time they left junior school.
This activity was designed to elicit current and future concepts of competency in
reading. The analogue scale was used as it was felt to be important that children were
not comparing themselves directly to other children, but to their ideal of a competent
reader.

Children repeated the self-concept exercise for their favourite school activity (e.g.
Maths, games). This test was designed to elicit information about overall school
confidence. Children who gave reading as their favourite activity were asked to
nominate a second favourite and to rate that instead.

3.7.6 Rating of Enjoyment for Reading & Writing

Children were asked to rate their enjoyment of reading and writing at school on a
simple 1-5 Likert scale. Children were given descriptive categories (e.g. "I love reading",
"I like reading, but I don’t really love it") which were recorded by the interviewer as a
score. Children were also prompted for more detailed responses to reading and writing,
see below for reading.

3.7.7 Reading Preferences

Children were asked about their reading preferences:
1. Do you have a favourite book at the moment?
   — What is it?
   — Where is it from?
   — Why is it your favourite?

2. When you take your books home in your book folder every night, do you read them at home?
   — How often?

3. Who do you read with at home?

4. Do you like reading to yourself best or reading to someone out loud?
   — Who do you like reading with?
   — Why do you like reading with x/reading by yourself best?

3.7.8 Generalised Mood Score (GMS)

Children were asked to rate how happy they felt in different circumstances on a 1-5 scale with descriptions (e.g. very happy with a smiling face graphic). A total composite score was calculated from these 8 questions. These questions were:

   — How happy do you feel today?
   — How happy do you feel usually when you’re at school?
   — How happy do you feel at home?
   — How happy do you feel when you’re reading?
— How happy do you feel when you’re singing?
— How happy do you feel when you’re drawing?
— How happy do you feel when you’re doing maths?
— How happy do you feel when you’re playing?

3.7.9 Print Awareness

Children’s print awareness was assessed in order to ascertain sensitivity to text in their environment. A count of all appropriate answers was recorded. Children were asked to think of any things they could “read but weren’t books”. If the child could not produce any examples, then an example was given (“I mean things like magazines”). Only examples different to this were recorded if items were named after this prompt.

3.7.10 Strategy Use and Naming

Recording reading strategy knowledge presents some considerable problems. One approach has been to conduct a miscue analysis (e.g. Arnold 1984; Pearson 1988). This involves a close observation of children while they are reading text and recording the types of mistakes they make according to a descriptive system. These mistakes are interpreted, according to the type of error made, to determine the strategy used by the child when trying to decipher the word. For instance, Arnold suggests classifying whether errors made are consistent with grapho-phonic, syntactic, semantic or no levels.

Children were asked to rate reading of school books to avoid distinguishing simply between those who had access to books at home and those who did not. It was school policy that all children took home a book every evening.
Running records used in Reading Recovery are another example of this method of strategy identification, (Clay 1989), but are used to aid the RR teacher to plan teaching. Running records have been used to record strategy use in RR children (e.g. Schmitt 2001). Errors made in the NARA have also been used for miscue analysis (Moore and Wade 1998). Both of these strategies were attempted on the samples from this study. In the former case, using running records, it was found that the samples taken from the classes were not suitable for two reasons. Firstly it was difficult to identify any texts common between children in different groups thus making between child comparisons difficult. Secondly texts were chosen, according to the principles of RR, at a reading level where children could read 95% of words correctly on their first attempt, resulting in very few words being misread and a small sample of miscues for analysis. The strategy used by Moore and Wade to overcome these problems, that is the use of the NARA texts in this context, was then attempted. However, a proportion of children in the sample group could not read enough text on the NARA to be able to derive a sample for analysis.

In conjunction with this, we were interested in children's awareness of their own strategy use and their ability to recall these strategies. This was particularly relevant to the intervention in question, since an explicit naming of strategies was encouraged as part of the routine tuition. In the semi-structured interview children were asked to name strategies that help them to read if they got stuck. Children were prompted for alternative answers twice, using standard phrases (see Appendix IV for exact text). All appropriate answers were recorded, and these results were analysed as a count of strategies named. Any sensible strategy that might help was recorded, for instance re-reading the sentence, including asking someone else for help, but variations on a
strategy were not counted, for instance asking a teacher for help, asking a friend for help, asking my mum for help would all be recorded but only counted as one strategy.

3.7.11 **Parent Questionnaires**

Home literacy experiences has been thought to have an effect on literacy acquisition and attitudes toward reading (e.g. Senechal et al. 1998; Greaney 1986). In particular the involvement of parents has been shown to have a positive impact on reading outcomes for children from disadvantaged backgrounds (Miedel & Reynolds 1999). The reading teacher was keen for carers to be involved in the project in order to support their children and to feel part of the process. In view of this stance, it was important to try to understand the parents’ and carers’ view of the intervention. Parents’ views were also sought of their children’s reading skills and enjoyment of school in order to have an additional perspective on these.

Questionnaires were sent home to parents in order to obtain their evaluations of the progress of their children, both at home and at school. This proved to be a very difficult avenue of research to pursue. It was clear at the outset that parents were going to be unlikely to complete questionnaires, but it was decided that to attempt to interview all parents would draw resources from the core of the evaluation. The response rate to parent questionnaires was very poor (below 20% overall) so this data source was reluctantly abandoned.

3.7.12 **Reading Diaries**

Children who entered the intervention groups completed reading diaries, carried between school and home to facilitate communication. These were collected as part of
the overall project evaluation and the number of entries made by someone at home was recorded.

3.8 Conclusion

In summary 18 assessments were used as baseline and outcome measures in this study. These measures were used with all pupils at baseline (pre-intervention) and post immediate intervention. They were repeated again for delayed intervention group only at the end of the second intervention period so as to ascertain change in this group over the period of the intervention. A selection of them was completed in the spring term of each year for previous cohorts. This meant that two follow-up data sets were available for cohort one children, one follow-up data set was available for cohort two children and none for cohort three. At all points every effort was made to complete all measures with all children but poor attendance sometimes made this difficult.

This exhaustive data was used in order to reduce the chances of change being missed. Since change was to be initially observed in the short term (over the period of the intervention for RCT phase) it was important to look for subtle changes and change in underlying factors that might facilitate later progress. Measures were selected to observe specific and global changes in both academic and affective modes.
Chapter 4  Results and discussion of baseline data

4.1 Introduction

The aim of this project was to investigate the effect of an intensive literacy intervention on a group of poor readers entering junior school. In order to establish change over the period of the intervention children’s skills and attitudes must be assessed at the outset of the project. A thorough baseline data set was collected and used to assess the characteristics of the school population and the characteristics of children in the study sample. These data was then used to compare the research sample to the overall sample and to explore the relationships between the variables measured.

Initially the data was examined in terms of the overall performance on all standardised tests of children in this sample in comparison to population norms. This established the level of need in the sample, and gave reference points for future gains. A second comparison was made between the children selected for the intervention and their peers in attitudes and performance on tests.

Performance in the sample as a whole was also used to look at some specific factors that might be supposed to affect performance. The two factors explored in this case were gender and home language. Specifically, there has been a great deal of concern around the increasingly poor performance of boys relative to girls (Ruddock & Gray, 2002). Baseline data was analysed in order to find out whether this pattern was apparent in this sample.
Home language is a particular concern in this school because of its ethnically diverse nature, and a large proportion of children speak a language other than English at home as either their first or an additional language. The group of children, described here as those with a home language other than English, comprise those whose families recorded a language other than English spoken at home on school records. Some of these children may be new to English, whilst others may speak both English and another language regularly at home. Nevertheless, there is an underlying assumption that exposure to English is a key element in advancing English vocabulary and literacy. The proportion of children from each of these groups in intervention groups was examined. A group of poor comprehenders was also derived and their representation in the intervention groups examined. The implications of the over or under-representation of these different groups in the intervention sample will be discussed in later chapters.

Finally, relationships between the variables at baseline were examined. These results allow a comparison of the efficacy of the different measures used according to how well they agree with other measures. This data can be used to explore relationships between different aspects of performance, and between performance and attitudinal factors.
4.2 Sample

The whole cohort baseline consists of all eligible children entering Year 3 (aged 7/8 years) over three consecutive school years in the study school for whom consent to participate was granted (n=204). Additional measures were taken for a subgroup i.e.; those children selected for participation in the intervention and a comparison group of class peers. These groups are shown in bold boxes below and are together referred to as the study sample. The selection process for intervention and comparison groups is discussed in full in section 3.3 Selection of Subjects. The number of children available for study at each stage of this process is summarised below (Figure 4:1). Groups making up the cohort have boxes outlined in bold. Whole sample data refers to all children eligible for the study for whom consent was granted (boxes 4.1, 5.3 and 5.4) and is reported to discuss the levels of achievements and attitudes of children in the school in general. Study sample data refers to all children selected for intervention (box 4.1) and those peers selected for comparison (box 5.3) and is reported to discuss the levels of achievement and attitudes in children in the intervention groups compared to peers, and when comparing groups within the sample. Study sample children completed individual assessments in addition to whole class assessments completed in the whole sample. Immediate intervention and waiting list control groups (boxes 5.1 and 5.2) will be compared in Chapter 5.
4.3 Results & Discussion

Results will be presented together with brief discussion of their implication for this project.

A detailed critical discussion of findings and methods will be presented in Chapter 8 –

70
Critical analysis of evaluation and findings.

4.3.2 Response Rates

Multiple attempts were made to complete tests with all children. The only exception to this was writing sample collection, which could only be completed in class at one time. As low attendance is associated with poor achievement it was particularly important to ensure that all children in the sample had the opportunity to complete the tests to ensure that this important sub-group was not missed. Nonetheless, one hundred percent response rates were seldom achieved. The tables below (Table 4:1 and Table 4:2) show the response rates for each test and the proportion of eligible children that this represents. Thus Table 4:1 shows the proportions for the whole cohort (n=204), and Table 4:2 shows the proportions for the study sample (n=126). Response rates were particularly poor for the writing sample both because only one attempt could be made to collect this data and for reasons discussed in section 3.7.1. The total number of children who were interviewed is reported, although for individual questions this number may be lower because the interview was deliberately informal and if a child was reluctant to answer some questions may have been omitted.

### Whole Sample Response Rates

<table>
<thead>
<tr>
<th>Test</th>
<th>Total</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young Group Reading Test</td>
<td>196</td>
<td>96%</td>
</tr>
<tr>
<td>Young Groups Math’s Test</td>
<td>194</td>
<td>95%</td>
</tr>
<tr>
<td>British Picture Vocabulary</td>
<td>196</td>
<td>96%</td>
</tr>
<tr>
<td>Raven’s Matrices (Non-Verbal IQ)</td>
<td>193</td>
<td>95%</td>
</tr>
<tr>
<td>Reading Attitude Survey</td>
<td>162</td>
<td>74%</td>
</tr>
<tr>
<td>Writing Sample</td>
<td>157</td>
<td>76.9%</td>
</tr>
</tbody>
</table>

1 This response rate is lower because in cohort 1 this was administered to study sample only.
Table 4:1 Whole Sample Response Rates

<table>
<thead>
<tr>
<th>Study sample (n=126)</th>
<th>Total</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neale Analysis of Reading Ability</td>
<td>123</td>
<td>97.6%</td>
</tr>
<tr>
<td>Phonological Assessment Battery</td>
<td>122</td>
<td>96.8%</td>
</tr>
<tr>
<td>Pupil Interview</td>
<td>110</td>
<td>87.3%</td>
</tr>
</tbody>
</table>

Table 4:2 Study Sample Response Rates

### 4.3.3 Standards of reading and other academic skills

#### 4.3.3.1 Whole Sample Results for Achievement Tests

**Whole Sample - Young's Group Reading Test**

The graph below (Figure 4:2) shows the distribution of scores from the whole cohort on Young's Group Reading Test. This shows age adjusted, standardised scores such that the mean figure expected is 100 and the standard deviation is 15. This allows a comparison between children regardless of chronological age.
In this sample the mean falls at 95.7, somewhat below the mean expected, showing poor performance on this measure. The distribution appears to be normal and the standard deviation in this sample is similar to that of the normed population sample.

**Whole Sample- Young's Group Maths Test**

The graph below (Figure 4:3) shows the distribution of scores from the whole cohort on Young's Group Maths Test. As with the Young Group Reading Test this is based on age adjusted, standardised scores such that the mean figure expected is 100 and the standard deviation is 15.
In this sample the mean falls at 91.4, somewhat below the mean expected, showing poor performance on this measure. The distribution appears to be normal and the standard deviation in this sample is similar to that of the normed population sample.

**Whole Sample - British Picture Vocabulary Scale**

The graph below (Figure 4:4) shows the distribution of scores from the whole cohort on the BPVS. Again this is based on age adjusted, standardised scores such that the mean figure expected is 100 and the standard deviation 15.
In this sample the mean falls at 88.1, below the mean expected, showing poor vocabulary in this group. In fact, the BPVS mean falls at the 20th centile for the normative population, which is very low indeed. This may be because of the high numbers of children in this group for whom English is an additional language. The distribution is normal and is without skew.

Whole Sample - Non-Verbal IQ (Raven's Matrices)

The graph below (Figure 4:5) shows the spread of scores achieved in the sample population for Raven’s Matrices. A vertical reference line is given, representing the mean score of Raven’s standardisation sample of British children of the same age.
The mean score in the sample population is slightly below the expected mean; the average score expected in this age group is 22, this sample has a mean score of 20.9. There is also a large negative skew in the data set, with far more children than would be expected scoring poorly. The modal score in this sample is 15, and this falls somewhere below the 25th centile in the normed population. However, 20 children (10.4%) achieved scores that would place them in the top 10 percent of the population (scoring 35 or above), and this proportion is that which would be expected. This suggests that it is the mid ranging scores that are missing in this population. An appropriate number of children are doing very well, but the number achieving average scores are depressed and more children scored poorly.

Whole Sample – Discussion
With the exception of Raven's Matrices, it is of note in these results that mean scores are typically depressed in this sample, that distributions remain normal and that standard deviations do not differ greatly from the normed population samples. If the depressed performance in this group was due to a sub-group with very poor performance, in a group with average performance one would expect to see a downward skew in the distribution and an increase in the standard deviation. The lack of these implies that the group as a whole is not scoring as well as expected on measures of skill. In this context it is particularly interesting that the same does not appear to be true of measure of non-verbal IQ. In this case it appears that the central part of the distribution has been shifted downward, while the tails of the distribution remain unaffected. For some children (those at the upper end of the distribution) this implies that their ability (as assessed by the Raven's Matrices) is above their skill level in the areas tested. These results are drawn from a small sample of children, and it possible that the distribution differences are an artefact of the small sample. It would be instructive to know how these findings compare with others where a range of tests has been applied to a similar population. A restricted search of the published literature did not find a paper reporting similar comparisons.

4.3.4 Study sample standards of reading and reading related skills

Study sample Results for Neale Analysis of Reading Ability (NARA)

The mean and distribution of reading accuracy, comprehension and rate as assessed by NARA are shown in Figure 4:6 below. This graph shows as a point the mean difference
between reading age and chronological age for each measure. The enclosed line shows one standard deviation either side of this mean and minimum and maximum scores are shown as crosses. The sample consists of those children chosen for the intervention plus the comparison group (study sample, n=126). This means that this sample constitutes the poorest achievers in this population and it might be expected that the scores would fall below age expected levels. Scores are presented as the difference between reading age and chronological age, in order to allow for sensitivity to chronological age. This means that a child reading at an appropriate level for his/her age will score 0, one month behind will score −1, one month ahead will score +1 and so forth. There were a number of children who could not reach basal level on the NARA (e.g. 9 children on the Accuracy scale) and a basal level was set at 48 months, this being the youngest age that a child can enter full time school. If this adjustment had not been made the data from these children would be lost. Means and standard deviations are shown alongside the graph.
Mean scores for accuracy, comprehension and rate all show that children in this sample are considerably behind their chronological age for reading. Children’s reading ages are on average over 15 months behind their chronological age for reading accuracy and more than 20 months behind for reading comprehension. For both accuracy and comprehension there is at least one standard deviation between the mean difference and the expected difference (i.e. zero difference). Children are not as far behind in reading rate, but the standard deviation on this measure is very large, indicating a wide range of scores. This data also shows that there are children who are performing well, with maximum scores more than 12 months ahead of their chronological age for all three measures.
The mean and distribution of each sub-test of the PhAB are shown in Figure 4:7 below. This shows the scores adjusted for age and standardised such that the mean expected score is 100 with a standard deviation of 15. The mean quotient for each test is shown as a point, with one standard deviation either side of the mean shown as an enclosed line. The results are from all children in the study sample, that is children chosen for the intervention plus the comparison group. Outliers are not shown for this group, since they were defined by the ceiling and basal scores for the standardised distribution. Mean scores and standard deviations are shown alongside the graph.

Children in the sample were performing at below average levels for spoonerism, rhyme fluency and nonword reading. Children performed at expected levels for semantic fluency.
and alliteration fluency. Semantic fluency is not a measure of phonological processing, but of vocabulary and cognitive processing, and the children in the sample showed that they were not deficient in these areas. We can also see by comparing results between subtests that semantic fluency did not correlate significantly with nonword reading, rhyme fluency or spoonerisms (see Table 4:3 below). The fact that children performed at expected levels for both alliteration and semantic fluency suggests that their poor performance at rhyme fluency is not accounted for by a lack of understanding of the task itself, but they were unable to make associations between rhyming words. Since awareness of rhyme appears after awareness of onset (Goswami & Bryant, 1990) it might be concluded that rhyme fluency is a more difficult or advanced task. These data would support this view, showing that children were performing at a similar level on the spoonerisms, nonword reading and rhyme fluency tasks. As with the NARA, a group of children could also be identified who were performing above expected age levels on these tasks. Performance on each of the subtests was in all cases except semantic fluency positively correlated with all other subtest measures (see Table 4:3 below) showing that these tasks were clearly related to each other. The tests also seemed to follow a developmental progression, with children tending to acquire fluency tasks first, followed by nonword reading and finally the spoonerisms task.
Correlation Matrix - Subtests of Phonological Assessment Battery

<table>
<thead>
<tr>
<th>Test</th>
<th>Nonword Reading</th>
<th>Alliteration Fluency</th>
<th>Spoonerisms</th>
<th>Semantic Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhyme Fluency</td>
<td>Pearson Correlation 0.32</td>
<td>0.19</td>
<td>0.45</td>
<td>0.13</td>
</tr>
<tr>
<td>Semantic Fluency</td>
<td>Pearson Correlation 0.14</td>
<td>0.30</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Spoonerisms</td>
<td>Pearson Correlation 0.56</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alliteration Fluency</td>
<td>Pearson Correlation 0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword Reading</td>
<td>Pearson Correlation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 Phonological Assessment Battery; Correlation

Whole Cohort Writing Sample

Table 4:4 presents the data from the children within the whole cohort who completed a writing sample. Since this was not a standardised test there are no norms available to compare the sample data. What can be seen is that overall the standard of spelling is fairly poor, with over 30% of all words written being incorrectly spelt.

<table>
<thead>
<tr>
<th></th>
<th>Mean Scores - Whole Sample (n=167)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Length</td>
<td>85.7 words</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>46.9 words</td>
</tr>
<tr>
<td>% Correct Spelling</td>
<td>67.6%</td>
</tr>
<tr>
<td>Number of Story Concepts</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Table 4.4 Writing Sample; Whole Sample Data

4.3.5 Study sample data – discussion

The data from the study sample is drawn from those identified by teachers or by standardised assessments as poor achievers on at least one of the measures used in assessment. Performance in this group would be expected to be poorer than average expected for the age group, and this was found to be the case. The sample included
children who were scoring at basal levels on all tests, and were struggling particularly in reading and writing tasks. This means that these children were finding it difficult to complete the simplest reading tasks. We can assume that children with this level of literacy difficulty are likely to have great difficulty functioning in class since they are unlikely to be able to read written tasks without considerable support.

4.3.6 Attitudes expressed toward books and reading, Self-esteem and general mood score (whole sample and study sample)

Whole sample data for scores on Reading Attitude Survey (RAS)

The distribution of scores from the sample for the RAS is shown in Figure 4.8 below. The mean score expected for the age group is shown as a vertical reference line, this score (58) is drawn from the original US data set (McKenna & Kear, 1990). The mean and standard deviation from this sample are given to the right of the graph. The data is taken from the whole cohort, that is all eligible children entering Year 3 who completed the survey.

![Histogram of RAS scores](image)

**Total Score on Reading Attitude Survey (Max 80, Min 20)**

1 This response rate is lower because in cohort 1 this was administered to study sample only.
The sample mean is somewhat below the score we might expect from children at this age indicating that these children have a less positive attitude toward reading than the original sample. What this means for the children in this sample is not clear, this may reflect a difference in response between American and English children rather than a difference in attitudes. This measure will be of most interest tracking changes over time and between children.

**Study sample self esteem and general mood**

The table below (Table 4:5) shows the mean scores and standard deviations for four elements from the pupil interview. They are the Battle self esteem inventory, a general mood score generated from the group of “happiness” questions, and reports of how much children liked reading and writing. These are all closed scales and their limits are also shown. In all four cases the higher the score the more positive the attitude expressed by the children.

<table>
<thead>
<tr>
<th>Study sample self esteem and general mood scores</th>
<th>n</th>
<th>mean</th>
<th>Standard Dev.</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battle Self Esteem Inventory</td>
<td>109</td>
<td>6.4</td>
<td>1.81</td>
<td>0-10</td>
</tr>
<tr>
<td>General mood score</td>
<td>110</td>
<td>32.7</td>
<td>5.07</td>
<td>8-40</td>
</tr>
<tr>
<td>Like reading</td>
<td>107</td>
<td>3.9</td>
<td>1.23</td>
<td>0-5</td>
</tr>
<tr>
<td>Like writing</td>
<td>106</td>
<td>3.9</td>
<td>1.35</td>
<td>0-5</td>
</tr>
</tbody>
</table>

Table 4:5 Self Esteem and General Mood; Study Sample Data

The children in this sample show a generally high level of self-esteem and positive attitudes toward reading and writing. When compared to data from age-matched US children
(Grade 3) the Battle Self-Esteem Inventory mean was higher than in this sample (sample mean=6.4; US mean=5.3). The mean score given to questions on the general mood score and to liking reading and writing were above the mid-point (i.e. 24 and 2.5 respectively) indicating a positive attitude. These results may not have been expected of a sample that shows depressed performance in performance at school across the board.

It is notable that this poor achievement in the school as a whole did not appear to be associated with less positive attitudes to books or to low self-esteem in children. Children in the sample reported higher than average levels of self-esteem and attitudes to reading. When asked, the majority said that they enjoyed school and could name the subject that they enjoyed most. This is contrary to evidence (Fontana, 1997) that children with poor academic performance become disenchanted with schooling. It is possible that the results in this case are the result of inaccurate assessments or artefact, but the replication of good self-concept and self-esteem across these measures suggest that they are a true representation of the way the children feel about their school experience. Children in this sample were younger than those used in other samples where the relationship between self-concept and achievement has been shown (Hay, Ashman & van Kraayenoord, 1997; Muijs, 1997). This may indicate that the children in this sample have yet to feel the effects of their poor achievement. Change over time in this sample will provide evidence as to whether this might explain this finding.

The positive attitude of the children is a credit to the school, which has clearly succeeded in supporting the confidence of children despite their apparent disadvantages. On the other hand, this raises the question of the extent to which a lack of perception of their own poor
achievement might create a new set of problems. If the children in this sample feel that their skill level is good, when in fact it is on average considerably below average, will they recognise the need for additional efforts in order to improve their skill. Is it necessary for children who are falling behind in reading and writing to know that they need to work extra hard in this area? Once again later data from this project may illuminate this question, where we can examine changes in attitudes and performance over the period of the intervention.

4.3.7 Relationships between variables

The relationship between variables at baseline was examined. It was hoped that key markers of achievement (with good correlation with other variables) and of discrimination could be identified for use in later analysis. Table 4.6 shows a bivariate correlation matrix of baseline measures, those that were shown to be significant are marked.

<table>
<thead>
<tr>
<th></th>
<th>Young Reading</th>
<th>BPVS Young</th>
<th>Raven's Matrices</th>
<th>NARA Acc</th>
<th>NARA Rate</th>
<th>NARA Comp</th>
<th>Non-word Reading</th>
<th>Total Words Written</th>
<th>Reading Attitude Survey (RAS)</th>
<th>Battle SEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battle</td>
<td>-0.05</td>
<td>0.18</td>
<td>0.08</td>
<td>0.08</td>
<td>-0.02</td>
<td>-0.17</td>
<td>-0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.00</td>
</tr>
<tr>
<td>RAS</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.23</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.06</td>
<td>0.13</td>
<td>-0.15</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Writing Total words</td>
<td>0.28</td>
<td>-0.08</td>
<td>0.12</td>
<td>0.04</td>
<td>0.30</td>
<td>0.14</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-word Rdg</td>
<td>0.55</td>
<td>-0.06</td>
<td>0.37</td>
<td>0.02</td>
<td>0.63</td>
<td>0.19</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Comp</td>
<td>0.53</td>
<td>0.08</td>
<td>0.18</td>
<td>-0.15</td>
<td>0.65</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Rate</td>
<td>0.40</td>
<td>-0.13</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.46</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Acc</td>
<td>0.74</td>
<td>-0.15</td>
<td>0.35</td>
<td>-0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven's Matrices</td>
<td>0.28</td>
<td>0.15</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maths</td>
<td>0.58</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP Vocab Scale</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6 All Baseline Measures; Correlation Matrix

In this analysis the Young Reading Test showed a significant positive correlation with all
other measures of academic performance (Maths, IQ, NARA accuracy, NARA comprehension, NARA rate, length of writing sample and nonword reading). We might have expected that non-verbal IQ test (as assessed by Raven's Matrices) would have been the measure that showed most correlation with other measures since IQ purports to be a measure of underlying ability. However, IQ only correlated significantly with the Reading, Maths and vocabulary scales. Although all the phonic and reading tests may be expected to relate to the Young Reading Test, correlations with the IQ test and the maths test may have more to do with other factors, for instance ability to perform in a test situation, ability to concentrate on one task for a long period of time or ability to follow instructions. This may mean that the measures that were chosen reflect aspects of academic performance rather than ability as measured by the IQ test, and so a general test of academic performance, such as the group reading test, was more predictive than IQ in this case. Alternatively, verbal and non-verbal abilities (as measured here) are closely related to each other.

Self-esteem did not correlate significantly with any other variable. This is contrary to expectations that poor academic performance results in disaffection and low self-esteem as discussed above (4.3.6). Attitude to reading was significantly correlated only with writing sample length. A positive relationship was shown between score on the reading attitude scale and story length in the writing sample (P=0.231, sig=0.006). Children with a more positive attitude toward reading tended to write longer stories. This is interesting because the same pattern is repeated when data is examined by gender and groups (see 4.3.9 below); groups that showed a significantly poorer attitude to reading also tended to write significantly shorter stories.
Few strong or consistent relationships between variables at baseline were shown. The Young Reading test seems the most representative as it correlated with most other baseline measures. This may be because it provides a rough and ready composite reading score—children must read accurately and comprehend the questions to perform well. It will favour those children who perform well in a test situation (although none of the measures were explicitly presented as a test) to the same extent as other measures. The Young Group Reading test was an unpopular choice for use because it was seen by teachers as being difficult to complete and unrealistic in nature. The strong positive correlation with NARA accuracy ($r=0.74$) and NARA comprehension ($r=0.54$) would argue that this was not the case. Despite its lack of appeal to adults, children were not uncomfortable with it and it provided a fair estimate of reading skill and was far quicker to administer than NARA. Bivariate correlations also confirmed that reading attitude and self-esteem were not associated with achievements on any measure of academic achievement.

4.3.8 Characteristics of Intervention Group

Baseline results of children selected for the intervention groups were compared to results from the remainder of the sample. Since some data was collected from study groups only, this does not always represent the remainder of the whole sample. Although it might be argued that it would have been desirable to collect all data for all children in the cohort, a decision was made that a restricted sample would be used for any measure conducted individually with a researcher. In part this was a pragmatic decision to focus data collection on key outcomes, but this strategy has an advantage if comparisons to group norms are to
be made. By weeding out the highest performers from some of the analysis and thus may make a more meaningful comparison to this poor performing sample. In particular it was hoped that if the intervention were successful, this is the group that those within the intervention group could be reasonably expected to “catch up” the typical (rather than mathematical average) performance in the group.

The tables below (Table 4:7 and Table 4:8) are divided into those measures for which whole sample data was available and those applying to study sample only. Referring back to the participation chart Figure 4:1, these refer to all boxes 5.1 to 5.4, and boxes 5.1 to 5.3 respectively. The number of children for whom data was available, the mean score and the standard deviation for each measure are shown along with the results of analyses to determine whether differences existed between intervention subjects and their peers. Where significant differences were found, results are shown in bold for clarity.

**Comparison between intervention groups and peers, whole sample data**

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP Vocab Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>55</td>
<td>86.40</td>
<td>14.90</td>
<td>t=-0.94, p=0.35</td>
</tr>
<tr>
<td>Remainder</td>
<td>136</td>
<td>88.87</td>
<td>17.05</td>
<td></td>
</tr>
<tr>
<td>Raven’s Matrices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>57</td>
<td>17.30</td>
<td>7.51</td>
<td>t=-3.93, p&lt;0.001</td>
</tr>
<tr>
<td>Remainder</td>
<td>131</td>
<td>22.61</td>
<td>8.93</td>
<td></td>
</tr>
<tr>
<td>Young Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>58</td>
<td>84.78</td>
<td>7.19</td>
<td>t=-12.12, p&lt;0.001</td>
</tr>
<tr>
<td>Remainder</td>
<td>132</td>
<td>100.97</td>
<td>8.99</td>
<td></td>
</tr>
<tr>
<td>Young Maths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>56</td>
<td>83.45</td>
<td>9.69</td>
<td>t=-6.93, p&lt;0.001</td>
</tr>
<tr>
<td>Remainder</td>
<td>132</td>
<td>95.08</td>
<td>10.87</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>48</td>
<td>59.02</td>
<td>43.67</td>
<td>t=-4.20, p&lt;0.001</td>
</tr>
<tr>
<td>Remainder</td>
<td>106</td>
<td>96.48</td>
<td>54.34</td>
<td></td>
</tr>
<tr>
<td>Reading Attitude Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>52</td>
<td>55.34</td>
<td>11.39</td>
<td>z=-0.05, p=0.96^</td>
</tr>
<tr>
<td>Remainder</td>
<td>106</td>
<td>55.23</td>
<td>12.24</td>
<td></td>
</tr>
</tbody>
</table>

Table 4:7 Baseline Measures, Comparison of Intervention Groups to Class Peers

^ Non parametric tests, Mann Whitney U Test15.44
Intervention and remainder groups showed very similar results for the BPVS and the RAS and no significant differences between the results of the two groups were found.

On the Young's group Maths and Reading tests, Raven's Matrices (Non Verbal IQ) and the length of story sample, the intervention group score significantly lower than the remainder group. Although the intervention group were selected on the basis of their literacy skills alone, it would appear that they are also struggling in maths and have a lower average IQ.

### Comparison between intervention groups and peers, study sample data

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NARA Accuracy</td>
<td>Intervention</td>
<td>59</td>
<td>-26.73</td>
<td>9.45</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>64</td>
<td>-5.27</td>
<td>11.67</td>
</tr>
<tr>
<td>NARA Comprehension</td>
<td>Intervention</td>
<td>59</td>
<td>-25.58</td>
<td>8.13</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>64</td>
<td>-15.13</td>
<td>9.93</td>
</tr>
<tr>
<td>NARA Rate</td>
<td>Intervention</td>
<td>59</td>
<td>-20.31</td>
<td>14.37</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>64</td>
<td>2.72</td>
<td>26.31</td>
</tr>
<tr>
<td>PhAB Spoonerism</td>
<td>Intervention</td>
<td>59</td>
<td>90.17</td>
<td>11.80</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>99.3</td>
<td>8.83</td>
</tr>
<tr>
<td>PhAB Semantic Fluency</td>
<td>Intervention</td>
<td>59</td>
<td>98.81</td>
<td>15.44</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>99.48</td>
<td>13.90</td>
</tr>
<tr>
<td>PhAB Alliteration Fluency</td>
<td>Intervention</td>
<td>59</td>
<td>98.61</td>
<td>13.62</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>102.87</td>
<td>14.96</td>
</tr>
<tr>
<td>PhAB Rhyme Fluency</td>
<td>Intervention</td>
<td>59</td>
<td>89.15</td>
<td>16.47</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>100.95</td>
<td>15.12</td>
</tr>
<tr>
<td>PhAB Nonword Reading</td>
<td>Intervention</td>
<td>59</td>
<td>87.85</td>
<td>9.57</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>101.56</td>
<td>8.96</td>
</tr>
<tr>
<td>Battle Self Esteem Inventory</td>
<td>Intervention</td>
<td>55</td>
<td>6.31</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>54</td>
<td>6.49</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Table 4.8 Baseline Measures; Comparison of Intervention Group to Class Peers

Significant differences were shown between the intervention and comparison groups on all three NARA subtests, with the intervention group scoring lower on average. The same was found to be true of PhAB spoonerisms, rhyme fluency and non-word reading where intervention group performed significantly poorer than comparison. Intervention and control groups did not differ significantly from comparison group for alliteration fluency, semantic fluency or self-esteem. The groups’ score did not differ significantly on the PhAB

---

1 Non parametric tests, Mann Whitney U Test
semantic fluency, replicating the BPVS data. This suggests that the intervention group and comparison group had an equally wide vocabulary.

It was expected that the intervention group would show poor reading and literacy skills, and this was found to be the case. This confirms that a group of low-achieving children have been identified from the sample as a whole for additional literacy support.

4.3.9 Analysis by Gender

Baseline results for girls and boys were compared to see if differential performance was observed in this sample. Typically boys have been found to score lower on most tests than girls (Ruddock & Gray, 2002). As previously, the results have been grouped into those for which whole sample data was available (Table 4:9), and those for which study sample data was available (Table 4:10). Sample size, mean score and standard deviations are shown for each test along with the results of appropriate statistical analysis. Variables that showed a significant difference between girls’ and boys’ scores are shown in bold for clarity. For score ranges please refer to Chapter 3, section 3.6.

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPVS</td>
<td>Male</td>
<td>96</td>
<td>91.47</td>
<td>14.85</td>
<td>t=2.86, p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>100</td>
<td>84.89</td>
<td>17.22</td>
<td></td>
</tr>
<tr>
<td>Raven’s Matrices</td>
<td>Male</td>
<td>93</td>
<td>22.45</td>
<td>9.57</td>
<td>t=2.42, p=0.02</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>100</td>
<td>19.42</td>
<td>7.81</td>
<td></td>
</tr>
<tr>
<td>Young Reading</td>
<td>Male</td>
<td>96</td>
<td>96.83</td>
<td>11.84</td>
<td>t=1.32, p=0.19</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>100</td>
<td>94.70</td>
<td>10.74</td>
<td></td>
</tr>
<tr>
<td>Young Maths</td>
<td>Male</td>
<td>94</td>
<td>93.55</td>
<td>12.32</td>
<td>t=2.49, p=0.01</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>100</td>
<td>89.42</td>
<td>10.82</td>
<td></td>
</tr>
<tr>
<td>Writing Sample</td>
<td>Male</td>
<td>79</td>
<td>72.72</td>
<td>49.36</td>
<td>t=-2.746, p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>78</td>
<td>95.86</td>
<td>56.06</td>
<td></td>
</tr>
<tr>
<td>Reading Attitude Survey</td>
<td>Male</td>
<td>76</td>
<td>53.02</td>
<td>10.10</td>
<td>Z=-2.56, p=0.01</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>86</td>
<td>56.98</td>
<td>13.05</td>
<td></td>
</tr>
</tbody>
</table>

Significant differences were found between girls and boys for five out of six of these
measures. Boys were found to be scoring higher than girls on average on BPVS (Vocabulary), Raven’s Matrices (Non verbal IQ) and Young’s Group Maths. Girls were found to be scoring on average higher than boys on the reading attitude survey (indicating a more positive attitude to reading) and wrote significantly longer stories than did boys.

Comparisons by gender, study sample data

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NARA Accuracy</td>
<td>Male</td>
<td>58</td>
<td>-14.52</td>
<td>14.35</td>
<td>t=0.73, p=0.46</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>-16.52</td>
<td>15.86</td>
<td></td>
</tr>
<tr>
<td>NARA Comprehension</td>
<td>Male</td>
<td>58</td>
<td>-18.90</td>
<td>9.77</td>
<td>t=1.24, p=0.22</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>-21.25</td>
<td>11.03</td>
<td></td>
</tr>
<tr>
<td>NARA Rate</td>
<td>Male</td>
<td>58</td>
<td>-2.02</td>
<td>30.63</td>
<td>t=2.42, p=0.02</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>-13.95</td>
<td>23.87</td>
<td></td>
</tr>
<tr>
<td>PhAB Spoonerism</td>
<td>Male</td>
<td>57</td>
<td>98.14</td>
<td>11.01</td>
<td>t=3.08, p=0.003</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>92.03</td>
<td>10.84</td>
<td></td>
</tr>
<tr>
<td>PhAB Semantic Fluency</td>
<td>Male</td>
<td>57</td>
<td>100.35</td>
<td>15.39</td>
<td>t=0.37, p=0.84</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>98.11</td>
<td>13.39</td>
<td></td>
</tr>
<tr>
<td>PhAB Alliteration Fluency</td>
<td>Male</td>
<td>57</td>
<td>101.98</td>
<td>14.16</td>
<td>t=0.84, p=0.40</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>99.78</td>
<td>14.69</td>
<td></td>
</tr>
<tr>
<td>PhAB Rhyme Fluency</td>
<td>Male</td>
<td>57</td>
<td>97.14</td>
<td>17.36</td>
<td>t=1.17, p=0.24</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>93.14</td>
<td>16.24</td>
<td></td>
</tr>
<tr>
<td>PhAB Nonword Reading</td>
<td>Male</td>
<td>57</td>
<td>96.32</td>
<td>11.82</td>
<td>t=1.25, p=0.21</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65</td>
<td>93.71</td>
<td>11.17</td>
<td></td>
</tr>
<tr>
<td>Battle Self Esteem Inventory</td>
<td>Male</td>
<td>51</td>
<td>6.68</td>
<td>1.82</td>
<td>Z=-1.41, p=0.16</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>58</td>
<td>6.16</td>
<td>1.78</td>
<td></td>
</tr>
</tbody>
</table>

Boys and girls were performing at equal levels for seven out of nine of these measures. The exceptions to this were spoonerisms, where boys scored significantly higher than girls, and reading rate, where girls were shown to read significantly slower than boys on average.

Contrary to expectations it was found that boys and girls in this sample were either performing at an equal level or boys were out-performing girls. The exceptions to this pattern were that girls had a more positive attitude to reading and wrote longer writing samples. It could be argued that story length is particularly affected by attitude toward
books and stories, so perhaps girls’ better performance on this measure is an aspect of their more positive attitude to books. This reflects the correlation found in section 4.3.7 where positive attitude was associated with longer stories. The areas were boys and girls where performing equally well were those associated with literacy (e.g. NARA accuracy and comprehension). Other findings have suggested that boys may be particularly at a disadvantage in some aspects of literacy (Ruddock & Grey, 2002) and also that boys are more likely to score at the extremes of the distribution, whilst girls are clustered around the middle (Davies & Brember, 1996). This latter finding is of particular interest in this case, where this may be closer to the situation in this school. This may indicate the existence of a good proportion of girls who do not fall at the lower extreme of the distribution, but who are none-the-less struggling. This moderately poor performing group may have the effect of lowering the group mean (as in this case) but may be under recognised as a group since they do not constitute the poorest performing group. This observation is discussed in 4.3.11 below.

### 4.3.10 Home language effects on performance

Pupils’ home language was taken from class registers and data compared between those children whose home language was recorded as English and those children whose home language was recorded as other than English. Within the group of children whose home language was other than English the level of English fluency was not assessed, so fluency in English was variable. No children who were new to English were included in the sample. Again, data is divided into those measures for which data is available from the whole cohort (Table 4:11) and those for which data is available for a restricted sample (Table 4:12).
Sample size, group mean and standard deviation for each variable are presented, together with the results of an appropriate statistical test. When analysis revealed significant differences between groups results are shown in bold for clarity.

<table>
<thead>
<tr>
<th>Home Language</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BPVS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>103</td>
<td>91.39</td>
<td>15.45</td>
<td>t=3.3, p&lt;0.01</td>
</tr>
<tr>
<td>Other</td>
<td>80</td>
<td>83.43</td>
<td>17.07</td>
<td></td>
</tr>
<tr>
<td><strong>Raven's Matrices</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>102</td>
<td>19.42</td>
<td>8.22</td>
<td>t=-2.55, p=0.01</td>
</tr>
<tr>
<td>Other</td>
<td>79</td>
<td>22.77</td>
<td>9.43</td>
<td></td>
</tr>
<tr>
<td><strong>Young Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>104</td>
<td>94.42</td>
<td>10.98</td>
<td>t=-1.26, p=0.21</td>
</tr>
<tr>
<td>Other</td>
<td>79</td>
<td>96.56</td>
<td>11.81</td>
<td></td>
</tr>
<tr>
<td><strong>Young Maths</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>102</td>
<td>91.41</td>
<td>11.30</td>
<td>t=0.26, p=0.78</td>
</tr>
<tr>
<td>Other</td>
<td>80</td>
<td>90.96</td>
<td>12.37</td>
<td></td>
</tr>
<tr>
<td><strong>Writing Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>86</td>
<td>81.51</td>
<td>52.55</td>
<td>t=-0.98, p=0.33</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>90.46</td>
<td>57.16</td>
<td></td>
</tr>
<tr>
<td><strong>Reading Attitude Survey</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Only</td>
<td>82</td>
<td>54.98</td>
<td>11.38</td>
<td>Z=-0.74, p=0.46</td>
</tr>
<tr>
<td>Other</td>
<td>67</td>
<td>56.31</td>
<td>12.37</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11 Baseline Measures; Comparison by Home Language

Children who spoke only English at home were found to score significantly higher than those who spoke more than one language on the measure of vocabulary in English (BPVS). It might have been expected that children who only speak English would have a wider vocabulary in English and would fare better on this measure and this data would confirm this assumption. On a measure of non-verbal IQ (Raven's Matrices) children who spoke only English scored significantly lower, showing a lower average IQ. No significant differences found demonstrated between groups on the other four whole cohort measures.

1 School record keeping changed in the final year of the project so that home language was more difficult to verify from records. In a few cases home language could not be confirmed, and data from this group are omitted from the data set. Whole sample data n is therefore slightly reduced.
Comparison by home language, study sample data

<table>
<thead>
<tr>
<th></th>
<th>Home Language</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NARA Accuracy</strong></td>
<td>English Only</td>
<td>66</td>
<td>-18.89</td>
<td>13.38</td>
<td>t=-2.36, p=0.02</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>58</td>
<td>-12.67</td>
<td>16.02</td>
<td></td>
</tr>
<tr>
<td><strong>NARA Comprehension</strong></td>
<td>English Only</td>
<td>66</td>
<td>-21.8</td>
<td>9.95</td>
<td>t=-1.39, p=0.17</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>58</td>
<td>-19.24</td>
<td>10.63</td>
<td></td>
</tr>
<tr>
<td><strong>NARA Rate</strong></td>
<td>English Only</td>
<td>66</td>
<td>-15.53</td>
<td>23.27</td>
<td>t=-3.02, p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>58</td>
<td>-0.91</td>
<td>30.45</td>
<td></td>
</tr>
<tr>
<td><strong>PhAB Spoonerism</strong></td>
<td>English Only</td>
<td>65</td>
<td>94.89</td>
<td>12.06</td>
<td>t=0.10, p=0.92</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>57</td>
<td>94.68</td>
<td>10.46</td>
<td></td>
</tr>
<tr>
<td><strong>PhAB Semantic Fluency</strong></td>
<td>English Only</td>
<td>65</td>
<td>100.35</td>
<td>14.88</td>
<td>t=0.28, p=0.98</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>57</td>
<td>97.73</td>
<td>14.33</td>
<td></td>
</tr>
<tr>
<td><strong>PhAB Alliteration Fluency</strong></td>
<td>English Only</td>
<td>65</td>
<td>97.54</td>
<td>14.72</td>
<td>t=-2.79, p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>57</td>
<td>104.71</td>
<td>13.30</td>
<td></td>
</tr>
<tr>
<td><strong>PhAB Rhyme Fluency</strong></td>
<td>English Only</td>
<td>65</td>
<td>96.72</td>
<td>16.89</td>
<td>t=0.99, p=0.32</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>57</td>
<td>93.68</td>
<td>16.81</td>
<td></td>
</tr>
<tr>
<td><strong>PhAB Nonword Reading</strong></td>
<td>English Only</td>
<td>65</td>
<td>92.71</td>
<td>11.09</td>
<td>t=-2.21, p=0.03</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>57</td>
<td>97.27</td>
<td>11.56</td>
<td></td>
</tr>
<tr>
<td><strong>Battle Self Esteem Inventory</strong></td>
<td>English Only</td>
<td>61</td>
<td>6.25</td>
<td>1.80</td>
<td>Z=-0.95, p=0.35</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>47</td>
<td>6.59</td>
<td>1.85</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.12: Baseline Measures, Comparison by Home Language

Of the nine measures considered in the restricted sample, four significant differences between the groups were found. These were NARA accuracy and rate, PhAB Alliteration Fluency and nonword reading. In all of these cases the children who spoke another language as well as English performed better than their English only speaking peers.

Overall it would seem that children who speak another language as their first language or in addition to English are performing better in this sample than their English only speaking peers. It is perhaps interesting to note that the advantage that these children seem to have was not also apparent in reading comprehension, where the difference between groups did not reach significance. Their skills are particularly pronounced in areas of decoding (NARA accuracy and nonword reading).
Children who speak English exclusively appear to be at a disadvantage in this sample. Children who spoke a language other than English at home performed significantly better than those who spoke English only on Raven's Matrices, NARA Accuracy, NARA rate, PhAB Spoonerism, alliteration fluency and nonword reading. Children who were new to English were not included in this sample and many of the children had been in the school since nursery (aged 3 or 4 years). It might be assumed that there were a large proportion of children in the group who had extended exposure to English, and who may be classified as bilingual, rather than having English as their “additional” language. It was found that children who spoke English in addition to another language did have a smaller vocabulary than English only speaking children, but outperformed them on other literacy measures. The findings that there is a minority of poor, white children who are doing less well than their peers is confirmed by previous HMI findings from the same area, where pupils classified as white and on free school meals were the lowest scoring group on a NARA assessment in Year 6, and constituted a “significant proportion of those stuck at or below a reading age of 8 years” (HMI, 1996). In other words there seems to be additional resources available to the poorest non-white children, which enables them to achieve better than their white peers. To suggest reasons why this may have been the case in this sample would be conjecture.
4.3.11  Population trends as reflected in intervention groups

4.3.11.2  Gender Trends

As discussed above (see section 4.3.9) in this sample, girls were performing worse than boys, although this was less pronounced in their reading skills. The proportion of girls and boys in the overall sample and in the study sample was examined to determine whether either gender was over or under represented in each of the groups (see Figure 4:9). No significant differences were found between the expected and observed proportions of boys and girls in any group \( \chi^2 = 1.8 \). There was no gender effect on group selection with proportions equal to that in the general population.

![Distribution by Gender](image)

Figure 4:9  Selection for Intervention; Distribution by Gender

Given the findings shown in 4.3.9 above we might have expected to see a greater proportion of girls being selected for the intervention group than boys, since we know that girls are doing less well in this sample. However, it seems that girls have not been referred
for additional support in greater numbers. This could be as a result of the decision to take those with worst performance, i.e. that although the average score of girls were lower they did not constitute the greatest of those with very low scores. It may also be explained by the girls' equal better attitudes toward books. Those girls who do not do as well in literacy but enjoy books and have a positive attitude may be seen as having low ability rather than low achievement and therefore less in need of additional support. It may also be the case that the differential performance observed was an artefact of the assessment measures used, that boys responded better to the test situation than did the girls and their performance in class was, in fact, equal and classroom performance was a more salient assessment of suitability for selection for additional literacy support.

4.3.11.3 Language Trends

As reported in section 4.3.10 children who spoke only English were performing at a lower level than those who spoke another language as well on literacy based tests. Once again the proportions of children in the sample population and in the study sample were compared. The distribution is shown in Figure 4:10 below. This table shows that the proportion of children whose home language is English is greater in the intervention group than in the sample as a whole. When a Chi-squared statistic was calculated this observation was confirmed, children with English as a home language were more likely to be included in the intervention group ($\chi^2=7.7$, p=0.02). The proportions of children in the study sample reflected the differential performance observed.
In this case it would seem that there was good agreement about the level of need between standardised assessments and group allocation. The standardised assessments indicated a greater need for support in children who spoke only English and this greater need was reflected in the proportions of children allocated places for additional literacy support.

4.3.11.4 Poor comprehenders represented in intervention groups

As discussed in Chapter 2 section 2.4.3 there is a school of thought that suggests some children are particularly poor at comprehension, and that the needs of this group are less often recognised as being in need of remediation than their peers with poor all round reading skills. To establish whether such a trend existed in this group, a group of children
with poor comprehension was derived and compared to those who were selected for intervention. Those children with NARA comprehension scores of less than or equal to 22 months behind their chronological age were identified. This cut off was selected as that which achieved a sample size approximately equal to that of the hypothetical intervention group (that is n=66, see 3.3). The distribution of children in this group is shown in Table 4:13 below.

Distribution of poor comprehenders in study groups

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n)</th>
<th>Comparison (n)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor comprehenders</td>
<td>39 (67.2%)</td>
<td>19 (32.8%)</td>
<td>58</td>
</tr>
<tr>
<td>Peers</td>
<td>20 (30.3%)</td>
<td>46 (30.3%)</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>59 (47.2%)</td>
<td>65 (52.4%)</td>
<td>124</td>
</tr>
</tbody>
</table>

It can be seen that around 1/3 of those with poor comprehension scores were not allocated to intervention. When the NARA accuracy scores of poor comprehenders was compared according to their allocation to intervention or not, it was found that those in the poor comprehenders group that were not allocated to intervention had significantly higher average accuracy scores than those who were allocated to interventions. The mean average difference between comprehension age and chronological age for the non-intervention poor comprehenders was —14 months, and the mean average difference for the intervention poor comprehenders was —30 months. This may imply that teachers were less sensitive to comprehension skills, and thus they were under-represented in this group. The nature of teacher selection as a strategy will be discussed further in Chapter 8 – Critical Analysis, as will the outcome for these non-intervention poor comprehenders.
4.4 Conclusions

The performance of the whole cohort is depressed with reference to population norms. This sample performed at below average levels for Young Group Reading Test, Young Group Maths Test, Raven Matrices, NARA accuracy, comprehension and rate, PhAB spoonerisms, rhyme fluency and nonword reading. This is in line with the Key Stage 2 SATs results in this school; fewer children than average are achieving level four in English.

It would also seem that while the distributions of results for some tests are depressed, with fewer children than expected achieving high scores, there remains a proportion of children in the school who performed very well in these tests. Currently the school is targeting its resources toward the children who are struggling to achieve, but it could be argued the there is also need at the other end of the spectrum. Children who are performing well may find themselves unchallenged in the context of the low average achievements of their peers.

As might be expected the literacy skills of the intervention group were found to be significantly worse than their peers.

Poor performance in reading and reading related skills were clearly established in this sample. Some anomalies were discovered since boys were not at a disadvantage relative to girls, and children whose only language was English seemed to be at a disadvantage. Overall the children in this cohort were found to be positive about school and reading.

The baseline data presented here was used to establish change over time in the intervention and comparison groups.
Chapter 5  Results of Randomised Controlled Trial

5.1  Introduction

This chapter examines the evidence for change found in the sample as a result of the intervention across measures of literacy skill and pupil attitudes. In order to attribute any changes to the impact of the intervention and not maturational or chance changes data from the intervention group will be compared with data from the waiting list control group (referred to as control group throughout). Data from standardised tests will be examined, as will attitudinal measures and teacher assessments.

The use of randomised controlled trials to establish effectiveness if intervention in educational settings has not been taken up with the enthusiasm that it has received within medicine (New Scientist 28/02/02; www.cochrane.org). It has been argued that it is unethical either to identify children as at risk and not address their needs, or to deprive children of a new advance that may be beneficial to them. This argument relies on an a priori assumption that intervention or new teaching approach will benefit children. This may not be the case. It is important to ascertain not only whether progress can be seen in a group involved in a teaching/intervention programme that is better than current practice, but also to establish causality where change is observed. The use of randomised controlled trials is the best way to infer causality and establish effect size. Such a methodology has historically been used in social interventions including educational interventions (Oakley, 1998), but is difficult to implement. There are features of the educational system that make such trials difficult to run. Since every school operates independently of all others, differences between schools can be large, even between schools that are geographically and sociologically similar. In addition
there is evidence across disciplines (Wilson, Lipsey, & Derzon, 2002) that suggests that being involved in new initiatives may have effects on the school beyond the immediate ones. For instance, being involved in a new teaching scheme may boost staff morale, may introduce a subject specialist to a staff team, or provokes an interest in that area for those not directly involved. In these ways comparing change in a sub sample (such as poor readers) of those in treatment and control schools may not always provide fair comparison. In order to overcome these problems, this study employed a waiting list control method in one school in which children were randomly allocated to either first or second intervention blocks. Thus the second intervention group provided control data during their waiting period.

This chapter reports an evaluation of the impact of the intervention on pupils' performance and mood during the controlled trial phase. This represents data collected over the shaded period in the evaluation summary plan below (Table 5:1). As described in section 3.2 this represents the period of the controlled trial where the first intervention groups take place and the waiting list control group are in a no treatment control period.
### Overall Plan for Evaluation

<table>
<thead>
<tr>
<th>Year</th>
<th>Autumn Term</th>
<th>Spring Term</th>
<th>Summer Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 1st intervention</td>
<td>End 1st intervention Post-measures</td>
</tr>
<tr>
<td>2</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 3rd intervention</td>
<td>End 3rd intervention Post-measures</td>
</tr>
<tr>
<td>3</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 5th intervention</td>
<td>End 5th intervention Post-measures</td>
</tr>
</tbody>
</table>

**CONTROL PERIOD**

Table 5.1 Overall Plan for Evaluation

Results of intervention and control groups were compared at baseline and again post intervention. In this way changes that are found in the intervention group that are not found in the control group can be attributed to the effects of the intervention.

### 5.2 The use of age adjusted scores

After children were selected for the intervention groups they were randomly assigned to either first or second intervention. Those assigned to the second intervention group constituted the control group over the period of the first intervention. A small number of children (6 out of 59 over the three years) were prioritised to the first intervention groups because of internal issues within the school. This has the potential to introduce inaccuracies in the data for any statistical test that assumes group independence, as most parametric tests do. For this reason it was considered preferable to analyse the results whenever possible, using baseline scores as a predictive variable rather than assuming that group means can be compared as independent samples. If one assumes that two groups are randomly assigned, when in fact there is even a small downward pressure on scores in one group (for instance if some children with poorer skills are deliberately
assigned to one group) then evidence for significantly different group means may be inaccurate. Although the number of children taken out of the randomisation was small, and they were not withdrawn from randomisation on the basis of skill level alone, the possibility of this affecting the results was thought important enough to select analyses with this issue in mind.

The decision to use baseline score as a predictive variable has a second advantage in this case. For all tests of academic achievement an allowance for chronological age must be made. Skills are assumed to progress with age and population distribution will reflect this. That is to say, reading at the same level whether aged 7 years or 9 years will place a child in a different position relative to his or her peers. Raw scores on standardised tests are therefore typically converted to age related scores using data from the test production. However, in this case we are examining the scores of poor readers identified on reading and reading related tests. Three problems arise from this:

1. A large number of children may fall below the minimum (basal) score for which population interpretation data is available and their data will be lost.

2. The accuracy of age-adjusted data at the tail ends of the distribution deteriorates. Since these tail ends contain necessarily fewer children in the sample from which population norms are derived, age adjustments are likely to be based on a small number of children and may estimate reading age less accurately.

3. The standardisation process reduces variance in the sample. Very small changes in raw score at the ends of the distribution can result in disproportionately large changes in converted score. Changes made at the extremes will therefore disproportionately alter mean change and mean score.
The reduction of variance in the sample by the use of standardised scores has been discussed as a problem with reference to the variance in a sample (Bast & Reitsma, 1998). Although in this study variance is not the key variable in consideration in itself, reduction in variance may make distortion of results from those at the extremes of the distribution more likely. Taking these problems into consideration it was decided that it would be preferable to use raw scores in the analysis, and chronological age as a second predictive variable.

Thus, for measures which met parametric assumptions and which contained an age-related component, a multiple regression analysis was conducted to examine post intervention scores controlling for baseline score and chronological age. Each outcome measure was used as the criterion measure in an analysis using the corresponding baseline measure and chronological age at baseline as predictor variables, followed by group in an ordered analysis. This means that the predictive effects of group can be explored over and above the effects of baseline score and age. Final scores between groups were not, therefore compared, but instead change in score relative to baseline. Baseline scores were compared between intervention and control groups and no significant differences were found (data not reported). Since this method of analysis assesses relative change rather than absolute change, however, baseline group equivalence does not impact on the validity of the findings.
5.3 Post Intervention Results and Discussion

5.3.1 Neale Analysis of Reading Ability

NARA Accuracy
The scattergram following (Figure 5:1) shows post intervention NARA Accuracy plotted against baseline NARA Accuracy. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. However, regression line for the intervention group appears to have a steeper gradient than that shown in the control group, indicating a faster rate of progress in the intervention group. This regression line, however, has not been adjusted for chronological age at the time of testing.

![Figure 5:1 NARA Accuracy; Scattergram showing change over time](image)

The descriptive statistics shown in Table 5.3 show that although the intervention group began with a slightly lower average score for NARA Accuracy, post intervention the groups have very similar scores.
NARA Accuracy: Intervention and control group pre and post intervention mean and standard deviation

<table>
<thead>
<tr>
<th>NARA Accuracy</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>11.03</td>
<td>14.27</td>
</tr>
<tr>
<td>S.D.</td>
<td>6.10</td>
<td>7.72</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>19.61</td>
<td>19.82</td>
</tr>
<tr>
<td>S.D.</td>
<td>7.62</td>
<td>9.06</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.2 NARA Accuracy; Pre/Post Intervention Data

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2 = 0.61$, $F(2,53) = 42.51$, $p<0.001$. This was largely due to the predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did predict post intervention score over and above these, $R^2_{change} = 0.06$, $F(1,52) = 8.99$, $p=0.04$. The effect reached statistical significant, although the size of effect was small accounting for just 6% of the variance between groups.

It can be concluded that participation in the intervention had a modest, positive effect on progress in reading accuracy as measured by NARA.

NARA Comprehension

The scattergram following (Figure 5:2) shows post intervention NARA Comprehension plotted against baseline NARA Comprehension. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. However, this relationship appears less strong than for Accuracy, the scores are spread further away from the regression line implying greater variability in change. Both regression lines have positive gradients, showing improvement over time. This is confirmed by
descriptive data (see Table 5:3). However, in both groups it appears that greatest change is seen in those with poorest scores, and little change is shown in those with higher scores at baseline.

The descriptive statistics shown in Table 5:2 show that although the intervention group began with a slightly lower average score for NARA Comprehension, post intervention the groups have very similar scores. This data includes individual pupils who made large losses and gains relative to the mean change, demonstrated by the comparatively large standard deviations.

<table>
<thead>
<tr>
<th>Comprehension</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>3.79</td>
<td>4.97</td>
</tr>
<tr>
<td>S.D.</td>
<td>2.68</td>
<td>3.07</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>6.29</td>
<td>6.07</td>
</tr>
<tr>
<td>S.D.</td>
<td>2.73</td>
<td>3.11</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5:3 NARA Comprehension; Pre/Post Intervention Data
The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.24$, $F(2,53)=8.37$, $p=0.01$. This was largely due to predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.31$, $F(1,52)=2.22$, $p=0.14$. Although there did appear to be greater change in the intervention group, the variability of change in both groups means we cannot be sure this is not a chance observation.

This investigation failed to demonstrate that the intervention had a significant, positive effect on progress in reading comprehension as measured by NARA.

This lack of effect may be because the intervention did not promote change in comprehension, but it is worth considering the possibility that improvements in reading comprehension necessarily lag behind improvements in other areas of reading, such as accuracy. The intervention group was a little behind the control group for both accuracy and comprehension at baseline, but while accuracy can proceed regardless of comprehension skills, comprehension skills can only develop once text can be successfully read. Had the control group had continued for some time after the end of the first intervention period, allowing a longer period for comparison, evidence may have emerged to confirm or refute this notion. Since the control group was a waiting list control group and immediately received the same intervention themselves, this data was not available.

**NARA Rate**

The scattergram following (Figure 5:3) shows post intervention NARA rate plotted against baseline NARA rate. This suggests a relationship between baseline and post
intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. This scattergram is more difficult to interpret by eye, since an outlier of the intervention group, who scored very highly at baseline and poorly post intervention, may have a disproportionate affect on average change.

![Scattergram showing change over time](image)

**Figure 5:3** NARA Rate; Scattergram showing change over time

The descriptive statistics shown in Table 5:4 show that the intervention group had lower average scores for NARA Rate than the control group both at baseline and post intervention. However, this average change in both groups includes individual pupils who made large losses and gains relative to the mean change, demonstrated by the large standard deviations, in each case of around a half of total score.
NARA Rate; Intervention and control group pre and post intervention mean and standard deviation

<table>
<thead>
<tr>
<th>Rate</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>27.84</td>
<td>35.25</td>
</tr>
<tr>
<td>S.D.</td>
<td>22.49</td>
<td>22.03</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>29.76</td>
<td>35.41</td>
</tr>
<tr>
<td>S.D.</td>
<td>14.98</td>
<td>16.89</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5:4 NARA Rate; Pre/Post Intervention Data

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.42$, $F(2,53)=18.81$, $p<0.01$. This was largely due to the predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.004$, $F(1,52)=0.32$, $p=0.57$.

This investigation failed to demonstrate that the intervention had a positive effect on progress in reading rate as measured by NARA.

Reading rate is a problematic measurement of reading skill. While it is true that a successful reader will read rapidly, a rapid reader is not always successful reader. In fact a reader may slow down his or her rate in order to pay better attention to accuracy or comprehension. Indeed, some poor readers employ a strategy of reading text at great speed to disguise their lack of understanding. This is reflected in the fact that of all three parts of the NARA, assessment reading rate showed the greatest variation between children, with some of the poor readers in this group managing to read at the rate expected of a 13 year old. In turn, this great variation of speed and the positive aspects of unhurried reading make changes in this measure difficult to interpret.
5.3.2 Phonological Assessment Battery

**PhAB Spoonerisms**
The scattergram following (Figure 5:4) shows post intervention PhAB Spoonerism plotted against baseline PhAB Spoonerism. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. The regression lines for the intervention and control groups would suggest that progress was largely similar between the groups.

![Figure 5:4 PhAB Spoonerisms; Scattergram Showing Change Over Time](image)

The descriptive statistics shown in Table 5:5 show that although the intervention group began with a slightly lower average score for PhAB Spoonerisms, post intervention the groups have very similar scores. This data mirrors that in the scattergram above, showing slightly better progress in the intervention group.
PhAB Spoonerisms: Intervention and control group pre and post intervention mean and standard deviation

<table>
<thead>
<tr>
<th>Spoonerism</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>5.14</td>
<td>6.27</td>
</tr>
<tr>
<td>S.D.</td>
<td>4.51</td>
<td>5.82</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>8.93</td>
<td>8.14</td>
</tr>
<tr>
<td>S.D.</td>
<td>6.22</td>
<td>5.21</td>
</tr>
<tr>
<td>n</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.5 PhAB Spoonerisms; Pre/Post Intervention Data

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.33$, $F(2,52) = 12.78$, $p<0.01$. This was largely due to predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.02$, $F(1,51)=1.94$, $p=0.17$.

It cannot be concluded that participation in the intervention had an effect on progress in spoonerisms.

PhAB Semantic Fluency
The scattergram following (Figure 5.5) shows post intervention PhAB Semantic Fluency plotted against baseline PhAB Semantic Fluency. This scattergram suggests that most change in both groups was shown in those children with lowest baseline scores and that overall there is a trend to score lower post intervention than pre-intervention.
The descriptive statistics shown in Table 5:6 show that the intervention and control groups began and ended the intervention period with very similar scores for PhAB Semantic Fluency, and that very little change was observed in this measure.

<table>
<thead>
<tr>
<th>Semantic Fluency</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>16.59</td>
<td>16.93</td>
</tr>
<tr>
<td>S.D.</td>
<td>5.41</td>
<td>6.24</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>16.89</td>
<td>16.64</td>
</tr>
<tr>
<td>S.D.</td>
<td>3.83</td>
<td>6.21</td>
</tr>
<tr>
<td>n</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.20$, $F(2,52) = 6.39$, $p<0.003$. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.0001$, $F(1,51)=0.008$, $p=0.93$. 
This data failed to demonstrate a positive effect of the intervention on progress in semantic fluency.

**PhAB Alliteration Fluency**

The scattergram following (Figure 5:6) shows post intervention PhAB Alliteration Fluency plotted against baseline PhAB Alliteration Fluency. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups, but it also appear that the rate of change may be higher for those with lower baseline scores. The two regression lines suggest very similar rates of change between the two groups.

![Figure 5:6 PhAB Alliteration Fluency; Scattergram Showing Change Over Time](image)

The descriptive statistics shown in Table 5:7 show that the intervention and control groups began with very similar average scores for PhAB Alliteration Fluency. Post intervention the intervention group shows a slightly higher average score than the control group. This data mirrors that in the scattergram above, showing slightly better progress in the intervention group.
PhAB Alliteration Fluency: Intervention and control group pre and post intervention mean and standard deviation

<table>
<thead>
<tr>
<th></th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>9.72</td>
<td>9.50</td>
</tr>
<tr>
<td>S.D.</td>
<td>3.92</td>
<td>3.79</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>10.37</td>
<td>9.36</td>
</tr>
<tr>
<td>S.D.</td>
<td>3.62</td>
<td>3.29</td>
</tr>
<tr>
<td>n</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.7 PhAB Alliteration Fluency; Pre/Post Intervention Data

The results of an ordered regression analysis confirm that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.21$, $F (2,52) = 6.96$, $p=0.02$. This was largely due to predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.01$, $F (1,51)=0.80$, $p=0.38$.

It cannot be concluded that participation in the intervention had a positive effect on progress in alliteration fluency.

PhAB Rhyme Fluency

The scattergram following (Figure 5.7) shows post intervention PhAB Rhyme Fluency plotted against baseline PhAB Rhyme Fluency. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. The regression lines for the intervention and control groups would suggest that progress was largely similar between the groups.
The descriptive statistics in Table 5:8 show that the intervention group began and ended the intervention period with a slightly higher mean score for PhAB Rhyme Fluency than the control group. However, the relatively large standard deviations indicate that a number of pupils made gains or losses very different from the mean change observed.

PhAB Rhyme fluency: Intervention and control group pre and post intervention mean and standard deviation

<table>
<thead>
<tr>
<th>Rhyme Fluency</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>4.66</td>
<td>3.33</td>
</tr>
<tr>
<td>S.D.</td>
<td>4.05</td>
<td>3.62</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>6.56</td>
<td>5.00</td>
</tr>
<tr>
<td>S.D.</td>
<td>3.38</td>
<td>3.92</td>
</tr>
<tr>
<td>n</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5:8 PhAB Rhyme Fluency; Pre/Post Intervention Data

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.48$, $F(2,52) = 24.44$, $p<0.01$. This was largely due to predictive effect of baseline score. The
second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.01$, $F(1,51)=1.34$, $p=0.25$.

It cannot be concluded that participation in the intervention had a positive effect on progress in rhyme fluency.

**PhAB Nonword Reading**

The scattergram following (Figure 5:8) shows post intervention PhAB Nonword Reading plotted against baseline PhAB Nonword Reading. This suggests a relationship between baseline and post intervention scores, with higher baseline scores being associated with higher post intervention scores for both intervention and control groups. There are a number of pupils scoring zero both at baseline and at post intervention on this measure.
The descriptive statistics shown in Table 5.9 show that although the intervention group began with a slightly lower average score for PhAB Nonword Reading, post intervention the intervention group has a higher average score. This data also demonstrates the wide range of scores recorded in the large standard deviations observed.

**PhAB Nonword reading: Intervention and control group pre and post intervention mean and standard deviation**

<table>
<thead>
<tr>
<th>Nonword Reading</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>3.52</td>
<td>4.03</td>
</tr>
<tr>
<td>S.D.</td>
<td>3.18</td>
<td>3.11</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>6.30</td>
<td>5.50</td>
</tr>
<tr>
<td>S.D.</td>
<td>4.87</td>
<td>4.53</td>
</tr>
<tr>
<td>n</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.9 PhAB Nonword Reading; Pre/Post Intervention Data

The results of an ordered regression analysis confirmed that there was a large and significant effect attributable to baseline score and chronological age, $R^2=0.43$, $F(2,52) = 19.84$, $p<0.01$. This was largely due to predictive effect of baseline score. The second analysis, examining change over and above this effect, found that group did not predict post intervention score over and above these, $R^2_{\text{change}}=0.08$, $F(1,51)=0.73$, $p=0.40$.

It cannot be concluded that participation in the intervention had a positive effect on progress in nonword reading.

**Discussion of PhAB results**

Within the five elements of the Phonological Assessment Battery used none showed a significantly different rate of progress between intervention and control groups. For the test of semantic fluency this is not surprising, it is a test of vocabulary and cognitive
processing rather than a test of phonological skills. The test which the intervention might have been expected to have had the most direct effect upon, was the non-word reading test. Essentially children are being asked to decode phonologically regular unknown words. Not only will children have spent additional time practising this skill on real words during the intervention period, but also regular patterns in words (such as “an” or “at”) were explicitly pointed out to them. The reasons why this anticipated effect was not seen in these children are unknown. It was also noted that phonological skills did not increase in a study where deficiencies present in poor readers were not ameliorated by RR teaching (Chapman, Tunmer & Prochnow, 2001).

5.3.3 Writing skills

Writing sample length

The total lengths of writing samples were calculated and compared pre and post intervention. A comparison of the data from the intervention and control groups is shown in the scattergram below (Figure 5:9). Both groups appear to have improved over time, but children in the control group appear to have made slightly more rapid progress, indicated by the rate of incline on the regression line.
The data in the table below (Table 5:10) confirms that data indicated in the scattergram, showing progress in both groups, but greater change in the control group than in the intervention group. However the very large standard deviations shown indicate wide variation in change scores in both groups.

### Writing Sample story length: Intervention and control group pre and post intervention mean and standard deviation.

<table>
<thead>
<tr>
<th>Total number of words</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>59.33</td>
<td>58.71</td>
</tr>
<tr>
<td>S.D.</td>
<td>44.92</td>
<td>43.34</td>
</tr>
<tr>
<td>n</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>91.31</td>
<td>102.32</td>
</tr>
<tr>
<td>S.D.</td>
<td>62.37</td>
<td>62.57</td>
</tr>
<tr>
<td>n</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

The results of an ordered regression analysis showed that no significant effect was demonstrated either over time adjusted by baseline score and chronological age, $R^2=0.09$, $F (2,23)=1.09$, $p=0.35$. The second analysis, examining change over and
above this effect, found that neither did group predict post intervention score, $R^2_{\text{change}}=0.03$, $F(1,22)=0.70$, $p=0.41$. Thus we cannot conclude whether length of writing sample tends to increase over time, or whether the intervention had any affect on change.

**Writing sample vocabulary**

The number of different words used in the writing sample was compared between groups over time. These data are represented in the scattergram below (Figure 5:10). As with previous scattergrams, it should be borne in mind that these results are not adjusted to age at time of testing. A relationship between pre and post intervention score in the intervention group is not apparent. In comparison the control group did appear to show improvement over time, at least for those with poorer scores at baseline.

![Figure 5:10 Writing Sample Vocabulary; Scattergram Showing Change Over Time](image-url)
The data shown in Table 5:11 below, giving mean scores and standard deviation pre and post intervention, explain the data in the scattergram above. Whilst average scores are shown to increase in both groups over time, the standard deviations in both groups also increased. Since the data set post intervention is not complete, average score does not accurately represent the average change. Average change scores were therefore also calculated. Mean score change in the intervention group was shown to be 48.75 (s.d.=68.69), compared to 21.79 (s.d.=64.02) in the control group. The results of an ordered regression analysis showed that no significant effect was demonstrated either over time adjusted by baseline score and chronological age, $R^2=0.04$, $F (2,22)=0.40$, $p=0.68$. The second analysis, examining change over and above this effect, found that neither did group predict post intervention score, $R^2_{\text{change}}=0.63$, $F (1,21)=0.63$, $p=0.44$. Thus we cannot conclude whether length of writing sample tends to increase over time, or whether the intervention had any effect on change.

**Writing Sample Vocabulary Used: Intervention and control group pre and post intervention mean and standard deviation.**

<table>
<thead>
<tr>
<th>Total vocabulary used</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>31.71</td>
<td>32.13</td>
</tr>
<tr>
<td>S.D.</td>
<td>19.59</td>
<td>20.50</td>
</tr>
<tr>
<td>n</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>48.94</td>
<td>52.74</td>
</tr>
<tr>
<td>S.D.</td>
<td>25.42</td>
<td>24.66</td>
</tr>
<tr>
<td>n</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 5:11 Writing Sample Vocabulary; Pre/Post Intervention Data

**Writing sample spelling**

The proportion of correctly spelt words was compared between groups pre and post intervention. This comparison is represented in the scattergram below (Figure 5:11). The spelling in both groups appears to have improved over time, but rates of progress seem very similar between the groups.
The mean and standard deviation in each group (Table 5:12) reflects the pattern shown in the scattergram. Intervention and control groups have similar pre and post intervention scores, and both show an increase over time. This pattern is confirmed by an ordered regression analysis, where post intervention score adjusted for chronological age was found to vary significantly with respect to baseline score; $R^2=0.42$, $F(2,23)=8.20$, $p=0.002$. Groups were found not to have a significant effect over and above this; $R^2=0.001$, $F(1,22)=0.05$, $p=0.82$. In other words both groups made gains over time, but the intervention was not demonstrated to have influenced this change.
Writing Sample Correctly spelt words per 100: Intervention and control group pre and post intervention mean and standard deviation.

<table>
<thead>
<tr>
<th>Percentage Correct Spelling</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>45.66</td>
<td>49.45</td>
</tr>
<tr>
<td>S.D.</td>
<td>19.13</td>
<td>16.25</td>
</tr>
<tr>
<td>n</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>71.51</td>
<td>74.36</td>
</tr>
<tr>
<td>S.D.</td>
<td>16.35</td>
<td>14.62</td>
</tr>
<tr>
<td>n</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 5.12 Writing Sample Spelling: Pre/Post Intervention Data

Writing sample content analysis

A graph comparing pre and post intervention story content scores for both groups (Figure 5.12) shows a similar pattern to that shown in writing vocabulary. It seems that progress was made on average in both groups, but that the pattern of change in the intervention group was difficult to interpret on the basis of this data alone.

Figure 5.12 Writing Sample Content; Scattergram Showing Change Over Time
The mean scores of both intervention and control groups before and after the intervention period are shown in Table 5:13 below. The intervention and control group shows a lower mean score than the control group at baseline, while intervention group post intervention scores are roughly equal to the control group mean. Standard deviations are also similar at all times between groups. Because of missing data, the sample size at post intervention is smaller, so mean change score needs to be examined to highlight any pattern that may emerge from this data. Average change in the intervention group (mean=5.92, s.d.=7.85) was larger than that in the control group (mean=1.36, s.d.=5.10) but the larger standard deviation in the intervention group suggests that there was more variation in change seen. Examining individual data points on the scattergram (Figure 5:12) shows that individual children made marked gains and losses on this measure over the period of assessment.

<table>
<thead>
<tr>
<th>Story Content</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Mean</td>
<td>5.08</td>
<td>6.42</td>
</tr>
<tr>
<td>S.D.</td>
<td>4.10</td>
<td>4.22</td>
</tr>
<tr>
<td>n</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Post Intervention Mean</td>
<td>10.06</td>
<td>10.26</td>
</tr>
<tr>
<td>S.D.</td>
<td>6.35</td>
<td>6.07</td>
</tr>
<tr>
<td>n</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 5:13 Writing Sample Content; Pre/Post Intervention Data

When this data was analysed using ordered linear regression analyses, the mixed pattern of the results was further confirmed. No evidence was found that baseline score predicted post intervention score (adjusted for chronological age); \( R^2=0.07, F(2,23)=0.90, p=0.42 \). Neither was a significant relationship between post intervention score and group (adjusted for chronological age and baseline score) found; \( R^2=0.09, F(1,22)=2.47, p=0.13 \).
Discussion of Writing Skills Results

The results of the analysis of writing samples are mixed, but overall improvement of writing skills were not demonstrated for those children in the intervention groups. Compared to children asked to complete a similar task in Moore and Wade's follow up of reading recovery students (Moore & Wade 1998) the children in this sample wrote far longer stories (ex-RR mean=68.23 words versus post-intervention in this sample mean=91.31 words). The children in this sample also made far more spelling mistakes (ex-RR mean no. mistakes=6.01 per 100, post intervention in this sample mean no. mistakes=28.49 per 100). The children in the Moore & Wade sample were approximately 2 years older than the children in this sample. It is possible that the older children were better able to monitor their writing, and restrict themselves to words they knew how to spell, resulting in shorter stories with fewer spelling errors. Further the children from the Moore & Wade sample were asked to invent a story, which may also have given them more scope for avoiding words they found difficult to spell. The differences may also reflect differences in teaching practice, where the school in this study put an emphasis on story content and quality, rather than spelling. Average improvements were made in story content and vocabulary, but the results for individual children varied substantially. Perhaps an unforeseen cost of the method of writing assessment is that responses depended more on motivation than they would in a spelling or dictation test. Story quality may be influenced more by enjoyment of the particular story, or mood on that day than an assessment using standard texts. Thus, particularly for these fairly young children, standards of writing on any particular occasion may not necessarily be indicative of writing skill. In all cases it seemed that those children who performed worst at the first writing task made greater progress than those who performed very well at baseline, this observation may support this theory of
variability in performance confounding results. A comparison of the story samples of children not included in the study groups, but collected at the same time, shows that the same pattern of variable performance is not apparent in among class peers. Although the largest change was observed in those with poorer baseline performance, overall there was marked improvement across all measures in this group. It would seem, then, that variability in performance, including decline in quality between samples, was associated with those children selected for the intervention classes.

5.3.4 Reading Strategy Recall

Pupils were asked to recall as many strategies for tackling reading problems as they could. The maximum number recorded was 3. The figure below (Figure 5:13) shows the mean number of strategies recalled by each group at baseline and post intervention.

![Figure 5:13 Reading Strategy Recall; Change Over Time](image)

There was very little change in the number of reading strategies that children could name over the period of the intervention. Both groups showed a slight decrease in number, but no significant differences were found between them.
It is unclear why children were less able to recall specific reading strategies later in the year than earlier. Children from the intervention group will have practised these skills whilst reading and also had them explicitly named and recalled for them during the intervention. It is possible that strategy use had increased in practice, but that children could not recall them out of context. For example, a fully integrated skill may be more difficult to name, as readers no longer explicitly use it. It is not possible to know the strategies used during their reading, so this explanation cannot be discounted, but the short period between baseline and follow-up does not support it.

5.3.5 Pupil Attitudes

Reading Attitude Survey

To establish whether there were differences between the intervention and control groups on attitudes to reading as expressed by the Reading Attitude Survey before and after the intervention a Mann-Whitney U test was conducted. The groups were found to differ significantly in their attitudes to reading at baseline ($z=-2.07$, $p=0.04$), with the control group showing a less positive attitude to reading than the intervention group. The post-intervention test established that there was no longer a significant difference between the groups on this measure, $z=-1.26$, $p=0.21$. Indeed when we examine the distribution of the data (see Figure 5:14) it can be seen that the mean score of the intervention group is lower than mean control group score post intervention. The pattern observed at baseline has been negated by a move towards a less positive attitude in the intervention group and a more positive attitude in the control group.
From this data we can conclude that the intervention group became less positive about reading during the period of the intervention, while the control group became more positive.

It would seem from this data after that the intervention the children were measurably less positive about reading. On the whole children become less positive about reading and school as they grow older (Literacy Trust 1999a; Literacy Trust 1999b; Literacy Trust 2000), but the differential change between the intervention and control groups
needs to be explained. It appears that the intervention produced some unexpected, negative results.

There may also be a design error that would account for this difference; there is a possibility that children's assessments were affected by their knowledge that they were to attend reading classes. Although every effort was made to ensure that baseline tests were conducted before any decisions had been made about who would be taking part in the intervention, it is likely that some of the children or their parents would have discussed the matter with the reading teacher. This might result in an anticipatory effect on reading attitude, that children might become excited by the prospect of attending the classes, or may be aware that it is an important issue for teachers and consequently rate enjoyment of reading higher. In turn, all children in the intervention group knowing that they were due to end their teaching sessions soon may have resulted in lower ratings. How likely this explanation is should be considered in the context of how likely it was that the children had knowledge of the classes when they completed the questionnaire. In this case it can be assumed that it is far more likely that children in the second intervention group knew in advance. By that time in the year the classes were well established for parents, teachers and pupils alike and parents and children were often keen to know who would be next. This was less likely to be a factor at the outset of term when no-one was getting lessons and many children were assessed who did not go on to receive lessons. The fact that the positive effect was greater at the beginning of the first intervention means that it is less likely that the issue of prior knowledge is the main effect being observed.
**Self Esteem Inventory (SEI)**

The results of the Battle SEI are in the form of a score on a discrete scale between 0 and 10, which requires the use of a non-parametric test for analysis. A Mann-Whitney U test was conducted to establish whether there was a difference between the groups on a measure of self-esteem (Battle, 1981). No significant differences were found at baseline ($z=-0.11$, $p=0.91$). No significant difference between the intervention and control groups at post intervention was found either, $z=-0.71$, $p=0.48$. Mean scores are shown in Figure 5:15 below. Little change was seen in either group over the period of the intervention.

**Battle SEI scores in Intervention and Control Groups**

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>S. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battle</td>
<td>Intervention</td>
<td>29</td>
<td>6.29</td>
<td>2.05</td>
</tr>
<tr>
<td>Baseline</td>
<td>Control</td>
<td>26</td>
<td>6.33</td>
<td>1.46</td>
</tr>
<tr>
<td>Battle</td>
<td>Intervention</td>
<td>27</td>
<td>6.33</td>
<td>2.63</td>
</tr>
<tr>
<td>Post</td>
<td>Control</td>
<td>25</td>
<td>6.90</td>
<td>1.84</td>
</tr>
</tbody>
</table>

From this data we can conclude that no effect was seen on pupils’ self-esteem.

**General Mood Score (GMS)**
To establish whether there was difference between groups on the measure of reported GMS in the 10 items recorded in the pupil interview a Mann-Whitney U test was used. No significant differences were found at baseline ($z=-0.06, p=0.95$). It was found that no significant differences existed post intervention on the GMS measure, $z=-0.73, p=0.46$. This data is represented in Figure 5:16 below showing mean score at baseline and post intervention for each group. A marked decrease in reported GMS can be observed in both groups.

![General Mood Score in Intervention and Control Group over Time](image)

This is no evidence to conclude that the intervention had a significant effect on these reports of GMS.

**Print Awareness**

Children were asked to name as many items as they could that contained print, but that weren’t books, in order to obtain an assessment of their awareness of print in their
environment. The mean number of items named by intervention and control groups at baseline and post intervention are shown in Figure 5:17 below.

Both groups were more aware of print in their environment post intervention than at baseline. When the mean change was examined, it was found that the mean increase in the intervention group was significantly larger than in the control group (t=2, p=0.05). The intervention group showed a significantly greater improvement in this skill than the control group (p=0.05).

**Self Concept Scores**

Self concept was measured on a visual analogue scale where children were asked to assess their own knowledge or skill across four parameters; current reading, reading by the end of junior school, favourite school activity now and their favourite school
activity by the end of junior school. It was found that these scores were not normally distributed and that there was a skew towards the maximum possible score on this closed distribution so a non-parametric test was chosen for analysis.

Baseline differences between the intervention and control groups were not proven. When this analysis was repeated post-intervention no significant differences were found between intervention and control groups on any of the four self-concept measures. Figures below (Figure 5:18 through Figure 5:21) show mean scores for each parameter at baseline and post intervention and report statistics for the post intervention analysis.

![Reading Self-Concept Now](image)

There was very little change observed in self-assessment of current reading knowledge over the period of the intervention. No significant difference between groups was found at baseline ($z=-0.52$, $p=0.61$) or at post intervention ($z=-0.34$, $p=0.74$). The intervention does not appear to have had an effect on this measure.
There was very little change observed in self-assessment of future reading knowledge over the period of the intervention. No significant difference between groups was found at baseline (z=-0.27, p=0.79) or post intervention (z=-1.23, p=0.79). The intervention does not appear to have had an effect on this measure.
Both groups showed a slight increase in self-assessment of skill at the child's favourite school activity over the period of the intervention. No significant difference between groups was found at baseline ($z=0.80$, $p=0.43$) or at post intervention ($z=-1.24$, $p=0.21$). The intervention does not appear to have had an effect on this measure.
There was very little change observed in self-assessment of skill at the child’s favourite school activity in the future. No significant difference between groups was found at baseline ($z=-0.80, p=0.43$) or at post intervention ($z=-0.59, p=0.56$). The intervention does not appear to have had an effect on this measure.

It can be concluded that the intervention had no significant impact on these four parameters of self-concept.

The evidence from the Battle self-esteem inventory, the questions regarding GMS and the children’s self-concepts provide a clear picture of a happy group of children. Contrary to evidence that difficulties with education in general, and reading in particular, impinge on mood state, we found that these children did not seem affected. It was impossible to examine evidence for improvement in these areas brought about by
the intervention. The little change that was observed was present in both intervention and control groups and certainly mediated by other elements in the children's lives.

**Reported Enjoyment of Reading and Writing**

Children were asked to say how much they liked reading and writing. These responses were recorded on a 1-5 Likert scale. The analyses of these two questions where responses constitute categorical data are shown in Table 5:14 below. Modal responses and statistical probabilities are recorded for each group at baseline and post intervention. No significant differences were found between the responses of the children in each group at the outset of the project, or at post-intervention for any of these questions. Children from different groups gave largely similar responses.

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>Pre intervention</th>
<th>Between groups Pearson's $\chi^2$</th>
<th>Post intervention</th>
<th>Between groups Pearson's $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td>Intervention</td>
<td>Like it, but don't love it</td>
<td>$\chi^2=3.5$</td>
<td>Like it, but don't love it/Love it (equal)</td>
<td>$\chi^2=4.4$</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Like it, but don't love it</td>
<td>$p=0.32$</td>
<td>Like it, but don't love it</td>
<td>$p=0.36$</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td>Intervention</td>
<td>Love it</td>
<td>$\chi^2=0.2$</td>
<td>Love it</td>
<td>$\chi^2=606$</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Love it</td>
<td>$p=0.99$</td>
<td>Love it</td>
<td>$p=0.16$</td>
</tr>
</tbody>
</table>

Table 5:14 Modal Rating of Reading and Writing Enjoyment

Children were very positive about reading and writing at baseline and this high opinion continued in both groups. This doesn't reflect the findings of the Reading Attitude Survey where the intervention group became less positive about reading over the period of the intervention. This may imply that this single item approach to attitude was not appropriate, and did not give scope to accurately reflect the complexity of the children's feelings towards reading.
Reading Habits

Children were asked how often they read their school books at home. Their responses were recorded on a 1-5 Likert scale. At baseline children in the control group reported reading much more often than the intervention group (see modal response reported in Table 5:15). Post intervention both groups report reading only "sometimes" despite the fact that the intervention group had been taking books home every night as part of the intervention scheme. However, despite the difference in modal responses there were no statistically significant differences between groups.

<table>
<thead>
<tr>
<th>How often Children Report Reading at Home – Average Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Post-intervention</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The fact the children are not reporting reading school books at home despite the intervention is discouraging. Children may not accurately report their home reading habits, they may be reporting the level of reading they think is acceptable to the interviewer rather than the amount of reading they actually do. Other studies and have found that self-report is an acceptable method for assessing reading frequency (Balding, Regis, & Wise, 1998; Daly & McGuiness, 1998; Hall & Cole, 1999). It may be that children in this sample were too young to make discriminatory estimates of frequency. This would explain the lack of variation between groups and over time. Children were asked an open question (“How often do you read at home?”) and responses were then categorised. It maybe have been preferable to offer explicit choices (every night, every other night etc.)
Discussion of Assessment of Change in Attitudes and Mood

There are some inconsistencies in the results of measures assessing change in mood and attitudes. For most measures, little change over time was seen. In a meta-analysis, other studies which have examined changes in self-concept (academic and attitudinal) have supported the view that academic interventions do have the potential to have a positive impact on self-concept (Elbaum & Vaughn, 2001). In this meta-analysis the "length and intensity" of intervention needed to have an effect on mood was stressed. The authors of the paper suggest that this is because changes in self-perceptions follow from changes in academic achievement; intensive instruction is first needed to achieve change in academic skill that is a prerequisite of change in self-perception. Under this interpretation, the study considered in this thesis was unlikely to affect mood since little change in academic skills was seen, and length of intervention and follow up may be not be considered sufficient (although it was an intensive programme). An alternative interpretation of the findings of the Elbaum and Vaughn study is possible, however. The meta-analysis found far wider variation between effect sizes in individual studies than might be expected, and notably the differences between studies within an intervention type grouping by intervention type was far greater than the differences between the groupings. In other words, there was an unusually wide spread of effect sizes between studies with similar interventions. In the methods of the review, little attempt to locate unpublished studies is reported. This strategy alone might explain the results, where studies with null or negative findings are far less likely to be reported (Sterne, Egger, & Davey Smith 2000), thus excluding these studies is likely to find a group of studies all with positive findings, but with wide variation including results close to zero effect. In this way, the findings of the meta-analysis may misrepresent the pool of studies examining the effect of academic interventions on mood. Close examination
of the sub-group analyses conducted supports this view. For example, little association between type of intervention and domain of self-concept where change was observed was shown. It seems unlikely that interventions focussing on particular aspects of self-concept would not show larger effects in those areas of focus. Missing studies with negative or null results would explain this finding. Thus the findings of this RCT, where little effects were seen and some negative effect observed may be representative of a body of unpublished studies. A new meta-analysis including unpublished work would be beneficial.

5.3.6 Attendance

Attendance data was taken from school records in bi-weekly samples representing the proportion of the week that children attended (i.e. 0.9 represents 4.5 days present out of 5). The weekly data fluctuated greatly and group averages rise and fall sharply. Figure 5.23 below shows this data summarised into half term averages for control and intervention groups, and also for a group of class peers (comparison). Class peers are all those in the cohort which were not included in intervention or control groups (n=141). Referring back to the project plan (Table 5:1), it can be seen that half term sections coincide with the timing of the intervention groups. The first half term is the pre-intervention phase; the second through fourth half term period covers the first intervention period. The fifth half term represents the first post-intervention period for the intervention group and the first intervention period for the control group.
From this data it can be seen that both the intervention and control groups showed poor attendance relative to their peers at nearly every time point.

Groups were compared for average attendance during each half term period, thus spanning the course of the first intervention and the commencement of the second intervention period. Since we are looking for change in attendance patterns, rather than differences between the groups per se, average attendance in the first half term (baseline attendance) was calculated and used as a co-variate in analysis. It was found that there was a significant relationship between attendance in the first half term and later attendance ($p=0.001$), but not between group and attendance independent of baseline attendance ($p=0.24$). In other words, any apparent differences in attendance were explained by group differences prior to the intervention rather than the intervention itself.
Attendance changed considerably over different periods of time, but this was not affected by participation in the intervention. In this age group it is unlikely that children have a great deal of control over school attendance. Children aged 7/8 years will go to school when they are taken, and stay home when they are not. However, even at this age children who are disaffected with school can make their opinions felt and make it more difficult for parents to take them to school or feign illness to miss time at school. If children were happier at school this might in turn improve attendance. It was also considered at the outset that the provision of a specialist teacher assigned to specific children could have had an impact on parents, marking the particular importance of school at this time, and in turn increasing attendance. The fact that no changes in attendance were observed means that the reasons for poor attendance in this group were not addressed by giving additional in school learning support.

5.3.7 Teacher Assessment

Reading Achievement

Teachers were asked to rate all children on the PLR Reading Fluency Scale (Barrs, 1998) at baseline and again post intervention. Class teachers were asked to rate pupils on a 1-5 scale of reading fluency for current achievement, potential for achievement and anticipated achievement. Mean assessment for intervention and control groups are shown in Figure 5:1 below.
Teacher assessment of reading achievement; Intervention and control groups pre and post intervention

<table>
<thead>
<tr>
<th>Achievement</th>
<th>Intervention</th>
<th>Mean</th>
<th>S. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>18</td>
<td>1.58</td>
<td>0.46</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>1.63</td>
<td>0.57</td>
</tr>
<tr>
<td>Potential</td>
<td>18</td>
<td>3.42</td>
<td>0.91</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>3.68</td>
<td>0.95</td>
</tr>
<tr>
<td>Anticipated</td>
<td>18</td>
<td>2.92</td>
<td>0.91</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>3.11</td>
<td>0.83</td>
</tr>
<tr>
<td>Current</td>
<td>29</td>
<td>2.16</td>
<td>0.60</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>2.24</td>
<td>0.75</td>
</tr>
<tr>
<td>Potential</td>
<td>29</td>
<td>3.78</td>
<td>0.56</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>3.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Anticipated</td>
<td>29</td>
<td>3.53</td>
<td>0.82</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>3.79</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 5:16 Teacher Assessment of Reading Fluency; Pre/Post Intervention Data

For both intervention and control groups the average teacher assessment of reading fluency, both now and in the future increased over the period of the intervention. Both groups showed very similar changes and no significant differences between the two groups were found. It can also be clearly seen that teacher assessment for both groups is low both before and after the intervention.
School Adjustment

Teachers were asked to rate all children on a 1-5 scale across three parameters of school adjustment: reading confidence, disruptiveness in class and group participation. Mean assessments are shown in Figure 5.24 below.

![Teacher Assessment of School Adjustment](image)

**Figure 5.24 Teacher Assessment of School Adjustment**

Teacher assessment of school adjustment: Intervention and control groups pre and post intervention

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>S. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention Baseline</td>
<td>18</td>
<td>2.56</td>
<td>0.86</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>2.79</td>
<td>1.13</td>
</tr>
<tr>
<td>Disruptiveness Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>18</td>
<td>1.78</td>
<td>1.06</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>2.45</td>
<td>1.34</td>
</tr>
<tr>
<td>Participation Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>18</td>
<td>2.39</td>
<td>1.58</td>
</tr>
<tr>
<td>Control</td>
<td>19</td>
<td>2.42</td>
<td>1.26</td>
</tr>
<tr>
<td>Reading Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Intervention</td>
<td>29</td>
<td>2.86</td>
<td>0.88</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>2.69</td>
<td>1.17</td>
</tr>
<tr>
<td>Disruptiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Intervention</td>
<td>29</td>
<td>2.38</td>
<td>1.45</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>2.24</td>
<td>1.48</td>
</tr>
<tr>
<td>Participation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Intervention</td>
<td>29</td>
<td>2.69</td>
<td>0.85</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
<td>2.69</td>
<td>1.28</td>
</tr>
</tbody>
</table>

**Table 5.17 Teacher Assessment of School Adjustment; Pre/Post Intervention Data**
Teacher assessment of reading confidence of children from the intervention increased over the period of the intervention, while that of children in the control group lessened. This result did not reach significance.

Teacher assessment of children's class participation increased in both groups, and this may well reflect the children's increased ease in a class from the beginning of the school year to later in the year. No significant differences between groups were found.

Children from the intervention groups were rated as more disruptive after the intervention than before, whilst the reverse was true for children from the control group. Once again this difference between the groups did not reach significance.

When interpreting the results of the teacher assessment of children it should be noted that all teachers knew who had been attending literacy classes and who had not. Importantly they knew which children were controls and were awaiting the intervention. This may be of particular importance to teachers because delaying the start of the intervention for some children was particularly problematic for them. Of the nine teachers questioned at the end of the intervention periods, two said they felt that there were particular children for whom having to wait had been a problem. This may have made teachers more sensitive to problems in the control groups and/or to progress in the intervention groups. Nevertheless, the two interview timings were some time apart, and teachers did not have their original assessments to refer to, so although some caution is needed the teacher assessments retain validity.

Teachers tended to assess all children higher on indices of reading skill later in the year than they had at the beginning. If these assessments are accurate then they demonstrate
the value of using equivalent control groups. All children can be expected to progress over time, the important question is whether an intervention is able to accelerate this process. There were few differences between teacher assessments of children for their school adjustment either over time or between groups. It would seem that teacher assessments of these factors are stable over time, and that the intervention didn’t have an effect that was obvious to class teachers.

5.3.8 Qualitative Feedback on interventions from pupils and from reading teachers

Forty children were interviewed following the intervention classes, 25 girls and 15 boys. These interviews, as described in Chapter 3 (section 3.7.4) were loosely structured, depending on how willing each child was to talk about each topic. Responses were coded according to categories derived from responses themselves and these are reported grouped by theme. The reading teacher was also interviewed at the end of the first intervention period. The purpose of this unstructured interview was to discuss the progress of individual children, discuss how the process had worked to identify whether changes needed to be made, and to use the opportunity to obtain her views on the structure and content of the lessons. Elements of this interview are reported here where they covered issues also discussed with the children for comparison. In the teacher discussions the names of children have been changed to alternative, same gender names to protect anonymity, in other cases names are not reported.

Content of lessons

Thirty-four children responded to this question, and the vast majority said they read in the classes (n=32) with many making explicit reference to reading books or a particular book (n=10). For example one girl said they “read a lot, [she’s got] some good books”.
Some children reported tasks that were peripheral to the classes (e.g. getting their photographs taken to go on the wall) or didn’t take part (e.g. doing plays). The table below (Figure 5:25) records all responses given (categories are not mutually exclusive) with counts of their occurrence. Writing and spelling were important tasks, but so were drawing and making things.

![Pie chart showing pupil feedback of lesson content](image)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read</td>
<td></td>
</tr>
<tr>
<td>Talk</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td></td>
</tr>
<tr>
<td>Draw</td>
<td></td>
</tr>
<tr>
<td>Spelling Make cards, books etc</td>
<td></td>
</tr>
<tr>
<td>Vowels &amp; alphabets</td>
<td></td>
</tr>
<tr>
<td>Phonics</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Do sounds</td>
<td></td>
</tr>
<tr>
<td>Say jokes</td>
<td></td>
</tr>
<tr>
<td>Getting photo taken</td>
<td></td>
</tr>
<tr>
<td>Testing our reading</td>
<td></td>
</tr>
<tr>
<td>Did plays</td>
<td></td>
</tr>
<tr>
<td>Have fun</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5:25 Pupil Feedback of Lesson Content**

Overall, the children had reasonable memory of the most important tasks included in the lessons. We were also interested to know what children felt was the result of the lessons, so asked some (n=24) of the children what they had learnt in the classes. Response fell into three camps, those mentioned a general skill learnt, those who mentioned a particular skill, and those who mentioned something unconnected to the main aims of the lessons. Thus;

**General skills** – learnt reading, spelling, writing neatly and poems.
Specific skills — vowels, reading to an audience, “That you never know if, if you can’t see the word and you have to spell it and it’s and ‘ed’ word but it sounds like a ‘t’, we learnt that.” (boy)

Unconnected skills — time, colouring, drawing, making things, maths and being kind to people.

One child said he’d learnt nothing, and another couldn’t say. The most common answer was that their reading had improved, which 15 children mentioned, for example ‘I’ve got a bit better. Like I know words like “suddenly”’ (boy). It is difficult to interpret the answers of children who said that they had learnt, for example, maths or time. A developing theme throughout some of the interviews was that a small number of children had misunderstood questions asked in interview, or misunderstood the nature and content of the lessons.

Enjoyment of lessons

Children were asked about whether or not they liked the lessons. This question was phrased in different ways at different points in the interview. Initially children were asked whether or not they liked the lessons, in addition children were asked about how they felt about the lessons finishing and finally they were asked if they knew who had been chosen to go to the lessons now and whether they felt these children were “lucky” or not to be chosen. Each of these questions addresses different aspects of how children felt about the lessons, and the was also an attempt to ensure children were given more than one opportunity to express any less positive feelings they might have. Despite the fact that children were told that what they said would be in confidence and they could say what they wanted to, it was recognised that it may be difficult to admit to
an adult that they didn't like the lessons. Talking about feelings at the end of the lessons, or how others might feel might be easier for children.

When simply asked how they felt about the lessons, most of those who responded said they liked the lessons (33 out of 38 who responded). Only two said they didn't, for example a boy reported that “...I hate reading. When I'm at home instead of reading I play outside or play on my computer”. One said it was “alright” and two said they “sometimes” liked going as another boy reported “sometimes it gets boring cos we read too much but sometimes I like it”. One girl said she had liked the lessons and when prompted further said “At first you thought it was rubbish – just sit there. I thought it'd be boring, I'd be moody.” “and now?[interviewer]” “not now... but it's important – reading and your education.” It is hard to know whether this represents ambivalence to her own feelings about the lessons, or that she never grew to “like” the lessons but valued them none-the-less. This reaction is interesting if we want to consider what might motivate children to engage in classes, that understanding the need for education was important for this child. Children were also asked later in the interview whether they were sad or glad the lessons had finished (to ensure all children were given more than one opportunity to express any negative feelings). One girl gave an interesting response, that she felt “in between cos I can spend time with my friends and I'm not [glad] cos I miss going, I liked that place”. This notion of enjoying the classes, but being aware of what was missed in the main classroom is important when we come to consider the overall value to children and to school of the lessons. Most of the children had a positive attitude toward the classes and talked with enthusiasm about the particular features they enjoyed (reported later).

Similarly 14 of the 18 children who were asked about how they felt at the end of the lessons told the interviewer they felt sad that classes were over. For example one girl
said "I feel upset that I'm not going any more" and another girl that "cos she does good work with me". Similarly only 1 of 13 asked about whether they felt being chosen to go to the lessons was lucky or unlucky only one thought it was unlucky.

What was the best/worst thing about the lessons?

Thirty children were asked what they liked most about the lessons, and again reading was the most common answer (11 children) and several (6 children) mentioning particular books or the books in the classroom particularly. For example one girl said "we have our own folder and clipboard and paper thing and we've got a book called the hungry cat and he eats everything . . .". This child was one of two who enjoyed getting their own folders etc. to look after. Children’s relationship with the teacher was also important for some, two said she was the best thing about going, and a third child (a girl) said "Reading to her [Ms g]. But at class when I read I get angry sometimes when they laugh at me when I get it wrong.". This response is interesting since it emphasises not only the enjoyment of reading and of the teacher, but indicates that in other circumstances (namely in class) reading isn’t a fun thing to do. The teacher’s efforts to make the lessons un-threatening for children she often described as “fragile” readers is clearly working for this child. The pie chart below (Figure 5:26) provides a count of all categories mentioned by children as the best thing that happened in the lessons.
What was the best thing about Ms G's lessons?

![Pie Chart: Best Things in Lessons]

Looking at this data we can see that things were mentioned 11 times that could be viewed as being peripheral to the lessons – making things, drawing, having own folder, having a party, colouring in, getting a photo taken and getting stickers and certificates. This is the same number who mentioned reading as their favourite thing. It seems that the children value the classes very much, but they enjoy the overall process as much as any particular activity within the classes. This may not be a surprising observation, teachers use parties and stickers because they recognise how much children enjoy receiving them, but it underlines the importance of creating an enjoyable and rewarding atmosphere for these children.

Children (22 in total) were also asked about the worst thing in the lessons, or to names things they didn’t like about the lessons. Half of the children volunteered that there was nothing they didn’t like about the lessons. Two thought the lessons were, or could be, boring. Two children said they didn’t like other children – either a particular child or that they didn’t like waiting for others. All answers given are summarised in the pie...
chart below (Figure 5:27). No child mentioned more than one thing they didn’t like about the lessons.

What was the worst thing about with Ms G’s lessons?

- Writing on board when hogs get cross
- Drawing
- Particular other child
- Reading
- It’s boring
- Missing other activities
- When someone’s going slow
- Spellings
- Remembering to bring things every day
- Nothing

Figure 5:27 Pupil Feedback of Worst Things in Lessons

Overall the children’s responses to the lessons were very positive. Once again a number of responses were concerned with activities or events peripheral to the lessons (another child, teacher getting cross, missing other events in school and waiting for others).

How would you feel about carrying on with the lessons?

1060 Children (n=37) were asked how they would feel about carrying on with the lessons. By far the majority (n=29) said they would like to carry on going, reflecting their positive attitudes toward the lessons and the importance they placed on them. For example when asked why she wanted to carry on one girl said, “I think I haven't learned enough”. Six children didn’t want to carry on going, for example one girl told the interviewer “I’ve had enough and I read all the books before and they’re little, little thin ones.” This could be interpreted that she found the classes tedious. It could be that these 6 children felt more able to say they didn’t want to carry on attending classes, than to say that they
didn’t enjoy the classes. One girl said she didn’t know whether she’d like to or not and one boy did not respond at all. In another part of one interview in which a boy was telling the interviewer that he preferred visiting once a week to change books rather than going every day he “so I can have more time to work in my class”. This reflects an earlier comment by a different child that the worst aspect of the lessons was missing out on other activities in the school.

The 29 children who said they’d like to carry on going were then asked how long they’d like to carry on going to the classes for. Sometimes children volunteered the information without being asked, for example one girl said “yeah, [I’d like to carry on] for ages. Until I’m grown up”. The most common answer was until they left primary school (see Figure 5:28 below). Other answers varied between one week more and until the child grew up. Three children gave answers that didn’t make sense. This implies they might not have understood the question, or that their wish to carry on wasn’t very considered.

![Pie chart showing pupil feedback on desire for more lessons](image)

**Figure 5:28 Pupil Feedback of Desire for More Lessons**
Again these answers represent overall positive feelings toward the lessons.

**How do you feel about Ms G?**

Children (n=29) were asked about their feelings toward the reading teacher. When children gave very positive response prompts would be given for them to talk about any negative or ambivalent feelings they might have (e.g. does she ever get cross? Do all the children feel the same way?). Responses to these prompts are not recorded here, as no child used this prompt to say anything new about their teacher. All children used positive language to describe her, most simply aid she was “nice” (n=9) or kind (n=4), other terms used to described her were “friendly”, “she was funny, she told us jokes” (boy), “happy”, “excellent”, “helps us”, “cute”, “alright”, “likes children”, “considerate” and “I like her lots”. Other children mentioned particular things that the teacher did to make her appealing; giving children pens or stickers, she “She learns people the easier way if they don’t know” (boy), she made special events (Easter and end of classes), she’s “not bossy” and “[doesn’t] shout when she gets angry”.

**How were you chosen to go?**

Thirty-six children were asked how they thought they were chosen to attend the lessons. The most common reply was the child didn’t have any idea how they were chosen (n=15). Some children were very sure of their poor reading skills; “I don’t know, cos I couldn’t read or I couldn’t write.” (boy) or “I think it was because my reading wasn’t good. It was appalling” (girl). The second most common was simply that the teacher had chosen them, but they didn’t know why (n=7). For example, one girl responded “she just takes me to go” and when prompted as to why she was taken shrugged and replied “she just picks me up”. Alternatively some children had inaccurate impressions of how they were
chosen, “she [Ms G] asks Ms S to tell her who needs help with things.” “What sort of things?” “maths”, or “I read a book and I was reading it nice and I never got a word wrong” (girl), or “Ms G said I’d done so well maybe I could help some other people in my group” (boy).

These impressions are likely to have been informed by what the reading teacher had told them at the time they were assessed and selected for the lessons. When asked about what the children knew about the lessons and whether they understood why they’d been chosen she responded;

“Well I told them that they’d been chosen by their teachers because they were working hard and we were going to do some more work on their reading and writing. And I’d ... saying that to their parents as well. But I tried to make it a very positive thing.”

This strategy had clearly worked in helping the children not to feel stigmatised by going, but it left the curious phenomena that few of them understood why they were going, or, therefore, what the aims of the classes were.

Would you rather go alone or in a group?

The use of group teaching in this context is somewhat controversial, so it seemed important to determine what the children’s views of group or individual teaching was. Of the seven children who went to individual lessons and were asked whether they thought they’d prefer to go alone or with others 5 felt they enjoy going in a group more and two were happy to go alone. Thirty one children who had attended group lessons were also asked which they would have preferred, and most (n=15) preferred to attend in a group, two couldn’t say which they preferred and fourteen would have preferred to
attend alone. Overall, then there was a slight preference for attending in a group but many children preferred attending alone.

Children were asked to explain their choice, and as previously these answers were categorised and a count of reasons made. For the most part children wanted to attend in a group because of the company they received. Reasons given were friends in the group (n=4), the others were funny (n=2), talking to people (n=2), it's more fun (n=1), that you “don't feel so happy on your own”. Some children mentioned that the children helped each other (n=3), or shared out the work (n=1), for example “cos if I'm stuck on a word I could ask them.” (boy). Two children said that they liked group reading activities, “I like reading with a lot of people... it, it gives you company.” (boy), and one said the others provided emotional support “when I got upset they cheered me up” (boy).

In contrast to these statements implying that children enjoyed attending in a group, the responses of those who preferred to attend alone were most often about avoiding things children didn’t like. Three children mentioned it was quieter alone, “cos it's not a bit of help having noise” (girl), or “won't get distracted by other people when you're reading” (boy). Two children mentioned not having to wait for others, one that no one could “copy your work”. Several children mentioned negative reactions from other children, that the others told you to hurry up, or that “sometimes I get scared that they might laugh at me” (girl). Another girl was unable to say why she had a preference for attending alone, but when prompted with questions about other children in her group told the interviewer that “sometimes L and S say things about me and that”. There three positive reasons given for preferring to attend alone, that you get more help, that the work is less hard and that you get to use chairs instead of sitting on the floor.
These responses reinforce a model of the social importance of school life in general and of reading in particular. However, if the social situation were uncomfortable children would rather avoid it. Only one child mentioned a change in relationship with the teacher in group work, in that you might get more help from the teacher if you attend alone. Those that preferred to attend in a group clearly saw each other as a resource, whether of company, entertainment or support.

The reading teacher was asked about group dynamics, and how the group or individual lessons had succeeded for individual children. She had been asked about group dynamics and had been describing the difficulties that one group had compared to the other in settling into the classes. The conversation continued:

Interviewer “Why was it more of a problem for that group than the other group?”

Ms G “...I think because of the individuals in the group. ...And they got very very good at helping each other. And then as the gap widened between those who were making a lot of progress and those not making so much progress ... that became quite difficult then and I could feel that the group dynamics were going, and towards the end it was quite hard Erkhan and Mia had just lost interest by the time others had managed to work the word out, so just getting that balance was quite difficult.

... But the pace, the speed, I mean that was a hard thing to get established and at a certain point we started to lose it again I mean we had to start over again. They didn't particularly, I mean once we'd started reading as a group they didn't particularly like reading individually.”
I “And what about the differences between the groups and the individuals?”

Ms G “… They miss out hugely on the response to text, hugely because there’s such a rich, well shared experience really from what they all bring, the individuals bring to the text. So for example Ant and Femi with Yasmin and Ann, that was a particularly interesting example actually because Ant’s a 2nd language learner is very focused on getting it right and on looking at the visual side of reading - decoding, that was his strategy that was his strength. And Femi completely different, wealth of language huge experience to draw on, … very rich and provided all sorts of opportunities to develop different strengths and strategies. So it is a bit harder with the 1-2-1 on that side but you can be more focused, obviously, on the individual. I couldn’t, I don’t think that Isaac or Joe would have coped well in a group”.

This conversation illuminates the comments of some of the children in the groups, and also shows the concerns of the teacher with regard to teaching children alone or in groups. She clearly felt children were a valuable resource to each other, but that this involved difficulties either for those children who couldn’t function in a group and also for those who progressed at different rates. We know from pupils’ comments that some felt others slowed them down, and those seen alone got “more help”. So clearly these difficulties were recognised by pupils and teacher alike. The views of children in these circumstances should be taken into account when planning intervention, that they do value each other for friendship and support, but tensions within groups can develop and not children will enjoy this atmosphere. The teacher also speaks about the individual skills and knowledge that each child brought to the lessons. For a teacher this was clearly an important feature of the groups, which enabled a modelling of different reading strategies and styles that others in the group could see and use.
Other questions

Since the interviews were led by the interest and enthusiasm of the children in preference to being strictly structured comments and questions arose which were unique to each interview and didn’t fall within any of the categories above. The overall impression from the children was that very nearly all had enjoyed the lessons and wanted to continue going to them. Many felt a sense of achievement one girl said she liked going every day “cos every day I get better and better at reading” and another girl proudly told the interviewer “Now I’m really good [at reading]. I can read books with 200 pages. I thought I’d never make it but I have”.

Several children were asked about their mum’s (or other carer’s) opinion of their attending classes. The relationship already developed between the children and the interviewers was particularly important here, because the interviewers knew who each child lived with and who were the important figures in their lives and could ask an appropriate question. Only children who lived with their mums had a response to this question, and on the whole they felt that their mum’s opinion was positive. One boy even reported that “my mum bought me a big bike two weeks before Christmas for going to see Ms G”. Children whose parent or carer was not very involved in their education were not asked this question. One girl had misunderstood the nature of the classes in an interesting way, she told the interviewer that she wanted to go to the lessons because “it wouldn’t be fair on my mum paying money and me not going”. This response recalls confusions about the purpose and reasons for the lessons, but also conveys the fact that this child clearly felt there was a duty attached to attending the lessons. If more children had understood the purpose of the lessons this might have been a more common response.
5.4 Conclusions

The only areas where progress was significantly improved in the intervention group compared to the control group, was in reading accuracy and print awareness. No impact of the intervention on reading and reading related skills was demonstrated. Changes are difficult to achieve in this hard to treat group. The fact that change was not observed does not necessarily mean that the intervention was ineffective. It may be that too few lessons were included and additional time in the teaching groups would have improved progress. Alternatively, lesson time may have been adequate, but insufficient time was available for skills learnt to become internalised before follow-up assessment. At the point of assessment however, gains were modest, and the likelihood of gain in other circumstances and compared to other treatments needs to be considered in interpreting the usefulness of this intervention.

The additional change observed in the group was a tendency to a less positive attitude toward books on quantitative assessment tools. There was a general tendency in both groups to report a less positive attitude toward books, school and themselves over the period of the intervention. This could well be due to age-related changes affecting children at this time. The intervention began soon after the move from infant to junior school and it could be that a generally less positive attitude prevails in this age group that infected our group as they became accustomed to it. In other studies (Balding, Regis & Wise, 1998) older children have been found to have a less positive attitude toward books than younger children. However, although both groups showed a tailing off, this was significantly greater for the intervention group than for the controls. It could be argued that an intensive intervention period poses two risks to pupils' attitudes to reading; that children may become bored through over exposure, and that children
may become more aware of their problems with reading. Either of these might result in disaffection with reading. In comparison to these findings, the qualitative data showed that most children enjoyed the lessons themselves and liked their teacher. There were many things they valued about the lessons and many would like to have carried on attending lessons for a longer period.

In total change was assessed across seven domains (reading skills, phonological skills, writing skills, reading strategies, attitudes toward reading and self, attendance and teacher assessment), comprised of 30 individual data points compared pre and post intervention. Conducting repeated analysis in this way increases the likelihood of producing anomalous or chance findings. These can lead to inaccurate conclusions. In this case, the number of iterations increases the likelihood that the lack of effect observed is a true reflection of the situation. It is very unlikely that the intervention had an effect in an area that wasn’t assessed. Comparing this finding to others, it is difficult to know whether it is typical. The study most comparable to it in the UK running at a similar time was the Catch-up Project (Clipson-Boyles, 1997). Findings from this project have not been published, but Brooks (2002) reports unpublished data that shows an apparent improvement for intervention children (average gain of 11.5 months reading age compared to 8 months). However, we do not know whether this result is statistically significant, nor are we sure of the equivalence of intervention and control groups. Fawcett et al. (2001) claim to show more favourable results with the same age group, taught in pairs for half an hour twice a week. However, although they showed a statistically significant interaction between time and group (p<0.01), the changes observed were not large enough to result in a difference between groups for post-test scores. Although change observed in the groups was different, their actual performance on tests at the end of the intervention was not different. This change effect was not shown at follow-up 6 months later. In this study,
then, a small effect was suggested in the short term (effect sizes are given for reading and spelling, but not confidence intervals so these are difficult to interpret) but this had no functional significance.

There is evidence that “standard” reading recovery is particularly ineffective at developing phonological processing skills in children (Chapman, Tunmer & Prochnow, 2001; Iversen & Tunmer, 1993) and that this constrains progress. The study that is the focus of this thesis took place in the context of the NLS where more direct phonological training should be provided for this age group (although there was no system to check that this took place), but additional tasks were provided where the teacher felt it was appropriate.

Iverson and Tunmer contrasted standard and modified RR (that is with additional phonological input) in the context of whole language classrooms and the introduction of the NLS may make this comparison less appropriate here where direct phonological work should have been taking place in class. The findings of Iverson & Tunmer that phonological processing skill predicted response to RR was not found in this sample. The analysis of factors association with change (Chapter 7) shows that none of phonological measures used (phonological manipulation, nonword decoding, rhyme and alliterative production) was correlated with outcome across the intervention sample.

The evidence from this evaluation would be that, whilst most children enjoyed the intervention classes, on balance this intensive provision of literacy support has not been shown to improve literacy skills and attitudes to reading in the intervention group. Although change was observed in the intervention group, this did not differ enough from changes observed in the control group to be able to conclude that changes were due to the intervention. This lack of effect was observed across a range of standardised, skill related, non-standardised and attitudinal measures. Two questions arise from this finding; first is this a result of short follow-up time so that change may be observed in the longer term, and second do observations regarding average change obscure change
in a sub-set of children who did improve significantly? These questions are addressed in the next two chapters.
Chapter 6 Long-term literacy achievement in intervention groups

6.1 Introduction

If we are to know about the efficacy of intervention we need to consider its effects in the longer term, as well as examining effects immediately after the period of the intervention, as in Chapter 5. In this study the short-term evaluation of the controlled trial of a reading intervention did not find evidence of change in the intervention groups in response to the intervention. This may be as a result of not allowing sufficient time for children to internalise the skills they learnt during the intervention. In order to examine this suggestion children were followed up in successive years to determine whether long term change was achieved. The results of this data set are presented in this chapter.

The long term aim of intervention is to bring the performance of those within intervention groups to a level where their skills are comparable to those of their peers, such that they can access lessons appropriate to their age group. A comparison group was constituted at the outset of this project (see 3.3 selection of subjects) whose achievements were used as a marker of average achievement among class peers. Assessments in the intervention group in the long term can be compared to assessments in this comparison group to see whether the gap between these groups did, indeed, decrease over time. Over the period of the intervention it was possible to observe change in the intervention group in comparison to that in the control group. This control group was a waiting list control group, comprising children waiting to receive the intervention. A waiting list control group was used because it enabled the school to fulfil its responsibility to the children in its care, by addressing the problems they had, whilst
enabling the research team to have access to an equivalent control group during the
period of the first intervention. The cost of this strategy is that the control group is lost
immediately the first intervention is complete, since the children in this group go on to
receive the intervention at this point (see methods section 3.2 Experimental Design).
Whilst this enabled the core study to proceed, it compromised the long term data, since
change cannot be compared to a true control group. The comparison group was created
in anticipation of this feature of the research design. Although this group is therefore less
similar than the control group, we were interested to know whether children could be
brought to a level where they are able to function at similar levels to their peers. The top
performing children were screened out of this group in order to provide a comparison
not to mean change in score, but to change in what could be considered the “average”
child in the class. Progress is reading is likely to be non-linear, and the most or least
skilled readers may make disproportionately small or large increases in achievement. Such
unusually large or small change scores will affect mean change, and has the potential to
over or under-estimate typical change in the group. Since we are interested in comparing
change in the poorest achievers to typical, rather than mean, change the high achievers
were removed from the sample in an attempt to better estimate typical change. Of
course, in the randomised controlled phase of this trial such issues do not come into play
since we are comparing change in groups drawn from the same population, but since
control group data is not available in the longer term this adjustment is important for face
validity of comparisons made.

At the final follow up point a new measure of functional reading was introduced. Whilst
the aim of most literacy interventions is to bring the reading levels of the poorest readers
closer to the average, there is a debate as to whether or not this is realistic. The argument
for changing the goals of intervention has been argued thus:
"...as stated by Rutter (1983) and Snow (1989), the only way to substantially reduce individual differences in scholastic attainment is to severely restrict and impair the schooling of the most advantaged students. If this is true, educational attempts to decrease differences between readers must be the wrong objective. A better goal would be, of course, to lead less able students to levels of (functional) reading skill that are minimally required in present-day society." (Bast & Reitsma, 1998 p1388)

Although part of the role of education in literacy is to ensure a high level of literacy for most of the population, there must be an imperative to ensure competent levels of literacy in all (Browder & Xin, 1998). The standardised assessments used in this study have compared progress to average age-appropriate levels, and have demonstrated literate skills for their own sake. They do not assess the extent to which children are functionally competent readers, or the extent to which children can or do use written information. Children may not achieve average reading skills, but the intervention may aid non-readers to become readers who, while their skills may be limited, can use written information to guide their actions.

The question of functional literacy is an important one. Children themselves recognise the importance of functional reading (Greaney & Neuman, 1990). An additional assessment of functional literacy was therefore developed at the final follow-up period and is described in section 6.2.5 below. This data was not collected at baseline, but was developed in response to the results indicating that progress compared to age-appropriate averages was modest. The data from this assessment can be viewed as a pilot of a novel method for assessing functional reading and a discussion point but, since pre-intervention data is missing, conclusions about the impact of the intervention on functional reading cannot be made.
6.2 Method

6.2.2 Sample

Data from comparison and intervention groups are presented in this section:

1. **Comparison children** At the beginning of the project a group of class peers was selected to compare to the intervention groups. As discussed in section 3.3 these constituted those children whose scores fell in the bottom quartile on any standardised test but who were not selected for intervention. This group therefore excludes a high scoring sub-sample of class peers and provides an appropriate peer comparison group.

2. **Intervention children** Data from both immediate and delayed intervention groups were pooled to form a larger sample of all those who completed the intervention. Some changes were made to the second intervention group in response to attrition, or revised teacher assessment. Thus the delayed intervention group contained largely, but not exactly, the same children as the waiting list control group. In response to these changes the baseline data was re-examined to see if any systematic differences were introduced by this alteration from those initially selected for intervention and those who actually received the intervention. No significant differences were found between immediate and delayed interventions groups at baseline. This implies that they were drawn from the same population and it is justifiable to pool their results. The changes made to the waiting list control group when entering the intervention are of interest in their own right, and a discussion can be found in section 8.2.2.
Data collection

Longer term data was collected in the years following intervention for years 1 and 2. As a result of the successive year-cohorts of children coming into the target year group, data was collected in school for two consecutive years following the completion of cohort 1 intervention and one year following cohort 2 intervention. The plan below (Figure 6:1, taken from section 3.2) shows the timing of intervention, with the longer-term follow up time periods shaded.

<table>
<thead>
<tr>
<th>Autumn Term</th>
<th>Spring Term</th>
<th>Summer Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 1st intervention, End 1st intervention Post-measures</td>
</tr>
<tr>
<td>Year 2</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 3rd intervention, Year 1 1 year follow up End 3rd intervention Post-measures</td>
</tr>
<tr>
<td>Year 3</td>
<td>Whole year screen. All study groups Pre-measures</td>
<td>Begin 5th intervention, Year 1 2 year follow up Year 2 1 year follow up End 5th intervention Post-measures</td>
</tr>
</tbody>
</table>

Thus, in year two of the study, cohort 1 was followed up during the spring term, allowing 18 months post baseline (which was 12 months after the immediate intervention finished and 6 months after the delayed intervention finished). In year three of the study, cohorts 1 and 2 were followed up during spring term. The cohort 2 data at this time point can be pooled with the cohort 1 data, providing a larger sample followed up 18 months post baseline. The year 3 follow up of cohort 1 data provides a limited sample of data 30 months post baseline (which was 24 months after the immediate intervention finished and 18 months after the delayed intervention finished).
Follow up data was available for two year-cohorts 18 months post baseline and for one year-cohort 30 months post baseline.

6.2.4 Selection and use of measures

A number of measures were selected for collection in the longer term. These were:

1. Raven’s Progressive Matrices
2. Young Group Reading
3. Young Group Maths
4. Reading Attitude Survey
5. Neale Analysis of Reading Ability
6. Phonological Assessment Battery - Nonword Reading
7. Writing Sample
8. Battle Self Esteem Inventory (shortened)

These measures were selected on the basis of three criteria:

- That they should be appropriate for the children in the age group at follow up. For this reason the Young group math and reading tests were only used at first follow up, as they were not appropriate for children at second follow up.
- That they assessed the core skills and attributes we wished to track over time
- That they had been demonstrated to discriminate between children (e.g. children were not performing at floor or ceiling levels).
Score, rather than rate of change, was compared between groups, since in this case the issue was not the amount of progress made by any group, but whether the gap between intervention and comparison children could be seen to close.

6.2.5 Functional Literacy assessment

A search of the literature failed to locate an assessment of functional reading in context suitable for children in this age group. An assessment was therefore created. The aim of the assessment was to investigate firstly whether children could use text instructions to complete a simple task, and secondly whether they paid attention to text instructions when they were accompanied by pictorial instructions. The task was designed to be engaging and simple to read and complete, but to assess whether and how adequately children followed text instructions. The task set was to make chocolate rice-pop cakes.

Children were provided with instructions and a choice of the ingredients and utensils needed to complete the task. Children completed the task individually using one of three parallel sets of instructions. The instructions varied slightly in the ingredients and utensils to be used. The choices presented are given in the instruction list below (Figure 6:2). Whether or not the children followed the instructions precisely they would produce chocolate cakes to eat, but researchers recorded which choices they made and compared these to the specifications on the instruction sheet. Black and white pictures illustrating the next step in the recipe accompanied some of the instructions. These were introduced in order to ascertain whether children were more or less likely to follow text when a pictorial instruction could be used instead. Children completed the task in groups of six (two using each instruction sheet) with two supervising researchers. Equipment was available in surplus, so that no child would have a restricted choice. School records were consulted to ensure that all of the children included had normal colour vision.
Instructions for making chocolate crispy cakes

1. First take the (blue/red/yellow*) bucket and the (green/yellow/blue*) cup from the equipment table.

2. Choose the (Cornflakes/Rice Crispies/Special K*) from the equipment table.

3. Take one and a half cups full of cereal and pour it all into the bucket.

4. Ask an adult for the melted chocolate to be poured onto the cereal.

5. Next, get a (yellow/blue/red*) plastic spoon and stir the cornflakes until all of the cereal is completely covered with chocolate.

6. Collect some (plain gold/spotty/stripy*) cake cups from the equipment table.

7. Using the spoon, fill eight cake cups with mixture until the bucket is empty.

8. Decorate each cake with some (white/milk/multi-coloured*) chocolate chips.

*Instructions for one choice presented to children.

Figure 6.2 Instructions for Making Chocolate Crispy Cakes

Researchers recorded choices made by children during preparation, took photographs of the completed cakes and recorded equipment left on tables at the end of preparation. This enabled researchers to focus on helping children knowing a permanent record of their workings was available for later examination. The children also enjoyed this very much.

Scores for each stage of the process were allocated according to the scale below (Table 6.1). Total score, text only and text & illustration scores were compared across the three parallel forms. No significant differences between forms were found and the equivalence of the three forms of the task was accepted.
Possible Choices and Scoring System for Functional Literacy Task

<table>
<thead>
<tr>
<th>Instruction</th>
<th>0 Point</th>
<th>1 Point</th>
<th>2 Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bucket colour (1 of 3)</td>
<td>Incorrect colour</td>
<td>-</td>
<td>Correct colour</td>
</tr>
<tr>
<td>Cup colour (1 of 3)</td>
<td>Incorrect colour</td>
<td>-</td>
<td>Correct colour</td>
</tr>
<tr>
<td>Cereal type (1 of 3)</td>
<td>Incorrect cereal</td>
<td>-</td>
<td>Correct cereal</td>
</tr>
<tr>
<td>Amount of cereal (1.5 cups)</td>
<td>Incorrect amount</td>
<td>-</td>
<td>Correct amount</td>
</tr>
<tr>
<td>Spoon colour (1 of 3)</td>
<td>Incorrect colour</td>
<td>-</td>
<td>Correct colour</td>
</tr>
<tr>
<td>Style of cake cups (1 of 3)</td>
<td>None of correct style</td>
<td>Some of correct style</td>
<td>Only correct style</td>
</tr>
<tr>
<td>Number of cakes (8)</td>
<td>Incorrect number</td>
<td>-</td>
<td>Correct number</td>
</tr>
<tr>
<td>Decorating cakes (1 of 3)</td>
<td>None of correct decorations</td>
<td>Some of correct decorations</td>
<td>Only correct decorations</td>
</tr>
</tbody>
</table>

Table 6.1 Scoring System for Functional Literacy Task

This task was completed by intervention and comparison children at second follow up with cohort one children, and researchers received permission to return to school for collection of this data for cohort 2. Thus data comes from a sample of intervention group children (n=30) and comparison group children (n=34) from cohorts 1 and 2 at 30 month follow up.

Since a validated measure was not available from other sources, this measure should be viewed as a piloting of this measure of functional assessment and an indication of proficiency among the children.
6.3 Results and Discussion

6.3.1 Baseline differences

The baseline differences between the comparison group and those selected for the intervention group have already been explored in section 4.3.7. This analysis was repeated for the actual intervention group (i.e. including only those who actually received the intervention). The table below Table 6:2 presents this data, showing the mean and standard deviation for each measure and an appropriate analysis of the differences between groups. This analysis was repeated for those measures for which longer term data was collected. As previously reported in baseline characteristics, scores reported here are adjusted for age wherever standardisations exist.
### Baseline differences - all intervention and comparison groups

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven's Matrices</td>
<td>Intervention</td>
<td>54</td>
<td>17.22</td>
<td>7.02</td>
<td>t=1.35, p=0.18</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>62</td>
<td>18.53</td>
<td>8.46</td>
<td></td>
</tr>
<tr>
<td>Young Reading (mean 100)</td>
<td>Intervention</td>
<td>54</td>
<td>84.63</td>
<td>7.02</td>
<td>t=-10.4, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>65</td>
<td>98.60</td>
<td>7.98</td>
<td></td>
</tr>
<tr>
<td>Young Maths (mean 100)</td>
<td>Intervention</td>
<td>53</td>
<td>84.02</td>
<td>9.95</td>
<td>t=-3.6, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>90.65</td>
<td>9.77</td>
<td></td>
</tr>
<tr>
<td>Reading Attitude Survey (20-80)</td>
<td>Intervention</td>
<td>49</td>
<td>55.36</td>
<td>11.77</td>
<td>z=-0.42, p=0.67</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>62</td>
<td>55.94</td>
<td>11.59</td>
<td></td>
</tr>
<tr>
<td>NARA Accuracy (months behind chron. age)</td>
<td>Intervention</td>
<td>55</td>
<td>-27.36</td>
<td>9.50</td>
<td>t=-11.92, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>-4.60</td>
<td>11.04</td>
<td></td>
</tr>
<tr>
<td>NARA Comprehension (months behind chron. age)</td>
<td>Intervention</td>
<td>55</td>
<td>-25.89</td>
<td>8.15</td>
<td>t=-6.51, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>-14.94</td>
<td>9.88</td>
<td></td>
</tr>
<tr>
<td>NARA Rate (months behind chron. age)</td>
<td>Intervention</td>
<td>55</td>
<td>-24.15</td>
<td>20.13</td>
<td>t=-6.19, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>63</td>
<td>3.00</td>
<td>26.50</td>
<td>t=-8.13, p&lt;0.001</td>
</tr>
<tr>
<td>PhAB Nonword Reading (mean 100)</td>
<td>Intervention</td>
<td>55</td>
<td>87.75</td>
<td>9.74</td>
<td>t=-3.77, p&lt;0.0011</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>61</td>
<td>101.85</td>
<td>8.95</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Total length</td>
<td>Intervention</td>
<td>46</td>
<td>59.07</td>
<td>43.26</td>
<td>t=-3.77, p&lt;0.0011</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>56</td>
<td>96.16</td>
<td>56.03</td>
<td></td>
</tr>
<tr>
<td>Writing Sample % correct spelling</td>
<td>Intervention</td>
<td>45</td>
<td>48.50</td>
<td>18.21</td>
<td>t=-7.7, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>56</td>
<td>74.55</td>
<td>15.72</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Vocabulary Count</td>
<td>Intervention</td>
<td>45</td>
<td>32.07</td>
<td>19.33</td>
<td>t=-4.85, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>57</td>
<td>51.51</td>
<td>20.73</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Content</td>
<td>Intervention</td>
<td>46</td>
<td>5.67</td>
<td>4.12</td>
<td>t=-3.89, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>57</td>
<td>9.11</td>
<td>4.70</td>
<td></td>
</tr>
<tr>
<td>Battle Self Esteem Inventory (0-10)</td>
<td>Intervention</td>
<td>52</td>
<td>6.31</td>
<td>1.83</td>
<td>z=-0.45, p=0.65</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>52</td>
<td>6.45</td>
<td>1.88</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2 Baseline Differences - All Intervention and Comparison Groups

On ten of the thirteen measures examined here the comparison group scored significantly higher than the intervention group. The exceptions to this were Raven's Matrices (non-verbal IQ), reading attitude and self-esteem where no significant differences existed between the groups. These results confirm that the children in the intervention group were struggling compared to their peers across a number of assessments of literacy.

1 Equality of variance not assumed
If the figures for NARA are examined, we can see that the intervention group are at least 10 months behind their peers on all three features of reading. It is also worth noting that the comprehension reading age for the comparison group is also low, being 15 months behind the average chronological age. This suggests that the text comprehension may be particularly challenging for this group, and that this comparison group probably serves as a more equitable marker of expectations for children in this school than does the standardised average. The wide standard deviations on all scores also suggest that the variation in performance in this sample is very large. This wide variation would tend to militate against finding significant differences between groups, although in the case of baseline measures the differences appear to be large enough to overcome this factor despite the small sample size.

6.3.2 Post intervention differences

6.3.2.2 First follow up - Eighteen months post baseline

The table below (Table 6.3) gives the sample size, mean and standard deviation of assessments taken 18 months post baseline together with the results of an appropriate analysis comparing intervention and comparison groups. Mean scores reported are age adjusted, thus higher scores generally indicated better performance. NARA scores represented the difference in months between reading age and chronological age.
Assessments completed at 18 months follow up, intervention and comparison group

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravens' Matrices</td>
<td>Intervention</td>
<td>35</td>
<td>24.06</td>
<td>9.16</td>
<td>t=-0.76, p=0.45</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>25.67</td>
<td>9.10</td>
<td></td>
</tr>
<tr>
<td>Young Reading</td>
<td>Intervention</td>
<td>34</td>
<td>82.65</td>
<td>7.55</td>
<td>t=-6.96, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>96.64</td>
<td>9.35</td>
<td></td>
</tr>
<tr>
<td>Young Maths</td>
<td>Intervention</td>
<td>37</td>
<td>84.19</td>
<td>10.69</td>
<td>t=-3.36, p=0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>40</td>
<td>91.53</td>
<td>8.42</td>
<td></td>
</tr>
<tr>
<td>Reading Attitude Survey</td>
<td>Intervention</td>
<td>36</td>
<td>52.92</td>
<td>13.75</td>
<td>z=-0.13, p=0.9</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>38</td>
<td>52.92</td>
<td>14.14</td>
<td></td>
</tr>
<tr>
<td>NARA Accuracy</td>
<td>Intervention</td>
<td>37</td>
<td>-25.76</td>
<td>11.36</td>
<td>t=-7.85, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>-1.38</td>
<td>15.31</td>
<td></td>
</tr>
<tr>
<td>NARA Comprehension</td>
<td>Intervention</td>
<td>37</td>
<td>-28.27</td>
<td>14.26</td>
<td>t=-2.2, p=0.03</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>-20.38</td>
<td>16.81</td>
<td></td>
</tr>
<tr>
<td>NARA Rate</td>
<td>Intervention</td>
<td>37</td>
<td>-30.70</td>
<td>20.29</td>
<td>t=-7.1, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>4.51</td>
<td>22.74</td>
<td></td>
</tr>
<tr>
<td>PhAB Nonword Reading</td>
<td>Intervention</td>
<td>36</td>
<td>93.83</td>
<td>6.17</td>
<td>t=-6.18, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>103.82</td>
<td>7.67</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Total length</td>
<td>Intervention</td>
<td>28</td>
<td>135.46</td>
<td>75.47</td>
<td>t=-2.65, p=0.01</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>23</td>
<td>195.26</td>
<td>85.60</td>
<td></td>
</tr>
<tr>
<td>Writing Sample % correct spelling</td>
<td>Intervention</td>
<td>27</td>
<td>87.30</td>
<td>13.94</td>
<td>t=-2.06, p=0.04</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>23</td>
<td>93.76</td>
<td>6.00</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Vocabulary</td>
<td>Intervention</td>
<td>28</td>
<td>62.21</td>
<td>23.41</td>
<td>t=-2.91, p=0.005</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>23</td>
<td>83.48</td>
<td>28.74</td>
<td></td>
</tr>
<tr>
<td>Writing Sample Content</td>
<td>Intervention</td>
<td>27</td>
<td>12.00</td>
<td>9.19</td>
<td>t=-1.82, p=0.08</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>23</td>
<td>16.74</td>
<td>9.12</td>
<td></td>
</tr>
<tr>
<td>Battle Self Esteem Inventory</td>
<td>Intervention</td>
<td>34</td>
<td>7.19</td>
<td>1.83</td>
<td>z=-0.73, p=0.46</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>39</td>
<td>7.41</td>
<td>1.77</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3 18-Month Assessments

On nine of the thirteen measures examined here the comparison group scored significantly higher than the intervention group. The exceptions to this were Raven’s Matrices (non-verbal IQ), reading attitude, self-esteem and content analysis of the writing sample where no significant differences existed between the groups. This result suggests that in the longer term children from the intervention group may be “catching” up with their peers in terms of the complexity and completeness of their story writing.

Mean scores on NARA and nonword reading, our core reading assessments, show how far below “average” achievement for this age group the intervention group remains. The mean reading age of children in this sample is on average more than 24 months.
behind their chronological age, whilst their peers are on average less than two months behind in reading accuracy and more than four months ahead for reading rate. Interestingly the intervention children have improved their age adjusted score in nonword reading, and are approaching the average expected for their age. However they still lag behind their peers and this difference reached significance.

The sample size for these tests is smaller than pre-post changes, as discussed in 6.2.2 above, because this sample is taken from cohorts one and two only. The lack of significance in the difference in the writing sample content may be a result of these reduced numbers, since if we look at the actual figures there is a clear difference between the average number of concepts included in stories by intervention children (14) and by comparison children (17) and the difference was approaching significance. However, the wide variation suggests that the average scores may be misleading. Since scores on the writing sample are confined to natural numbers it may be informative to consider the median scores, which were 9 for intervention and 20 for comparison, where the apparent differences between groups is greater. A larger sample with similar score distributions would have shown a statistically significant difference. The change in writing sample scores over time will need to be considered carefully at subsequent follow-up to examine whether or not this change persists.

In addition to this analysis of mean score, a simple comparison of change over time is warranted. If we wish to decide whether the intervention group is “catching up” with their peers it is useful to know whether they show greater change scores over time as well as whether the difference between intervention and comparison children has disappeared. Thus a graph comparing mean change score was prepared (see Figure 6:3). We can see from this graph that for NARA Comprehension and Nonword reading the
intervention group appear to have made better progress than the comparison group. A t-test showed that neither of these rates of change reached significance (Comprehension t=0.10, p=0.33; Nonword reading t=1.50, p=0.14). Therefore there is no evidence to support the view that the gap between intervention and comparison children is closing on these scores.

First Follow-up Mean Group Change Scores

<table>
<thead>
<tr>
<th></th>
<th>Comparison</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story Vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven's Matrices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Battle SEI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARA Accuracy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6:3 18-Month Follow-Up; Mean Group Changes

290 It is also important to note that from figures given in Table 6:3 there is evidence to support a generalised effects of intervention to performance in the Maths test, the mean score of the intervention group remain significantly below that of the comparison group (84.19 and 91.53 respectively, p=0.001). The poor performance in their maths tests were not taken into account when selecting the intervention group, and their difference at baseline (see Table 4.8) probably reflects their poor overall school
adjustment. Although this intervention was not intended to address maths skills directly, it could be proposed that a generalised effect might be seen with regard to school adjustment or achievement. The data collected for this study do not provide support for this interpretation.

6.3.2.3 Second follow up — 30 months post baseline

The table below (Table 6:4) gives the sample size, mean and standard deviation of assessments taken 30 months post baseline together with the results of an appropriate analysis comparing intervention and comparison groups.

| Assessments completed at 30 months follow up, intervention and comparison group |
|---|---|---|---|---|
| Reading Attitude Survey | Group | n | Mean | Standard Dev. | Analysis |
| Intervention | 13 | 55.69 | 6.49 | Z=-0.42, p=0.67 |
| Comparison | 18 | 54.28 | 13.57 | |
| NARA Accuracy | Intervention | 15 | -25.20 | 16.57 | t=-3.52, p=0.001 |
| Comparison | 22 | -3.68 | 19.34 | |
| NARA Comprehension | Intervention | 15 | -35.60 | 11.56 | t=-1.21, p=0.24 |
| Comparison | 22 | -29.32 | 17.72 | |
| NARA Rate | Intervention | 15 | -16.93 | 26.78 | t=-1.81, p=0.08 |
| Comparison | 22 | -0.23 | 28.04 | |
| PhAB Nonword Reading | Intervention | 15 | 97.07 | 6.31 | t=-3.59, p=0.001 |
| Comparison | 21 | 103.71 | 4.81 | |
| Writing Sample Total length | Intervention | 9 | 141.44 | 74.36 | t=-1.51, p=0.15 |
| Comparison | 14 | 180.79 | 50.85 | |
| Writing Sample % correct spelling | Intervention | 9 | 91.34 | 6.59 | t=-2.7, p=0.01 |
| Comparison | 14 | 96.52 | 2.40 | |
| Writing Sample Content | Intervention | 9 | 70.67 | 24.93 | t=-2.3, p=0.03 |
| Comparison | 14 | 91.00 | 17.65 | |
| Writing Sample Vocabulary | Intervention | 9 | 8.11 | 5.23 | t=-0.83, p=0.41 |
| Comparison | 14 | 9.50 | 5.39 | |
| Battle Self Esteem Inventory | Intervention | 15 | 7.27 | 1.91 | z=-1.31, p=0.19 |
| Comparison | 21 | 7.81 | 2.34 | |

Table 6:4 30-Month Assessments

On five of the ten measures examined here the comparison group scored significantly higher than the intervention group. The exceptions to this were reading attitude, self-esteem, total length and content analysis of the writing sample and NARA comprehension where no significant differences existed between the groups.
At this two-year follow up numbers of children available for the sample are greatly depleted. Thus, for instance, it could be argued that the lack of significant difference in total length of writing sample is due to small sample size and wide variation in the story length within each group. It is also notable that on almost every measure the standard deviation is greater in the intervention group than the comparison group. This suggests that the intervention group contains children with a wider variety of skill levels than does the comparison group. This difference is important when considering the next analysis undertaken on those in the intervention group in the following chapter, to explore whether a differential effect on a sub-group of children can be identified, suggesting characteristics of those who did make significant progress in response to the intervention.

This data can be compared to the scores presented in Table 6:2, from baseline data. Since most results presented are age adjusted we can simply compare the average scores achieved by each group at each time point. It can be seen that overall progress toward the mean has been achieved for both groups on the nonword reading test, and that writing skills have increased across the board. Although the writing test is not standardised, because samples were taken equivalent to current class work, we can expect that the difficulty of the writing task increased as time went on. We can therefore conclude that on some measures of literacy (namely nonword reading accuracy and writing skills) both groups appear to be making progress.
6.3.3 Functional reading task

The results of the functional reading task are given below (Table 6:5), comprising the average total score achieved by each group, together with the average scores on the text only and text and illustration items.

Functional reading assessments completed at 30 month follow up, intervention and comparison group

<table>
<thead>
<tr>
<th></th>
<th>Intervention n=30</th>
<th>Comparison n=34</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean total score (0-16)</td>
<td>10.87</td>
<td>10.21</td>
<td>t=0.81, p=0.42</td>
</tr>
<tr>
<td>Mean text instructions only</td>
<td>2.80</td>
<td>2.79</td>
<td>t=0.02, p=0.99</td>
</tr>
<tr>
<td>score (0-6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean text and illustrated</td>
<td>8.07</td>
<td>7.41</td>
<td>t=1.00, p=0.32</td>
</tr>
<tr>
<td>instructions only (0-10)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6:5 Functional Reading Task Scores

Intervention and comparison group scores are shown not to be significantly different.

We cannot show that this is attributable to their participation in the intervention, but it seems that a high proportion (9/30) of children who two years previously struggled with reading and were not able to read at basal level on NARA, were now able to use their reading skills to extract information from text.

Constructing measures for a group whose skills are considerably behind their chronological age peers presents some challenges. The aim of the measure of functional reading developed herein was to design a task in which poverty of reading skill would not restrict the completion of the activity, but which also provided challenges for those who could read fluently. To this end, the outcome was not accuracy of reading, but accuracy of task completion. This task was shown to be achievable by all in the group across skill levels, but a larger sample would be needed to demonstrate whether such a measure could discriminate between children. It might be
particularly relevant to attempt the task with young, reading-able children to ensure that results do not simply differentiate developmental level but reflect literacy.

6.4 Conclusions

The clear conclusion that can be drawn is that these children remain struggling readers. If the results on NARA alone are inspected we can see that, in the long term, these children are still falling a long way behind their age group. Their competence at reading will have increased over this period, but they have not been able to "catch up" with their age group, remaining more or less two years behind their chronological age. These results are shown in Table 6:4.

The results of this chapter, alongside those from the RCT, can be taken as a clear indication that little improvement is seen in the literacy achievement of these children. This lack of effect needs to be considered in the light of the impact of withdrawing children from class. The half hour teaching time for this intervention was allocated in addition to compulsory curriculum subjects (including one hour per day literacy instruction). A simple calculation, removing breaks, registration and compulsory curriculum subjects (daily act of worship, P.E., literacy and numeracy hours), indicates a maximum of around 12.5 hours per week for other aspects of the curriculum. Half an hour per day represents 20% of this time, a significant proportion for these children to miss. One of the underlying assumptions of much innovation in education seems to be that it cannot do harm. Proof of group gain has been enough to prove efficacy. This study would suggest that very little gain has been bought, at the cost of some loss of positive attitude and of curriculum coverage.
The cost of this intervention, because it was a research funded project, was an additional £1,152.54 per child that had intervention classes. This does not take into account research resources (which were additional and not taken from school funding) nor costs to the school associated with the lessons (for example the provision of a room for the classes to take place in, or additional photocopying). This estimated cost is therefore slightly low, but does include the costs of additional books and other teaching materials. Cost-effectiveness is not often used as a measure of the relative success of reading interventions (Nicolson et. al. 1999, Sesenbaugh, 1995). Costs per child to deliver an intervention need to be offset against estimated avoided costs per child through gains in education or other areas in the short and long term. Thus a more expensive programme can be cost-effective if gains produced are longer or more enduring than another, cheaper programme. In this case where little change was seen cost-effectiveness is minimal. Although no evidence of harm was observed in intervention children, the lack of progress, together with decreased curriculum time and the cost of such an intensive teaching programme, means that such a programme may well not be either appropriate or cost effective.

The possibility of generalised effects cannot be discounted, whilst they do not appear to have been seen in the intervention group itself (see discussion of Maths scores lines 285-295 above), there remains the possibility of generalised effects of the intervention in those not included in it. These might include benefits to class members when particularly needy or demanding members are removed for part of each day, or improvements in class teachers' teaching because of the provision of a resident reading expert in the school. Whilst it was known that class teachers consulted with the reading teacher with regard to the progress of those in the intervention groups, we do not know if there was a process of “leakage” either from the reading teacher or the children in the
literacy groups of expert practice to the classroom. This would benefit all children in the year group, and may even mask any improvements made by the intervention group by encouraging equal improvements in class peers. These leakage or diffusion effects were not assessed as part of this project, so no data is available to comment on them. Alternative sources of data such as this, and alternative strategies for assessment of effectiveness are discussed in Chapter 8. What we do know is that, although the literacy classes included some of the most demanding children, they were not selected on the basis of their classroom disruptiveness. In informal discussions with teachers, while some certainly spoke of relief at having some children removed from class, on the whole those removed for literacy were not the most disruptive members of the class, and in some cases teachers complained that children were more confident and therefore more disruptive as a result of time out of class.
Chapter 7  Determinants of progress in intervention groups

7.1  Introduction

The examination of the data thus far has revealed little change in the intervention group. However, as was recognised in the previous chapter, one feature of the intervention group at all times, and one likely to militate against the likelihood of finding significant differences between group means, was the wide distribution of scores on the assessments used. What we do not know is whether this is due to a sample of children within the group who have either improved greatly or who have made very little progress thus distributing the scores across a wider range. If there were such a group of children it would be important to ascertain whether they were defined by certain characteristics. If such a high or low performing group could be identified it would inform future provision; if we knew which children were likely to respond well and which would need ongoing support, planning of services and design of intervention could be tailored to their needs. In this chapter, therefore, the effect of fixed factors on the progress of children was examined.

Factors both internal and external to the child could contribute to differential progress. For instance, teaching style, school environment, classroom experience, family influence, phonological skills, vocabulary, cognitive style, attitudes toward the classes or reading or IQ. The mechanism for contributing to change will vary according to the factor under consideration. For example, we know from research into reading acquisition that phonological processing plays a key role in the development of literacy. One could sensibly
ask, therefore, whether those children with stronger phonological skills at the outset of this
project had an advantage over their peers, and made better progress. In this particular
instance, it doesn't matter whether the stronger phonological skills are as a result of better
erly literacy or inherent ability; we are interested in whether these skills have the potential to
predict likelihood of progress. In the same way other factors could be considered to
determine whether they are associated with better or worse progress following an
intervention. The factors we chose to consider in this study are discussed below in Methods
2 – Within child factors associated with change, p. 223 below.

As a result of the single school design in this study, most external factors were constant (e.g.
school setting, intervention teacher), but one important feature was varied between children,
namely timing of the intervention. Did the timing of the start of the intervention have an
effect on final outcome? The strategy of delayed start for some children to create a waiting
list control group was designed to meet the needs of the RCT design balanced with the
school's need, but it enables the data to be examined to determine the effect of this delay on
the children involved. One could propose several theories as to how timing might affect
outcome in either direction: early intervention success would suggest that the earlier the
intervention began the better progress would be made, the fact that the second intervention
group spanned the summer holidays (a period associated with learning "fall off") may also
have given the first group an advantage. In contrast, the move from infant to junior school
may be a stressful period for the children and delaying intervention until they were settled in
class may be advantageous. Alternatively the skills of the individual teacher's planning a
small group intervention may improve progressively over time giving an advantage to those
who come later in the teaching programme. As well as contributing to the literature
considering the optimum timing of intervention, this question is of particular importance from a methodological point of view. If we are to use experimental methodologies in applied settings, it is important to ask whether these methods in themselves have an effect on the children involved. The use of control groups in educational settings is, as discussed in Chapter 2, controversial and it would be highly informative to consider how appropriate this strategy is to children and schools. This factor will be considered first and at some length, rather than grouped in the overall analysis of factors affecting change.

7.2 Methods 1: The effects of waiting for intervention

Fifty-nine children were initially chosen for the intervention – 29 in the immediate group and 30 in the waiting list group. All 29 children in the first intervention group received the intervention; this group will be referred to as the immediate intervention group. Of the 30 waiting list children, one left the school in the interim, a further six did not receive the intervention, and four additional children were chosen. Therefore, in total 27 children attended the second intervention group of whom 23 came from the original waiting list group; this group constitutes the delayed intervention group. Including the immediate and the delayed groups, a total of 56 children completed the intervention.

As reported in Chapter 6, Long-term progress, the sample size diminishes over time. The numbers of children at each time point and which year cohort they are drawn from is described in full in section 6.2.2.
Results were analysed in order to assess whether either group (immediate or delayed intervention) showed differential progress on any of the assessments used either during the course of the intervention or during the follow up period.

Data from the following assessments have been used in this analysis:

1. Neale analysis of reading ability
2. Phonological assessment battery
3. Writing sample
4. Raven’s Matrices
5. Young Group Maths Test
6. Young Group Reading Test
7. Battle Self Esteem Inventory
8. Self-concept
9. General Mood Score
10. Reading attitude survey

Data is examined for effects over the period of the intervention and in the longer term. The results of these analyses are presented in Results 2: p. 226 considering short and long-term effects for each assessment in turn. Long-term data is presented for those assessments that were completed at 18 and 30 month follow up; see Chapter 6, section 6.2.4 for description.
The usual course of action for establishing factors associated with change (a regression analysis) was not thought to be appropriate for all factors because the small sample size (n=57) would make it likely that a consideration of all potential factors may mask change associated with individual factors. A series of factors thought likely to contribute to change was established, informed by the literature on reading acquisition and school failure and the data available for analysis. Once defined each factor was correlated with change over time on three key measures of change – NARA accuracy, NARA comprehension and non-word reading. Once a relationship between individual factors and outcome was established in this way, then a further analysis was carried out using just these to establish the association conclusively. In this way the rigour of the regression analysis is achieved, but the possibility of an iterative analysis missing a factor of potentially important impact was reduced. The ten factors thought to potentially affect change were:

1. Gender: In response to evidence for the poor performance of boys relative to girls in early years, and in literacy in particular, it was felt that gender might affect subjects’ response to the intervention.

2. Home Language: As observed in baseline results, children whose home language was other than English were in general performing better than their English only speaking peers. It is relevant to ask what the effect might be on performance during the intervention group.
3. Parent Involvement: Diaries were sent home regularly by the reading teacher with questions and comments for parents/carers to complete when they were reading with their children. A count of the number of entries made by someone other than the child was made, as an indication of the extent of additional home support a child was receiving. Comments were often not recorded by parents, but by older siblings or other relatives. No distinction was made between such entries. This measure of family impact may indicate the extent to which the family offers support to the child, or values the intervention/reading. Whilst it would be difficult to conclude what aspect of family functioning is at work in this assessment, it enables some estimate of family factors to be made in the absence of data from families since this was not available from parental questionnaires (see Chapter 3 section 3.7.5). Data from thirty-seven families was available. A histogram showing the distribution of numbers of entries recorded is shown below (Figure 7:1). Most families complete no, or less than five entries and a large number of diaries were not returned.

**Numbers of Entries in Reading Diaries**

![Histogram showing the distribution of numbers of entries recorded](image)
4. Number of sessions attended: A count of the number of sessions attended would give an indication of the relative absenteeism of children, and also the amount of exposure to teaching within the intervention each child received. Some children (n=9) were taught individually rather than in small groups (n=50). There were too few children taught individually for a statistical comparison between these groups to be valid. For this analysis the total number of either individual or group lessons attended will be considered, but further discussion of the possible importance of group teaching will follow in Chapter 8, section 8.3.1.

5. Free school meals: It has been observed (Hurry & Sylva 1995) that Reading Recovery has been particularly successful with those children who are in receipt of free school meals. It may be particularly important to achieve change in this group, as it is likely to include those who are most disadvantaged.

6. Baseline spoonerism test: It was essential to include a measure of phonological ability in response to the literature on the relationship between reading acquisition and phonology. The spoonerisms test was selected from the PhAB because it was found to differentiate more between children at baseline than other tasks on the PhAB. It requires children to be able to manipulate phonemes in complex ways as well as recognising them and was the most challenging of the sub-tests.
7. Reading Attitude Survey: This was chosen as a representative measure from those assessing attitudes towards books and reading.

8. Non-verbal IQ – Raven’s Matrices: To assess whether non-verbal IQ affects progress during the intervention.

9. Young Reading Score: This was selected from the baseline reading measures, since this had the strongest correlation with other measures at baseline (see section 4.3.7). It was included as a single indicator of baseline achievement.

10. British Picture Vocabulary Scale: This was selected as a proxy measure for exposure to language. It might be supposed that greater exposure to language, either spoken or read, would increase vocabulary. Moreover, vocabulary is one aspect of competence in language. Since those with English as an additional language were under-represented in the intervention groups (see section 4.3.10) the effect on this measure of unfamiliarity with the language should be minimised in this group.

7.4 Results 2: The effects of waiting for intervention

7.4.2 Organisational Effects of Waiting

As described in Methods 2 above, all 29 children in the first intervention groups received the intervention while only 23 of those assigned to the delayed intervention went on to receive the programme. Of the 30 waiting list (or control) children, one left the school before the intervention began. Teachers decided that a further six children from the delayed group had
made so much progress over the course of the two terms that they were no longer in need of
the intervention, and four additional children who had not been initially identified were
chosen to replace them. This is summarised in Table 7:1 below. The first question asked
was whether this “dropping out” effect was, in itself an important one. There was a
significantly greater probability of not receiving the intervention among children in the
delayed group than among the immediate group.

Initial selection and final inclusion: comparison of immediate and delayed intervention
groups

<table>
<thead>
<tr>
<th></th>
<th>Number of Children Initially Selected</th>
<th>Number of those who received intervention</th>
<th>Percentage</th>
<th>Fisher’s Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate Intervention</td>
<td>29</td>
<td>29</td>
<td>100%</td>
<td>p=0.01 (two tailed)</td>
</tr>
<tr>
<td>Delayed Intervention</td>
<td>30</td>
<td>23</td>
<td>77%</td>
<td></td>
</tr>
</tbody>
</table>

Table 7:1 Selection and Inclusion in Intervention Groups

The delayed group were randomly selected from the same pool of children as the immediate
intervention group. In other words they were determined by teachers to be in need of literacy
support at the beginning of the year. It may be that, given time, the “refinement” process that
happened to the delayed intervention group enabled the school to filter out those who could
make progress without support in exchange for those who were now struggling. Even if those
who did not receive the intervention didn’t “need” it as much as those who did receive it, they
were children the school had identified as struggling with reading and writing at the beginning
of the year. For the individual children involved it did mean that being randomly assigned to
waiting meant they did not receive the support planned for them. The data considered here
cannot determine whether first or later impressions of need were more accurate (this will be
discussed in Chapter 8, section 8.3.3) but it is clear that in the event of a successful
intervention this result demonstrates a clear disadvantage to those in the delayed group.

7.4.3 **Post Intervention and follow up Results**

To compare progress made during the intervention, change in scores over the period of the intervention was examined. That is, scores at pre and post intervention for each intervention group were compared. For those measures that met parametric assumptions, a regression analysis was conducted. Chronological age was used as a co-variate in this analysis. Pre-intervention scores were also used as co-variates when analysing changes over the period of the intervention, and baseline scores were used when analysing longer-term change. Results for each measure are presented and discussed in turn below; statistically significant findings are shown in bold.

7.4.3.2 **Neale Analysis of Reading Ability**

The NARA was completed for all children at 18 and 30 months post intervention. Mean scores reported below (Table 7:2) show change in raw score since baseline. The number of children still present at school who completed the tests are shown, with only 15 of the original sample of 21 from the first cohort still attending the school two years post intervention. Results of appropriate analyses comparing progress between groups are also shown.
NARA mean change, comparison of immediate and delayed intervention groups

<table>
<thead>
<tr>
<th></th>
<th>Mean Change in Raw Score</th>
<th>Immediate Intervention Group</th>
<th>Delayed Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nara Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td></td>
<td>n=28</td>
<td>n=26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=9.21</td>
<td>mean=8.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.=4.92</td>
<td>s.d. = 6.51</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.007, F (1,48)=1.11, p=0.30</td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td></td>
<td>n=19</td>
<td>n=16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=16.21</td>
<td>Mean=18.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D. = 7.85</td>
<td>S.D. = 6.74</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.002, F (1,31)=1.84, p=0.67</td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td></td>
<td>n=8</td>
<td>n=7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=28.50</td>
<td>Mean=28.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D. = 14.19</td>
<td>S.D. = 11.95</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.007, F (1,11)=0.10, p=0.75</td>
<td></td>
</tr>
<tr>
<td><strong>Nara Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td></td>
<td>n=28</td>
<td>n=26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=2.54</td>
<td>mean=1.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.=2.90</td>
<td>s.d. = 2.76</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.002, F (1,48)=0.16, p=0.69</td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td></td>
<td>n=19</td>
<td>n=16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=4.74</td>
<td>mean=4.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 3.84</td>
<td>s.d. = 4.69</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change&lt;0.001, F(1,31)=0.002, p=0.97</td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td></td>
<td>n=8</td>
<td>n=7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=6.88</td>
<td>mean=4.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 4.09</td>
<td>s.d. = 4.54</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.02, F(1,11)=0.27, p=0.61</td>
<td></td>
</tr>
<tr>
<td><strong>Nara Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td></td>
<td>n=28</td>
<td>n=26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=2.57</td>
<td>mean=6.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 4.17</td>
<td>s.d. = 24.59</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.04, F (1,48)=2.42, p=0.13</td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td></td>
<td>n=18</td>
<td>n=16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=19.56</td>
<td>mean=3.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 15.48</td>
<td>s.d. = 17.57</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.01, F(1,30)=5.78, p=0.02</td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td></td>
<td>n=8</td>
<td>n=7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=33.48</td>
<td>mean=28.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 16.28</td>
<td>s.d. = 23.65</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td>R^2 change=0.004, F(1,11)=0.11, p=0.75</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2 NARA; Comparison of Immediate and Delayed Intervention Groups

No significant differences were found between groups on progress on any measure of the
NARA over the period of the intervention.

In longer term follow, five out of six of the follow-up analyses showed no significant differences between groups in progress since the outset of the project. A modest effect was seen in NARA Rate at 18-month follow-up, where immediate intervention children showed a significantly greater increase in NARA Rate than delayed group. However the effect size was small, with group accounting for only about 1% of the variance observed. Since this difference had disappeared at 30-month follow-up it may be that this was an anomalous result.

No “waiting” effect in either direction was observed for reading as assessed by NARA. The implications of this finding will be considered in section 7.4.4 below.

7.4.3.3 Phonological Assessment Battery

Changes in score over the period of the intervention are shown for all parts of phonological assessment battery, but only nonword reading was used at follow up. Mean change scores are shown for each measure, together with the results of a regression analysis comparing pre and post intervention scores between groups; these are shown in Table 7:3 below.
**Phonological Assessment Battery mean change, comparison of immediate and delayed intervention groups**

<table>
<thead>
<tr>
<th></th>
<th>Mean Change in Raw Score</th>
<th>Immediate Intervention Group</th>
<th>Delayed Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spoonerisms</strong></td>
<td>Over period of intervention</td>
<td>n=27</td>
<td>n=25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=3.96</td>
<td>Mean=3.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.=5.56</td>
<td>S.D.=4.62</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.001, F (1,46)=0.09, p=0.78</td>
<td></td>
</tr>
<tr>
<td><strong>Semantic Fluency</strong></td>
<td>Over period of intervention</td>
<td>n=27</td>
<td>n=25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=0.04</td>
<td>Mean=1.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.=4.64</td>
<td>S.D.=7.05</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.01, F (1,46)=0.63, p=0.43</td>
<td></td>
</tr>
<tr>
<td><strong>Alliteration Fluency</strong></td>
<td>Over period of intervention</td>
<td>n=27</td>
<td>n=25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=0.67</td>
<td>Mean=0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.=3.70</td>
<td>S.D.=2.87</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.007, F (1,46)=0.49, p=0.49</td>
<td></td>
</tr>
<tr>
<td><strong>Rhyme Fluency</strong></td>
<td>Over period of intervention</td>
<td>n=27</td>
<td>n=25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean=1.89</td>
<td>Mean=1.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.=2.95</td>
<td>S.D.=3.45</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.002, F (1,46)=0.18, p=0.68</td>
<td></td>
</tr>
<tr>
<td><strong>Nonword Reading</strong></td>
<td>Over period of intervention</td>
<td>n=27</td>
<td>n=25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=2.74</td>
<td>mean=1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.=4.17</td>
<td>s.d.=2.85</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.02, F (1,46)=1.28, p=0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 months follow up</td>
<td>n=18</td>
<td>n=15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=4.06</td>
<td>mean=4.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.=4.47</td>
<td>s.d.=2.00</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.006, F(1,29)=0.13, p=0.72.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 month follow up</td>
<td>n=8</td>
<td>n=7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=7.75</td>
<td>mean=8.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d.=2.96</td>
<td>s.d.=3.34</td>
</tr>
<tr>
<td><strong>Regression Analysis</strong></td>
<td></td>
<td>R²change=0.002, F(1,11)=0.10, p=0.76</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3 PhAB; Comparison of Immediate and Delayed Intervention Groups

Rates of change over the period of the intervention were not shown to be significantly different between immediate and delayed intervention groups on any measure of the PhAB over the period of the intervention, or for nonword reading in the longer term.

7.4.3.4 **Writing Sample**

Change in score over the period of the intervention and a regression analysis of the change over the period of the intervention and in the longer terms are show in Table 7:4 over.
<table>
<thead>
<tr>
<th></th>
<th>Mean Change in Raw Score</th>
<th>Immediate Intervention Group</th>
<th>Delayed Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total length</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>n=12</td>
<td>mean=48.75 s.d. 68.6</td>
<td>n=6</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.12, F (1,23)=3.62, p=0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>n=16</td>
<td>mean=60.25 s.d. 64.39</td>
<td>n=8</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.006 F(1,20)=0.17, p=0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td>n=5</td>
<td>mean=34.2 s.d. 26.13</td>
<td>n=3</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.13 F(1,4)=2.93, p=0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Percentage correct spelling</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>n=12</td>
<td>mean=22.96 s.d. 13.44</td>
<td>n=15</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.001, F (1,23)=0.007, p=0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>n=16</td>
<td>mean=43.09 s.d. 30.56</td>
<td>n=8</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.001, F(1,20)=0.01, p=0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td>n=5</td>
<td>mean=34.96 s.d. 18.68</td>
<td>n=3</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.50, F(1,4)=17.54, p=0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Content Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>n=12</td>
<td>mean=5.92 s.d. 7.84</td>
<td>n=15</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.02, F (1,23)=0.59, p=0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>n=15</td>
<td>mean=8.93 s.d. 15.12</td>
<td>n=8</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.04, F(1,19)=1.07, p=0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 month follow up</td>
<td>n=5</td>
<td>mean=-1.4 s.d. 4.67</td>
<td>n=3</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.35, F(1,4)=3.96, p=0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>n=12</td>
<td>mean=22.67 s.d. 31.05</td>
<td>n=15</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.002, F (1,23)=0.6, p=0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/12 mo post intervention</td>
<td>n=16</td>
<td>mean=25.63 s.d. 21.34</td>
<td>n=8</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.004, F(1,20)=0.13, p=0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18/24 mo post intervention</td>
<td>n=5</td>
<td>mean=22.6 s.d. 15.66</td>
<td>n=3</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.27, F(1,4)=3.86, p=0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.4 Writing Sample; Comparison of Immediate and Delayed Intervention Groups
No significant differences between groups in rates of progress over the period of the intervention for percentage correct spelling, story content or vocabulary used were observed. There was shown to be a significantly greater increase in total length of story in the delayed intervention group compared to the immediate intervention group.

There were no significant differences between groups for progress in measures of writing skill in the longer term for seven out of eight comparisons. The one statistically significant result showed that the delayed intervention group had increased their correct spellings significantly more than the immediate intervention group at 30-month follow-up. The difference between groups accounts for half of the variance. The sample size at this point is very small.

7.4.3.5 Other Standardised Measures of Academic Achievement

Two other baseline measures of academic achievement (Young’s Group Maths Test and Young’s Group Reading Test) along with Raven’s Matrices were repeated at 18-month follow up. Mean change since baseline and appropriate statistical analyses are shown in Table 7:5 below.
Measures of academic achievements mean change, comparison of immediate and delayed intervention groups

<table>
<thead>
<tr>
<th></th>
<th>Mean Change in Raw Score</th>
<th>Immediate Intervention Group</th>
<th>Delayed Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven’s Matrices</td>
<td>18 month follow up</td>
<td>n=18</td>
<td>n=16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=5.06</td>
<td>mean=7.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 7.38</td>
<td>s.d. = 4.77</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.01, F(1,29)=0.72, p=0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young’s Reading Test</td>
<td>18 month follow up</td>
<td>n=17</td>
<td>n=15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=8.35</td>
<td>mean=8.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 6.51</td>
<td>s.d. = 6.78</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change&lt;0.001, F(1,27)=0.002, p=0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young’s Maths Test</td>
<td>18 month follow up</td>
<td>n=16</td>
<td>n=18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean=14.06</td>
<td>mean=11.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>s.d. = 5.64</td>
<td>s.d. = 6.58</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R^2 change=0.04, F(1,29)=4.40, p=0.045</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7:5 Academic Achievement; Comparison of Immediate and Delayed Intervention Groups

No differences were observed between immediate and delayed intervention groups on Raven’s Matrices or on the reading test at 18-month follow up. The immediate intervention group was shown to have progressed significantly more since baseline than the delayed intervention group on the maths test, although the effect size was small (4% of variance).

7.4.3.6 Affective Measures

Mean change in scores over the period of the intervention and in the longer term on affective measures is shown in Table 7:5 below, along with an appropriate statistical analysis of change. Change scores are reported over the period of the intervention and in the longer term according to data available, thus the reading attitude scale (completed in whole class settings) was not collected at the end of the second intervention period and therefore change over the period of the intervention can not be reported, nor were self-concept scales used at follow up.
### Attitudinal and Affective Measures: Comparison of Immediate and Delayed Intervention

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean Change in Raw Score</th>
<th>Immediate Intervention Group</th>
<th>Delayed Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Esteem (Battle) (Range 0-10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>0.06</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>0.25</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-0.68, p=0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 months follow up</td>
<td>0.19</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-1.20, p=0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Attitude Survey (Range 0-10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>-9.63</td>
<td>-2.35</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 months follow up</td>
<td>-12.6</td>
<td>-1.21</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-2.44, p=0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favourite Self concept now (Range 0-15.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>0.99</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favourite Self concept future (Range 0-15.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>0.2</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Regression Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Self concept now (Range 0-15.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>0.14</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R² change=0.01, F(1,45)=0.55, p=0.46 n.s.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Self concept future (Range 0-15.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>-0.4</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>R² change=0.01, F(1,49)=0.50, p=0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Mood Score (GMS) (Range 10-40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over period of intervention</td>
<td>-2.52</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-2.61, p=0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 months follow up</td>
<td>-1.39</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-1.70, p=0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 months follow up</td>
<td>-1.25</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>-0.84, p=0.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.6 Attitudinal and Affective Measures; Comparison of
Immediate and Delayed Intervention Groups

The only measure showing a statistically significant difference in change score over the period of the intervention was the general mood score, where the immediate intervention group showed a fall in reported happiness, compared to a slight gain in the delayed intervention group.

No difference between groups was seen at 18-month follow up, and only one difference at 30-month difference. The immediate intervention group became more negative in their attitude to books than the delayed groups.

7.4.4 The effects of waiting for intervention; Discussion of results

The decrease in General Mood Score over the period of the intervention and in Reading Attitude Score in the long term, combined with evidence from the RCT phase of the evaluation, indicates that the immediate intervention group were less positive about reading following intervention. This might indicate that children in the immediate intervention group were less happy following the intervention than before. We know that at baseline groups were not different from each other on mood scales. It might be that the difference is explained by sensitivity to negative effects present in the immediate intervention group, which was absent in the delayed intervention group. For instance, the second group may be protected by personality or age against feeling a stigma associated with additional teaching support. Alternatively, this could simply be a feature of timing of assessment, that perhaps children feel more positive at the beginning of the school year (as at baseline, and post delayed intervention)
than they do later in the school year.

There was no evidence of harm to the waiting list control group, in that the children who had to wait for their intervention and received the intervention did seem to benefit at least as much as their peers who received the intervention immediately. Indeed, there is some evidence of limited benefit in the waiting group in longer-term progress in reading accuracy and spelling. However, these slight benefits to the waiting list group need to be balanced with the evidence that those who waited were far less likely to actually receive the intervention. Teachers re-assessed the needs of that child or their peers such that others were chosen in their stead. In most cases these “replacement” children could have been identified at baseline as falling behind their peers, but these revisions did not always reflect the previous or current performance of children. From a policy standpoint the question of timing, waiting and reassessment of selection are of less importance than the fact that minimal change was seen in response to the intervention overall. From a research standpoint the findings are more important, they demonstrate that randomising children to a waiting list control group in order to achieve a rigorous research design was not shown to disadvantage either group in terms of their response to that intervention. This is important given the reluctance of many to use such techniques in educational settings. This project does not randomise to no intervention, so every child is given the opportunity to benefit, but using a temporary control group was an effective method for ensuring true control data was available.

7.5 Results 2: Factors associated with change

As reported in methods section (7.2 p223) a group of ten factors was generated which were considered to be potential variables affecting change. These were:
Initially these were examined to see if there was evidence for a relationship between any of them and pre/post change score over the period of the intervention. Three outcome measures were selected from the whole data set: NARA Accuracy and comprehension and nonword reading. These measures are markers of change in key competencies. The table below (Table 7.7) shows mean change scores for NARA accuracy and comprehension and non-word reading (raw scores) and the results of an appropriate analysis of their relationship to likely factors influencing change. Dichotomous factors (such as gender) were analysed by comparing mean change in each group, continuous factors (such as number of sessions attended) were analysing by correlation with outcome variables. Factors with relationships with change scores reaching or approaching significance at the 0.05 level on any of the three outcome measures are shown in bold.
Four factors were shown to have a relationship with change scores reaching or approaching significance with one or other outcome measure: gender, spoonerisms, non-verbal IQ and vocabulary. These four factors were used in a regression analysis to determine whether they had a significant effect on outcome.

It was not appropriate to group the four factors since the three continuous factors were not significantly correlated with each other at baseline (see 4.3.6 Relationships between variables). Moreover they showed a mix of positive and negative correlations with outcome suggesting
effects in different directions. Similarly gender, being a dichotomous variable, must be analysed separately. Linear regression analyses were conducted for each factor, adjusting for baseline scores and chronological age to ascertain whether they did, indeed, have a significant effect on progress during the period of the intervention. The dependent variable in all cases was change in score over the period of the intervention. This analysis was ordered such that baseline score and chronological age were analysed in the first step and the factor under consideration was introduced in the second step. Part and partial correlations were calculated, such that the model showed the regressions statistics for baseline and chronological age and regression statistics for the effect of the factor over and above this underlying effect.

The results of this analysis are shown in Table 7:8 below. This table shows the effect size ($R^2$) and significance of these four factors, over and above the effect of baseline score and chronological age, for each of the three outcome variables. The effect of pre-intervention and chronological age and baseline score alone are not given.
Table 7.8 Linear Regressions using identified likely factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Outcome measure</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoonerisms</td>
<td>Nara Accuracy Change</td>
<td>$R^2\text{change}=0.04$, $F(1,48)=2.13$, $p=0.15$</td>
</tr>
<tr>
<td></td>
<td>NARA Comprehension change</td>
<td>$R^2\text{change}=0.00$, $F(1,48)=0.04$, $p=0.85$</td>
</tr>
<tr>
<td></td>
<td>Nonword Reading change</td>
<td>$R^2\text{change}=0.12$, $F(1,46)=0.45$, $p=0.51$</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Nara Accuracy Change</td>
<td>$R^2\text{change}=0.01$, $F(1,45)=0.73$, $p=0.40$</td>
</tr>
<tr>
<td></td>
<td>NARA Comprehension change</td>
<td>$R^2\text{change}=0.09$, $F(1,45)=4.7$, $p=0.035$</td>
</tr>
<tr>
<td></td>
<td>Nonword Reading change</td>
<td>$R^2\text{change}=0.03$, $F(1,43)=1.35$, $p=0.25$</td>
</tr>
<tr>
<td>Non Verbal IQ</td>
<td>Nara Accuracy Change</td>
<td>$R^2\text{change}=0.04$, $F(1,46)=2.12$, $p=0.15$</td>
</tr>
<tr>
<td></td>
<td>NARA Comprehension change</td>
<td>$R^2\text{change}=0.04$, $F(1,46)=1.70$, $p=0.20$</td>
</tr>
<tr>
<td></td>
<td>Nonword Reading change</td>
<td>$R^2\text{change}=0.08$, $F(1,44)=4.53$, $p=0.04$</td>
</tr>
<tr>
<td>Gender</td>
<td>Nara Accuracy Change</td>
<td>$R^2\text{change}=0.06$, $F(1,48)=3.3$, $p=0.08$</td>
</tr>
<tr>
<td></td>
<td>NARA Comprehension change</td>
<td>$R^2\text{change}=0.06$, $F(1,48)=0.28$, $p=0.60$</td>
</tr>
<tr>
<td></td>
<td>Nonword Reading change</td>
<td>$R^2\text{change}=0.00$, $F(1,46)=0.15$, $p=0.70$</td>
</tr>
</tbody>
</table>

Of these twelve analyses, two reached significance (the effect of vocabulary on progress in comprehension and the effect of IQ on Nonword reading) and one approaches significance (the effect of gender on progress in reading accuracy), indicating a trend toward progress associated with this factor. Each of these findings will be discussed in turn.

A significant relationship was found between baseline vocabulary score on the British Picture Vocabulary Scale, and change in comprehension score over the period of the
intervention. The scattergram below (Figure 7:2) shows the regression line between change in comprehension and baseline BPVS score. This figure does not adjust for baseline score on comprehension or chronological age as the analysis above does, however it serves to represent the relationship observed between comprehension and vocabulary. It would seem that those children with a better vocabulary at the outset of the project made more progress than those with a more restricted vocabulary as measured by the BPVS.

This relationship is particularly interesting in what it may tell us about the role that vocabulary plays in comprehension development and vice versa. We cannot know about the direction of causality in this case; either those with slightly better comprehension skills may be better at acquiring vocabulary, or those with greater vocabulary may improve comprehension. Either way, these two variables have an intuitive association in that knowledge of words and their understanding would seem to be linked, and this seems to be supported by this finding. Theoretically, a semantic lexicon would be involved in both the acquisition of word
understanding and the processing of written language (Temple, 1993; Treiman, 1993).

A significant relationship was found between baseline non-verbal IQ, and change in nonword reading score over the period of the intervention. The scattergram below (Figure 7:3) shows the regression line between change in nonword reading and baseline Ravens' Matrices score. As with the figure above (Figure 7:2), this figure does not adjust for baseline score on comprehension or chronological age as the analysis above does. It would seem that those children who scored higher on the Ravens' Matrices (and thus had higher non-verbal IQs) made more progress in nonword reading over the period of the intervention.

Scattergram showing relationship between change in nonword reading and non-verbal IQ test

A non-significant relationship was found between gender, and change in reading accuracy over the period of the intervention. One needs to show particular caution interpreting results that approach, but do not reach statistical significance. This trend is discussed here.
for completeness, but we cannot be certain that this does not represent a chance finding. The scattergram below (Figure 7:3) shows the regression lines by gender for change in reading accuracy by baseline reading accuracy score. Since gender is a dichotomous variable it is possible to represent the results of the linear regression analysis in this way, and this figure more closely represents the analysis conducted above than the previous two figures. It would seem, according to this scattergram, that boys progressed at a more rapid rate than girls. The figures for mean change confirm this, with boys' score improving by 2.05 points on average, compared to girls' average improvement of 1.9 points. If we refer back to section 4.3.8, analysis by gender, we can see that there the baseline scores of boys and girls did not differ significantly by gender in the sample overall. Although this may suggest there was an advantage to boys in this respect, this trend in the results is not proven.
Scattergram showing relationship between change in reading accuracy and baseline reading accuracy by gender

Figure 7:4 Relationship Between Accuracy Change and Baseline Vocabulary

As well as considering the extent to which these three factors were associated with change in this sample, it is also worth noting the implications of the finding that the other seven factors failed to show an association with outcome. These were home language, parent diaries, number of sessions attended, free school meals, reading attitude survey, spoonerisms and Young group reading test. The failure to find factors associated with change may, in part, be attributable to the low rate of change overall. Nonetheless a notable feature of change observed in this group was the large standard deviations, in other words different levels of change were observed between children. It does not appear as if this variation is attributable to any of the potential factors identified. It would suggest that this intervention was equally ineffective for all children.
No single factor was seen to have significant relationship with all outcome measures, and conversely no outcome measure was seen to be associated with more than one potentially predictive variable. However, the direction of significant effect was shown to be the same, for both vocabulary and IQ: those children who showed greater achievement at the outset seemed to have benefited more from the intervention than their peers. This means that those who are at an advantage at the outset seem to build on that advantage, whilst those who are at less of an advantage remain so.

We know from the RCT results that, relative to controls, very little change was seen in the intervention group and that the intervention group did not appear to “catch up” with their peers in the longer term. It would also appear that there is little evidence for a sub-group of children who did particularly well in the intervention. This latter conclusion is, of course, limited to those factors considered in this analysis and collected within the scope of this study.

7.6 Conclusions

There are two main conclusions that we can draw from the data presented in this chapter. Firstly that being placed in a waiting list control group is disadvantageous for children, not because of the direct effects of waiting for about six months for the intervention, but because children in the waiting group were less likely to go on to receive the intervention. In one case this was because the child left the school in the interim, but for the others this is because their needs were reassessed over this period compared to other children and it was
decided that they were not in as great need of the intervention. This conclusion is particularly important when considering planning future interventions or evaluations and will be considered in the following chapter. The second conclusion is that it was not possible to identify a single characteristic associated with “success” in the intervention, however it does appear that those with better baseline skills were at a slight advantage over their peers and made more progress over the period of the intervention in some areas.
Chapter 8  Critical analysis of evaluation methodology and findings

8.1 Introduction

This project was designed to explore the impact of a literacy intervention on children who were struggling with their reading and writing. The method selected was a randomised controlled trial of intervention for those children with poorest performance entering Year 3 in a large London primary school. The results of the intervention were analysed between groups and over time. The main conclusions that can be drawn from these analyses were:

1. The children in the intervention group did not make significantly different progress on most measures of reading skill or attitude than a control group who did not receive the intervention. A modest difference in reading accuracy was found, with the intervention group showing slightly greater progress. Balanced against this was some evidence that attitude to reading might have become less positive.

2. No variables tested were found to predict change in response to the intervention.

3. Long-term progress was seen in the intervention groups, but we cannot know to what extent this was attributable to the intervention. This progress was not enough to close the gap between themselves and their class peers.

4. The children had a positive response to the intervention, which they enjoyed very much and would have chosen to continue.
The evaluation was designed to incorporate a randomised controlled trial of the intervention (see Chapter 2 for rationale). Whilst this methodology is thought to be preferable, there are considerable difficulties with executing experimental methodologies in an applied setting, where many factors out of the experimenter’s control are likely to impact on the findings. Emerging from the execution of this study is an example of the tension created by placing a rigorous experimental design within a naturalistic educational environment. A balance must be sought which enables the needs of both the participants (school, teachers and children) and the research programme to be met. In order to achieve this both groups must recognise those areas where compromise is possible and, equally, those areas where compromise would be counter-productive. This process of compromise and negotiation requires all stakeholders to recognise the importance of key issues to others in the research group.

In the first instance, this chapter will examine critically the methodology employed, possible alternative strategies, and the implications of all of these for the findings of this project. This will constitute an exploration of both the robustness of the methodology and of the findings. It is also intended to document the difficulties encountered and strategies taken to overcome the differing needs of the experimental design and the educational setting demonstrating that the use of experimental methodologies in school settings is an achievable goal. The second part of the chapter will discuss the conclusions that can be drawn from this and other similar projects and ask whether data from such research projects can be used in planning services for children.
8.2 Discussion of Methodology

The difficulties of executing experimental methodologies in an applied setting are well recognised (Robson 2002). The practical and ethical barriers that one encounters in such an endeavour are numerous. There has been a vein of literature that suggests that such experimental methodologies should be avoided in the social sciences, where ethical considerations may be particularly important. For example, whilst it is ethically acceptable to randomise assignment to a treatment when it cannot be introduced simultaneously to all (as in this group), where this is not the case and where participants clearly prefer treatment it might be thought of as unethical to use no-treatment controls. The difficulty with this ethical reasoning is in the assumption that the treatment in consideration (regardless of participant preference) will not do harm. That is, such a justification for not using experimental methodologies is underpinned by the assumption, in advance of the findings of the research, that the intervention “works”. Moreover, alternative methodologies (such as non-experimental and quasi-experimental designs) in turn have difficulties that make them inappropriate in particular settings. If intervention studies are to attribute the causality of outcomes observed the accepted methodology is that of a controlled, preferably randomised, trial. This need not underestimate the value of different sources of data for determining other factors. For example, controlled trials with their focus on outcome data may not contribute information on the process by which information is internalised by children or the ways in which different teachers may execute the intervention. Data on such aspects of intervention may provide invaluable information about the implementation of the intervention, but cannot establish whether the intervention was “successful” in changing outcome. That is, it cannot establish whether any changes seen in the
intervention group can be attributed to the effects of the intervention itself, and not to other factors.

Accepting the preference for randomised controlled trials, we must consider the difficulties associated with applied settings. In order to accept that the findings of the trial are not an artefact of poor design we must examine the extent to which the trial conforms to experimental standards and whether the process of conducting trials in a real-world setting introduces methodological problems which would lead to a rejection of the findings. To do this a model of experimentation must be adopted, to which any trial may be compared. For this purpose, the seminal description of experimental design in social research described by Campbell and Stanley (1966) was selected. The authors provide a description of experimental and quasi-experimental methodologies, along with a description of internal and external factors that may jeopardise experimental validity. Internal validity refers to the extent to which features of the trial design or execution affect the likelihood that the findings are representative. External validity refers to the extent to which the findings are generalisable to those outside the trial — i.e. to the outside world. The extent to which the study under consideration here has accounted for these factors provides an assessment of the quality of the trial design. Each factor defined by Campbell and Stanley as contributing to external and internal validity will be examined in the context of this trial.
8.2.1 Factors Jeopardising Internal Validity

1. History — "the specific events occurring between the first and second measurement in addition to maturation" (Campbell & Stanley, 1966).

Historical factors outside an experiment may affect outcome, thus for example, in a study looking at attitudes toward emergency services, the beginning of a strike by fire fighters might affect outcome. This supports the case for the use of a control group, where historical events might be supposed to have the same effects on outcome for each group, so difference between groups could be attributed to intervention. In this trial, one significant "historical" event might be thought of as affecting outcome. The trial was going on in school at a time when there was a heightened awareness of the teaching of reading in their school as a whole due to the introduction of the National Literacy Strategy. Since both intervention and control groups were equally as likely to be affected by this change, it should not jeopardise results. If there was, indeed, an historical effect from this, or other unknown events, both intervention and control groups would have been exposed to it and the effects are likely to be similar for both. This does not hold true for the comparisons made between early and late intervention groups discussed in section 7.4. Thus, as mentioned previously, results from this comparison are less robust than those made during the controlled phase of this trial.

2. Maturation — "processes within respondents operating as a function of the passage of time" (Campbell & Stanley, 1966)

This factor includes both developmental or ageing effects over longer periods and more immediate effects, such as tiredness, thirst or hunger.
Since groups ran concurrently, the aging effects can be assumed to be the same for both intervention and control groups. As before, for analyses conducted at other time points (for example early vs. late comparison) this was not the case. In part, this has been controlled by using chronological age at the time of testing as a co-variate throughout. This enables a comparison to be made between children of slightly different ages. This practice, however, relies on an assumption that the effects of time and/or development are linear; i.e. the maturational difference between a child of 7 years 0 months and 7 years 6 months is the same as that between a child of 7 years 6 months and 8 years 0 months. We have reason to suppose that this is not the case; models of development do not typically follow a linear course, but instead show periods of rapid gain, internalisation and then slow growth until a skill is mastered and where experience is an important modifying factor. Using pre-test scores as an additional co-variate will lessen this effect, such that maturation in chronological age and stage of learning will be accounted for. This is a problem particularly in working with children, who are likely to go through stages of rapid developmental change rather than have steady progress. It is difficult to account for the non-linear nature of learning growth within the confines of measurable outcomes beyond adjusting for existing developmental level or skills.

The immediate effects of the passage of time — tiredness, hunger etc. can be accounted for by ensuring that all participants are in a similar physiological state at the time of testing. Thus, for instance, children should always be tested immediately on arrival at school. Such a strategy is not compatible with the execution of research in practice. At the very least, the time taken to test all participants individually would be prohibitive. This problem is particularly acute in school settings, where access to children is already difficult, and disruption of school timetabling needs to be minimised. Even within the testing routine, delivery of the test battery should ideally be systematised so that every
child receives the tests in the same order with the same period between each one and so forth. In this trial, where a number of tests of differing length were used over a minimum of two occasions, this ideal proved impossible to deliver. In order to avoid disruption, testing could only take place in particular parts of the day for particular classes and therefore pragmatic, on the spot, decisions were made as to which tests there would be time for. Added to this, pupil absence made it more difficult to test particular children who were frequently absent. There were, however, no systematic differences in the way intervention and control groups were tested, thus any maturational effects were equally likely to affect results in either condition. This is an example of the ways in which experimental methodology must account for validity constraints (by randomisation and by non-systematic variation) without jeopardising the execution of the trial as a whole.

3. Testing — “the effects of taking a test upon the scores of a second testing” (Campbell & Stanley, 1966)

Whenever possible parallel forms of tests were used to minimise re-test effects. Careful records were kept so that no child received the same forms consecutively. For tests where parallel forms were not available and where exact repetition was necessary, re-test affects may have been an issue. The use of a control group drawn from the same population sample as the intervention group means that any re-test effects seen in the intervention group would also have been apparent in the control group. In comparing differences between groups, rather than absolute scores, any such effect occurring in both groups would be cancelled out.
In this trial, the relevant question is whether or not changes in observers would produce different results, and whether this potential effect was accounted for in the design. This is a beguilingly simple question. On the one hand, it would seem that this trial by using a number (four in total) of different testers and observers over the course of the trial took account of this, but observers were not blinded to the group of the participants. Where possible (for instance in the writing sample) data was anonymised before scoring took place and inter-observer reliability was calculated and found to be adequate (e.g. see section 3.7.1). This was not the case for face-to-face assessments. Both participants and interviewers knew which group the participant belonged to and, therefore, it could be argued that a set of expectations as to differential progress could affect the behaviour of the interviewer. If this had been a trial of a new medicine, this fact would have been sufficient to render the trial invalid. However, it is important to consider the implications of the assumption of blindness in this case. The purpose of blinding observers to the status of the participant is to ensure there is no bias in the recording of outcome measures. In contrast to this, there is a question of the effects of being observed per se in children. Good practice guidelines on research with children suggest that the relationship of trust and communication may be important in establishing the rights of the child in the research process (Harker, 2002). “Issues such as where and when research is conducted, power relationships between the researcher and the child, who is present and who will be informed of the findings, are all likely to impact upon what children are willing to talk about and the candour of their responses.” (NCB & Barnardo’s, 2002). The development of a relationship between researcher and child may be important in other respects too, where the performance of a child is likely to be affected by how comfortable he or she
feels in the situation. Children are likely to feel more relaxed when being tested by a
known researcher and this is likely to have implications for both the test results achieved,
and the ethical considerations of asking children to be involved in such research. This is
not the only sphere in which the advantage of the independence of the researcher has
been called into question (see for example Oakley, 1981). In the context of social
research, there is a balance to be struck between the need for objectivity and the reality of
the interaction between researcher and subject. It is certainly arguable that a child would
be more likely to be honest about negative feelings toward school or another adult, if that
child felt comfortable that the person they are speaking to was a friend. Unlike the earlier
discussion about the possibility of executing idealised research designs in applied settings,
this is a question of the appropriateness of design to particular settings. Educational
studies are not always explicit about whether observers are blind to grouping. In this case,
a particular decision was made about the relative importance of these two conflicting
needs; that it was more important that the researchers develop a relationship with the
children and school staff at the risk of surrendering objectivity.

5. Statistical Regression — "operating where groups have been selected on the basis of their
extreme scores" (Campbell & Stanley, 1966)

Simply put, this is the observation that, by chance, in repeating a test there is likely to be
a regression towards the mean, which may result in the appearance of change. By
assigning a control group derived from the same population sample as the intervention
group this potential confounding factor is accounted for, in that the same factor will be
acting equally on each group. Regression is more probable when samples are chosen
from the extreme ends of the range.
Biases resulting in differential selection

As discussed at some length by Campbell and Stanley, randomisation is the preferred method for ensuring that no biases exist in the selection for each group. This process has been particularly unpopular in educational research, where matching has been preferred. In fact, matching of control and intervention groups does not deliver the hoped-for lack of bias and random assignment of group was chosen for this trial.

However, as discussed in 3.4 and 5.2 there was some deviation from this strategy in Years 2 and 3 of the trial. Due to the wishes of the school management team, six children were withdrawn from the randomisation process and assigned to an intervention group. While this decision might be seen as jeopardising the validity of the groups as a whole, it should be remembered that it allowed the trial as a whole to go ahead. Had this pragmatic compromise not been made it is unlikely that the randomisation of the remaining 52 children would have been possible. In a larger trial, it might have been advisable to remove these children from the data set because of their non-randomisation. This was not done because of the small total sample size in the study. The impact that this decision may have had is impossible to know, although we do know that the two groups remained equivalent on all measures of ability despite the changes. Whilst it is important to bear this factor in mind when considering the results of this trial, it is also important to recognise the contribution that this flexibility made to the possibility of completing the work.

Experimental Mortality — “differential loss of respondents from the comparison groups” (Campbell & Stanley, 1966)

In many studies, differential attrition between groups can be very problematic. The dropout rates of those receiving an intervention they find unpleasant, or of those not perceiving any relief of their problem, can severely affect outcome. Those with most
severe problems might be most likely to drop out of control groups, or those for whom intervention is not effective to drop out of intervention groups. This problem can be dealt with by including the outcome measures of all those in either group, regardless of whether the intervention was complete, or by assigning an outcome score to those lost to the trial entirely. This was not a problem in this trial, where all but one child present at the outset of the trial were present at the end of trial period.

8. Selection-maturation interaction

This question of the interaction between selection for groups and maturational change is only relevant for multiple treatment designs and not applicable here.

8.2.2 Factors Jeopardising External Validity

Before embarking on a consideration of the external validity of this study it is worth remembering that it is probably the case that inductive reasoning is never fully justified. Philosophically speaking the notion of generalising from one case to another based on their similarity was rejected by David Hume in the 18th Century. However, whilst it cannot be proved that a causal link in one scenario will be valid in another, we can examine how likely it is that the relationship will be valid in both cases. This is interpreted as understanding how unique the research scenario is. If A causes B the in scenario X, A is likely to cause B in scenario Y if X and Y are similar, and there are no factors unique to X that might have caused B. For example, findings from a US study might be thought to be less relevant to the UK, since US-ness, rather than A, might have caused outcome B. The consideration of factors affecting “typicality” within study design is external validity. The criteria for external validity set out by Campbell and Stanley are as follows;
1. Reactive or Interaction effect of testing — "pretest might increase or decrease an individual's sensitivity to the experimental variable."

This interaction might make the findings not relevant to an un-tested population. The use of a control group does not demonstrate whether this is true; we could only assess it by using a post-test only group. In this instance, this was not possible, since change in skill levels was the key outcome rather than absolute post-test score. It is hard to imagine how this particular factor might operate in this case, especially considering that the abilities tested were ones which would be routinely tested within school, albeit in a less formalised fashion. It seems unlikely that there would be an interactive effect of testing in this case.

2. Interaction of selection bias and experimental variable

Are there factors operating at the level of selection that would, in interaction with the effects of the intervention, change the outcome so that the results are unlikely to be replicated in other groups? Are the conditions operating or individuals selected for this trial unlikely to be represented elsewhere such that the effects of the intervention such that results are unlikely to generalise?

Children selected for this trial were those with poorest levels of literacy (as assessed by their teachers). Was this group unusually bad at reading and writing, and are other children of similar ages and ability likely to be found elsewhere? Data from other sources would suggest that, while not typical of general ability levels in literacy, children with reading problems are relatively common, given that currently 25% of children do not achieve level 4 in KS2 SATs (DfES 2002 figures). We can also note that estimates of children with special educational needs has assessed in 1978 as 18% (Gipps, Gross & Goldstein, 1987) and current estimates are 22% (Audit Commission, 2002). This represents a stable and significant number of children who difficulties at school, although these problems are across a wide spectrum. There is difficulty interpreting these figures since definition of SEN varies between schools and Local Education
Authorities (Dickins, 2003) and severity of difficulties will vary greatly. However, they suggest that this group is not atypical.

We turn to considering, then, whether there might be an interaction between selection and intervention that would make this group atypical. The children in this sample were chosen by teachers rather than by standardised selection, and the question must be asked as to whether either of these two possible methods is likely to produce a group biased in some way. In order to do this it is necessary to examine the sample actually selected for inclusion, and compare these to a theoretical sample that might have been chosen by baseline measures.

The method of selection used here was that the specialist reading teacher, in consultation with class teachers, undertook the selection of children for the intervention groups. These selections can be compared to recommendations derived from baseline assessments. The extent to which these two methods of selection agree and any differences in outcome between these groups can inform a debate about their relative merits. Three key markers were chosen from the data set. These were NARA accuracy, NARA comprehension and the non-word reading test. These were chosen as the most accurate assessments of reading and decoding. The teaching time allocated allowed for a maximum of 22 children per year group to be included in the intervention groups and these were to be selected as the poorest performers in that year group. As a comparison, an alternative group of children whose scores fell into the bottom 22 in the year group on at least two of the three markers was derived. These groups were compared to those who were actually chosen for the intervention in each cohort.

Seventy-two children were chosen using either selection method (59 by teacher selection, 63 by standardised selection). Of these, the majority (65.3%) were selected by both methods. Of the remainder, nine children (12.5%) were chosen only by teacher selection and 13 (18.1%) only by standardised selection. A further three children were initially chosen both by teacher and standardised selection, but did not go on to receive the intervention. This means that either method seems to capture, largely, the same group of children. A comparison of the characteristics of children selected by only
One method and their final group allocation is summarised in Table 8.1 below. Each child was assigned an anonymous identification number at the outset of the project that is given alongside its gender and baseline scores. Where baseline score was low enough to qualify for allocation to intervention by standardised measure in their year group, the score is shown in bold.

**Source of allocation to group and final outcome — children for whom methods disagree**

<table>
<thead>
<tr>
<th>Selection method</th>
<th>Gender</th>
<th>ID No.</th>
<th>Accuracy Score*</th>
<th>Comprehension score*</th>
<th>Nonword reading score*</th>
<th>Final group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher</td>
<td>Male</td>
<td>29</td>
<td>-30</td>
<td>-8</td>
<td>97</td>
<td>1st intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Female</td>
<td>30</td>
<td>-16</td>
<td>-15</td>
<td>94</td>
<td>Moved from 2nd to no intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Female</td>
<td>44</td>
<td>-21</td>
<td>-7</td>
<td>98</td>
<td>1st intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Female</td>
<td>46</td>
<td>-14</td>
<td>-14</td>
<td>80</td>
<td>2nd intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Male</td>
<td>65</td>
<td>-15</td>
<td>-16</td>
<td>105</td>
<td>2nd intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Female</td>
<td>142</td>
<td>-20</td>
<td>-25</td>
<td>95</td>
<td>1st intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Male</td>
<td>143</td>
<td>-21</td>
<td>-20</td>
<td>90</td>
<td>Moved from 2nd to no intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td>Female</td>
<td>234</td>
<td>-11</td>
<td>-20</td>
<td>104</td>
<td>2nd intervention</td>
</tr>
<tr>
<td>Teacher</td>
<td></td>
<td>216</td>
<td>-16</td>
<td>-25</td>
<td>87</td>
<td>Moved from 2nd to no intervention</td>
</tr>
<tr>
<td>Teacher/Stand</td>
<td>Male</td>
<td>124</td>
<td>-26</td>
<td>-18</td>
<td>91</td>
<td>Moved from 2nd to no intervention</td>
</tr>
<tr>
<td>Teacher/Stand</td>
<td>Female</td>
<td>135</td>
<td>-30</td>
<td>-32</td>
<td>92</td>
<td>Moved from 2nd to no intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>7</td>
<td>-17</td>
<td>-17</td>
<td>97</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>8</td>
<td>-25</td>
<td>-23</td>
<td>92</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>47</td>
<td>-3</td>
<td>-29</td>
<td>97</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>67</td>
<td>-23</td>
<td>-25</td>
<td>100</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>170</td>
<td>-24</td>
<td>-25</td>
<td>93</td>
<td>Moved to 2nd intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>171</td>
<td>-31</td>
<td>-32</td>
<td>92</td>
<td>Moved to 2nd intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>206</td>
<td>-19</td>
<td>-21</td>
<td>97</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>209</td>
<td>-13</td>
<td>-24</td>
<td>94</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>221</td>
<td>-26</td>
<td>-31</td>
<td>94</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>250</td>
<td>-12</td>
<td>-1</td>
<td>89</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Male</td>
<td>253</td>
<td>-9</td>
<td>-22</td>
<td>95</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>264</td>
<td>-21</td>
<td>-35</td>
<td>105</td>
<td>No intervention</td>
</tr>
<tr>
<td>Stand. Assess</td>
<td>Female</td>
<td>265</td>
<td>-26</td>
<td>-25</td>
<td>94</td>
<td>No intervention</td>
</tr>
</tbody>
</table>

*All scores falling into “band” (i.e. lowest 22 scores in year group) are shown in bold.*

Table 8.1: Comparison of Selection Methods
Firstly we can note that, as highlighted in Chapter 6, whilst all those identified by the teacher for 1st intervention received it, three out of five selected for the 2nd intervention by teachers and not by standardised assessments did not go on to receive intervention at time two, plus three initially identified by both methods. Two children who were selected by standardised selection only were moved into second intervention group. Only one other child was moved into the second intervention group, and this child was a late starter not present at the beginning of the year. Comparing the scores of those identified by standardised assessments alone they do appear to be lower than those for children selected by teacher selection alone. We might fairly ask then if there are criteria other than performance that teachers are using to assess need. This might either be another skill that the assessment tests were not sensitive to, or other criteria not made explicit by the teacher assessment method. For example, a child that asks for help more often may be perceived differently than a child with similar skill levels that does not. The more vocal child may be either seen as needing more help, or alternatively as able to garner support independently and therefore less needy of additional teacher support. It may also be the case that teachers interpret the importance of difficulties in different areas differently. So rather than placing equal weight on all reading skills, teachers may be particularly sensitive to certain markers of skill. To examine whether this was the case the proportions of children identified by each individual measure were compared. This information is summarised in Figure 8:1 below.
Where children were performing poorly across all three measures, standardised assessment and teacher selection agreed. In contrast, the teacher seemed to be less sensitive to poor reading comprehension or nonword reading accuracy. Other factors may have influenced teacher selection, on which data is not available such as behaviour in class, parental factors or performance in other subject areas. This difference between standardised and teacher assessments provides limited evidence for the sort of biases that may be operating in this selection procedure. This implies that the selection criteria in this study were not explicit and would therefore be difficult to replicate. This, in turn, may limit the generalisability of these findings to other studies with selection criteria that may appear to be similar.
In order to address the possibility of bias and generalisability we would recommend that selection by standardised measure is preferable in research evaluations of reading interventions. As well as this research-led preference, there is also reason to suppose that standardised selection might have been better for the school. The time spent by class and specialised teacher on selecting, assessing and re-assessing children could have been spent on additional teaching time. Saving time by using standardised assessments would only be a true saving if the standardised assessments were at least as good as teacher selection. The data from this project suggests that this was, in fact, the case and is supported by other workers in the field who have suggested that selection by a “transparent assessment process” is preferable (Hurry, 2000). Among the examples of controlled trials cited in Chapter 2 Section 2.5, which selected children at the individual, rather than class level three used published standardised assessments (Fawcett et al., 2001; Hatcher, Hulme & Ellis, 1994; Nicolson et al., 1999); two in-school standardised assessments (Clipson-Boyles, 2000; Mackay, 1999); two (as in this study) used teacher-based reading recovery assessments (Center et al., 1995; Moore & Wade, 1998); and one used teacher nominations alone (Douglas & Willatts, 1984). The use of teacher-based assessments, as opposed to standardised assessments, is common. As with deviation from randomisation above, allowing teachers to select children for the intervention was an important concession to school management in order to allow the trial as a whole to go ahead. Although the alternative seems preferable, mostly the same group of children was selected by both methods and it is hoped therefore that teacher selection did not have a significant impact on the findings.
3. Reactive effects of experimental arrangements — "preclude generalizations about the effects upon persons... in a nonexperimental setting"

This problem of generalisation to non-experimental settings is a difficult one for social interventions. What it means is that processes shown to be effective in experimental circumstances are not always transferable to real world circumstances. In physiological experimentation, this implies that what is shown in vitro or in animal studies cannot be assumed to transfer to in vivo or human biology. In the case of social interventions this refers to the extent to which interventions are context dependent (e.g. programmes designed for use in a hospital setting transferring to the community) or, more often, social and educational programmes delivered in experimental circumstances tend to be better resourced than routine programmes and thus the findings affect the transferability of demonstration programmes to routine practice. For example, additional resources may be directed to encouraging attendance that would be unlikely to take place in everyday practice. The potential difference between demonstration and routine practice programmes has been addressed in the literature on school-based interventions (for example Stage & Quiroz, 1997; Wilson, Gottfredson & Najaka, 2001) and it is recognised that demonstration programmes consistently produce better results. In dealing with this effect, we are asked to be cautious in extrapolating the effects of demonstration programmes and to ensure that evaluation of routine practice is undertaken. In the study under discussion here, the use of a teacher already in place in the school and the co-operative relationship with the school means that the results are likely to be closer to those achieved had the research trial not been attached to the programme. However, because of the efforts of the school and the research study, the specialist teacher was better resourced than would usually be the case. We do not know how important this might be, but it is certainly true that several children referred to the
"belongings" (files etc.) that they received as part of the study as being a positive element of the teaching programmes.

In this particular study, there was a second element to the consideration of the generalisability of findings, and that is whether within or between school control groups are desirable (Pinnell et al., 1994). One of the difficulties of working within educational settings is that it is very difficult to standardise conditions between schools. Schools differ in their pupil and teacher intake, their teaching and learning cultures, their classroom geographies and so forth, so it is problematic to assume that two schools are necessarily comparable. For these reasons it is sometimes thought preferable to use in-school controls where all such factors are necessarily controlled for within the study. However, if within-school control groups are used, one risks underestimating the effect of the intervention. This is because we can suppose that it is possible that the introduction of the intervention will have generalised effects (see 6.4 on “leakage” or diffusion effects) beyond the intervention groups, and these will contribute to change in the control group limiting the potential for difference between groups. In other words, the likelihood of making type 2 errors (concluding no effect because the change does not lie in areas examined) is increased. This might also limit the extent to which it is possible to conclude that findings will be relevant to other schools. In order to address these issues it is probably preferable to use both between and within school control groups for every trial (e.g. Sylva & Hurry, 1995). However, this is clearly a resource intensive strategy, where each intervention child would be represented by two control children (one for each control group). In this study, such a strategy was not possible. In the absence of this, it was felt that a within school group was preferable since it ensured a better control of known variables.
4. Multiple-treatment interferences — "prior treatments are not usually erasable"

This factor is not relevant to the experimental design in itself because a multi-treatment design was not used. However, since all children in the UK are part of an existing intervention structure (compulsory schooling) there may be interaction with previous and current learning experiences with other teachers. Since most of the children in these classes had been attending the school since nursery, such effects were minimised between children. However, for individual children this "interference" could pose an important confounding variable. Little can be done to avoid the effects of such personal educational history, although a larger sample size would have the potential to minimise the effects of individual differences on whole sample change.

8.2.3 Other Factors Jeopardising Validity - Outcome Assessment

There exists another aspect to validity, not considered within questions of the validity of the study design per se, that is the validity of the outcome measures used. As discussed in Chapter 3 considerable effort was put into deciding on the most appropriate measures to use in outcome assessment, and a repetition of this discussion would not be productive. Suffice to say that the outcome assessments selected are bound to contribute to the findings, in that their strengths or weaknesses will be reflected in the detail available in any area. Overall, the decision to use a wide range of measures would protect against some of these effects. However, in one area there may be merit in discussion of the selection of measures in the light of outcome data, which is in the assessment of attitudes. The lack of appropriate measures of attitude and mood for a group of poor readers was discussed in section 3.6.5 and thus an American measure
(Reading Attitude Scale) was employed along with a straightforward question about enjoyment of reading in the pupil interview and the use of an analogue scale for reading self-concept. It is particularly important to verify the suitability of the reading attitude survey, since one of the findings of this study was that a decrease was seen on this measure in the intervention group (see 5.3.5). The table below combines the data from previous chapters giving mean scores on each measure over time for intervention, control and comparison groups into a single score including all those assessed at each time point. Sample size is given for each measure at each time point. It would be wrong to interpret differences in score over time given here as change, since the sample decreases at each time point because of the limited opportunity for follow-up. Comparisons should be made between measures at each time point, where samples are largely the same.

<table>
<thead>
<tr>
<th>Mean Scores for Attitudinal measures, all participants</th>
<th>Baseline</th>
<th>Immediate Post Intervention</th>
<th>12/18 months Post Intervention</th>
<th>18/24 months Post Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Attitude Survey (20-60)*</td>
<td>Mean (s.d.)</td>
<td>55.61 (11.44)</td>
<td>56.26 (12.72)</td>
<td>53.36 (14.02)</td>
</tr>
<tr>
<td>n</td>
<td>116</td>
<td>115</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Reading Self-Concept Now (0-15.5)*</td>
<td>Mean (s.d.)</td>
<td>9.21 (5.41)</td>
<td>8.91 (4.5)</td>
<td>10.32 (3.89)</td>
</tr>
<tr>
<td>n</td>
<td>110</td>
<td>116</td>
<td>79</td>
<td>38</td>
</tr>
<tr>
<td>Reading Self-Concept Later (0.15.5)*</td>
<td>Mean (s.d.)</td>
<td>13.61 (3.63)</td>
<td>13.61 (3.4)</td>
<td>14.30 (2.54)</td>
</tr>
<tr>
<td>n</td>
<td>108</td>
<td>115</td>
<td>78</td>
<td>38</td>
</tr>
<tr>
<td>Reading enjoyment (0-5)*</td>
<td>Mean (s.d.)</td>
<td>3.65 (1.35)</td>
<td>3.69 (1.4)</td>
<td>3.89 (1.33)</td>
</tr>
<tr>
<td>n</td>
<td>115</td>
<td>79</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Self Esteem (0-10)*</td>
<td>Mean (s.d.)</td>
<td>6.4 (1.81)</td>
<td>6.75 (2.06)</td>
<td>7.28 (1.78)</td>
</tr>
<tr>
<td>n</td>
<td>109</td>
<td>112</td>
<td>77</td>
<td>38</td>
</tr>
</tbody>
</table>

Table 8.2 All Attitudinal Measures

Across all measures a generally positive mood is recorded, which would suggest that all these measures are an accurate reflection of mood and attitudes toward books within

* Minimum and maximum scores. In all cases, higher scores imply more positive attitude.
the group. A meta-analysis of research into changing self-concept in students with learning disability was published some time after the beginning of this trial (Elbaum & Vaughn, 2001) in which 64 intervention studies were found which addressed self-concept in this population. This implies that the initial searching of the literature in this case to identify measures of self-concept was probably not sufficient and it may have been possible to identify an existing measure that could have been utilised. However, the overall findings of this meta-analysis would agree with findings from this research, that changes in self-concept are small with a modest, although significant, mean effect size reported (ES=0.19). There are difficulties in accepting the overall conclusion of this meta-analysis since it is not clear what quality criteria were used as inclusion criteria for the studies. Thus, the conclusions from the meta-analysis may have been altered by studies whose findings may be questionable. In the context of this study, the meta-analysis does demonstrate that large changes in individual self-concept are unlikely to be seen over time. The initial preference for measures using alternatives to written questions has not been explored in literature published since the beginning of this study. This consideration remains an important one when considering outcome measures for groups with reading difficulties.

8.2.4 Methodological Critique Conclusions

The strength of conclusions that can be drawn from any research study depend on the methodological rigour employed within the experimental design. When significant effects are seen it is important to ensure that these are not as a result of systematic differences in treatment in the intervention or control group that may have affected change over and above, or instead of, the intervention under consideration. In
contrast, when no (or little) effect is seen it is important to consider type 2 errors and systematic features of the evaluation of the intervention which may have made it less likely to demonstrate statistically significant change that did actually occur.

In order to contribute to the body of research, a study needs have an explicit and rigorous design (whether that design is quantitative or qualitative) and be seen to conform to that design as closely as possible. This trial has demonstrated the possibility of accomplishing a research design that is acceptable to a research audience by co-operation with practitioners. Achieving both of these ambitions is important if we hope to influence practice. Not only must researchers deliver research data of sufficient quality for their conclusions to hold water, it is also necessary for these results to speak to service planners. By conducting research in real-world settings and including discussion of practitioner concerns, it is possible to address both these audiences.

8.3 Discussion part two – Implications of findings

Having established that the study in question was conducted within the bounds of a rigorous research methodology, we can accept the validity of the findings in this case. As discussed in section 8.2.2 above, whether these findings can be extrapolated to other circumstances will be less clear. Few statistically significant effects were observed, and while this does not prove absence of change, it seems unlikely that educationally significant changes were made that were not discovered. If this demonstrates that this intervention, for this group of children was not effective we must consider what implications this finding will have for other groups, interventions and research programmes. If we accept that the merit of the intervention in this study has not been
proven, what can we conclude about the value of this intervention and what recommendation for practice can we make?

8.3.1 Instructional features of the Intervention

Instructional Content

The intervention in this study comprised an intervention that conformed to the key content identified as important in intervention studies (Hurry, 2000) delivered by a skilled teacher. The intervention was typical of an approach likely to be taken by skilled UK teachers using good practice recommendations and their own experience and is an example of similar interventions, as opposed to a prescriptive programme. The backbone of the intervention was techniques derived from Reading Recovery (RR) and, although this intervention was not RR, a consideration of some criticisms of RR is necessary. There is a particular need to examine the literature considering the effectiveness of adapting RR (a one-to-one intervention) to its use in small groups (as used here). There is evidence that “standard” reading recovery is particularly ineffective at developing phonological processing skills in children (Chapman, Tunmer, & Prochnow, 2001; Iverson & Tunmer, 1993), and thus might be less effective at changing phonological skills or in achieving rapid progress than direct instruction in phonics. The work of Iverson and Tunmer contrasts standard and modified (with additional explicit phonological teaching) RR within the context of whole language classrooms. This study took place in the context of the NLS where direct phonological training should be provided (although not necessarily for this age group), and additional phonological work was provided where the teacher felt it was appropriate. This may suggest that the findings of Iverson & Tunmer are less applicable here, indeed findings that phonological processing skill predicted response to RR did not hold in this sample. An additional analysis (not reported here) found that none of the phonological measures used (phonological
manipulation, nonword decoding, rhyme and alliterative production) was correlated with outcome across the whole intervention sample.

The intervention was designed to be responsive to the needs of the children rather than following a clearly prescribed programme. This is thought to be good teaching practice, but it makes clear conclusions about the strength of the particular intervention difficult to make. The absence of a definitive inventory of topics and teaching techniques included results in a lack of evidence for the adequacy of the programme in any one area (such as explicit phonological training). An alternative strategy would have been to record the content of a sample of the lessons using close observation techniques. This approach would not have provided an exhaustive list of lesson content, but would have enabled a reasonable assessment of content. Whilst this approach would be labour intensive, a future study in this area would benefit from the inclusion of some such data.

Intervention programmes assessed in research more typically assess demonstration programmes introduced for researching their effects (for examples see Chapman, Tunmer & Prochnow, 2001; Hatcher, 2000; Iverson & Tunmer, 1993; Stuart, 1993; Stuart, 1999; Torgesen et al., 2001). A defining feature of this programme was that it was a standard practice programme in a real world context. Findings in other areas of work have shown that demonstration programmes typically show larger effects than standard practice programmes (Wilson, Lipsey & Derzon, 2002). Thus, it was felt to be important that this programme was closer to standard practice than to demonstration. The lack of conclusions concerning the content of the programme can be balanced with the evidence for the relevance of the intervention in a real world context. The findings in this case were not that effects were smaller than that seen in similar demonstration
programmes, but that very little effect was shown at all. The context in which an intervention takes place is crucially important, and this study would support the conclusion that standard practice programmes similar to this one are less successful than their demonstration counterparts might suggest.

620 Instructional Context - Teacher effects
The relationship between teachers and pupils is an important one. Therefore, in addition to the consideration of this intervention as an example of everyday good practice, a focus on an individual intervention may have the potential to build on this relationship. Experimentally, a relationship between the attachment relationship between pupils and teachers and adjustment to school has been shown (Pianta, 1994; Pianta, 1997; Pianta, Steinberg & Rollins, 1995). Indeed, it is possible that some of the difficulties that boys experience at school may be predicted by early temperament and behaviour problems, in turn affecting school achievement (Sanson, Prior & Smart, 1996). The supposed impact of a single inspirational teacher is one that has much currency in the popular media (for example the weekly column in the Times Educational Supplement “My Best Teacher”) and has been prominent for many years (MacKaness, 1928). In this context, the importance of the relationship between children and a particular teacher could be seen as crucial. The assessment of the impact of individual teachers may contribute to the research literature. A comparison of the effectiveness of different teachers trained on and delivering the same material would be interesting, although demanding the delivery of a standard programme may interfere with the teacher attributes that a study would hope to assess. An alternative approach would be to pursue the relationship between the characteristics of student/teacher relationship and school outcomes. This approach has been adopted by some (e.g. Pianta, 1996) and shows some potential for predicting school adaptation.
Instructional context — Group teaching

There is some discussion as to the benefits of teaching children alone or in small groups (see 2.4.4). An analysis of the difference in progress between the children in the group or individual lessons could provide a definitive answer to whether either group received a benefit associated with teaching context. Such an analysis was not possible in this case. Firstly, because the children allocated to group and individual teaching were not allocated randomly, but were selected for particular characteristics (e.g. behaviour, stage of reading) and are thus not equivalent comparisons. Secondly, the size of the sample does not allow for a statistical analysis. Whilst it is tempting to compare the progress of the individually and group taught children in this study by giving change scores, it would be inadvisable to do so since the bias in the selection and the small sample size means that confidence could not be placed in the validity of the findings.

An improvement of the trial design would have been to allow all children to receive group teaching. Random assignment to group or individual teaching would have required larger numbers in the sample, which was not possible within the constraints of this project. In fact direct comparisons between group and individual instruction are uncommon; in a meta-analysis of the effect of individual training programmes only two studies were found which made such a comparison, both of which were unpublished US dissertations (Elbaum et al., 2000). It is unlikely that a large body of literature lies outside that found in this review, and was not found within the review of literature undertaken herein. We cannot yet make policy recommendations concerning this, and further research in this area would be recommended.
Instructional Context — Intensiveness of Intervention

The intensity of this intervention can be compared to other interventions to consider whether the intervention in this case did not last long enough, or did not occur frequently enough to have significant effects on children in the intervention group. Certainly, compared to RR the intensity of this intervention was acceptable. In their study comparing standard RR to an adapted RR regime, Iverson & Tunmer (1993) reported an average number of 57.31 (s.d.=11.22) lessons attended. In this study students attended an average of 50 (s.d.=6.64) lessons. These numbers appear similar, but most children in this study received group instruction. One could argue that teaching in groups of four is a reduction in teaching time per child, and therefore the children in this intervention received less teaching time on average. Alternatively, we can examine other, successful group teaching programmes, such as reported by Nicolson et al. (1999). In their study students received two half-hour lessons per week for 10 weeks, meaning a maximum of 40 lessons per student (allowing for absences fewer would actually be attended, but mean attendance is not reported). The study under discussion in this thesis provided more lessons, more intensively, over a similar period. We cannot prove that additional lessons would not have improved performance in the study considered in this thesis, but comparison with these other studies suggests that it is unlikely that the intensity or duration of the intervention was responsible for its lack of success.

8.3.2 When is intervention appropriate?
Having considered particular features of this intervention that may have contributed to the success or failure of this programme to achieve change for children, we also need to consider what we can learn about the value of intervention in general. What is clear is that the children in this group constitute a group where change is hard to achieve. This may be due to the intractable nature of reading difficulties, or because many children in this sample had complex and multi-faceted problems, of which difficulty with reading is only one. The link between what has been termed “social exclusion” and poor educational outcomes is well established (Schools Plus Action Team, 1999; Smith & Noble, 1995). The utility of any intervention focussing on one problem may be questionable. The sophistication of the response should mirror the complexity of the problems. For many of the children in this sample, and many more across the country, difficulties with reading and writing are symptoms of their difficult lives as much as they are a cause of them. This may be a bi-directional relationship where global problems cause specific problems that in turn exacerbate global problems. In order to address such cyclical difficulties it is likely that sophisticated responses are needed. Policies such as the introduction of Sure Start may address some of these issues in the longer term, although the efficacy of delivery and outcome of this programme is currently under question. Whilst long-term solutions are needed, literacy is undoubtedly a key skill that can enable change for individuals.

8.3.3 What criteria should be applied to selection for intervention?

This intervention, in line with RR and others, chose to address the needs of those children with the poorest performance in the school/class. It is likely that this decision may have contributed to the lack of effect since change was hardest to achieve in this
group. The question arising is whether intervention should be targeted at those whose need is greatest, or those who are most likely to make progress. These two conditions are not mutually exclusive, but in fact often constitute different groups. It could be argued that making a greater change for a smaller number of children is more important for their long term gain, but this inevitably implies that fewer children receive greater resources. Research cannot contribute to such a discussion of ethics, but can illuminate the choices made. In this case it seems that targeting a smaller number of children intensively did not seem to achieve the desired changes, and a programme benefiting more children might have been preferable. We can also examine some of the practical decisions made in running the project. In the case of selection, this study would confirm Hurry's (2000) preference for a "transparent assessment process." When the selections made by teachers are examined in detail as above in section 8.2.2, no advantage to teacher selection was apparent. Instead, the accuracy of standardised assessments seemed at least as good, and used far less teacher time to administer.

8.3.4 What are appropriate goals for intervention?

In the short term, the goal of this intervention was to significantly improve the reading skills of children in intervention groups relative to the progress of other poor readers with reading difficulties. In the longer term it was hoped that the intervention would enable some "catch up" with class peers – that is to close the skill gap between themselves and peers. Standard RR explicitly aims to raise the scores on reading measures of individual children to the class average by the point of discontinuation. Thus, progress in RR has been discussed in terms of full or partial "recovery" (that is achievement of class average) (Center, 1995). Center considers those to have reached or exceeded class mean to be fully
recovered, and those "near" class mean or with differential performance to be partially recovered. We could then consider a NARA score no more than one standard deviation below class mean (excluding those in intervention groups) to represent those in full or partial recovery. Over the course of the study three children passed this point for NARA accuracy, eight for NARA comprehension and five for nonword reading. Thus, few children were "recovered" by the programme. Additionally two, children in the intervention group were already above this point at the outset of the project on NARA accuracy, 12 for comprehension and 12 for nonword reading. Within this project, we instead asked whether the difference in score between those in the intervention groups and their peers was no longer statistically significant. Which of these measures of "gain" is reasonable and desirable as an outcome for intervention? Few studies ask that children achieve the "average" achievement demanded of RR, but the rationale behind this benchmark is important; that this is the point at which children might be expected to be able to effectively access class teaching. However, in order to "catch up" in this way their progress must be faster than that of their peers. If the children in these groups were likely to make slower progress than their peers without additional support, then achieving the same rate of progress would represent progress. The only way that this progress could be seen is in relation to progress in a control group.

An alternative strategy, explored in this thesis, is that the achievement desired should not be against average performance (whether nationally or within peers), but rather bottom-line, such as functional literacy. In this case, an increase in the number of children achieving the bottom-line marker would be seen as a success. Experimentally the measure that must be desired is to show increase relative to controls but in practice, this might not be desirable. A measure that refers to the numbers of children achieving a "successful" outcome is probably preferable in practice. The desired outcome of a
reading intervention is unlikely to be “improvement” which does not necessarily imply an educationally significant change, but rather “successful” reading or writing. The number of children achieving level 4 in their KS2 SATs is one prominent measure of such “success”. Reducing the outcome of intervention to a dichotomous measure of success will also collapse data and is thus likely to lose data. It would seem preferable that progress relative to a control group is used, but that this study would have benefited from setting an additional target performance against which individual performance could have been measured. In the long-term the clear preference for outcome should be one with educational significance.

8.4 Conclusions

The aim of this study was to determine whether change could be achieved in measures of reading in a group of Year 3 children with existing difficulties with literacy. In addition we asked whether such an intervention would have any impact on measures of mood or attitude to books and whether other variables emerged (such as delaying intervention) which affected outcome. In order to explore these effects measures were taken across a wide range of outcomes. These were; reading skills, writing skills, phonological ability, teacher assessments of achievement and school adjustment, school attendance, attitude toward books and school, self-esteem, awareness of print, reading strategies known, and home reading habits. The findings of the study were that little change was observed in the literacy skills of the intervention groups in the short term, and no evidence of long-term “catch up” growth was seen. Although the children enjoyed the intervention, the only significant change in measures of mood was a decrease in enjoyment of reading. None of the variables tested as co-variates predicted intervention outcome. It is
important to recognise that no harm was seen to result from this intervention, which cannot be assumed. It is also important that the intervention was popular with children, parents and teachers. However, the lack of impact demonstrated is critical when considering the lessons to be taken forward.

The importance of reading in the current policy climate continues. The number of children passing benchmarks at each level of schooling has increased (DfES, 2002). For primary school aged children the targets set are:

- "Raise standards in English and maths so that by 2004 85% of 11 year olds achieve level 4 or above and 35% achieve level 5 or above with level of performance sustained to 2006.

- Ensure that by 2006, the number of schools in which fewer than 65% of 11-year-olds achieve level 4 or above in English and maths is significantly reduced.

- Reduce to zero the number of local education authorities where fewer than 78% of 11-year-olds achieve level 4 or above in 2004, thus narrowing the attainment gap.” Objective one – Give children an excellent start in education so that they have a better foundation for future learning.

Objective 2 – Enable all young people to develop and to equip themselves with the skills, knowledge and personal qualities needed for life and work.”

In order to achieve this strong policy recommendation it is necessary to improve appreciably the skills of those children performing poorly. The role for applied research is to contribute to the processes by which such change is achievable, but also to contribute to a discussion of what is achievable. The consequences of this study could be seen to imply that requiring the level 4 benchmark for all children is not achievable. The focus needs to be on what can be achieved, and what the key characteristics of intervention are. Interpreting the results of studies with non-significant results is difficult. Given the importance of basic literacy skills in the modern world, it is important to argue that every attempt to improve literacy should be taken. The focus for research should be to systematically evaluate interventions such that best opportunities can be created to meet this need.
References


Adams, M. J. 1990, Beginning to read: Thinking and learning about print The MIT Press, Cambridge, MA, USA.


Barris, M., Ellis, S., Kelly, C., O'Sullivan, O., & Stierer, B. 1996, Using the Primary Language Record Reading Scales: Establishing Shared Standards Centre for Language in Primary Education/Southwark Council, London.


Battle, J. 1981, Culture-Free Self Esteem Inventory for Children & Adults Special Child Publication, Seattle, WA.


Brand, V. 1987, Remedial Spelling Egon.


Brooks, G. 2002, What Works For Children with Literacy Difficulties? The Effectiveness of Early Intervention Schemes., 2 Edn, DFES.


Clay, M. M. 1993, Reading Recovery Heinemann, Auckland, NZ.


DES 1990, The Teaching and Learning of Reading in Primary Schools, HMI, London.


http://www.economist.com/printedition/displayStory.cfm?Story_ID=1010802


Hurry, J. 2000, Intervention strategies to support pupils with difficulties in literacy during key stage 1. Review of research, Qualifications and Curriculum Authority, London.


Literacy Trust. 1999a, Children's Reading Habits.

Literacy Trust. 1999b, Reading habits update.

Literacy Trust. 2000, Attitudes to reading.

Literacy Trust 2001, Literacy and Prisoners.


Pianta, R. C. 1996, Manual and scoring guide for the student-teacher relationship scale, University of Virginia, Charlottsville, USA.


QCA 2000, Qualifications and Curriculum Authority, London.


Robson, C. 1993, Real world research Blackwell Publishers Ltd, Oxford UK; Malden MA, USA.


Ref Type: Personal Communication


Turner, M. Sponsored Reading Failure. 1990b. Warlington, Surrey, IPSET Education Unit.


difficult-to-remediate and readily remediated poor readers: More evidence against
the IQ-achievement discrepancy definition of reading disability", Journal of

Wade, B. & Moore, M. 1998, "Attitudes to Reading: a longitudinal study of the
33, pp. 95-106.

Wagg, S. 1996, "'Don't try to understand them': Politics, Childhood and the New
Education Market," in Thatcher's Children? Politics, Childhood and Society in the

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A.,
Burgess, S. R., Donahue, J., & Garon, T. 1997, "Changing Relations between
Phonological Processing Abilities and Word-Level Reading as Children develop
from Beginning to Skilled Readers: A 5-Year Longitudinal Study", Developmental

Reading Attitudes Along with Measures of Reading Achievement and Scholastic

17, no. 3, pp. 247-272.

intervention programs on aggressive and disruptive behaviour: A meta-analysis",


Appendix 1—Teaching Methods

To explain the content of literacy lessons the reading teacher supplied the following description, which is given verbatim.

1.1 Teaching framework

The children identified were withdrawn daily for individual or group sessions of thirty minutes.

The sessions took place in a quiet well-resourced room. The teaching sessions drew on a range of approaches to the teaching of reading and involved the application of some of the main tenets of Reading Recovery (Marie Clay) but adapted for children at Key Stage 2. Individual sessions followed a Reading Recovery style format.

The lessons were structured, as in Reading Recovery, to incorporate the following activities:

- practising high frequency words;
- re-reading familiar texts;
- taking a running record;
- focusing on word work (onset and rime, etc);
- writing a sentence;
- sequencing a cut up sentence;
- introducing a new book.

The framework for group sessions included:

- practising of high frequency words
- re-reading familiar books and poems
- board work
- word patterns, alphabet work, onset and rime, etc
- introducing a new text (book or poem)
- first reading of the new text

Reading approaches included:

- reading together
- taking turns
- reading in pairs
- reading individually

Once per week members of the group were heard individually so that a running record could be taken.

Books and poems were selected which encouraged the group's response to text and developed an awareness of authors and genre. See sample list of books at the end of this document. Teaching at word and sentence level occurred within the context of books and poems read. Specific elements of spelling and syntax were dealt with "on the run" or, more often, by returning to the text after a re-reading. This afforded opportunities for discussion in a shared context.

The weekly focus on a spelling pattern was reinforced as much as possible in the texts.

1.2 **Key points**

Key points were; setting the ground rules; increasing confidence; developing interest in books; and promoting effective reading strategies.

1.2.1 **Setting the ground rules**

The most important element in the initial stages was to create a positive climate for these fragile readers. A supportive atmosphere in the group was encouraged (taking turns, listening to each other and helping at points of difficulty). Attention was paid to developing the learners' independence and organisational skills through the organisation of resources and the establishment of routines (e.g. remembering folders, organising books and finding what was needed, getting to the sessions on time, etc). A sense of rigour and pace was developed in the sessions to encourage a more active approach to learning.

The baseline assessments were used to inform and plan. These provided the starting point for the teaching. The teaching aimed to consolidate what was known and to build on the knowledge and effective strategies that had already been acquired.
1.2.2 Increasing confidence

The careful selection of, and introduction to books ensured that new books were well within the readers' control and their confidence increased. Re-reading books played an important part in the teaching programme and favourites were quickly established. The children selected books to read at home from their familiar books in order to create more opportunities for success and praise.

The teaching sought to develop reflections about what was read, from direct observations about what was seen to inferences about what the characters were feeling and what the author intended. Some individuals were confident about making observations whilst some were able to model their responses on the others in the group and so develop their own "voice". These were children who would have lacked confidence in and been on the fringes of a larger group.

1.2.3 Developing interest in books

Emphasis was placed on the reflective process of reading. The children were encouraged to make observations about the books they read and relate what they observed to their own experiences and other books. Comments made by parents indicated that this reflection was extended to reading at home. Children were encouraged to develop their own awareness of books using the books introduced and read in the sessions and at home to provide a basis from which they could explore other books. They started to discuss authors and publishers, and to expand their view of reading. The choices they made in the library often reflected what they had read in the sessions and other books by the same authors were discovered. They enjoyed reading "their" books from the sessions to friends and family and would give this as a reason for choosing a particular book. All the comments from parents have highlighted this increased interest in books.

1.2.4 Promoting effective strategies

Effective reading strategies were introduced and developed in the context of reading books and poems. Much discussion took place about the strategies being used in
order to develop the children's own awareness of what worked and why. The children were encouraged to help each other at points of difficulty and to articulate their strategies. Opportunities for work on visual analysis were carefully planned and linked to the texts used. Books and poems were also selected to develop strategies and to provide specific opportunities to focus on print detail. Board work using magnetic letters was used to reinforce patterns in words and to encourage generating words by analogy. Each week there was a spelling pattern focus for the year group which was reinforced in the sessions through reading and writing.

Running records were taken once a week for each child in the group lessons and daily for individuals to monitor strategies being used in context.

1.3 The running record

A running record is similar to a miscue analysis. It records everything the child says and does as he or she reads a new book. In the individual sessions the running record was taken daily on the book introduced in the previous lesson. In the group a running record was taken weekly for each child on a newly introduced book. The running record provided an opportunity for observing patterns that characterised the reading behaviour. The teaching after the reading focused on reinforcing the use of effective strategies and developing emerging strategies. In the group it allowed for discussion on a one to one basis about the book.

Error analysis from the running records provided evidence of monitoring (eg re-running, cross-checking, self-correction and cues used in errors and self-correction). The number of times a word was prompted or told was also recorded. This information was used to plan new books for the following week to match the level of challenge needed by the group.
1.4 Group lesson

1.4.1 Organisation
Each child had a magazine file for his/her books and poem cards. Other resources included a dry wipe A4 boards, clip boards, folders, contact books, note pads and a white board with magnetic letters.

The balance of the lessons varied according to the focus but the framework of the lessons followed the same structure as outlined in the framework above.

1.4.2 Practising of high frequency words
Each week two high frequency words would be highlighted and practised daily. Simultaneous oral spelling, cursive script, quick writing and writing with eyes closed were all used. The words practised were re-visited regularly, sometimes in a dictated sentence.

1.4.3 Re-reading
The re-reading of books and poems provided an opportunity for children to practise reading strategies and to develop fluency and expression in their reading. Turn taking and group reading were both used.

1.4.4 Print focus
This area of the lesson fell into two categories: planned focus (e.g. work on onset and rime, syllables, letter work (vowel sounds and digraphs), phonemes, etc); and problem solving "on the run". The board work focussed on attention to print detail. Magnetic letters were used for onset and rime. Different coloured pens were used to highlight patterns, words within words and morphemes. Picture prompt cards were used, for example, to reinforce vowel sounds, consonant clusters and vowel digraphs.
1.4.5 **Introduction to new texts**

In the first weeks of the teaching new texts were given a full introduction. The group explored the structure of the story and any specific vocabulary together. The children would look through the book and discuss the story in order to enable them to become familiar with the structure and language/vocabulary of the book before reading. As the children became more confident the introductions became shorter, moving towards using the information readily available to them (e.g. on the jacket).

1.5 **Links**

1.5.1 **Parents**

Initial meetings with parents were set up to explain to parents that children were to be coming for literacy support on a regular basis. It gave parents an opportunity to ask questions and voice concerns, and the literacy support was described. Contact books were introduced to establish a link between home and school. Parents and other members of the family as well as the children themselves were encouraged to write in these books. Comments were elicited about the child's response to books and their sense of audience. Follow up meetings were arranged with parents to discuss progress at the end of the teaching programme.

1.5.2 **Class teachers**

A weekly diary was completed for class teachers showing the main focus of the work covered and the accuracy and self-correction rate of the running record with comments. Class teachers provided copies of their literacy planning and where possible links were made with the National Literacy Strategy. Formal meetings were arranged to discuss progress and individual education plans but additional informal discussions also took place.

1.5.3 **SENCO**

There was close co-operation with the Special Educational Needs Co-ordinator ("SENCO") who was involved in discussions about selection and monitoring and the provision, where necessary, for ongoing support.
Examples of Individual Titles Used (Not comprehensive)

Early Stages

King, P. *The Hungry Cat*, Walker (1986)
Dale, P. *Bet You Can't*, Walker (1986)
McKee, D. *Not Now Bernard*, Red Fox (1990)

Building Confidence

McNaughton, C. *Suddenly*, Picture Lions (1996)
Ross, T. *I'm Coming to Get You*, Puffin Books (1986)

Increasing Independence

Selway, M. *Wish You Were Here*, Red Fox (1994)
McNaughton, C. *Oops!*, Collins Picture Lions (1998)


Appendix 2, 3 & 4 - Pupil Interview Schedule and constituent parts

Greet and ensure confidentiality

Open with general question, e.g., do you like school?

1. What do you like doing best at school?

2. What about reading, do you like reading (1-5 where 1 is hates it, 3 is not bothered either way and 5 is loves it)?

3. And writing, do you like writing (1-5)? (Prompt: what do you like to write? What do you write at home?)

4. Do you have a favourite book? What is it?

5. How often do you look at the books you take home from school? (Prompt: Every night, most nights, sometimes, hardly ever, never)

6. When you do read school books at home - do you look through them on your own or with someone else? Who?

7. And at school who do you read books with?

8. Do you like to read with best alone or to someone else - who? Why's that?

9. Can you think of any things that you read that aren't books? (If necessary prompt, how about comics, magazines anything else?)

(Appendix 3, Strategy Use and Naming)

10. When you're reading and you get stuck, what do you do?

   What helps you?

   What can you use to help?
Thank you very much that was very helpful. We've got a couple more things to do. I'm going to ask you some questions about who you feel about things. First there are some sad and happy faces.................

1. very sad  
2. a little sad  
3. not sad  
4. a little happy  
5. very happy

Which face do you feel like today?  
1 2 3 4 5

Which face do you feel like when you are at school?  
1 2 3 4 5

Which face do you feel like when you are at home?  
1 2 3 4 5

Which face do you feel like when you are reading?  
1 2 3 4 5

Which face do you feel like when you are singing?  
1 2 3 4 5

Which face do you feel like when you are drawing?  
1 2 3 4 5

Which face do you feel like when you are doing maths?  
1 2 3 4 5

Which face do you feel like when you are playing?  
1 2 3 4 5
**Battle Self-Esteem Inventory**

Just a few more questions like that, this time I'd like to know whether the sentence I read is just like you or not like you at all.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>I never worry about anything</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most girls and boys play games better than I do</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing pictures or writing stories are fun things to do in school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would change many things about myself</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I often feel that I am no good at all</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am happy most of the time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I often feel like giving up school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can do things as well as other girls and boys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I usually give up when my school work is too hard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other boys and girls like to play with me</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I wish I were younger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>When I Play games with my friends I feel happy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3 – Self Concept (Reading and Favourite Activity)

Almost finished. See this line with the books on it\(^1\)? This is a learning to read line - you start off on the line by the small book when you are just about to start learning to read, but can’t read yet. As you learn more and more about reading you walk along the line until, when you reach the end with the big book, you’ve learnt all there is to know about reading. People sitting at that end of the line are really good readers and can read anything. Who’s the best reader that you can think of, and we’ll write their name down at that end of the line? (Prompt, it could be a grown-up or a child, who’s the best reader you can think of?)

Where on the line do you think you are? (Prompt are you closer to the people who are just starting to learn to read or the people who have already learnt?) Put a cross there.

On the next page is the same line, but this time I want you to imagine that you were in Year 6 and ready to leave Junior school. Can you imagine that? Where do you think you’ll be on the line then? Put a cross there. (Prompt, will you be in the same place, or while you be in a different place? If relevant, ask why.)

Here’s another line like the others, but this time, instead of being a reading line, it’s a favourite thing line. When I asked you before which was your favourite thing to do in school you said ............... so let’s draw/write that in the circles. Now, the people at the beginning of the line are just starting to learn how to........ and when you reach the end you know all there is to know about ........ Where do you think are on the line? Put a cross.

On the next page is you favourite thing for when you’re in year 6. Just like last time imagine that you are in year 6, where will you be on the line?

\(^1\) Only one example is given in following pages for illustration
Favourite Thing - Now