OUTCOME EVALUATION OF THE KIRKLEES PAIRED READING PROJECT

Keith James Topping

thesis submitted for the award of the degree of Doctor of Philosophy

Division of Education
University of Sheffield

March 1990

ABSTRACT

The impact of services to help schools to guide and support parents (and peer tutors) in the use of the Paired Reading technique for improving children's reading was evaluated. Compared to all studies previously reported in the literature taken together, the Kirklees research yielded more than double the volume of pre-post norm referenced outcome data, double the amount of control or comparison group data and triple the amount of follow-up data. Additionally, in Kirklees baseline data were compared with pre-post data to give a time series comparison. Although outcomes on reading tests were slightly less favourable than those selectively reported in the previous literature, the research suggested that an adequate level of effectiveness was possible in a large field study incorporating many schools in one Local Education Authority, representing a significant test of the generalisability and replicability of the technique. The Kirklees study also examined the influence of a number of organisational, demographic and within-subject factors as they related to outcomes. In addition a very large volume of subjective feedback from teacher, parent and child participants was collected in a systematic way, and proved extremely positive. The research also examined the inter-relationship of the various outcome measures deployed with a view to assessing their relative reliability and validity for this purpose. As very few process data were gathered it was not possible to demonstrate what proportion of participants actually utilised the Paired Reading technique in the way they were trained. It is thus difficult to partial out to what extent the positive outcomes are due to the impact of the technique and/or the service delivery support package. However, the technique and service delivery package combined are suggested by the data to be associated with improvements in children's reading skill and attitude to reading. The study provides a number of pointers to the probable success of the Paired Reading approach but conclusive evidence on this must await the findings of properly controlled studies.

ACKNOWLEDGEMENTS

The completion of the research programme involved a great deal of activity by very many people, to all of whom I am indebted.

Firstly, thanks must go to all the children, parents and teachers who made the Project work.

Secondly, thanks to those professionals who put a great deal of hard work into the programme: Marjorie Whiteley, Sue Hayward, Geoff Lindsay, Ken Preston, Linda Hardcastle, Fiona Lunn and Geraldine Joyce.

Thirdly, thanks to others who helped in a less direct way: Richard Atkinson, Lindsay Davies, Robin Hedderly, Peter Holling, Mary Grange, Graham Lawton, Avril McArthur (Bush), Andy Miller, Roger Morgan, Don Radcliffe, Professor Colin Robson, Joyce Townsend and the officers and elected members of the Kirklees LEA.

ACCESS TO RAW DATA

The raw data upon which this study is based are too voluminous usefully to be included here in hard copy. It is anticipated that other researchers may wish to re-analyse the raw data utilising different techniques, and to this end the raw data are to be made available in a variety of electromagnetic media from various different locations. Current information regarding formats and sources of availability may be obtained from the Kirklees Psychological Service, Oastler Centre, 103 New Street, Huddersfield HD1 2UA, United Kingdom, telephone 0484 422133 extension 3915.

Outcome Evaluation of the Kirklees Paired Reading Project

Abstract Page

Acknowledgements

Contents

List of Tables, Figures, Synopses and Appendices

1. The Paired Reading Technique 1
2. Previous Research - Pre-post, Control & Comparative Studies 27
3. Previous Research - Process, Reading Style and Follow-up 61
4. Previous Research - Effect Sizes and Summary 82
5. Method - Service Delivery 89
6. Method - Evaluation 118
7. Descriptive Statistics & Data Analysis Format 143
8. Results - Main Outcomes and Effect Size 164
9. Results - Effect of Organisational Factors 174
10. Results - Effect of Within-Pair Factors 184
11. Results - Subjective Feedback 194
12. Results - Inter-relationship Between Outcome Measures 206
13. Discussion 215
14. Summary and Conclusions 236

References 252

Appendices 290

List of Tables

		Page
1.	Number of Schools, Projects and Data Yielders 1984-7	147a
2.	Frequency of Data Yielding Projects by School Type	148a
3.	Frequency Distribution of Length of Intensive Period of Projects	148b
4.	Frequency Distribution of Length of Inter-test Period of Projects	149a
5.	Frequency Distribution of Length of Baseline Period of Projects	149b
6.	Frequency Distribution of Length of Follow-up Period of Projects	150a
7.	Distribution of Project Participants by Type of Tutor, All Projects	150b
8.	Frequency of Supportive Home Visiting in Data-Yielding Projects	151a
9.	Mean Reading Frequency and Total Reading Time	151b
10.	Frequency of Finding of Statistical Significance, All Analyses	157a
11.	Relationship Between Pre-test and Post-test Reading Ages	159a
12.	Relationship Between Pre-post Reading Gains and Pre-test Scores	159b
13.	Relationship Between Pre-test Scores and Follow-up Gains	160a
14.	Relationship Between Pre-test Scores and Follow-up Ratio Gains	160b
15.	Mean Pre-post and Baseline Gains and Ratio Gains, All Projects	164a
16.	Comparison of Baseline Ratio Gains with Pre-post Ratio Gains	165a
17.	Comparison of Pre-post Gains in Participant and Control Groups	165b
18.	Comparison of Pre-post Ratio Gains in Participant and Control Groups	166a
19.	Pre-post and Short-term Follow-up Gains and Ratio Gains	167a
20.	Pre-post and Long-term Follow-up Gains and Ratio Gains	167b
21.	Pre-post and Follow-up Ratio Gains, by Project According to Tutor Type	167c
22.	Short-term and Long-term Follow-up Gains in Control Group Projects	171a
23.	Short-term and Long-term Follow-up Ratio Gains in Control Group Projects	171b
24.	Pre-post Effect Sizes, Control Group Projects	172a
25.	Pre-post "Effect Sizes", All Projects	173a
26.	Ranking of School Types in Relation to Size of Pre-post Gains	174a
27.	Pre-post Gains as a Function of Type of School	174b
28.	Pre-post Gains as a Function of Time of Year	176a
29.	Relationship of School Size, Project Length and S.E.S.to Gains in Reading	178b
30.	Relationship between Pre-post Gains, SES and Home Visiting Frequency	179a
31.	Ranking of NNV Categories According to Size of Pre-post Gains	179b
32.	Statistical Significance of Differences in Pre-post Gains between NNV Categories	179c
33.	Relationship between Pre-post Gains and Chronological Age and Degree of Retardation	184a
34.	Mean Pre- & Post-test Reading Ages for Male and Female Subjects	185a
35.	Pre-post Gains as a Function of Sex of Tutee	185b
36.	Effect of Sex of Tutee on Difference between Baseline & Pre-post Ratio Gains	185c
37.	Difference between Baseline & Pre-post Ratio Gains for Male & Female Participants	186a
38.	Differences in Pre-post Gains between Participant & Control Groups by Sex of Tutee	186b
39.	Mean Pre- & Post-test Reading Ages for Asian & White Subjects	187a
40.	Pre-post Gains as a Function of Ethnic Origin of Tutee	187b
41.	Differences in Pre-post Gains between Participant & Control Groups by Ethnic Origin	189a
42.	Mean Pre- & Post-test Reading Ages According to Tutor Type	189b
43.	Pre-post Gains as a Function of Type of Tutor	190a
44.	Effect of Tutor Type on Difference between Baseline & Pre-post Ratio Gains	191a
45.	Pre-post Gains of Tutors and Tutees in Peer Tutor Projects	192a
46.	Comparison of Baseline & Pre-post Ratio Gains (Accuracy) for Peer Tutees	192b

continued ....

Page

47. Mean Pre-post Gains for Differing Sex Combinations of Peer Tutoring Pairings 192c
48. Significance of Comparisons, Pre-post Gains for Differing Sex Combinations 193a
49. Aggregate Responses to Items in Parent Questionnaire 194a
50. Aggregate Responses to Items in Teacher Questionnaire 195a
51. Aggregate Responses to Items in Child Questionnaire 197a
52. Aggregate Responses to Items in Peer Tutor Questionnaire 198a
53. Aggregate Responses to Items in Peer Tutee Questionnaire 200a
54. Subjective Feedback from Parent Tutors and Classteachers 200b
55. Subjective Feedback from Parent Tutors and Peer Tutors 201a
56. Subjective Feedback from Parent Tutees and Peer Tutees 201b
57. Subjective Feedback from Age-peer and Cross-age Peer Tutees 203a
58. Subjective Feedback from Age-peer and Cross-age Peer Tutors 205a
59. Relationship between Reading Gains & Frequency of Reading & Time Spent Reading 207a
60. Relationship between Positivity of Feedback from Parents, Teachers & Children 208a
61. Relationship between Reading Gains & Positivity of Parent, Teacher & Child Feedback 209a
62. Relationship between Reading Gains, Reading Time & Frequency, & Positivity of Feedback, According to Reading Test Used 209b
63. Ranking of Reading Tests by Size of Mean Pre-post Gains 211a
64. Comparison of Mean Pre-post Gains by Reading Test Used 213a
65. Significant Differences in Pre-post Gains between Reading Tests 213b
66. Relationship between Reading Gains on Main & Secondary Tests 213c

List of Figures

1. Pre-post Gains (Accuracy) as a Function of Organisational Factors 177a
2. Pre-post Gains (Comprehension) as a Function of Organisational Factors 177b
3. Pre-post Gains (Comprehension) as a Function of School's Experience 178a

List of Synopses

1. Participant and Control Mean Ratio Gains - Literature 45a
2. Mean Ratio Gains - Literature 82a
3. Effect Sizes - Literature 84a
4. Comparison of Mean Ratio Gains - Literature & Kirklees 244a
5. Comparison of Mean Ratio Gains in Control Group Projects 245a
6. Comparison of Distribution of Effect Sizes in Control Group Projects 245b

List of Appendices

1. How To Do It Leaflet for Parents 290
2. Paired Reading - What Are The Advantages? 294
3. Home Reading Record Sheet 296
4. Beyond Paired Reading 297
5. Teacher Feedback Questionnaire: Evaluation Checklist 298
6. Parent Feedback Questionnaire: "What Do You Think?" 299
7. Peer Tutor Feedback Questionnaire: "What Do You Think?" 300
8. Tutee Feedback Questionnaire: "What Was It Like?" 301

CHAPTER 1

THE PAIRED READING TECHNIQUE

Parental participation in the formal education of children is a relatively new phenomenon. At the turn of the century, there was some societal expectation that teachers would impress school values upon the home, but it was not until the 1930's, and the establishment of Parent Teacher Associations, that any dialogue really began. With the advent of the post-war welfare state, professional encouragement of parents in basic areas of health and hygiene was less needed. However, in 1967 the Plowden Report introduced a new concept: a "partnership" between parents and children to directly influence children's educational progress (Wolfendale, 1983).

The previous emphasis on the strong correlation between socio-economic status and levels of attainment, by some teachers fallaciously taken to imply that parents were "to blame" for children's reading failure, began to be replaced by a more pro-active view of the positive contribution parents had made and could make to the educational progress of their children. Research findings began to confirm the "new wisdom". Moon and Wells (1979) reported that attainment in reading at age 7 was strongly predicted by knowledge of literacy on entry to school, and this in turn was predicted by parental interest in literacy and quality of verbal interaction with the child in the pre-school years. Walberg (1984) reported: "school-parent programmes to improve academic conditions in the home have an outstanding record of success in promoting achievement. The 'curriculum of the home' is twice as predictive of learning as is family socio-economic status. In 29 controlled studies, 91% of the comparisons favoured children in such

programmes over non-participant control groups. Although the average effect was twice that of socio-economic status, some programmes had effects 10 times as large" (referring to the USA).

Blatchford et al. (1985) reported on the literacy and numeracy skills of a sample of black and white United Kingdom children just prior to entry into 33 ILEA infant schools. There were few differences in children's test scores attributable to ethnic origin; variation was more attributable to parental teaching of literacy and numeracy at home and mothers' educational achievement. Greaney (1986) reviewed research concerning the influence of parents on the reading skills and habits of their children, and concluded that measures of intellectual interaction between parents in the home environment were more closely related to children's subsequent abilities and achievements than were traditional measures such as socio-economic status. Blatchford and Farquar (1988) followed through the children from their earlier study, and found that there were large differences between schools in the children's subsequent progress, to which home factors were now much less related, but with which teacher expectations were highly correlated. Teacher expectations did not correlate well with the children's skill levels on entry to school.

Discrepancy between the educational impact of the home and that of the school was also highlighted in the earlier Dagenham study (Hewison and Tizard, 1980), which set out to answer the question: what factors characterise successful readers in an area of low socio-economic status? A large number of factors were found to be correlated with reading success in the sample of children studied, including parental attitudes and parental language and child scores on "intelligence" tests. However the factor which

was most strongly associated with subsequent reading success was whether or not the parents reported that they regularly heard the child read at home. Direct parental help with reading conferred a greater advantage on children with respect to reading skills than the stimulation of a generally more favourable language environment in the home. If children did not receive help with their reading at home, their IQ made very little difference to their reading performance. Despite being conducted on a sample of only 300 children in one geographical area, this study aroused great interest in the UK, and led directly to the Haringey Reading Project, which was to have a major impact on public consciousness at the start of the 1980's and be associated with a surge of growth in schemes to involve parents in children's reading development.

In the Haringey Project (Tizard, Schofield and Hewison, 1982) parents of children in selected top infant classes were asked to listen to their children read aloud for a short period, several times a week, from reading material selected and sent home by the child's class teacher. Comparable children in other classes in the two schools taking part in the project acted as comparison groups and the 'experimental' and 'control' classes were chosen at random. Both schools were in socio-economically disadvantaged areas and had reading standards below the national average. The project intervention ran for two years and standardised reading tests were used to assess the performance of experimental and control subjects before, during and after the intervention period. Most experimental parents agreed to help their children, and the majority who helped did so constructively, so far as observations made during home visits could determine.

At the end of the intervention period, experimental children were reading at a substantially higher level than control children, with a marked reduction in the proportion of experimental children performing at the lowest levels on the reading tests. Even more impressively, when the children were followed up three years after the end of the intervention, the proportion of children reading below their chronological age level was very much smaller in the experimental group than in the control group (Hewison, 1987). Throughout the Haringey Project, comparisons were also made with another group of children who were given extra reading tuition in school. During the two years of the experimental period, children receiving extra tuition at school made no greater gains than control group children, and this situation was the same at long-term follow-up.

However, not all projects involving parents in their children's reading development were to yield such positive norm-referenced results. The Belfield Reading Project (Jackson and Hannon, 1981; Hannon and Jackson, 1987) was associated with a good deal of enthusiasm, high take-up rates and much media publicity, but the differences in reading test scores between project participants and children who had not yet participated in a project were very small and failed to reach statistical significance, although one of the two group reading tests used was also used in the Haringey research (Hannon, 1987).

In the United Kingdom, the Haringey project was particularly influential in encouraging more teachers to actively involve parents in their children's reading development. This movement was neither new nor confined to the United Kingdom, however. An evaluation of a tutoring programme in reading for parents was reported by Tannenbaum in the USA in 1968. By 1970,

Niedermeyer was reporting on his Parent-Assisted Learning Programme - "a system to help parents effectively instruct their kindergarten grade children in basic skills at home." The approach was highly structured and emphasised sight vocabulary and phonic skills. An even more highly structured approach, strongly founded on behaviour modification, was reported in the USA also in 1970 by Staats et al. and Ryback and Staats.

Nor were these early developments confined to the United States. A method of tutoring suitable for non-professionals and also heavily influenced by behaviour modification principles was being developed in New Zealand (Fry, 1973), the same country later originating the "Pause, Prompt and Praise" technique, orientated towards oral reading error analysis and discriminatory correction (McNaughton et al., 1980). A substantial early Australian project is reported by Richardson and Brown (1978). One experimental group of children were allocated to a condition wherein their parents received "educational counselling", another group received a daily 40 minute period of specialised help in reading from a remedial teacher, a third group received both treatments and a control group no special help other than ordinary classroom teaching. Results showed that there was no significant difference in any school arising from the method of treatment, although numbers in sub-samples were very small and considerable variation between schools was evident. The authors did however conclude that "parental educational counselling can be as effective as withdrawal remedial classes".

As interest in the field of parental involvement developed, the volume of associated literature grew, and annotated bibliographies and reviews of research began to be produced. In the United States, an early annotated bibliography was produced by Quisenberry (1977). 1981 saw the publication

of selected papers on parental involvement in reading from the seventh International Reading Association World Congress on Reading (Sartain, 1981), and an annotated bibliography on parental involvement from Friedlander. By 1985 Boehnlein and Hager had compiled and published an annotated bibliography on "Children, Parents and Reading" for the International Reading Association, while at around the same time the ERIC Clearinghouse on Reading and Communication Skills was producing a digest on "Resources for Involving Parents in Literacy Development" and one on "Parental Involvement in Reading" (ERIC 1984, 1985).

In the United Kingdom, an annotated bibliography was produced by Hannon et al. in 1985, and the same year saw the production of a "briefing paper" reviewing the field, authored by Robinson for the National Children's Bureau. A substantial collection of papers edited by Topping and Wolfendale was also published in 1985. A section of this book on "Parent Listening" included papers on the Dagenham and Haringey research, the Belfield project, the PACT project in ILEA, community education programmes in Coventry and a number of smaller initiatives where parents had been encouraged to hear their children read but had not been given specific instruction in how to set about this. Another section of the book included a number of papers on the Paired Reading technique, which will be referred to in more detail later. A section on "Behavioural Methods" included papers on the use of a home-based token economy, the Pause Prompt and Praise technique, parental involvement in precision teaching and parental support of Direct Instruction techniques. A fourth section of the book reviewed variations on the methods listed and combinations thereof, and also featured papers on the "Workshop" approach (in which parents come into school regularly to make materials and receive instruction on their use at home and

school), and a project concerned with the teaching of parents to teach reading in order to improve the language skills of their Down's Syndrome children.

The field of parental involvement in reading had thus by the middle of the decade shown a great expansion and proliferation of different strategies, and professionals had begun to debate about which technique was the "best" in which circumstances. Nomenclature was beginning to become rather muddled, with, for instance, some teachers asserting that their school "did" Paired Reading when in fact what they did was unrecognisable by the inventor of the technique, and the term "Shared Reading" acquired a multitude of different meanings, very few of which were precise (Topping, 1986). In the midst of this apparent increase in complexity, many parents fortunately continued doing what they had always done, reading and listening to and with their children at home irrespective of whether any encouragement or guidance was offered by the school.

Hannon et al. (1986) reported on the basis of an intensive study of a small sample of 52 children aged 5-7 years at school and at home that there were considerable similarities between the strategies of parents and teachers when "hearing" children read. Parents were somewhat more likely than teachers to intervene in response to child error, but the responses to errors made by parents and teachers were similar. Parental concern for comprehension was somewhat more restricted than that of teachers, and there were some differences in the pattern of positive feedback, praise and criticism, but it was clear that "no justification exists for considering parents incompetent in hearing their children read" even without specific training. The question "Does It Matter What Parents Do?" was inevitably

subsequently raised in the literature (Loveday and Simmons, 1988). These authors evaluated three different approaches to parental involvement, but as is the case with many such small scale comparative studies, failed to find any statistically significant differences between the three groups on norm-referenced reading tests.

The issue of "is parent training necessary?" and the associated question of "which specific technique is most effective?" continued to be debated by professionals and became entangled with a third major issue, regarding which technique and associated format of service delivery was most cost-effective. The evaluative comparison of the Paired Reading technique with other forms of parental involvement in reading will be discussed further later, but there is as yet very little adequate research concerning the relative impact of different techniques on different children of different ages with different levels of reading ability and different specific difficulties. The hope has been expressed that practising teachers will not assume there is a "best" technique of universal applicability, but familiarise themselves with a number of techniques and build these strategically and sequentially into a "Whole-School Policy" on parental involvement in reading (Topping, 1989; Wolfendale, 1989).

In the second half of the decade, there has been a proliferation of manuals and handbooks for teachers and other professionals, of varying size, complexity and clarity, offering guidance on the establishment of parental involvement in reading projects (Robinson et al. 1984, Wolfendale and Gregory 1985, Fredericks and Taylor 1985, Long 1986, Branston and Provis 1986, Bloom 1987 and Reid 1987). The 1980's have also seen a succession of books of even more variable quality intended to give parents advice about

how to help their children with reading at home (Baker 1980, Cutting 1982, Jeffs 1984, Young and Tyre 1985, Morgan 1986, Root 1988 and Heald and Eustice 1988.) Videos have been produced, largely for teachers to use when working with parents, and a variety of books and schemes purporting to be particularly suitable for parents to use to help their children with reading have appeared in the shops.

Although parental involvement in children's reading has obviously expanded considerably throughout the 1980's, the number of schools operating some form of coherent organised scheme still varies enormously from area to area, often within the same local education authority. Hannon and Cuckle (1984) reported an early survey of a sample of 16 Infant and First schools. It was found that while there was general support for the idea of parental involvement in the teaching of reading, this stopped short of helping parents hear their own children read at home. In the schools studied, at that time comparatively few children regularly took school reading books to read at home. However, Weinberger et al. (1986) reported high rates of take-up and persistence in the Belfield Project in a disadvantaged area. Hancock (1988) reported the results of a survey in the London Borough of Brent, where 55% of primary schools claimed to have structured home-based reading programmes of some sort. All but 3 of the remainder of the schools reported that they "encouraged" children to take their reading books home to read to their parents. Hancock notes that these figures are very similar to those reported in the London Borough of Hackney in 1984, but much less good than those reported in the london Borough of Havering in 1985.

Williams (1987) reported that of 219 primary schools in Shropshire, 106 were using Paired Reading as of December 1986. By December 1989, a very similar

proportion of schools in Kirklees had operated at least one Paired Reading project (51%). In both Shropshire and Kirklees, many more schools had expressed interest in, or the intention of, initiating a parental involvement in reading project. High concentrations at schools operating parental involvement in reading projects can likewise be found in parts of Coventry, Cambridgeshire and Cleveland, but it seems likely that elsewhere in the United Kingdom the incidence of teacher initiatives is very patchy. It is as well that many parents help their children with reading at home anyway. Hewison and Tizard (1980) found in Dagenham that about half of the children aged 7-8 in this disadvantaged area were regularly heard read by their parents, and only about 10% were never helped. This is not dissimilar to the proportions reported in the 1960's in Nottingham by Newson and Newson (1977).

Description of the Paired Reading Technique

What then is the need for a specific structured technique? The Paired Reading technique was designed by Roger Morgan to meet two basic criteria: firstly, general applicability through inherent flexibility and a capacity to adapt to individual and changing reading performance, and secondly, sufficient simplicity to be used effectively by a child's own parents at home with a minimum of professional training and supervision. It first appeared in the literature in the mid-seventies (Morgan, 1976), and a follow-up article by Morgan and Lyon appeared in 1979. Morgan has subsequently written a number of descriptive articles and a book for parents and teachers (1986), together with a recent article reporting further outcome data (Morgan and Gavin, 1988).

Although the technique is fundamentally simple, and much less open to misinterpretation from written information than many other educational

initiatives, it has proved subject to a small amount of modification by practitioners over the years, whether conscious or unconscious. Thus, for instance, Topping (1987) notes that Morgan's own description of the error correction procedures for use in Paired Reading showed some change between his 1976 paper and his 1986 book. Morgan's own most recent description of the Paired Reading technique (in Morgan and Gavin, 1988) runs as follows:

> "tuition is undertaken using a book of the child's choice, suitable for his or her interests and chronological age rather than being restricted to his or her reading age. Sessions begin with parent and child reading simultaneously and aloud, the parent thus providing a continuous prompt or model for the child's reading. When the child is sufficiently confident to read a few words or a passage alone, he or she signals this by knocking on the table. The parent praises this, stops reading with the child and the child continues reading aloud alone. While the child is reading alone correctly, the parent reinforces the correct reading by frequent praise or 'feedback that you are right', using positive comments suitable for the individual child. When the child becomes 'stuck' or makes a mistake while reading alone (this is pointed out by the parent if the child does not realise that a mistake has been made), the parent allows approximately 4 seconds for further attempts. If the child is unable to resolve the problem in this time, the parent supplies the correct word, the child repeats it with the parent, and simultaneous reading is resumed until the child again 'knocks'." (page 201)

This brief recent description is actually very similar to the original 1976 version, and it may be that the extra space available in the 1986 book led the inventor of Paired Reading to over-elaborate the technique.

Subsequent to the first two papers on the Paired Reading technique, the first large-scale application within a local education authority setting was carried out in Derbyshire at the initiative of a group of educational psychologists. It was the Derbyshire group who were instrumental in introducing the Paired Reading technique to Kirklees via an in-service training session prior to the project leader's arrival in 1983. Through another channel, the project leader's own training in the Paired Reading

technique was also a transmission of the "Derbyshire version", and this undoubtedly influenced the way the technique was subsequently disseminated in Kirklees. However, scrutiny of the papers published by the Derbyshire group reveals only the sparsest descriptions of the technique, often omitting features which both Morgan and the current writer consider crucial, such as the four-second pause (Bushell et al. 1982, Robson et al. 1984, Miller et al. 1986, Miller 1986 and Miller 1987).

The basic verbal instructions given to parents and children at Paired Reading training meetings in Kirklees during the period under study will be summarised below, and may be found in further detail in Appendix 1, Paired Reading: How To Do It). During the course of individual training meetings, additional information and advice would be given to the group in response to questions raised by parents, as well as supplementary guidance given to individual pairs during the course of the practice session and subsequent home visits where incorporated, but these idiosyncratic additional subtleties were far too numerous and varied to be incorporated here.

> "Participating children choose their own high motivation books, from school, home, public libraries or elsewhere, which can be above the child's independent readability level. If children become bored with a book, it is their own fault, and they can change it. Paired Reading should be done for a minimum of 5 minutes 5 days each week, but not for more than 15 minutes unless the child insists. More than one member of the family can help, but all must use the same technique. Try to find a place that is quiet and comfortable where both tutor and tutee can see the book easily. Discuss the book before and during the reading at natural pauses, both to show interest and to check on comprehension and prediction. When a child makes an error, the parent merely tells the child what the word says - then the child repeats it after you. Do not make the child struggle or 'break it up' or 'sound it out'. When your child reads well, smile, show you are pleased and praise the child. Particularly praise for: correct reading of difficult words, self-correction and increasing span of correct reading. On text which is difficult for the child, both parent and child read all the words out loud together. Parental speed must be adjusted to the natural speed of the child. The child must read every word.

> Finger-pointing, preferably by the child, can help ensure visual as well as auditory attention. When the child feels confident to read alone, the child makes a non-verbal signal to silence the parent. This can be a nudge, a knock or anything agreed and easy for the pair. The parent stops reading out loud immediately and praises the child for signalling. During reading alone, if the child struggles on a word for more than 5 seconds, or struggles on a word and gets it wrong, the standard correction procedure is applied and the pair revert to reading together."

In practice, parents tended to deal with the situation where the child made an error and carried on reading past the error word as an indication for application of the correction procedure and return to reading together. Morgan's 1986 version emphasises the role of parental fingerpointing and other ancillary prompts here, but these formed no part of the Kirklees version of the instructions, although some parents doubtless adopted this practice spontaneously.

A number of variations on Paired Reading have been described, departing much more substantially from Morgan's original conception. A number of these are briefly reported in Topping (1986). Some of these variants involve repeated reading, by child or parent or both, and another variant intended for beginning readers leaves much more control with the parent. One of the many approaches which have labelled "Shared Reading" incorporates only the reading together aspect of Paired Reading and errors are ignored. Evaluations of these variations have been much less numerous and on a much smaller scale than those of the "original" technique, and none of them have been used in Kirklees to the knowledge of the current writer.

Theoretical Framework for the PR Technique

There is very little new in the Paired Reading technique. It does however combine a number of useful elements of parental practice into a coherent

"package" which is widely applicable and easily transmitted. Group reading of textual passages in synchrony was a more common feature of classrooms some decades ago than it is today. Likewise, speech therapists have for some time used the technique of "shadowing" with stammerers - providing a continuous model and prompt of fluent speech - and indeed it is this application which first gave Morgan the idea for the technique.

As early as 1968 Neville had noted that listening to a reading or recording of a text while following it visually helped increase fluency. A further study (Neville, 1975) of 180 children of normal reading ability showed that the slowest of 3 speeds of simultaneous listening with silent reading resulted in the highest level of comprehension. This "pacing effect" seemed even more marked in a small sample of "remedial" readers, and seemed to be more prominent among boys. In a developmental study of children in grades 2 to 6 in the USA, Wilkinson (1980) noted that the decoding by novice readers was poorer than their understanding of a comparable spoken text, and in skilled readers a similar loss of understanding occurred when accurate recognition was accomplished at a fast pace. However, when skilled readers listened to a text that they could simultaneously read, their understanding was better than when they merely read orally, although this did depend on text content. Johnson (1982) found a gender differential, boys having a very poor recall performance compared to girls on a silent reading task, in contrast with their equal recall performance after listening to and orally reading a message. Horowitz and Samuels (1985) also reported that little difference in <u>listening</u> comprehension is evident between good and poor readers.

It may be that simultaneous reading and listening, as in "Reading Together", frees the struggling reader from a preoccupation with laborious decoding and enables other reading strategies to come into play. If the "limited processing capacity" (Curtis, 1980) of the remedial reader is totally devoted to accurate word recognition or phonic analysis/synthesis, no processing capacity may be left to deploy other strategies, perhaps of a more psycholinguistic nature. A number of studies have explored the development and interaction of different components of reading skill as children moved through the school system. Curtis (1980) assessed efficiency in verbal coding and listening comprehension in skilled and less skilled readers in the second, third and fifth grades. Younger and less skilled readers differed from older skilled readers on both factors. However, as verbal coding speed increased, comprehension skill became the more important predictor of extraction of meaning. Curtis concluded that when verbal coding processes are slow or inefficient, they reduce the amount of attention available for other reading processes, thereby producing comprehension deficits.

Hutson et al. (1980) compared remedial readers aged 8-9 years with average readers of the same age and with younger children aged 6-7 years, in terms of their usage of a psycholinguistic approach versus a decoding approach to reading, and in terms of the strategies the children self-reported for word attack. Remedial readers were less willing to skip over letters in a word, showed a heavy emphasis on phonics and a narrower range of potential strategies than did other children of the same age. Similar findings were reported in the UK by Potter (1982), who was interested in whether good readers made better use of the succeeding context by using a better strategy or whether they did so simply because of their superior knowledge. A sample

of 121 7-8 year old children were found to be better at using the preceding context than the succeeding context, and the better the reader the better was use of the succeeding context. As the knowledge of the readers had been controlled by the structure of the measure used, Potter attributed this finding to a difference in strategies used by better and poorer readers. Potter did not however find that good readers made better use of any obvious self-correction strategy, as self-corrections were not found to be related to reading ability. It was thus unclear whether good readers used a different strategy than poor readers, or whether they merely used a different balance of the same strategies more skilfully and with greater efficiency.

Harding et al. (1985) examined the reading errors of 14 children with reading ages ranging from 5-11 years. At a reading age of about 8 years, expansion in sight vocabulary was associated with largely graphophonic and phonemic strategies, but beyond this there was a change in reading style to a more whole-word approach. The more proficient readers appeared to process larger units of information, and demonstrated the ability to use a wider range of strategies when necessary. However, as Harding et al. (1985) and Lees (1986) note, some previous studies yielded contrary evidence, although comparability between studies was hindered by the use of different age groups of children and different techniques of reading error analysis. Lees (1986) interprets 4 studies as indicating that "poor" readers are as capable of using context to aid word recognition as are "good" readers, when the context is appropriate to their reading age - i.e. if they can't decode it, they can't use it.

Mudd (1987) took this approach further by comparing the reading strategies used by adults and by children in the early stages of learning to read. The 72 adults and 96 children who participated in the study all had reading ages ranging from 7-8 years, although the chronological age of the children ranged from 7-8 years and the chronological age of the adults ranged from 19-45 years. Mudd reported that the novice adult readers showed strategies (and deficiencies therein) which were very similar to those of the less able child readers. Furthermore, although the comprehension of the adults was significantly better than that of the children, the difference was not as great as had been expected considering the presumed superior linguistic and background knowledge of the adults.

There is, then, some fairly strong evidence that weak and novice readers are usually less able, and consequently less willing, to use a wide range of strategies in reading, including the psycholinguistic, and they may tend to depend heavily on a single, often phonic, approach. Paradoxically, there is associated evidence that teachers tend to reinforce this over-dependence, not only by over-teaching analytic decoding skills and encouraging readers to process very small units of information consecutively, but also by interrupting weak readers to give phonic prompts far more often, in proportion to mistakes made, than they do with competent readers. This constant interruption further reduces the contextual clues available to the reader. Types of teacher interruption behaviour also differ as a function of reading ability level (Allington, 1980).

Vellutino and Scanlon (1986) randomly assigned poor and average readers in the second and sixth grades to treatments that simulated three general methods of teaching word identification: a whole-word/meaning-based method,

a phonic method and both combined. The whole-word/meaning-based method fostered a global processing strategy, even in the identification of single nonsense words, while the phonics method fostered an analytic strategy. However, the combined teaching method resulted in the use of both processing strategies, and subjects who received the combined treatment performed better than subjects who received only one or another. It was concluded that the use of only one of these methods of reading instruction to the exclusion of the other may create "processing bias" that could impair the acquisition of fluency and word identification.

The literature on the effect of teacher interruptions during oral reading instruction was reviewed by Shake (1986). It was concluded that teacher interruption during oral reading may have varied effects on the oral reading and comprehension performance of average, remedial and learning disabled students, and that delayed feedback and lower amounts of teacher interruptions were likely to increase students self-monitoring and self-correction, both desirable strategic reading behaviours.

A good deal of other research has implications for the structure of the Paired Reading technique. Tunnel and Jacobs (1989) offered a review of research findings on the effectiveness of reading instruction which was based on "real" books rather than reading schemes. Examination of 40 studies from 1937 through 1971 which compared the "core reading scheme" approach to reading instruction with the "individualised reading" approach resulted in the conclusion that 24 of the studies favoured individualised reading, while only one reported progress on "basal readers" was better, the remaining studies showing no significant difference. A study of 1,149

children in second-grade in 50 classrooms by Eldredge and Butterfield (1986) yielded 20 statistically significant differences, of which 14 favoured the literature-based approach coupled with instruction in de-coding skills, although the "literature-based-only" group was also highly placed. A number of other recent studies favouring the "real" book approach are cited in Tunnel and Jacobs (1989).

The effect on reading behaviour of the introduction of novel materials and modelling by adults was documented some years ago in a study by Haskett and Lenfestey (1974). The introduction of novel books into the classroom increased the reading related behaviour of some children, but adults who modelled reading by reading aloud produced larger and more stable increases in such behaviour. The influence on learning disabled children of teacher modelling was also documented by Smith (1979), but in both these studies sample sizes were small.

The issue of error correction procedures in parental involvement in reading techniques has been dealt with in some detail by Topping (1987). A study by Jenkins and Larson (1979) evaluated 6 different procedures for correcting oral reading errors. Isolated word drill on error words on flash cards subsequent to reading of the continuous text produced the highest word recognition scores at post-test, and by comparison, word supply had a much smaller effect on subsequent word recognition, little greater than that of no correction whatsoever. However, the experimental sample consisted of only 5 learning disabled junior high school students, and the post-test was applied one day after the original reading exercise. More recent studies (summarised in Singh and Singh, 1986) have suggested that word supply may be more effective than word analysis with learning disabled children. Other

studies have demonstrated some effectiveness of "positive practice over-correction" and "delayed attention to errors". Rosenberg (1986) has recently supported the original findings of Jenkins and Larson (1979), reporting that word drill was more effective than word supply or phonic rehearsal with learning disabled children of low intelligence aged 12 - 14 years - but again the sample size was very small (4). In any event, more complex error correction strategies may be fundamentally unsuitable for incorporation in any technique designed for use by non-professionals.

The role of self-image or self-concept as a factor in learning to read has received increased attention in recent years, and a classic study was reported by Butkowsky and Willows (1980). Children of good, average and poor reading ability were assessed on tasks in which success and failure were manipulated. Consistent with predictions, poor readers displayed characteristics indicative of "learned helplessness" and low self-concept of ability. These included significantly lower initial estimates of success, less persistence, greater attribution of failures to lack of ability, greater attribution of successes to factors beyond personal control, and greater reduction in expectancy of success after the experience of failure. The importance of the extent to which the Reading Together element of the Paired Reading technique helps weak readers to "believe" that they "can" read should not be under-estimated.

For many weak readers, low expectancy of success may be associated with dysfunctional levels of anxiety when confronted with a reading task. Sharpley and Rowland (1986) compared the effectiveness of bio-feedback and relaxation training to reduce anxiety with that of remedial teaching to directly act on the cause of stress, together with two control procedures,

with 50 elementary schoolchildren who had been referred for reading disability. Only the remedial teaching group, experiencing direct action on the cause of stress, showed significant improvement in accuracy and speed of reading. The Paired Reading technique reduces stress by giving support within a maximum period of 4-5 seconds, and by giving successful reading practice also acts directly upon the reading deficit.

However, it has been proposed that merely the effect of extra attention in a one-to-one relationship could be sufficient to produce gains in reading skill. This view is supported by the work of Lawrence (1972, 1973, 1975 and 1985), who deployed non-professionals to give "counselling" to reading retarded children along Rogerian lines. Lawrence's early studies claimed effectiveness in raising reading skills by this method, but had many methodological weaknesses and were widely criticised. Lawrence's more recent paper (1985) attempted to rectify these weaknesses and compared counselling with direct instruction and drama therapy and combinations of these. Lawrence (1985) concludes: "children retarded in reading made more gains in reading when remediation of the skills in reading was combined with a therapeutic approach designed to enhance self-esteem." A similar exercise is reported by Murfitt and Thomas (1983), who investigated the effects of peer counselling on the self-concept and reading attainment of secondary age slow-learning pupils. Results were somewhat erratic for primary age counselled pupils, but secondary age counselled pupils made more positive progress in reading attainment and a markedly higher gain in self-concept than the matched control group.

A variety of techniques and methods showing some similarities to Paired Reading have been reported from North America, variously known as "the Lap

Method", "Shadow Reading", "Reading-While-Listening (R.W.L.)", "Duet Reading", the "Neurological Impress Method", "Assisted Reading" and "Prime-O-Tec". An associated technique in the United Kingdom is "ARROW". Only some of these are reported in detail in the literature.

The Neurological Impress Method was invented and developed by Dr R G Heckelman, commencing in the late 1950's. NIM is intended to be a remedial reading method wherein the student and the instructor read aloud together in unison. The instructor leads the reading while a finger is slid along under the words of the sentence being read. The instructor sits a little behind the student and speaks directly into the right ear of the learner. There is no correction procedure and no corrections are made during or after the reading session. The method is intended to be multi-sensory and also furnish the student with a model of correct reading. The method is not intended for use by non-professionals - very little adequate evaluation evidence is presented. Specimen references are Heckleman (1966) and (1986).

Assisted Reading is a kind of supported reading which has distinct similarities to Paired Reading in conception and theory, but which appears to be very various in actual methods used by practitioners. The literature does not make very clear of what the method actually consists (Hoskisson et al. 1974, Hoskisson 1974.) In a more recent paper, Richek and McTague (1988) have taken Assisted Reading to mean "a joint reading of a book by teachers and students", very similar to Holdaway's (1979) method labelled "Shared Reading".

Reading While Listening is a development of NIM, and is a multi-sensory approach to practising reading while listening to taped material and pointing to the text at the same time with a finger or pencil. A major evaluation study of RWL is reported by Schneeberg (1977) and a more recent review of research will be found in Wisner (1988). Although better evaluated than NIM and Assisted Reading, RWL is not designed for non-professional use and requires the availability of expensive materials. Wisner (1988) notes some similarities to Assisted Reading and to the LAP Method described by Moffett and Wagner (1976). A similar method designed for use with adult disabled readers called Prime-O-Tec was developed by Jordan in 1965, who saw it as an adaptation of NIM, and is reported in Meyer (1982). Carbo has described similar work in an article entitled "Teaching Reading With Talking Books" (1978). Other workers in the US have advocated variations on "Repeated Reading" (e.g. Dowhower, 1989), but this has more to do with some of the variations on Paired Reading than with the original form, although the author notes that repeated reading can be assisted or unassisted and refers to Morgan and Lyon (1979).

A parallel method to RWL, NIM and Prime-O-Tec developed in the United Kingdom is ARROW (Lane and Chinn 1986, Lane 1987). ARROW stands for Aural-Read-Respond-Oral-Written and involves young children listening to their own recorded voices as a continuous prompt while reading, writing or responding orally. Again, expensive audio-visual equipment is involved and the method is primarily designed for use by teachers, although it is asserted that volunteer helpers can supervise ARROW and some children can undertake ARROW training on their own.

The original theoretical conception of Paired Reading was totally behavioural, based on the concepts of "participant modelling" and reinforcement. Later workers have felt that the technique has strong psycholinguistic overtones and it seems clear from evaluation results that Paired Reading does not just improve mechanical decoding skills.

However, Reading Together was designed as participant modelling, in which the child receives a model and a continuous prompt for correct reading during his or her own attempt to read the words. Independent Reading aimed to provide the opportunity for positive reinforcement by praise of correct reading responses. Praise for signalling the wish to read independently was introduced to reinforce the selection of Independent Reading, which then gave the opportunity to practise (and be praised for) responses acquired during simultaneous reading, and thus the opportunity for learning to be consolidated by successful and praised use. Free choice of reading material was intended to escape the aversive and inhibiting effects of simplified or special texts, and to increase motivation to read in the relatively unmotivated, through the intrinsic reward of reading something the pupil wanted to read. Avoiding being made to try any word for more than 4 or 5 seconds before being given the answer was a means of limiting anxiety, a known inhibitor of learning when excessive (Morgan, 1985).

Morgan had been much influenced by the earlier work of Rachman (1962) and Staats and his co-workers (Staats, 1973). McKerracher (1967) and Koven and Le Bow (1973) were also strong influences. Morgan later added: "drawing on experience rather than theoretical considerations, it is possible that Paired Reading works as much by lifting previous inhibitions upon learning reading (such as aversive pressure, failure, boredom, anxiety and

uninteresting material), as by introducing anything new." Paired Reading can be construed as "a period of minimised adverse influences on progress and maximised opportunity for, and reinforcement of, success." Morgan also comments: "its neutrality to the child's mode of word attack and concentration upon successful performance effectively selects and strengthens any successful decoding strategies the child might be using, while 'culling out' ineffective strategies" (In Morgan and Gavin, 1988).

As Morgan thus points out, irrespective of the behavioural origins of Paired Reading, the technique fits in well with "Top-Down" models of the reading process, which have gained increasing favour in recent years. In 1981 a small study by Heath indicated that the positive effects of the technique were not solely due to the reinforcement aspect. In 1982 Bushell et al. first noted that the effects of the technique were equally applicable within a conceptualisation of the reading process in which the child's self-esteem was crucial and capable of stimulating various psycholinguistic aspects of the reading process. Failure was eliminated, and tutor and tutee usually learned the technique together from scratch.

> "At the psycholinguistic level, the technique creates an opportunity for semantic and syntactic prediction to occur by reducing the amount of time and attention devoted to difficult words, thus both discouraging a single word decoding strategy from being seen as so important and maintaining a steadier flow of contextual clues. The fact that the parent and child are prepared for dealing easily with all unknown words and know there will be no points of anxiety-provoking decision-making probably also enables psycholinguistic processes to function more effectively" (Bushell et al. 1982, page 11).

Pursuing the same line of argument, Barrett (1987) went so far as to describe Paired Reading as "psycholinguistics in practice". Barrett continued: "Paired Reading satisfies virtually all the psycholinguistic criteria of good practice" specified by Frank Smith (1978), crystallized in a quotation from Richards (1942): "Read it as though it makes sense and perhaps it will."

A summary list of the advantages of the Paired Reading technique, derived from practical experience rather than theory, was often given to parents participating in school-based projects in Kirklees (see Appendix 2). Many advantages were listed, although of course some of these were common to other techniques for parental involvement in reading.

> "In Paired Reading, children pursue their own interests and read for their own purposes - just like adults. They become more in control of the proceedings. Failure is eliminated. The technique is highly flexible, able to accommodate not only to difficulty of text but also to current levels of interest, mood, tiredness, confidence and so on. There is a lot of emphasis on understanding and meaning-getting, and the emphasis on fluency and continuity facilitates this - Paired Reading eliminates stopping and starting and makes it easier for children to make sensible guesses at new words, based on the meaning of the surrounding words. Children are given a perfect example or model of how to read well, avoiding the learning of errors, and during Reading Together children are able to copy expressiveness, pacing, rhythm and other subtle reading skills from the tutor, as well as merely accuracy of word recognition. Not only do the children receive extra parental attention, which may itself be beneficial, they also obtain more sheer practice at reading. Not least, tutors and tutees have a clear, simple structure to follow - so neither is likely to become confused, worried or bad-tempered about reading. Even better than this, children receive a good deal of praise, which is likely to make them feel better about themselves and better about reading."

Objectives of the Current Study

At this point, a brief overview of the issues to be addressed in the following text may serve as a useful advance organiser.

The objectives of the Kirklees study were several:-

1. To synthesise the outcomes of many projects which showed greater organisational homogeneity than those in the previous literature.

2. To explore whether multiple field replications with a large number of schools in a single Local Education Authority yielded markedly worse outcomes than those selectively reported in the published literature.

3. To add to the pre-existent studies reporting control or comparison group data, to increase confidence in any conclusions regarding effectiveness.

4. To contribute a significant volume of baseline and follow-up data, hitherto very scarce in the literature, with the same purpose.

5. To contribute subjective "consumer satisfaction" data from the main participants (parents, children, teachers), collected systematically from a large number of participants using the same method, an important alternative approach to evaluation hitherto scarce in the literature.

6. To examine the influence on outcomes of a number of organisational, demographic and within-subject factors (e.g. type of tutor, degree of follow-up support, frequency of reading, length of project, reading test used; socio-economic status of school catchment area; age, gender, race,

and retardation of tutees), only some of which had been considered at all in the previous literature and then in a highly heterogeneous manner.

7. To consider whether the Paired Reading technique and service delivery methods deployed in Kirklees could be considered cost-effective.

In the next 3 chapters (2-4), the *pre-existent literature is reviewed* and summarised. Chapters 5 and 6 give details of what was done in Kirklees and how it was evaluated, respectively. Chapter 6 gives descriptive statistics concerning what was done and reports how the Kirklees data were analysed. Results are in Chapters 8-12, the main norm-referenced outcomes in Chapter 8 and the main subjective outcomes in Chapter 11. Discussion of the results in Chapter 13 considers methodological flaws and threats to validity of any conclusions. Chapter 14 comprises a summary of conclusions, compares the Kirklees data to the aggregate data from previous studies and gives recommendations for further research.

PREVIOUS RESEARCH - PRE-POST, CONTROL AND COMPARATIVE STUDIES

The literature on Paired Reading is substantial, a bibliography in Topping (1988a) listing over 140 items. For the purposes of this review, papers which were descriptive and included no numerical outcome data were ignored. Studies reporting Kirklees data which are subsumed in the current study were likewise not included, unless they had additional interesting features. Studies relating to the use of Paired Reading in Further Education and Adult Literacy, studies of variations on the Paired Reading technique and studies of Paired Reading with specialised groups such as children and adults with severe learning difficulties were all omitted, although relevant references will be found in the bibliography in Topping (1988a). Where two separate papers report the same data in full, reference is usually made to the briefest, most recent and/or most accessible source. Sometimes, it was necessary to search the original source for data of the requisite detail. In some cases, arithmetical errors and inconsistencies were found in the original source, and re-calculation from the raw data where available was necessary. In studies wherein more than one norm-referenced reading test was used, results on the most stable and appropriate test are emphasised in the following discussion.

"Comprehension" scores refer to scores on separate comprehension scales for those tests featuring these (mainly Neale); tests yielding only one score are all subsumed under the "reading accuracy" category. The tests used are discussed in full in a later section. In many studies, the data reported were inadequate for the purposes of the current review. Statistical significance of findings was not always given or calculable, and where supplied was occasionally derived from a statistical procedure

of doubtful validity in the circumstances. A minority of studies were published in refereed journals, many being published in bulletins of an in-house nature, or being entirely unpublished. The quality of studies was correspondingly extremely various.

The studies reviewed incorporated very various "intensive periods" of participation. In order to enable some approximate comparison of studies incorporating different lengths of intensive period, reference will be made to "Ratio Gains". Ratio Gain can be defined as the gain in reading age made by a subject on a reading test during a chronological time span, expressed as a ratio of that time span, i.e. Ratio Gain = reading age gain in months divided by chronological time span in months. Ratio Gains are sometimes construed as a multiple of "normal" rates of gain in reading, on the assumption that a "normal" gain is one month of reading age per chronological month elapsed. This fallacious assumption ignores the non-linearity of reading development and the non-equivalence of one month of reading age gain from differing reading age baselines. The validity of the use of Ratio Gains is discussed in greater detail later.

This review will initially discuss studies utilising a simple Pre-Test/Post-Test design. Then Control or Comparison Group studies will be considered, then studies comparing Paired Reading to other methods and techniques. Reports on the detailed process of Paired Reading projects will then be considered, followed by a review of studies which have examined the impact of Paired Reading on the reading style of subjects. Studies reporting post-project follow-up data will then be considered

and followed by a discussion of studies incorporating other interesting miscellaneous features. Finally, the Effect Sizes emerging from those studies where sufficient data is given to enable the calculation of this statistic will be summarised. In all of the subsections, the studies deploying parent tutors will be considered first, those deploying peer tutors considered second, and those deploying other tutors considered last.

Pre-Post Studies

The first published report on Paired Reading appeared in 1976 (Morgan, 1976). Morgan had himself tutored for one hour per week 3 subjects age 9 to 12 years who were between 2 and 4.5 years retarded in reading. Tuition continued for 18 weeks, and the Neale Test was applied with an inter-test period of 19 weeks. Data was available for only two subjects, who gained an average of 5.5 months of reading accuracy age and 16 months of reading comprehension age, yielding ratio gains of 1.26 and 3.65 respectively. Morgan also measured error rate per subject per session on consecutive samples of the text utilised for tutoring purposes, which was the same for all subjects.

In 1979, Morgan and Lyon reported a second study in which the parents of the children were deployed as Paired Reading tutors, in the constellation now most prevalent. Subjects were 4 children aged 8 to 11 years who were between 7 months and 3 years 8 months retarded in reading, with IQs ranging from 104 to 124. Tutoring was intended to occur for 5 or 6 periods of 15 minutes per week over 12 weeks, and weekly "supervision" meetings were held with the parents. The inter-test period was 27 weeks, and the mean gains on the Neale Test 11.75

months for accuracy (ratio gain = 1.88) and 11.50 months for comprehension (RG = 1.84).

By the early 1980's, the use of Paired Reading had begun to spread rapidly, and the volume of associated literature expanded accordingly. Bushell et al. (1982) reported on the deployment in Derbyshire of natural parent tutors with 22 subjects aged 8 to 11 years with mean reading retardation of 2 years. Two consecutive training meetings were held, Reading Together being trained and practised for a week before Reading Alone was trained at the second meeting, and home visits for support and monitoring were made every 2 weeks to all families. Project intensive period was 9 weeks, inter-test period 8 weeks, mean gains on Neale accuracy 5.8 months (RG = 3.14) and on comprehension 13.0 months (RG = 7.04). The most retarded children in the participant group made similar progress to the less retarded children.

Bush (1983) reported on several phases of the deployment of Paired Reading with natural parent tutors in a severely socio-economically disadvantaged area. In all, 65 children aged 8 to 11 years who were at least 2 years retarded in reading participated in projects of intensive period 8 weeks, each family being supported by 5 home visits during that period. Mean reading age gains on the Neale Test were 11.29 months in accuracy (RG = 6.12) and 16.77 months in comprehension (RG = 9.08).

Also in 1983, Kidd and Levey first reported the application of parent tutored Paired Reading to High School children. Subjects were 11 children aged 11 to 12 years and between 0 and 3 years retarded, involved in a project of intensive period only 5 weeks. Training was by

one group meeting and one session of individual consultation for each family. Another feature novel at the time was that no home visits were incorporated. Over an inter-test period of 5 weeks, gains in reading age on the Neale Test for accuracy averaged 6.00 months (RG = 5.20). Pre-post gains were greater than those noted in a 5 week pre-project baseline period. Participant children also made substantial gains pre-post on the Daniels and Diack Spelling Test.

The application of Paired Reading to younger children and in larger numbers was reported by Pitchford and Taylor (1983). A total of 64 subjects aged 7 to 8 years in several different schools were supported with two home visits during a project period of 8 weeks. Mean gains on the Neale Test were 6.3 months for accuracy (RG = 3.41) and 10.5 months for comprehension (RG = 5.69). Subjects made lesser gains on the Schonell Word Recognition Test (mean gain = 3.9 months, RG = 2.11). Pitchford and Taylor also elaborated further tutoring procedures for parents to use "Beyond Paired Reading".

A project carried out in 1983 with 13 subjects aged 6 to 11 years with reading difficulties is reported by Winter (1987). The project period was only 4 weeks and mean gains on the Neale Test of 5 months in accuracy (RG = 5.41) and 9 months in comprehension (RG = 9.75) are cited. Winter claims that participant group scores were superior to control group scores but no details are given. Parents were trained in the Paired Reading technique in 3 to 4 consecutive workshops during the four weeks, each workshop of one hour duration.

A project involving 8 high school children aged 11 to 12 years and between 2 and 3 years retarded in reading is reported by Evans (1984) and Morris (1984). The project incorporated one training and one follow-up meeting for parents, but no home visits. The project period was 6 weeks but the inter-test period was almost 11 weeks. Mean gains on the Neale Test were 6.25 months in accuracy (RG = 2.51) and 18.75 months in comprehension (RG = 7.52) (Evans' figures).

Evans (1984) also reported on the deployment of parent tutored Paired Reading with 6 subjects who had been diagnosed as "dyslexic" by the Dyslexia Institute and who attended the Institute for special tuition. The children were aged from 10 to 13 years and their retardation in reading ranged from 1.5 to 5.5 years. Over a project period of 7 weeks and an inter-test period of 8 weeks, mean gains on the Neale Test were 2.80 months in accuracy (RG = 1.52) and 1.35 months in comprehension (RG = 0.73). The first two weeks of the project involved only Reading Together, and families were supported by weekly home visits. The comprehension results were extremely erratic, and Rate of reading as indicated by the Neale Test reduced on average.

Gollop (1984) reported a 6 week project for 30 children aged 8 to 11 years most of whom were below average readers. On the Primary Reading Test (a group administered paper and pencil sentence completion test) mean gains of 12 months were recorded (RG = 8.67). No home visits were incorporated, and Gollop notes that 2 very capable readers included in the project made the greatest progress on the reading test.

Bush (1985) documented a subsequent phase of her P.R.I.N.T. Project, incorporating 21 subjects aged 8 to 11 years who were at least two years retarded in reading and socially disadvantaged. Mean gains on the Neale Test were 12.6 months for accuracy (RG = 6.83) and 14.7 months for comprehension (RG = 7.96).

Another report on what was becoming the standard format for High School projects, (parents tutoring first year "remedial" pupils, unsupported by home visits) was offered by Sweetlove (1985). Over an inter-test period of 13 weeks, 11 subjects of below average reading ability made mean gains on the Neale Test of 19.67 months in accuracy (RG = 6.56) and 17.44 months in comprehension (RG = 5.81).

Barrett (1986) reported on a 10 week project for five 8 to 10 year old below average readers, utilising the Salford and Burt Reading Tests. Participants were trained to spend the first 5 weeks of the project only Reading Together. Home visits were incorporated, at an average of 3 per family. On the Salford Test, mean gains of 6.4 months (RG = 2.77) were recorded, and on the Burt Word Recognition Test mean gains of 5.6 months (RG = 2.42). Barrett claimed that the Burt gains represented 3 times the rate of progress made by the subjects during a baseline period, but provided no details.

Yet another different reading test was deployed by Byron (1987), in a project for 47 below average first year high school pupils. No home visits took place, and the project period of 7.7 weeks ran over the Christmas school holiday. Nevertheless, mean gains on the Holborn Sentence Reading Test were 11.5 months (RG = 6.47).

Projects lasting 7 to 8 weeks incorporating 20 six to eleven year old subjects of mixed reading ability in 3 different schools in socio-economically disadvantaged areas are reported by MacMillan et al. (1988). Home visits were made to children in one school, but not those in the other two. Mean gains on the Neale Test were 7.10 months in accuracy (RG = 4.20) and 6.05 months for comprehension (RG = 3.58).

MacMillan et al. (1988) also report on a 6 week project for 9 eight year old pupils in a special school for children with moderate learning difficulties. No home visits were incorporated. Mean gains on the Neale Test were 4.89 months for accuracy (RG = 3.53) and 8.22 months for comprehension (RG = 5.94).

Turning to studies of Peer Tutored Paired Reading, the first published report was a paper by Winter and Low in 1984. Fifteen same-age tutorial pairs of 10 to 11 year olds used Paired Reading 3 times a week in their break time for six weeks on material of controlled readability. On the GAP Reading Test, mean gains of 3 months for tutees (RG = 2.17) and 5 months for tutors (RG = 3.61) are reported, although it was felt that there was a ceiling effect for the tutors on this test.

A similar project was reported by Crombie and Low (1986), again operating same-age peer tutoring on a whole- class mixed ability basis, but on this occasion with tutorial contact 5 times weekly for the 11 pairs. On the GAP Reading Test, mean gains of 10.03 months for tutees (RG = 7.24) and 6.83 months for tutors (RG = 4.93) are recorded.

Control or Comparison Group Studies

A total of 18 studies have included control or comparison groups in the research design. Thirteen of these were parent tutored and 5 peer tutored projects. The nature of the control or comparison group is very various between studies, this also being true of method of allocation to groups. Some of these studies fail to cite statistical significance of differences found, but with the small samples prevalent the finding of statistical significance is not likely.

Heath (1981, 1985) reported on a 13 week project for 7 to 9 year olds who were at least one year retarded in reading. Thirty-two subjects were randomly assigned to experimental and control groups. The resulting control group had a higher mean pre-test reading accuracy age than the PR group. The two schools involved already had a less structured form of home/school reading scheme. Parents were trained on a one to one basis in school. On the Neale Test, the PR group gained 6.8 months of reading accuracy age (RG = 2.27) and 10.7 months of reading comprehension age (RG = 3.57), compared to control children who gained 3.3 months of reading accuracy age (RG = 1.1) and 6.4 months of reading comprehension age (RG = 2.13). I.Q. was not found to be related to progress on reading tests.

A paper by Arora and Sheppard in 1982 reported the same data as Bush (1982), but as the former contains many more arithmetical inconsistencies than the latter it will be disregarded. Bush reported on one phase of her work with 9 to 10 year old pupils who were at least 2 years retarded in a socio-economically disadvantaged area. Seven

Paired Readers participated in the 8 week project and were supported by fortnightly home visits. Two of them were tutored by elder siblings rather than parents. A comparison group of 18 children who were "self-selected non-participants" were also assessed using the Neale Test. Participant children gained 12.57 months in reading accuracy (RG = 3.13) and 18.71 months in reading comprehension (RG = 4.66), while comparison children gained 6.11 months in accuracy (RG = 1.52) and 10.78 months in reading comprehension (RG = 2.68), differences for both accuracy and comprehension achieving statistical significance.

A complex project involving second year high school pupils aged 12 to 13 years in three different high schools who were up to 3 years retarded in reading is reported by Carrick-Smith (1982,85). Fifty six children were allocated to pairs matched approximately by chronological age and reading age, and one member of each pair allocated at random to the experimental group. Some of the 28 experimental subjects were tutored by their natural parents, some by cross-age peer tutors (VIth form volunteers) and some by teacher volunteers. Training was given one month before the project purportedly commenced. All participants were monitored by home or school visits three times during the 6 weeks of project participation. Arithmetical inconsistencies are evident in the data, and the figures cited here have been re-calculated from the raw data. On the Neale Test, participants gained 4.93 months of reading accuracy age (RG = 3.56) and 8.61 months of reading comprehension age (RG = 6.22), while the control group gained 1.21 months of reading accuracy age (RG = 0.87) and 4.11 months of reading comprehension age (RG = 2.97). The difference between experimental and control groups was

statistically significant at the 1% level for reading accuracy, but not for comprehension. Reading Rate as measured by the Neale was found to reduce.

Jungnitz et al. (1983) conducted a project for 7 year old non-readers in a highly socially disadvantaged area in which parental participation was sustained over a whole academic year (39 weeks). Twenty one children participated, and a comparison group of self-selected non-participant children numbered six. Participant families received an introductory home visit and an unspecified number of subsequent home visits as deemed necessary. Daniels and Diack Test 1 was used, and it was assumed that complete failure to score on this test could be considered equivalent to a reading age of 5.0 years, and thus the degree of reading progress made by the subjects could have been underestimated. Participant children gained 2.09 years of reading age (RG = 2.79) while control children gained 0.70 years of reading age (RG = 0.93). This difference was statistically significant at the 0.001 level.

Byron and Brock (1984) carried out a 12 week project for children aged 8 to 11 years who were below average in reading ability. Sixteen subjects were randomly selected from the available population, and then randomly assigned to experimental or control conditions. Training was by a group meeting followed by an individual meeting with families and supported by two home visits. The participant group achieved gains on the Neale Test of 6.3 months in reading accuracy age (RG = 2.28) and 9.7 months of comprehension age (RG = 3.50), while the control group gained 4.6 months in accuracy (RG = 1.66) and 3.4 months in reading comprehension (RG = 1.23).

A 28 week project with 11 to 12 year old remedial pupils in three high schools was reported by Spalding et al. (1984). Experimental and control groups both numbered 28 and were said to be matched according to chronological age and reading age, but it is not stated if allocation to groups was random. Supportive home visits were conducted fortnightly. On the Neale Test, the participants gained 6.5 months in reading accuracy (RG = 1.01) and 6.18 months in comprehension (RG = 0.96), while control children gained 6.3 months in accuracy (RG = 0.98) and 1.75 months in reading comprehension (RG = 0.27). The authors note that they suspected some contamination between experimental and control groups, but the gains cited are unusually low for both experimental and control groups.

A 13 week project with physically handicapped children aged 5 to 11 years with below average reading skills was reported by O'Hara (1985). Twelve experimental and 12 control children were matched by chronological age, reading age and gender. Three supportive home visits were made to each participant. On the Neale Test, participants gained 7.25 months in reading accuracy (RG = 2.42) and 12.4 months in reading comprehension (RG = 4.13), while controls gained 1.3 months in reading accuracy (RG = 0.43) and 5.9 months in reading comprehension (RG = 1.97). The Daniels and Diack Test 1 was also used in parallel, and on this the experimental group gained 7.2 months of reading age while the control group gained 3.0 months of reading age.

An alternative approach was exemplified by Simpson (1985) who operated an 8 week project for 26 junior age remedial children with a distal comparison group drawn from two other local schools. On the Neale

Test, participant children gained 5.76 months in reading accuracy (RG = 3.12) and 11.56 months in reading comprehension (RG = 6.26), while comparison children gained 2.16 months in reading accuracy (RG = 1.17) and 4.79 months in reading comprehension (RG = 2.59).

A controlled study based on their work in Derbyshire was reported by Miller et al. in 1986 although the work had been carried out some time before. The participant group of 33 children aged 8 to 11 years and at least 18 months retarded in reading were drawn from 13 different schools. Allocation to experimental or control groups was arranged to produce a balanced representation of the different schools supplying participants. The two groups showed no statistically significant differences in chronological age or scores on the English Picture Vocabulary Test, but the control group did have significantly higher pre-test reading ages. The project operated for 6 weeks and assessment on the Neale Test was carried out blind. The experimental group gained 2.43 months in reading accuracy (RG = 1.76) and 4.36 months in comprehension (RG = 3.14), while the control group gained 0.81 months in reading accuracy (RG = 0.59) and 1.69 months in comprehension (RG = 1.22). Experimental/Control differences achieved statistical significance for reading accuracy but not for reading comprehension. Subsequently, when "control" children participated in a second phase of the project, they achieved gains in reading accuracy of 4.85 months and in reading comprehension of 6.31 months, i.e. substantially better than the first experimental group.

Richardson (1986) conducted an 8 week project for junior aged children who were at least one year retarded in reading. The study incorporated

12 participant children and 11 comparison children, but it is not clear how comparison children were selected. Home visits were incorporated and the Neale Test used for assessment. The participant group gained 8.4 months in reading accuracy (RG = 4.55) and 12.6 months in reading comprehension (RG = 6.83). It is stated that the difference between participant and control children reached statistical significance for both accuracy and comprehension, but no figures are given for the control group.

A mixed ability project for 8 to 9 year olds in an area of low socio-economic status was reported by Gautrey (1988). In a six week project, the Hunter-Grundin Literacy Profiles Reading for Meaning Level II Test was utilised. The project incorporated weekly supervisory meetings for participant families in school and a points system for reinforcing regular reading. The control group constituted the whole of a similar class of children, but no information is given about numbers in either experimental or control groups. Participant children gained 11 months in reading age (RG = 7.94) while comparison children gained 5 months (RG = 3.61).

Morgan and Gavin (1988) drew their seven 9 to 11 year old participants of mean retardation 2 years 3 months from three different schools. No information is given as to how the comparison group of 8 children were selected, but assessment was blind. The families were trained on a one to one basis over the 13 week project. On the Neale Test the participant children gained 6.29 months in reading accuracy (RG = 2.10) and 9.29 months in comprehension (RG = 3.07), while the comparison group

gained 2 months in accuracy (RG = 0.67) and lost an average of 0.4 months in reading comprehension (RG = -0.13). Both differences were statistically significant at the 0.01 level. The comparison group were offered Paired Reading involvement at the conclusion of the initial phase of the project, and they subsequently recorded gains of 5 months in accuracy and 10.2 months in comprehension during a project period of 3.5 months.

Turning to Peer Tutored control group studies, Limbrick et al. (1985) reported in great detail on a project involving only 3 tutorial pairings. The tutors were 10 to 11 years old and 18 to 24 months retarded in reading, while the tutees were 6 to 8 years old and at least 18 months retarded in reading. A modification of Paired Reading was used, involving delay of correction until the end of the sentence and with a greater emphasis on discussion. At both age levels, 6 subjects were selected randomly from a pool of under-achievers and then randomly allocated to experimental or control conditions. Baseline process measures were taken during a preceeding period of untrained tutoring. Pairs participated for 6, 8, or 10 weeks respectively. This report contains some arithmetical inconsistencies, and the means cited here have been re-calculated from the raw data. On the Neale Test, the 3 tutees gained 11.0 months in reading accuracy (RG = 3.67) and 23.3 months in comprehension (RG = 7.77), while the tutee control group (n = 3) gained 6.3 months in accuracy (RG = 2.1) and 4.7 months in comprehension (RG = 1.57). The 3 tutors gained 19.0 months in accuracy (RG = 6.3) and 25.3 months in comprehension (RG = 8.43), while tutor controls (n = 3) gained 3.03 months in accuracy (RG = 1.01) and 6.7

months in comprehension (RG = 2.22). The difference between the tutees and the tutee control group achieved statistical significance, but this was not true for the tutors and tutor control group, presumably because of greater variability within this group (analysis of variance used).

Crombie and Low (1986) report a six week project for 12 tutors aged 10 to 11 years and 12 tutees aged 7 to 8 years. Subjects were allocated randomly to experimental and control groups subject to the consideration that numbers were balanced by gender, as same-sex tutoring was to operate. Tutor and Tutee control groups were involved in individual reading practice to help control for the "practice effect" of extra reading. Readability of materials used was controlled for all subjects. On the Neale Test, tutees gained 9 months in reading accuracy (RG = 6.50) and 9.96 months in comprehension (RG = 7.19), while tutee controls gained 2.64 months in accuracy (RG = 1.91) and 3.72 months in comprehension (RG = 2.69). Tutors gained 8.88 months in reading accuracy (RG = 6.41) and 12.84 months in comprehension (RG = 9.27), while the tutor control group gained 4.56 months in accuracy (RG = 3.29) and 9.84 months in comprehension (RG = 7.11). Overall, differences between tutoring and non-tutoring conditions achieved statistical significance for both accuracy and comprehension. The difference between tutors and tutor controls achieved statistical significance for reading accuracy but not for comprehension. Differences between tutees and tutee controls reached statistical significance for both accuracy and comprehension. Both boys and girls made statistically significant gains in accuracy and comprehension compared to their controls.

Grundy (1987) randomly allocated subjects to four groups stratified by gender; 10 pupils were cross-age peer tutored, 8 pupils were same-age peer tutored, 9 pupils received traditional extra remedial help on a withdrawal basis from trained teachers and 10 children formed a control group. Outcome data is only given for tutees, all of whom were aged 8 to 11 years and between 20 and 30 months retarded in reading. Random allocation resulted in the teacher-taught group being most retarded on average at pre-test and the cross-age tutees least retarded. On the Primary Reading Test, over a 17.3 week period, the cross-age tutees gained 7.2 months of reading age (RG = 1.80), the same-age peer tutees 3.75 months (RG = 0.94), the remedially taught group 1.33 months (RG = 0.33) and the control group 3.90 months (RG = 0.98). These outcomes are unusually low, except that for the cross-age tutor group. The author notes that the same-age peer tutor group was inadequately supervised. The remedial teacher-taught group performed exceptionally badly.

The work of Alan Low and his collaborators continued to extend, and Low et al. (1987) reported a cross-age tutor project, wherein 10 to 11 year old children tutored 6 to 7 year old children. Allocation to groups was random, stratified by gender. The project period was very short (four weeks), incorporating reading 5 times a week for 15 minutes. On the MacMillan Graded Word Reading Test, tutors gained nine months of reading age (RG = 9.75) and tutor controls 4.5 months (RG = 4.88), while tutees gained 7.0 months (RG = 7.58) and tutee controls 3.5 months (RG = 3.79). Similar differences were evident from error counts based on Informal Reading Inventories.

This work was then further extended by the deployment of six to seven year olds to tutor five year olds (Low and Davies, 1988). Twelve tutors and twelve tutees were each compared with a control group of similar size. Unsurprisingly, the "floor" of the MacMillan Graded Word Recognition Test proved not low enough to discriminate with a population of this kind. No norm-referenced data is given for the tutees, and subsequent comments are based on recalculation of the data excluding children who were "Below Scale" on both pre-test and post-test. The tutors gained 3.38 months (RG = 3.66) and the tutor controls 3.00 months (RG = 3.25), the difference failing to reach statistical significance. Both tutor and tutee groups were also assessed by error counts on an Informal Reading Inventory, and again no statistically significant differences were found. The authors note that the younger children had difficulties with the Paired Reading technique, but all reading was carried out in the children's own break time rather than during class-time. Readability of materials was controlled. The authors' conclusion that cross-age peer tutored Paired Reading was ineffective with children so young might be construed as premature in the absence of replication.

One further control group project deployed a professional as tutor (Lees 1986, 1987). Three groups of ten children were matched for reading age, chronological age and gender, but it is not known how allocation to conditions was carried out. The participant group in the 8 week project consisted of 10-12 year olds who were 1 - 4.5 years retarded in reading. The first comparison group was similar but a second comparison group of 8-9 year old children who were average readers was also established. On the MacMillan Analysis of Reading Ability, the participant group gained

9 months in reading accuracy (RG = 4.88), the similar control group lost 0.8 months (RG = -0.43) and the control group of younger average readers gained 4 months (RG = 2.17). Tuition occurred only twice weekly for 15-20 minutes.

A summary of results from control group studies is presented in Synopsis 1. In 13 parent tutored control group projects involving 212 experimental and 195 control subjects, the mean ratio gain for experimentals was almost 3 times that for controls. In reading comprehension, controls made greater than "normal" gains, but experimental mean ratio gain was still twice that for controls. In 5 peer tutor projects involving 46 experimental tutees and 45 controls, the mean ratio gain in accuracy for tutees was more than 3 times that for controls. Although numbers were very small for the comprehension comparison, and controls made much greater than "normal" gains, the mean ratio gain of experimentals was still almost 3 times that of controls. Aggregated results for peer tutors show much less difference between experimental and control groups - although experimental mean ratio gains in accuracy and comprehension are very high, they are also very high for control groups. It is not obvious why this should be so much more the case for peer tutors than it is for peer tutees. Overall, mean ratio gains in reading accuracy for all experimentals are almost 2.5 times those for controls, while in reading comprehension mean ratio gains for all experimentals are almost twice those for controls.

Studies Comparing Methods

A number of the studies comparing Paired Reading with other methods and techniques for non-professional tutoring of reading are in fact student

Synopsis 1 Synopsis of Participant and Control Mean Ratio Gains

			Accuracy		Comprehension	
Type of Project	Group	Projects	n	Mean Ratio Gain	n	Mean Ratio Gain
Parent Tutored	E	13	212	3.13	165	3.91
	C		195	1.19	153	1.93
Peer Tutees	E	5	46	4.63	15	7.31
	C		45	1.40	14	2.45
Peer Tutors	E	5	40	6.66	15	9.10
	C		39	3.63	14	6.06
Overall	E	18	298	3.84	195	4.57
	C		278	1.56	181	2.29

theses. Some of these are poorly constructed, and many involve very small samples. Several authors acknowledge the possibility of contamination between groups, and evidence is rarely offered that the techniques prescribed were actually followed in practice, thus rendering the comparison a true one. Thus, in addition to doubts about the validity of the purported comparison, it is hardly surprising that a number of studies fail to find differences between experimental treatments, and indeed some fail to find significant differences between experimental and control groups.

The first comparative study was reported by Heath (1981, 1985), who in a 13 week project with children aged 7-9 years who were 1 or more years retarded in reading randomly allocated 4 children to a Paired Reading group, 4 children to a "reinforcement" group and 4 to a control group. Parents were trained on a one-to-one basis in school in two half-hour sessions, followed by fortnightly monitoring meetings. In the reinforcement condition, parents were required to listen to the child reading and praise successful reading. The Paired Reading group made gains on the Neale Test of 8 months in accuracy (RG = 2.67) and 13 months in comprehension (RG = 4.33), the reinforcement group 5 months in accuracy (RG = 1.67) and 4 months in comprehension (RG = 1.33), while the control group gained 0.25 months in accuracy (RG = 0.08) and 0.5 months in comprehension (RG = 0.17). Statistical significance is unlikely with such small samples, and was not commented on by the author.

Small samples were also a feature of a study by Wareing (1983, 1985) to compare the progress on the Neale Test over 8 weeks of 5 Paired Readers,

5 children who were "heard" reading with error correction after 10 seconds, another group of 5 who used a similar "Listening" method but where the parents were non-English speaking, and a "Linguistic Method" group of 5 (where children were read to, discussed the book, re-told the story in their own words, then read aloud their version of the story written down by the parent). Four comparison groups, each of 5 children, were also established on a self-selected non-participant basis, albeit matched by chronological age and reading age. Each experimental group operated in a different primary school, the subjects being aged 9-10 years. Participating schools chose which method they preferred to be used in their schools. Training was conducted in one group session, followed up by one home visit per family. Very variable gains were subsequently evident, and arithmetical errors are evident in the results. The Reading Aloud (Non-English) group made by far the biggest gains in reading accuracy and the Paired Reading group made by far the biggest gains in reading comprehension. In three cases, results from control groups were negative i.e. there was an average reduction in reading age. Although differences were extremely erratic and various, none reached statistical significance. The author notes that the Paired Reading group did not in many cases actually do Paired Reading, but tended to drift towards a Reading Aloud method.

Grigg (1984) compared 14 Paired Readers to 14 children in a "Listening" group using a Reading Aloud method in which errors were not indicated until the end of the sentence and prompts were then provided from the context or phonic structure. During a 9 week project, on the Neale Test the Paired Readers gained 6 months in reading accuracy (RG = 2.88), the Listening group 7.5 months (RG = 3.61) and the control group 5.5 months

(RG = 2.65). In comprehension, all 3 groups showed gains, but the control group gained most. There were no statistically significant differences. Parents were trained in 3 coaching sessions totalling 1.5 hours, but it should be noted that parents were not volunteers, rather being persuaded to take part. Grigg raises the question of possible contamination between groups.

Jungnitz (1984, 1985) compared the progress of 7-9 year old "remedial" children in a multi-ethnic school in a 12 week project. Ten children who were already involved in the school's existing home-school reading scheme, modelled on the Belfield project, were compared with a group of 11 much weaker readers, mostly with non-English speaking parents, to whom Paired Reading was introduced by the experimenter. The Paired Reading group were all of Asian ethnic origin, the Listening group all of Asian origin with one exception, but from families where parents were far more competent in spoken and read English, and the self-selected "control" group included 3 subjects of Asian origin, one of Caribbean origin and 3 indigenous whites. In the Paired Reading group, only one parent could effectively tutor the child, and in other cases tutoring was carried out by siblings or members of the extended family, together with one subject who "did Paired Reading with himself" by taking home tape-text combinations and receiving generalised support and praise from his uncomprehending parents. Paired Reading families were involved in a training meeting in school, subsequent to a group meeting in school mid-project and received the support of 4 home visits from the experimenter.

Arithmetical inconsistencies are evident in this report, and re-calculations have been made based on the raw data. The Paired Reading

group gained 13.36 months in reading accuracy (RG = 4.14) and 15.82 months in comprehension (RG = 4.90), while the Listening group gained 7.10 months in accuracy (RG = 2.20) and 6.80 months in comprehension (RG = 2.10), the comparison group gaining 3.85 months in accuracy (RG = 1.19) and 0.14 months in comprehension (RG = 0.04). In reading accuracy, both the Paired Reading and the Listening group gains were statistically significantly better than those of the comparison group, while for comprehension the Paired Reading group did significantly better than the Listening group which in turn did better than the comparison group. The Paired Reading group were also substantially better than the other groups on a secondary test, the Schonell Word Recognition Test.

Dening (1985) compared the progress of 30 Paired Readers with 30 children experiencing the Pause Prompt and Praise technique, 30 children using a Listening method (tutor pauses at error, supplies word and praises) and 10 children in a control group. Subjects were aged 5-9 years and drawn from 3 schools in a middle-class area. They included both average and below average readers, with a mean retardation of 1.6 months in accuracy and 7.5 months in comprehension. In addition to project involvement, the children also had assigned reading homework. Training for all three experimental conditions was similar, with parents attending a group meeting and practising the technique on each other via role play. The amount of subsequent support and follow-up was also manipulated, some families receiving 4 home visits, some receiving telephone contact, and some receiving no contact.

The Neale Reading Test was used, the experimenter carried out all the testing herself, and one arithmetical anomaly is evident in the data given. The mean pre-test reading age of the Paired Reading group was lower than that of the other groups, but analysis of co-variance was utilised to take account of this. The Paired Reading group gained 7.48 months of reading accuracy age (RG = 3.0) and 7.62 months in comprehension (RG = 3.1), the PPP group gained 7.07 months in accuracy (RG = 2.8) and 11.74 months in comprehension (RG = 4.7), the Listening group 6.77 months in accuracy (RG = 2.7) and 10.97 months in comprehension (RG = 4.4), while the control group gained 3.10 months in accuracy (RG = 1.2) and 4.1 months in comprehension (RG = 1.6). In reading accuracy all three experimental groups were statistically significantly different from the control group, but not from each other. In reading comprehension, the Listening group and the PPP group were statistically significantly different from the control group, but the Paired Reading group was not. Variations in degree of follow-up support made no difference to gains in reading accuracy. In reading comprehension, for the PPP and Listening groups increased support was associated with higher gains, but this was not the case with the Paired Reading group. The author acknowledges the possibility of contamination between groups.

Lindsay et al. (1985) also manipulated degree of follow-up, some subjects receiving a weekly home visit and a follow-up group meeting while others merely received telephone contact support. The subjects were Middle school pupils aged 8-10 years, and among the least able readers in the school in question. They were allocated to groups on the

basis of stratification by age, gender, and a balance in mean retardation between groups. Ten subjects were involved in Paired Reading, and ten in "Relaxed Reading", a method designed to encourage parents to be warm and positive and rewarding when helping their children with reading. The emphasis in Relaxed Reading on positive feedback would appear to have implications more for the ethos of tutoring than for any specific technique involved, but during the course of training meetings families were given specific individualised advice appropriate to their own situation. In a 6 week project, assessed with the Neale Test, Paired Readers gained 9.7 months in reading accuracy (RG = 6.47) and 14 months in Comprehension (RG = 8.10), while Relaxed Readers gained 7.3 months in accuracy (RG = 4.67) and 13.6 months in comprehension (RG = 8.17). Differences between groups were not statistically significant. Home visits did not significantly interact with the size of reading gain in either technique condition. Reading Rate as measured by the Neale was noted to reduce for both groups.

Winter (1985) trained the parents of 8-11 year old below average readers in both Paired Reading and the Pause Prompt and Praise technique. Subsequently, 6 participants chose to use Paired Reading and 6 chose to use Pause Prompt and Praise, although Winter notes that all the Paired Reading group preferred Reading Alone, and some contamination between groups seems likely to have occurred. Norm-referenced tests were not used, but instead pre-post Informal Reading Inventories were analysed for reading rate, error rate, refusal rate and self-correction rate. Subjects also received additional tuition from cross-age peer tutors in school using the same technique as the parents. The parents were

trained in 4 x 1 hour sessions and were asked to read for 20 minutes 6 times each week. Over both experimental groups, reading rate increased 17%, errors decreased by 22%, refusals decreased by 28%, but self-corrections did not alter in frequency. There were no statistically significant differences between the two groups.

Burdett (1985, 1986a, 1986b) compared two different techniques with 2 groups of 16 8-11 year olds of below average reading ability drawn from two schools for Armed Forces children in Hong Kong. In one school the children were randomly allocated to experimental conditions, but in another the Head Teacher selected children for conditions according to unspecified criteria. The groups contained twice as many boys as girls. Parents were trained in a group meeting followed by a home visit, one group doing Paired Reading and another group pursuing a modification of Pause Prompt Praise which Burdett termed Individualised Reading. Half of each group received reading tutoring from adult volunteers in schools as well as from natural parents at home, and the other half received tutoring only from adult volunteers in school, the latter offering 3-5 sessions per week while the parents offered five sessions of 5-10 minutes per week. A non-participant comparison group of 16 children was also established for the duration of the 8 week project (inter-test period 12 weeks) and the Widespan test was utilised, groups being compared in terms of raw score gains on the test. Both the Paired Reading and the Individualised Reading groups showed raw score gains which were statistically significantly larger than those of the control group, but not significantly different from each other. The impact of parental involvement in addition to volunteer tutoring in school was statistically significant for the Individualised Reading group, but not

for the Paired Reading group. An additional analysis of error rate indicated that the Paired Reading group reduced its error rate by 25%, the Individualised Reading group reduced it by 9%, while in the control group the error rate rose by 6%.

Jones (1987), working with mixed ability children aged 6-7 years, randomly allocated 21 to a Paired Reading group and 21 to a "Listening" group for a 10 week project. The Paired Reading group was offered a training and a follow-up meeting, and two families were subsequently home visited. However, not all participating families attended the training meeting, and some parents were merely sent written information. On the Neale test, the Paired Readers gained 8 months of reading accuracy age (RG = 3.47) and the Listening group 6 months (RG = 2.60), the difference failing to reach statistical significance.

A 26 week project with 11-12 year old pupils retarded by at least 18 months in a high school is reported by Sweetlove (1987). Twenty-four children did Paired Reading, 12 children experienced both Paired Reading at home and the Corrective Reading programme in school, 9 children received only Corrective Reading in school and 22 children formed a comparison group. The author gives no information as to how children were allocated to conditions, but it is evident that children involved in Corrective Reading tended to be weaker at reading than the other children. No home visits were incorporated and testing on the Neale Analysis was carried out blind by outsiders. Eleven of the 36 participating Paired Reading families did not attend the training meeting and were merely sent written information about the technique. As with the previous study, there must therefore be grave doubts about

process variables in this study. The 24 children doing only Paired Reading made average gains of 11.08 months in reading accuracy (RG = 1.85) and 18.46 months in comprehension (RG = 3.08), while children receiving Paired Reading and Corrective Reading made gains of 13.08 in accuracy (RG = 2.18) and 18.08 months in comprehension (RG = 3.01), the difference between these two groups failing to reach statistical significance. Children receiving Corrective Reading only gained 5.44 months in accuracy (RG = 0.91) and 14.78 months in comprehension (RG = 2.46), while comparison children gained 7.32 months in accuracy (RG = 1.22) and 11.77 months in comprehension (RG = 1.96). Making valid comparisons between groups is difficult, but it seems clear that a combination of Paired Reading and Corrective Reading results in larger gains in reading accuracy than Corrective Reading alone. Additionally, the two Paired Reading groups performed better than the control children, but the author does not state whether this difference achieves statistical significance, although this seems doubtful for the comprehension comparison.

Paucity of information is even more problematic in the study reported by Loveday (1988). Three different experimental treatments were deployed, for each of which subjects were drawn from a different school. One group played language games, one group did some form of reading at home which could be either Paired Reading or "Listening", while the third group participated in language games and reading at home. It seems likely that this project operated with children aged 7-9 years of mixed ability, but this is not stated by the author. In a 6 week project period the Primary Reading Test was utilised on a pre-post basis, but

numerical data are not given. The author asserts that girls tended to do better with language games, while boys did better with reading only. No other information is given.

Turning to consider 'studies comparing techniques within the peer tutoring format, better quality of research becomes evident. Winter (1988) worked with children aged 10-11 years in two schools in Hong Kong, one school supplying 14 data yielding tutorial pairs and the other 16. Group training lasting one hour was given, in one school before identifying participants and another school after identifying participants. The training before identification of participants produced a markedly higher volunteering rate. The GAP test was used with the tutees only during a 6 week project period. The tutees in school 1 gained 3.48 months of reading age (RG = 2.50) and those in school 2 gained 6.36 months of reading age (RG = 4.59).

Winter collected process data on a number of participating pairs, and noted that there appeared to be very little relationship between gains in reading age and the extent to which the Paired Reading technique was actually conformed to, with the exception of a correlation between error rate and reading age. This caused Winter to raise the question of whether conformity to technique was actually important in producing gains in peer tutoring of reading. However, large differences in outcomes were evident between schools, and much more process data was drawn proportionately from the school yielding the worst outcomes. The process data indicates that the behaviour of the children departed grossly from the prescribed technique, raising the question of the

adequacy of supervision by the teachers involved. It should also be noted that tutorial pairs were not matched by reading age differential, which may have promoted departure from the prescribed technique. There are also doubts about the appropriateness of book choice and method of selection of pairs for process analysis. It was also unfortunate that a ceiling effect on the GAP test resulted in the discarding of tutor scores.

Two experiments touching on the same issues are reported by Joscelyne (1989). This author worked with three classes in two primary schools, maintaining a 12 month differential in reading ability between tutors and tutees. Paired Reading was compared with a Listening method involving no Reading Together and utilising material of controlled readability. Tutees were arranged in matched reading groups then randomly allocated to experimental conditions. The tutors could not be so matched. Over a project period of 7 weeks the 11 tutorial pairs in each condition read for 4 fifteen minute periods per week. On the Neale Test, Paired Reading Tutees gained 10.36 months in accuracy (RG = 6.41) and 16.45 months in comprehension (RG = 10.18), while Listening tutees gained 2.91 months in accuracy (RG = 1.80) and 6.55 months in comprehension (RG = 4.05). These differences reached statistical significance in both accuracy and comprehension. The Paired Reading tutors gained 6.64 months in accuracy (RG = 4.11) and 8.91 months in comprehension (RG = 5.52), while Listening tutors gained 3.91 months in accuracy (RG = 2.42) and 14.72 months in comprehension (RG = 9.11). Differences for tutors did not reach statistical significance, and the author expressed concern about the ceiling effect on the Neale Test.

In a replication, Paired Reading tutees gained 6.27 months in accuracy (RG = 3.88) and 8.55 months in comprehension (RG = 5.29), and Listening tutees 3.64 months in accuracy (RG = 2.25) and 5.73 in comprehension (RG = 3.55). The differences between groups achieve statistical significance (favouring the Paired Reading group) in reading accuracy only. The Paired Reading tutors gained 0.45 months in accuracy (RG = 0.28) and 11.18 months in comprehension (RG = 6.92), while the Listening tutors gained 0.19 months in accuracy (RG = 0.12) and 8.45 months in comprehension (RG = 5.23). No difference for tutors achieved statistical significance. Joscelyne (1989) also notes a gain in words read in context rather than purely from grapho-phonemic clues - the Paired Readers were 2.4 times better than Listeners in this respect, this difference achieving statistical significance. On the basis of these 2 studies, Joscelyne concludes "Paired Reading, when carefully monitored, is more useful in accelerating children's reading than simply listening". However, no detailed data are given, although pairs were "checked" twice weekly to "ensure" they were adhering to the trained method.

Thirkell (1989), working with two classes in the same multi-cultural school, deployed 18 tutorial pairs and compared them to a group of 15 children who occupied themselves with individual silent reading, thus controlling for the effect of extra reading practice. Reading occurred for 15 minutes 5 days a week over a project period of 6 weeks. On the MacMillan Analysis of Reading Ability, Paired Reading tutors gained 5.7 months in reading accuracy (RG = 4.12) and 11.9 months in comprehension (RG = 8.59), Paired Reading tutees 7.75 months in accuracy (RG = 5.60)

and 9.0 months in comprehension (RG = 6.50), while the Silent Readers gained 6.1 months in accuracy (RG = 4.41) and 8.2 months in comprehension (RG = 5.92). Differences between the groups did not reach statistical significance.

Data reported by Townsend and Topping (1986) and Townsend (1987) are also pertinent in this context, although the data is subsumed in a non-comparative way within the aggregated data collected in Kirklees. In this study, comparisons were made of the impact of parent tutoring and peer tutoring, and the two types of tutoring deployed in succession with small samples of children in a First school in a severely disadvantaged area. Unfortunately allocation to experimental conditions was not random: the parent tutored participants (n = 7) were self-selected, many of the rest of the class then being selected by the class teacher as suitable to be tutors or tutees (both n = 7), and the remainder constituted a non-participant control group (n = 9). Children who began as parent tutored were involved in peer tutoring later in the academic year. In reading accuracy, all participant pairs did equally well in the short term, while in reading comprehension parent tutees did better. However, in the longer run when all pairs experienced peer tutoring, the differential in comprehension gains favouring the parent tutees disappeared. In the long run, the peer tutors did best overall, although group sizes were small and differences did not reach statistical significance.

Finally, two comparative studies involving the use of Paired Reading by tutors who were professional teachers will be considered. Welch (1985) worked with 8-9 year olds, matching 18 pairs of children and

allocating them to Paired Reading or Listening conditions. In an 11 week project, outcome measures were based on error analysis in relation to Informal Reading Inventories. Differences in use of context, understanding, rate of reading, number of refusals and attitude to reading were all in favour of the Paired Reading group, but none achieved statistical significance.

Spiby (1986) matched 5 pairs of 11-12 year old remedial pupils in a high school then allocated them randomly to Paired Reading and Listening groups. The Listening group were "encouraged to use attack skills" and given prompts to do so. In a 20 week project with an inter-test period of 24 weeks, on the Neale Test the 5 Paired Readers made mean gains of 10.92 months in accuracy (RG = 1.97) and 10.78 months in comprehension (RG = 1.95), while the Listening group made gains of 10.06 months in accuracy (RG = 1.82) and 9.56 months in comprehension (RG = 1.73). Differences between the groups are clearly not statistically significant. Tutoring took place for only two 15 minute sessions per week, and all children read the same books in the same order in both conditions, being withdrawn from lessons for the purpose of tutoring. Spiby (1986) comments that she did not find Reading Together easy, and the quality of Paired Reading technique in operation may thus be in doubt.

Overall, of 16 studies comparing Paired Reading to other techniques, only 4 found Paired Reading markedly superior (to "Listening" in 3 cases, 2 of which were peer tutor projects), although in many cases

small sample sizes rendered statistical significance elusive and in some contamination between groups was suspected. No study found PR inferior. The general finding from comparative studies was that PR had similar effects to other methods. However, in the 11 studies yielding adequate norm-referenced data, the mean pre-post Ratio Gain for Paired Readers was 3.74 and for all other techniques 2.25 (peer tutor data included).

SUMMARY OF PREVIOUS RESEARCH - PRE-POST, CONTROL AND COMPARATIVE STUDIES

Section A = Pre-post Studies	There is little to differentiate these studies except participant group size and they are therefore listed alphabetically according to name of first author. These predominantly parent tutored studies are very varied in participant age (6-13 years), group size (n=2-65), project length (4-13 weeks) and outcomes (Ratio Gains from 1.26 - 8.67 on reading accuracy).
Section B = Controlled Studies	Reports incorporating control or comparison groups are considered to yield more valuable information. These are divided into 5 quality bands, low [5] to high [1] according to adequacy of research design and execution, listed in alphabetical order within bands. Generally, band 5 studies have more than 2 major flaws, band 4 studies have 2 major flaws, band 3 studies have 1 major flaw, band 2 studies have 1 minor flaw, band 1 studies are substantially flawless as reported or have minor flaws balanced by particular design strengths. Major flaws include: small size of experimental or control group (<10), doutbful comparability of control and experimental groups irrespective of method of allocation and groups, impurity of Paired Reading technique as trained and other atypical factors in project organisation such as poor monitoring, infrequent tutoring, overcontrol of reading materials and unusually short or long project periods. The absence of crucial data in a report automatically relegated the study to the lowest band (5). Positive compensating design strengths included: blind testing, equivalent extra reading practice for control groups.
Section C = Comparative Studies	Comparing Paired Reading to another alternative treatment. This section is banded like Section B, except that Bands 1 and 2 are combined.

Tutor type: PA = *own Parent*, PT = *same-age* Peer Tutor, XT = cross-age Peer Tutor, PAV = Professional Adult Volunteer, AV = non-professional Adult Volunteer.
RTD = Retarded in reading, MA = Mixed Ability in reading.
n = number of Paired Reading subjects yielding data.
? = data not given
ss = statistically significant
CG = relevant statistic for Control/Comparison Group. AT = relevant statistic for Alternative Treatment Group

Study First Author	Year	Tutor Type	Subject Age (yrs) & type	n (CG) [AT]	Project Period (weeks)	Reading Test Used	Reading Age Accuracy Gain (mnths) (CG) [AT]	Reading Age Accuracy Ratio Gain (CG) [AT]	Reading Age Comprehension Gain (mnths) (CG) [AT]	Reading Age Comprehension Ratio Gain (CG) [AT]	s.s. (where given) Nature of Alternative Treatment/Comments
A. PRE-POST STUDIES (in alphabetical order)											
Barrett	1986	PA	8-10 RTD	5	10	Salford Burt	6.4 5.6	2.77 2.42			Reading Together only for first 5 weeks. Baseline comparison claimed, no data given.
Bush	1983	PA	8-11 RTD	65	8	Neale	11.3	6.12	16.8	9.08	
Bush	1985	PA	8-11 RTD	21	8	Neale	12.6	6.83	14.7	7.96	
Bushell etc.	1982	PA	8-11 RTD	22	9	Neale	5.8	3.14	13.0	7.04	

Byron	1987	PA	11-12 RTD	47	8	Holborn	11.5	6.47			Project ran over Christmas holiday
Crombie etc.	1986	PT	10-11 MA	15 + 15	6	GAP	10.0 6.8	7.24 4.93	(tutees) (tutors)		
Evans	1984	PA	10-13 RTD	6	7	Neale	2.8	1.52	1.4	0.73	Subjects diagnosed "dyslexic"
Evans Morris	1984	PA	11-12 RTD	8	6	Neale	6.3	2.51	18.8	7.52	
Gollop	1984	PA	8-11 MA	30	6	Primary	12.0	8.67			
Kidd etc.	1983	PA	11-12 RTD	11	5	Neale	6.0	5.20			Baseline comparison also made. Pre-post gains on Spelling Test also.
Macmillan etc. (1)	1988	PA	6-11 MA	20	8	Neale	7.1	4.20	6.1	3.58	Subjects in 3 schools, home visits in only one school.

Macmillan etc. (2)	1988	PA	8 RTD	9	6	Neale	4.9	3.53	8.2	5.94	Subjects in MLD special school
Morgan	1976	PAV	9-12 RTD	2	18	Neale	5.5	1.26	16.0	3.65	
Morgan etc.	1979	PA	8-11 RTD	4	12	Neale	11.8	1.88	11.5	1.84	
Pitchford etc.	1983	PA	7-8 RTD	64	8	Neale Schonell	6.3 3.9	3.41 2.11	10.5	5.69	
Sweetlove	1985	PA	11-12 RTD	11	13	Neale	19.7	6.56	17.4	5.81	
Winter etc.	1984	PT	10-11 MA	15 + 15	6	GAP	3.0 5.0	2.17 3.61	(tutees) (tutors)		
Winter	1987	PA	6-11 RTD	13	4	Neale	5.0	5.41	9.0	9.75	Control group reported but no CG data given.

B. CONTROL/COMPARISON GROUP STUDIES (in quality bands, low [5] to high [1]; alphabetical order within bands)

BAND 5

Study First Author	Year	Tutor Type	Subject Age (yrs) & type	n (CG) [AT]	Project Period (weeks)	Reading Test Used	Reading Age Accuracy Gain (mnths) (CG) [AT]	Reading Age Accuracy Ratio Gain (CG) [AT]	Reading Age Comprehension Gain (mnths) (CG) [AT]	Reading Age Comprehension Ratio Gain (CG) [AT]	s.s. (where given) Nature of Alternative Treatment/Comments
Gautrey	1988	PA	8-9 MA	? (?)	6	Hunter-Grundin	11.0 (5.0)	7.94 (3.61)			Comparison group = parallel class
Jungnitz etc.	1983	PA	7 RTD	21 (6)	39	Daniels Diack 1	25.0 (8.4)	2.79 (0.93)			CG = self-selected non-participants. Participant/control difference ss.
Low etc.	1988	XT	tutees 5 MA	12 + 12	4	Macmillan GWRT	no results for tutees				All reading in subjects' *own* break time and readability of materials controlled. Difference not s.s., nor on error count analysis. Allocation presumed random within gender. 4 tutors & 4 controls below 'floor' of test. Conformity to technique suspect.
			tutors 6-7 MA	(12 + 12)			3.4 (3.0)	3.66 (3.25)	(tutors)		

Richard-son	1986	PA	8-12 RTD	12 (11)	8	Neale	8.4 (?)		4.55 (?)	12.6 (?)	6.83 (?)	Allocation to groups unclear. No data given for controls, although s.s. claimed for difference.
Spalding etc.	1984	PA	11-12 RTD	28 (28)	28	Neale	6.5 (6.3)		1.01 (0.98)	6.2 (1.8)	0.96 (0.27)	Groups matched by chronological and reading age, not known if allocation random. All results unusually low. Authors suspected contamination between groups.
BAND 4												
Bush	1982	PA	9-10 RTD	7 (18)	8	Neale	12.6 (6.1)		3.13 (1.52)	18.7 10.8)	4.66 (2.68)	CG = self-selected non-participants Participant/control differences ss for accuracy and comprehension.
Lees	1986	PAV	10-12 RTD CG 1 CG 2	10 (10) (10)	8	Mac-millan Analysis of RA	9.0 (-0.8) (4.0)		4.88 (-0.43) (2.17)			Groups matched for chronological and reading age and gender. Tuition 2 x 15 mins weekly. CG1 were 10-12 yrs old RTD. CG2 were 8-9 yrs old average readers.

Limbrick etc.	1985	XT	tutees: 3+3 6-8 (3+3) RTD tutors: 10-11 RTD		8	Neale	11.0 (6.3) 19.0 (3.0)	3.67 (2.10) 6.30 (1.01)	23.3 (4.7) 25.3 (6.7)	7.77 (1.57) 8.43 (2.22)	Modification of PR used. Random assignation. s.s. reached between tutees and their controls, not between tutors and theirs. Project period varied, 8 weeks average.
Morgan etc.	1988	PA	9-11 RTD	7 (8)	13	Neale	6.3 (2.0)	2.10 (0.67)	9.3 (-0.4)	3.07 (-0.13)	Unclear how comparison group were selected. Participant/control differences s.s. Control gains unusually low. Controls subsequently participating yielded similar gains to original experimentals.

BAND 3

Byron etc.	1984	PA	8-11 RTD	8 (8)	12	Neale	6.3 (4.6)	2.28 (1.66)	9.7 (3.4)	3.50 (1.23)	Random assignation.
Grundy	1987	XT PT	tutees 8-11 RTD	10 8 (10) [9]	17	Primary	7.2 3.8 (3.9) [1.3]	1.80 0.94 (0.98) [0.33]			Allocation random within gender. Alternative treatment was extra withdrawal remedial help from a teacher. AT group most retarded at pre-test & XT group least so. Outcomes unusually low except for XT. PT group poorly monitored. No tutor data.

O'Hara	1985	PA	5-11 RTD	12 (12)	13	Neale Daniels Diack 1	7.3 (1.3) 7.2 (3.0)	2.42 (0.43)	12.4 (5.9)	4.13 (1.97)	Groups matched by chronological and reading age and gender. Participants were physically handicapped.
Simpson	1985	PA	8-12 RTD	26 (30)	8	Neale	5.8 (2.2)	3.12 (1.17)	11.6 (4.8)	6.26 (2.59)	Distal comparison group from 2 other local schools.
BAND 2											
Heath	1981	PA	7-9 RTD	16 (16)	13	Neale	6.8 (3.3)	2.27 (1.10)	10.7 (6.4)	3.57 (2.13)	Random assignation; left CG more able.
Low etc.	1987	XT	tutees 6-7 MA tutors (13+ 10-11 MA	13+13 13)	4	Mac-millan GWRT	7.0 (3.5) 9.0 (4.5)	7.58 (3.79) 9.75 (4.88)	(tutees) (tutors)		Allocation random within gender. Similar differences on error count analysis.

BAND 1

Carrick-Smith	1982	PA/ XT/ PAV	12-13 RTD	28 (28)	6	Neale	4.9 (1.2)	3.56 (0.87)	8.6 (4.1)	6.22 (2.97)	Matched pairs allocated at random to experimental or control group. Participant/control differences ss for accuracy only.
Crombie etc.	1986	XT	tutees: 12+12 7-8 MA?		6	Neale	9.0 (2.6)	6.50 (1.91)	10.0 (3.7)	7.19 (2.69)	Control groups had extra individual reading practice. Allocation random within gender to ensure same-sex tutoring. Readability of materials controlled. Tutoring/non-tutoring differences s.s. for accuracy and comprehension, except for tutor comprehension.
			tutors: (12+12) 10-12 MA				8.9 (4.6)	6.41 (3.29)	12.8 (9.8)	9.27 (7.11)	
Miller etc.	1986	PA	8-11 RTD	33 (33)	6	Neale	2.4 (0.8)	1.76 (0.59)	4.4 (1.7)	3.14 (1.22)	Subjects drawn from 13 different schools. Allocation by balanced representation from all schools; left CG more able. Blind testing. Participant/control differences ss for accuracy only. Controls subsequently participating yielded much higher gains than original experimentals.

C COMPARATIVE STUDIES (in quality bands, low [5] to high [1/2]; alphabetical order within bands)

BAND 5

Study First Author	Year	Tutor Type	Subject Age (yrs) & type	n (CG) [AT]	Project Period (weeks)	Reading Test Used	Reading Age: Accuracy Gain (mnths) (CG) [AT]	Reading Age: Accuracy Ratio Gain (CG) [AT]	Reading Age: Comprehension Gain (mnths) (CG) [AT]	Reading Age: Comprehension Ratio Gain (CG) [AT]	s.s. (where given) Nature of Alternative Treatment/Comments
Loveday	1988	PA	?	?	6	Primary	no numerical data given				Each group from a different school.
Spiby	1986	PAV	11-12 RTD	5 [5]	20	Neale	10.9 [10.1]	1.97 [1.82]	10.8 [9.6]	1.95 [1.73]	Matched pairs, allocated randomly. AT = Listening + prompts. Infrequent tutoring, reading material specified, same for all subjects. No significant differences.
Sweetlove	1987	PA	11-12 RTD	24 (22) [12]1 [9]2	26	Neale	11.1 (7.3) [13.1] [5.4]	1.85 (1.22) [2.18] [0.91]	18.5 (11.8) [18.1] [14.8]	3.08 (1.96) [3.01] [2.50]	AT1 = Corrective Reading in school + PR at home. AT2 = Corrective Reading only. Non-random allocation, AT groups more retarded than PR & CG. 11/36 PR parents did not attend. training. Differences between PR and AT1 not s.s. Significance not given for other comparisons.

Winter	1988	PT	10-11 MA	14 16	6	GAP	3.5 6.4	2.50 4.59	tutees school 1 tutees school 2		Two schools involved, yielding very different results, with differential data loss and selection/training sequencing. Participants self-selected - no experimenter allocation. Process data indicate gross departure from PR technique; doubts re pair matching & ability differentials.

BAND 4

Heath	1981	PA	7-9 RTD	4 (4) [4]	13	Neale	8.0 (0.3) [5.0]	2.67 (0.08) [1.67]	13.0 (0.5) [4.0]	4.33 (0.17) [1.33]	Random allocation. AT condition = Parent listening to child read and praising.
Jungnitz	1984	PA	7-9 RTD	11 (9) [10]	12	Neale	13.4 (3.9) [7.1]	4.14 (1.19) [2.20]	15.8 (0.1) [6.8]	4.90 (0.04) [2.10]	Control group self-selected non-participant, AT group (Parent Listening) pre-existent, PR and AT groups E2L. PR and AT group gains significantly larger than control on reading accuracy; for comprehension PR > AT > CG.

Wareing	1983	PA	9-10 RTD CG1-4	5 [5]1 [5]2 [5]3 (4x5)	8	Neale	results extremely erratic and none s.s.	Each experimental group in a different primary school; schools chose their own method: Paired Reading, Parent Listening with error correction, ditto with E2L group, plus whole language method involving listening, discussion, re-telling. E2L group made biggest accuracy gains, PR group biggest comprehension gains.
Winter	1985	PA	8-11 RTD	6 [6]		Pre-post IRIs showed no ss differences between groups		Parents trained in PR and Pause Prompt Praise & self-selected to treatment condition - contamination likely. PR group all Read Alone. Subjects also had XT tutoring.
BAND 3								
Burdett	1985	PA+AV AV	8-11 RTD	16 (16) [16]	8	Widespan	Raw Score Gains were compared.	Two participant schools; random allocation to conditions in one, headteacher selection in another. AT condition = modification of Pause Prompt Praise. Each group half PA & AV, half AV only. PR and AT gains significantly larger than CG, but not from each other. Error rate analysis markedly favoured PR group.

Grigg	1984	PA	8-11 RTD	14 (14) [14]	9	Neale	6.0 (5.5) [7.5]	2.88 (2.65) [3.61]	all increased; largest gain by control group.		Allocation random. AT condition = Parent Listening with delayed error indication & varied prompts. No s.s. differences, control group also showed high gains. Possible contamination between groups acknowledged. Parents persuaded to participate, not volunteers.
Jones	1987	PA	6-7 MA	21 [21]	10	Neale	8.0 [6.0]	3.47 [2.60]			Random allocation. AT = Parent Listening. Not all "PR" group attended training. Differences not s.s.
Thirkell	1989	PT	tutees RTD tutors MA	18+18 [15]	6	Mac-millan Analysis of RA	7.8 5.7 [6.1]	5.60 4.12 [4.41]	9.0 11.9 [8.2]	6.50 8.59 [5.92]	AT = extra individual silent reading practice. No s.s. differences. Allocation by reading age differential; CG approximately matched by ability to participants. In a multi-cultural school. Large CG gains also.

BAND 2/1

Dening	1985	PA	5-9 MA	30 (10) [30]1 [30]2	11	Neale	7.5 (3.1) [7.1] [6.8]	3.00 (1.20) [2.80] [2.70]	7.6 (4.1) 11.7] 11.0]	3.10 (1.60) [4.70] [4.40]	AT1 = Pause Prompt Praise, AT2 = Parent Listening with word supply at error. All participant groups had other "reading homework". Subjects from 3 schools. Allocation age-stratified random. Possible contamination acknowledged. All 3 treatments significantly different from control group, but not each other, in accuracy. In comprehension, AT groups significantly different from control, PR not.
Joscelyne 1 1989		PT	10 tutees RTD; tutors able readers	11+11 [11+ 11]	7	Neale	10.4 [2.9] 6.6 [3.9]	6.41 [1.80] 4.11 [2.42]	16.5 [6.6] 8.9 14.7]	10.18 [4.05] 5.52 [9.11]	Allocation of tutees by matching groups then randomly to condition. Subjects drawn from 3 classes in 2 schools. AT = Listening. Differences between groups ss for tutees, not for tutors. ? Ceiling effect on tutor scores.
Joscelyne 2 (replication)	1989	PT	9-10 tutees RTD; tutors able readers	11+11 [11+ 11]	7	Neale	6.3 [3.6] 0.5 0.2]	3.88 [2.25] 0.28 [0.12]	8.6 [5.7] 11.2 [8.5]	5.29 [3.55] 6.92 [5.23]	Differences between tutee groups s.s. for accuracy, not for comprehension. No significant differences for tutors. PR group showed significant improvement in reading from context. Tutor accuracy gains unusually low.

Lindsay etc.	1985	PA	8-10 RTD	10 [10]	6	Neale	9.7 [7.3]	6.47 [4.67]	14.0 [13.6]	8.10 [8.17]	Allocation by stratification by age, gender & mean retardation. AT condition = "Relaxed Reading", parent listening & praise without criticism. No s.s. differences.
Welch	1985	PAV	8-9 MA	18 [18]	11	IRI error analysis					Matched pairs, allocated randomly. AT = Listening. All differences favoured PR group but none s.s.

Notes

1. There is a great difference between the importance and impact of a study at the time of its publication and the adequacy of its research design years later when seen in the context of a large volume of relevant literature.

2. The nature and intensity of training and follow-up in these studies were very varied, and in some studies it was a manipulated variable. However, these showed no consistent relationship to the size of outcome gains. Many other confusing variables were doubtless at play, given the heterogeneity of the background literature.

3. In control/comparison group studies, it is evident that peer tutor projects are more likely than parent tutor projects to show good research design. This is unsurprising, since peer tutor projects are typically contained on one site and therefore more experimentally controllable. However, results from high quality controlled studies, which are as likely to be cross-age peer tutored as parent tutored, cannot be assumed to be generalisable to parent tutored projects in general, or indeed to same-age peer tutored projects.

4. Allocation to conditions by matching pairs then allocating members of pairs randomly to conditions would appear to be the most satisfactory for this application.

5. In control/comparison group studies, there is little evidence that better quality studies were less likely to yield outcomes favouring Paired Reading. The mean Ratio Gain difference between experimentals and controls was very similar for bands 1,2 and 4, with bands 3 and 5 markedly lower. Glass et al. (1981) note that in the typical meta-analysis there is no strong relationship between the quality of the study and the average size of the effect obtained.

6. Design and execution quality of comparative studies (Section C) were generally less good - more flaws were evident, but arguably these workers were attempting something more difficult. Again, peer tutor studies tended to be of better quality than parent tutored studies, although here (interestingly) same-age peer tutoring very much pre-dominated, rather than cross-age. As in Section B, there was little evidence that quality of study related to outcomes, the mean Ratio Gain difference between experimental and alternative treatment groups proving very similar for bands 2/1 and 4 while being much lower for bands 3 and 5, while of course also being lesser throughout than the differences for the experimental/control comparison. The few (3) studies finding Paired Reading statistically significant better than the alternative treatment were found in bands 4 (n = 1) and 2/1 (n = 2).

7. Throughout, the estimation of the quality of a study was very dependent on the amount of detail and specificity in the written report. Studies lacking written detail might have been less likely to be relegated to a lower quality band as any design flaws were unreported - and arguably this is unlikely to be compensated for by any equivalent tendency towards under-reportage of design strengths.

8. Compared to the previous literature, the Kirklees study had the design strength of a higher degree of homogeneity between constituent projects, especially with respect to purity of technique, training, monitoring and other aspects of project organisation, together with the advantage of a very large aggregate number of participants. Additionally, the Kirklees project generated baseline and follow-up data hitherto very scarce in the literature, together with novel data of types not previously reported.

PREVIOUS RESEARCH - PROCESS, READING STYLE AND FOLLOW-UP

Process Studies

Relatively few studies of Paired Reading have reported detailed information on the behaviour of participants subsequent to training during involvement in projects. Thus what actually occurred during PR sessions is not known to the external enquirer. It cannot be assumed that participant behaviour was standard throughout. Morgan and Lyon (1979) collected detailed baseline and post-training data on the percentage of words read which were verbally reinforced by parent tutors. In the 4 participating pairs, percentage of words verbally reinforced rose from 0 during baseline to between 50 and 75% for participants subsequent to training, which took place on a one-to-one basis during several lengthy sessions amounting to between 3 and 4.5 hours in total.

Bushell et al (1982) in Derbyshire completed observational checklists relating to parent and child behaviour while observing pairs in action during follow-up home visits (this is reported in more detail in Miller, 1987). Observational checklists on elements of the Paired Reading technique covered Reading Together (synchrony, parental adjustment of pace, child attention to each word, parent allowing time for self-correction, parent re-modelling errors) and Reading Alone (child signals, parent responds, child praised, parent indicates minor errors, return to Reading Together after 4 seconds, child praised regularly), as well as checking whether reading material was chosen by the child and whether parents avoided negative and anxiety-provoking comments.

Checklists were completed subjectively and no data on inter-rater reliability is given. Checklists completed were grouped into overall "high" and "low" quality of Reading Together and Reading Alone.

For Reading Together, 44 checklists were rated as high quality and 10 as low quality. There was widespread difficulty indicated with parental praise for signalling and for independent reading, so the praise item was ignored in adjudicating "quality". For Reading Alone, 37 checklists were judged high quality and 17 low quality. However, comparison on specific checklist items between the high and low quality groups indicated differences reaching statistical significance in only one case (return to Reading Together after 4 seconds during Reading Alone). Only 4 aspects of the behavioural process of use of the Paired Reading technique correlated with reading accuracy gains on reading tests: quality of independent reading (+ 0.27), percentage of words read independently (+ 0.25), the quality of simultaneous reading (+ 0.10) and the total time spent on Paired Reading (less than 0.10). Statistical significance of coefficients is not given, but the last two are unlikely to be educationally significant. The Derbyshire workers were surprised by these results, as they had thought Reading Together would be the more important aspect of the process, and they subsequently speculated that Reading Together was important in the elimination of parental criticism, thereby having an impact on all aspects of subsequent parental behaviour.

A more detailed study of 10 participants in the Derbyshire Project was conducted by Toepritz (1982), who audio-taped 3 follow-up home visits

for each participant. The tapes were then analysed with respect to the Derbyshire "Checklist of Elements of Paired Reading", and an inter-rater reliability of 73% is cited. Over the time span covered by the 3 consecutive home visits, the percentage of time spent on independent reading within pairs rose, but this was not found to be related to reading age gains to a statistically significant degree. The quality of Reading Together was found to be very various, but this did not appear to be related to reading age gains either. No correlations achieved statistical significance, largely a function of the small n, the largest correlation of 0.44 being between time spent in independent reading and reading accuracy gains.

Elliott (1989) conducted post-hoc interviews with parents who had participated in Paired Reading projects and made audio recordings of some families, in a pilot project including nine interviewed subjects and a main project in a different school including thirteen interviewed subjects. The participating children were mixed ability 6 to 7 year olds. In the main study, 15 of 30 parents had been "Listening" to their children read prior to the Paired Reading project. Post-training, 17 of the 30 parents did not use the Paired Reading technique "perfectly". Two pairs did only Reading Together, two only Reading Alone, three had difficulties Reading Together and six tended to switch from Paired Reading to "Listening" as they went along. In four cases pairs did not continue because the child rejected the technique and in two cases because the parents rejected the technique. As time went on, there tended to be more reversion to "informal listening". Elliott (1989) concludes that in many cases the Paired Reading technique is

integrated with a pre-existing method, although the interview data supports the view that Paired Reading results in reduction of stress in the reading relationship and that the error correction proceedure does result in the retention of sight vocabulary. It should be noted that the degree of conformity to "pure" technique was much greater for participants in the pilot scheme.

Turning to process studies of peer tutored Paired Reading, Limbrick et al. (1985) collected very detailed process data on 3 pairs in which tutors were aged 10-11 years and tutees 6-8 years, utilising a minor modification of the Paired Reading technique. Pre-training baseline measures and post-training measures were made of: amount of discussion, praise for correct responses, praise for independent reading, attention to errors, supplying of unknown words, eliciting positive responses, and avoiding negative comments. Each pair was observed weekly, but no data on inter-rater reliability is given. Post-training, substantial increases in praise for both correct responding and independent reading were evident, together with increases in prompting to elicit the correct response from the tutee. Amount of attention to error showed some small increase, but amount of supplying unknown words and amount of negative comments stayed the much the same.

Winter (1988) conducted a process analysis based on audio-recordings with 18 pupils participating in projects in two schools. However, a disproportionate number of subjects were included from one school which showed substantially poorer outcome results on reading tests, and the selection of subjects for process data collection was far from random.

Inter-rater reliabilities cited ranged from 0.28 to 0.93, some of these being unacceptably low. Measures were taken of the number of errors corrected, the numbers of errors uncorrected and the amount of positive verbal reinforcement. Attempts were made to collect data with reference to other measures but it proved impossible to do this reliably. Winter (1988) reports that the mean use of praise was less than one in 200 words (less than twice in 5 minutes), and 6 pairs used none at all. It is also reported that uncorrected errors outweighed corrected errors in a ratio of 4:1. Pairs were however "uniformly conscientious" about using modelling for error correction and this method accounted for 98% of the error correction observed. Considerable consistency of participant behaviour across observational sessions is reported, and it is noted that correlations between process measures and reading age gains failed to reach statistical significance.

Joscelyne (1989) notes that in her peer tutored Paired Reading projects there was a tendency for "pairs (to) drift into other methods of reading", and close monitoring was necessary to ensure adherence to the Paired Reading technique. Joscelyne concluded that "Paired Reading - when carefully monitored - is more useful in accelerating children's reading than simply Listening". This finding, based on a different research design than that of Winter (1988), contradicts Winter's findings, which are obviously based on projects where detailed process data show monitoring of participant behaviour has been ineffective. Joscelyne (1989), however, gives no detailed process data. Winter's findings also conflict with those of Limbrick et al. (1985).

In both the parent tutored and peer tutored process data, many contradictory findings are evident and much of the data are so unalike as to be non-comparable. It is obviously possible for participants to manifest the required process behaviour but this would appear to be more likely in studies of smaller numbers of participants, especially when the training has been more detailed. In larger field studies of parent-tutored Paired Reading, conformity to good technique has been found in from 75% to 43% of participants, the higher figure being associated with home visits. Given the paucity of process research, the relationship between process (e.g. "purity" of technique) and outcome remains obscure. The vast majority of studies (including the Kirklees study) have evaluated on a crude input-output model. Output parameters may therefore more reflect the structure of service delivery (training and follow-up) than the impact of a particular technique which is assumed to have been applied. There are implications here for further research into the cost-effectiveness of various approaches to training and follow-up, as well as good quality studies of process x outcome interaction.

Reading Style Studies

A number of studies utilising the Neale Analysis have measured changes in Rate of reading on the test passages on a pre-post basis. In some of these studies (eg. Lindsay et al. 1985) a reduction in the rate of reading at post test after Paired Reading was found, although in other studies (eg. Winter, 1985) an increase in rate of reading is reported (of 17%, in this case). The measuring of the rate of reading using the Neale Test has thus yielded various results, in contrast to the

measurement of the rate of reading on samples of text specifically selected for the purpose from a variety of sources in individual studies, which latter will be referred to in greater detail below. Most reading style studies have applied some form of miscue or error analysis on a pre-post basis utilising parallel but different texts of similar readability on the two occasions.

Four studies report reading style change data from parent tutored projects. Bush (1982) applied the miscue analysis structure proposed in the Neale Test to 7 participant and 18 comparison children aged 9 - 11 who were at least one year retarded in reading. Miscues of the control group showed little change from pre- to post-test, while the miscues of the participant Paired Reading Group showed a reduction in refusals from 58% to 31% and an increase of 19% in substitutions. Paired Readers also showed substantial increases on the Daniels and Diack Tests of phonic skills, but the difference between participant and control children did not reach statistical significance. Differences between participant and control groups on tests of visual and auditory sequential memory likewise showed no statistically significant differences.

Four participants in the Derbyshire project were investigated in detail by Scott (1983), who utilised pre and post measures on the Aston Index together with a miscue analysis on a set passage. Only one subject showed any improvement on the Aston sub-tests, and his gain in reading age was minimal. The miscue analysis showed an increased tendency among the Paired Readers towards the use of contextual cues, although this was not regular or predictable.

Winter (1985) collected reading style data on 10 of 33 subjects involved in a project, all of whom were below average readers aged 9 - 11 years. The ten subjects were audio-taped reading a text of controlled readability on a pre- and post-test basis. In this study, parents were trained in both Paired Reading and Pause Prompt Praise. At post-test, rate of reading on a similar text had increased, errors as a percentage of words read decreased by 22%, refusals decreased by 28% and the proportion of self-corrections remained the same. There were no significant differences in changes in reading style between parents reporting having used Paired Reading and parents reporting having used Pause Prompt and Praise.

Similar methodology was used by Green (1987) on 18 of 44 children involved in a four week summer programme for the children of migrant farm workers in the United States, not all of whom were actually doing Paired Reading as prescribed. Only one reading style indicator was checked, namely, the semantic appropriateness of miscues. In 13 out of 18 cases improvements in this area were evident, and of these 13, 10 were definitely doing Paired Reading. However, the mean increase was modest (9%).

Eight studies have concerned themselves with changes in reading style in Peer Tutored Paired Reading projects. The first of these was that of Winter and Low (1984) who reported data only on their fifteen 10 - 11 year old tutees, in a same-age peer-tutored project based in one class. Different texts of similar readability were applied on a pre- and post-participation basis and students tape recorded performing upon them. The tutees' rate of reading rose by an average of 30%, error rate fell

by an average of 50%, percentage of self-corrections as a proportion of all errors rose by 70% and percentage of refusals as a proportion of all errors fell from 7% to 0. On the whole, these changes were less marked for the most retarded tutees, except that for this latter group the percentage of refusals dropped even more sharply.

A project in which 11-12 year old remedial pupils in a high school were cross-aged tutored by Sixth Form pupils and incorporating follow-up data is reported by Cawood and Lee (1985) and Lee (1986). Of 22 participant tutees, reading style change data are reported for 16, the same 2 passages of controlled readability being used on both occasions. For all 16, the percentage of errors reduced, for 12 of the 16 the percentage of refusals reduced, for 10 of the 16 the percentage of errors which were self-corrected improved and 10 tutees showed an increase in speed of reading in words per minute while 4 stayed the same and 2 became slower. Follow-up data one year later were gathered on 13 tutees, dropping the easier of the 2 original texts and adding a new one of much higher readability. Over the year since pre-test, the percentage of errors had reduced by 41% on average overall, the percentage of self-corrections had increased by 135%, and the percentage of errors which were contextually relevant had increased by 100%. Lee (1986) concluded that there was evidence that changes in reading style accruing from a brief cross-age peer tutor project showed no signs of wash-out at long-term follow-up even though no tutoring had occurred in the interim.

Limbrick et al. (1985) deployed 3 tutors aged 10 - 11 years with 3 tutees aged 6 - 8 years, all of whom were retarded in reading. The

tutors and tutees read graded passages taken from classroom reading materials and answered comprehension questions based upon them. Two tape recordings were made weekly and measures of reading accuracy and proportion of self-corrections taken from them. In addition, cloze comprehension exercises were completed weekly by participants, and the syntactic and semantic appropriateness of clozes was assessed. During the course of the participant period, there was evidence of a rapid rise in reading accuracy and self-correction, a rapid recovery from the impact of change to more difficult materials being apparent. The proportion of appropriate substitutions rose and the proportion of correct responses to comprehension questions did likewise. This was true for both tutors and tutees.

Results contrary to the general tendencies evident in other studies were reported by Lees (1986, 1987). Ten Paired Readers aged 10 - 12 years who were on average 2.8 years retarded in reading were compared to a similar non-participant group and to a non-participant group of 8 - 9 year old average readers on repeated sub-skill tests. Assessment was made of word pronunciation, non-word pronunciation, semantic appropriateness, lexical appropriateness, visual matching, phonological segmentation and use of context. Although the Paired Reading group showed the largest increase in reading age, there was no evidence of an increase in the use of context by this group. In fact, there was some evidence to suggest an improvement in decoding skills, by "phonic or direct visual access". The non-participant group of elderly retarded readers were found at pre-test to use context as much as the younger average readers.

Low et al. (1987) used 7 graded reading passages covering a wide range of readability with their 13 tutorial pairs and 26 control children. The error rate of the tutors reduced by 71% compared to a control group reduction of 59%, and the error rate of the tutees reduced by 50% compared to a control group reduction of 42%. The participant group changes were more pronounced for girls than for boys.

Two studies are reported by Joscelyne (1989), the second being a replication of the first, both comparing a group of Paired Readers with another group where the tutors merely "Listened". In the first study, both groups showed a small reduction in the number of errors. The Paired Reading group showed a 15% increase in the proportion of errors which were substitutions, while the Listening group showed no change, and this difference was statistically significant. The Paired Reading group showed a reduction of 6% in refusals while the Listening group showed an increase of 5% in refusals, and this difference reached statistical significance. There were no statistically significant differences between the two groups on the proportion of substitutions which were grapho-phonemically appropriate or contextually appropriate. Likewise, on a phonics test, the number of errors reduced for both groups but the difference between them was not statistically significant. In the replication with 11 Paired Reading pairs and a similar number of Listening pairs, two passages matched for readability were utilised, and subjects were tested on key words appearing in the passages first in isolation and subsequently in context. The difference between words read in isolation correctly and words read in context correctly was calculated for both groups, to constitute an index

of change in the use of contextual information. The Paired Readers showed an increase in the extent to which they read more words correctly in context than in isolation, but this was not true of the Listening group, and this difference reached statistical significance. These findings applied to both tutors and tutees.

One study has considered changes in reading style as a result of Paired Reading tutoring by teachers (Welch, 1984), and here again a Listening group was compared to a Paired Reading group. Measures were taken of the use of context, comprehension, rate of reading and a number of refusals. The Paired Reading group did better on all measures than the Listening group, but none of the differences reached statistical significance.

Considering parent, peer and teacher tutored studies together, in five of these studies error rates have been found to reduce in Paired Readers and in no cases have error rates increased. In six cases, Paired Readers showed decreases in refusal rates and in no case an increase. In six cases, use of context showed an increase, in one case no difference was found and in no case was there a decrease. In four cases the rate or speed of reading showed an increase and in no case was there a decrease, (but it should be noted that studies utilising the Neale Analysis as a measure of rate of reading have yielded increases and decreases). In four studies, self-correction rate showed an increase and in no case a decrease. In three cases the use of phonics showed an increase and in no case was there a decrease. Although many of the differences cited did not reach statistical significance and only a few

studies used control groups or comparison groups who were non-participant or utilised another technique, nevertheless strong consistent trends emerge from all these studies considered together. The general pattern is of Paired Reading resulting in fewer refusals (greater confidence), greater fluency, greater use of the context and a greater likelihood of self-correction, as well as fewer errors (greater accuracy) and better phonic skills.

Follow-up Data

Follow-up data gathered some time after the end of the intensive period of projects have been reported in 5 studies, 2 wholly parent tutored, one cross-age peer tutored, one deploying a combination of natural parent, cross-age peer and adult volunteer tutors, and one incorporating tutoring by professionals.

Bushell et al. (1982) reported six month follow-up data on an unspecified number of subjects in the pilot Derbyshire study. Considerable differences were evident between three different participating schools, but for reading accuracy on the Neale Test, the children from two schools appeared to plateau after the intensive period, while in a third school the participants continued to improve their reading test scores at the same accelerated rate as was evident during the intensive period. In reading comprehension on the Neale Test, subjects at one school had regressed on average at follow-up, although not back to the pre-test level, subjects at a second school maintained progress at normal rates (i.e. their intensive period gains showed no sign of wash-out), while a third group maintained accelerated

progress at less than the pre-post rate but at greater than normal rate. It was not known if families continued to do Paired Reading after the end of the intensive period - they were certainly not given any specific encouragement to continue. The authors conclude that, with respect to follow-up gains, the "nature of school involvement is an important factor".

A study of just five 10 year old weak readers by Lees (1985) also used the Neale Test at follow-up 13 weeks after the end of the intensive period. However, in this case the initial intensive period involved tutoring by a teacher, while during the 13 week follow-up period parents were trained to continue with Paired Reading for a minimum of 5 sessions per week during the 13 weeks. During the intensive 10 week period of teacher tutoring on a twice weekly basis, participants made average Ratio Gains of 3.2 in reading accuracy and 1.3 in reading comprehension. During the subsequent 13 week period of parental tutoring the subjects made mean Ratio Gains of 2.6 in accuracy and 2.1 in comprehension. Albeit with small numbers, this study thus demonstrated that continuing gains were possible with continuing input.

A longer term follow-up, 46 weeks after post-test and a full year after pre-test, was reported by Carrick-Smith (1982, 1985) in a project involving pupils from 3 high schools who at pre-test were 11-12 years old and up to 3 years retarded in reading. Tutoring was variously by parents, cross-age peer tutors and adult volunteers. Separate follow-up results are not given for the 3 tutor groups, and the composite follow-up sample was 27 subjects. On Neale Accuracy, over the 10.6 month

follow-up period, participant children made further mean gains of 8.1 months of reading age while controls made 6.6 months. In Comprehension mean participant gain was 10.2 months and mean control gain 6.8 months. Thus gains in both accuracy and comprehension for participant subjects were greater than for control or comparison subjects, but even participant gains during the follow-up period were less than "normal". However, there was great variance between subjects and between schools in the follow-up data, in both reading accuracy and reading comprehension. As with the Bushell (1982) study, one school contributed disproportionately to the experimental gains at follow-up.

Lee (1986) reported 12 month follow-up data on 13 of 22 participants in a cross-age peer tutored Paired Reading project in which the tutees were high school remedial pupils aged 11-13 years. On the Daniels & Diack Test 12, the participant tutees gained 1.2 years of reading age during the pre-post participant period and a further 0.5 years during the post-test to follow-up test period, while the control subjects gained 0.7 years from pre- to post- test and only a further 0.2 years during the follow-up period. Total gains from pre-test to follow-up test for participants were thus 1.7 years of reading age on average, and only 0.9 years for control children. As in the Carrick-Smith (1982, 1985) study, gains during the follow-up period were less for both groups than would normally be expected, although the differential favouring the participant group remained. Lee (1986) also reported follow-up data for participant tutees only on changes in reading style, based on error analysis of oral reading of passages of variously controlled readability. As previously mentioned, over the whole pre-test to follow-up test period, errors reduced by 41%, self-corrections increased

by 135% and contextually appropriate errors increased by 100%. Lee concluded that the changes in reading style among the participants evident at post-test endured and were consolidated through to follow-up one year later, even though no further tutoring had occurred in the interim.

Burdett (1985, 1986a, 1986b) gathered follow-up data just 4 weeks after the end of the intensive period, in a study in which children aged 8 to 11 approximately 1 year retarded in reading were tutored by professionals. The Widespan reading test and an analysis of error rate were the outcome measures, and sub-samples also experienced Paired Reading at home with natural parents, while other sub-samples did not. Burdett found that both experimental groups (parent-involved and not-involved) had retained highly significant gains over control groups, whether using the Paired Reading technique or the "Individualised Reading" approach based on Pause Prompt and Praise. On average, experimental subjects made 3 times more progress than control subjects. The error rate of the Paired Reading group at follow-up had shown further decrease by 25% while the Individualised Reading group showed a further decrease of 11%, compared to an increase in the control group of 9%. Concerning reading style, Burdett reported that the diagnostic indicators on the Widespan test suggested that Paired Reading resulted in improvements in both decoding and psycho-linguistic capabilities, while the Individualised approach resulted in increases in decoding skills only. The additional benefit of parental involvement became increasingly apparent at follow-up testing.

In summary, it seems clear that even within the same study follow-up gains may vary considerably from school to school. Continued acceleration at above "normal" rates is relatively rare, and indeed some follow-up gains cited are less than normal rates, while still remaining better than those of control or comparison groups. Follow-up periods have been very various, ranging from 4 weeks to 12 months, but the length of follow-up does not appear to consistently relate to the favourability of follow-up findings. The standard of the studies is not high; however, there is relatively little suggestion here of washout of experimental gains, this being reported for only a small number of subjects in one school in one study. ("Wash-out" is defined here as a decline in rate of acceleration to below pre-project levels, resulting in overall "normal" progress or less over the total period (baseline/project/follow-up). It is considered unrealistic to expect acceleration at above normal rates to continue indefinitely. Thus, herein "washout" refers to erosion of relative gain, not of relative acceleration, at follow-up). There is evidence that acceleration can be sustained and even increased with the deployment of different types of tutor consecutively, and that changes in reading style can also endure in the long-term.

Other Features

A small number of studies have incorporated an attempt to measure changes in self-concept by means of a paper-and-pencil instrument, but results have been extremely erratic and sometimes wholly implausible (e.g. Carrick-Smith, 1982, 1985). A larger number of studies have incorporated some information regarding client satisfaction as indicated

by recorded verbal responses or questionnaire completion. Unfortunately the many different formats in which this feedback is expressed renders it impossible to summarise. In many cases proportional feedback rates or questionnaire return rates are not given so it is impossible to gauge how representative of the whole participant population the cited feedback might be. In any event, informal verbal feedback notoriously suffers from the "grateful testimonial" effect, and cannot be considered a reliable outcome indicator.

If children "learn to read by reading", one factor in the effectiveness of Paired Reading (or any other parental involvement in reading scheme) might be expected to be the influence of extra reading practice alone. Thus, other things being equal, more time spent doing Paired Reading should be associated with greater gains in reading skill. Some workers have explored this relationship. Bushell et al.(1982) reported very small correlation coefficients between reading accuracy and comprehension and time spent reading during a PR project (Accuracy 0.15, Comprehension 0.33, n = 19), and this finding was repeated in their later and larger controlled study (Accuracy 0.008, Comprehension 0.142) (Miller et al., 1986). Carrick-Smith (1982), Wareing (1983) and Dening (1985) likewise found no statistically significant correlation between total time spent doing PR and test gains, although Dening (1985) did report a positive correlation (0.43) between frequency of tutoring and gains made, which was also true of the Pause Prompt Praise group but not the Listening group, but Morgan and Gavin (1988) found no such significant relationship. However, few of these studies utilised very adequate measures of "time on task", tending to rely on participant self-recording.

A few studies have included other measures in the attempt to demonstrate peripheral benefits from Paired Reading. One of the few successful attempts was that of Jungnitz et al. (1983) who noted increased scores on the English Picture Vocabulary Test when 21 non-reading 7-year olds were involved in PR over a long period. The mean standardised score rose from 95.9 to 99.9 (in a highly disadvantaged area), the difference attaining statistical significance. Evans (1984) similarly noted large gains on the British Picture Vocabulary Scale for 6 "dyslexic" High school subjects.

Some workers have utilised Paired Reading in combination with other approaches in the context of an intervention, and then been unable to demonstrate which aspect(s) of the project resulted in which elements of overall gains made. An example in the work of Young and Tyre (1983), who deployed a number of variations on Paired Reading sequentially and/or simultaneously according to the needs of individual participants together with parent-tutored writing and spelling activities and a "holiday school". Over one year, the control group advanced 0.8 years on a reading test, while a "dyslexic" experimental group advanced 1.8 years and a 'remedial' experimental group 2.0 years.

Cooknell (1985) involved parents in workshops where they were taught a simple "Listening" method, Pause Prompt and Praise and Paired Reading, as well as a more complex 'linguistic' approach focusing on comprehension. Cooknell notes 'it could be that we asked too much of the parents" and gains on reading tests were modest. Ripon et al. (1986) combined the Reading Together aspect of PR with Datapac Reading,

a precision teaching approach. Baseline, pre-post and follow-up data from the Spar and BAS reading tests showed marked acceleration from baseline during the project period which was sustained at short-term follow-up for 3 out of 4 data sets.

Holdsworth (1985, 1986) reported on workshops based in a special school for children with learning difficulties, in which parents were taught Reading Together, Precision Teaching and Direct Instruction (DISTAR) approaches, to mesh with the curriculum of the school as experienced by their children. A mean ratio gain of 2 over a long period of intervention is cited, but it is uncertain to what this improvement can be attributed. By comparison, Sweetlove's (1987) aforementioned study is considerably more informative.

Also of interest are attempts to apply Paired Reading to "special" populations. These have included the deployment of PR with pupils in special schools for children with moderate learning difficulties, severe learning difficulties, and behavioural difficulties, in Kirklees and elsewhere. Reports on the use of PR with families of South Asian origin, often where parental skills in speaking and reading English are limited, have been provided by Jungnitz (1984, 1985), Bush (1985), Welsh and Roffe (1985) and Vaughey and MacDonald (1986). A training video demonstrating Paired Reading on single and dual-language texts in English, Urdu and Gujerati has been produced (Topping and Shaikh, 1989). Bush (1985) reports interesting data on relative take-up rates from ethnic groups in a multi-ethnic school. Highest take-up was from Afro-Caribbean families and lowest from white families, with Asian families involving themselves just a little more (proportionately) than whites.

Among the reports of usage of Paired Reading in the USA is one by Ulmer and Green (1988), who trained parents who were migrant summer farm workers in rural areas of Vermont to use the technique with their children while supported by peripatetic teachers during a (moving) "summer school". No numerical data were cited, but this population was found challenging in terms of service delivery.

Interesting work based on very large numbers of children involved in reading activity with parents at home in Australia is reported by Rowe (1989). In a longitudinal study of the factors affecting literacy development in an aggregated sample of 3600 5-8 year olds in Melbourne, Rowe (1989) found that high levels of home reading activity impacted strongly on students' attitudes towards reading, and also were associated with a reduction in dysfunctional classroom behaviours (especially attention deficit), as well as impacting positively on criterion measures of students' reading achievement. Although much reference is made to "Paired Reading", it is uncertain whether the method(s) used constituted Paired Reading as defined in this review.

PREVIOUS RESEARCH - EFFECT SIZES AND SUMMARY

Effect Sizes

The summary of mean pre-post ratio gains for parent tutored and peer tutored projects is given in Synopsis 2, categorised according to whether the study merely compared pre-post gains of experimental groups, included control or comparison groups, or included other experimental groups to enable comparison of different techniques to be made. In all cases, only the gains of the Paired Reading participant group are given in the Synopsis.

In 20 pre-post design parent-tutored projects incorporating 333 Paired Readers, mean ratio gains in accuracy were 5.19 and in comprehension 6.82. In 13 parent-tutored control group projects incorporating 212 Paired Readers, ratio gains were considerably lower at 3.13 for accuracy and 3.91 for comprehension. In 11 parent-tutored projects comparing different techniques, incorporating 131 Paired Readers, mean ratio gains for accuracy were 3.07 and for comprehension 3.89, very similar to those for Paired Readers in control group studies. The overall mean pre-post ratio gain for the 676 Paired Readers involved in all types of parent-tutored study were 4.13 for accuracy and 5.32 for comprehension, when the ratio gains in each study were weighted by the number of subjects in each study.

The pattern of lower mean pre-post ratio gains in control and comparative than in simple pre-post studies is not repeated in the data for peer tutor projects. As yet much less data on peer tutor project outcomes are available, and numbers in some cells are still too low to

Synopsis 2 Synopsis of Mean Ratio Gains

Type of Study	Parent Tutored					Peer Tutored						
									TUTEES		TUTORS	
	Number of Reports	Number of Projects *	Number of PR Subjects with data	Mean Ratio Gain + Accuracy	Mean Ratio Gain + Comp'sion	Number of Reports	Number of Projects *	Number of PR Subjects with data	Mean Ratio Gain + Accuracy	Mean Ratio Gain + Comp'sion	Mean Ratio Gain + Accuracy	Mean Ratio Gain + Comp'sion
PRE-POST	15	20	333	5.19	6.82 (n = 245)	2	2	26	4.32 (n = 26)		4.17 (n = 26)	
CONTROL	13	13	212	3.13	3.91 (n = 165)	5	5	86	4.63 (n = 46)	7.31 (n = 15)	6.66 (n = 40)	9.10 (n = 15)
COMPARISON	11	11	131	3.07	3.89 (n = 96)	3	4	110	4.61 (n = 70)	5.15 (n = 38)	2.45 (n = 50)	4.61 (n = 50)
OVERALL	39	44	676	4.13	5.32 (n = 506)	10	11	222	4.56 (n = 142)	5.76 (n = 53)	4.29 (n = 116)	5.65 (n = 65)

* Summing all projects, Accuracy mean RG = 4.22 (n = 934), Comprehension mean RG = 5.39 (n = 624).

* Some projects covered more than one school, but results were not always given separately for different schools, so this statistic is somewhat arbitrary.

+ Weighted by the number of subjects in each study. Comprehension gains weighted by number of subjects with comprehension scores.

permit the drawing of valid meta-analytic conclusions. Mean pre-post ratio gains of Paired Reading tutees are very similar in reading accuracy irrespective of the type of study, averaging 4.56 overall for 11 projects incorporating 222 Paired Readers. For tutee comprehension the mean ratio gain is 5.76, but this is based on only 53 subjects. Considering the mean pre-post ratio gains of tutors in peer tutored projects, accuracy gains in pre-post studies are very similar to comprehension gains in pre-post studies, but accuracy gains in control group studies tend to be much higher while those in comparative studies tend to much lower. The tutor comprehension gains in control group studies also appear very high, but the number of subjects involved here is very small, and this finding may prove misleading as further data accumulate.

Considering mean pre-post ratio gains of Paired Readers in all types of study together, it is evident that on average outcomes in both reading accuracy and reading comprehension are as good for participants tutored by peers as those tutored by parents, indeed if anything the results from peer tutor projects are somewhat better. The outcomes for the peer tutors themselves are on average very similar to those for the tutees, just very slightly less good, and still better than the outcomes for those subjects who were parent tutored.

A summary of results in the literature for control group studies only has already been given in Synopsis 1 (above), in which participant and control group mean pre-post ratio gains are detailed, categorised in terms of results for parent tutees, peer tutees and peer tutors. This will be repeated here.

In 13 parent tutored control group studies, mean pre-post ratio gain in accuracy for Paired Readers was 3.13 (n = 212) while that for control subjects was 1.19 (n = 195). In comprehension, mean ratio gain for participants was 3.91 and for controls 1.93, the latter figure demonstrating again the need for caution in interpreting simple pre-post gain scores in comprehension on the Neale Analysis.

For peer tutees in 5 projects, mean participant ratio gains in accuracy were 4.63 (n = 46) and for controls 1.40 (n = 45). In reading comprehension, the difference between peer tutees and their controls was very large, but sample numbers were very small. Peer tutors demonstrated a mean ratio gain of 6.66 (n = 40) while their controls were markedly lower at 3.63, although still substantially above "normal" rates of gain. There was a similar pattern for peer tutor reading comprehension, but sample numbers were again very small.

A summary of Effect Sizes is given in Synopsis 3, categorised into parent-tutored projects, peer-tutor projects (with separate data for tutees and tutors where available), and studies comparing different techniques, almost all of which were parent-tutored. The effect size indicator used was Glass's Delta (Glass et al., 1981), which is discussed in greater detail later. This effect size indicator standardises the mean gain of the experimental group by subtracting from it the mean gain of the control group and dividing by the standard deviation of the control group. It is thus impossible to compute this Effect Size indicator if the standard deviation of the control group gains is not given in the published report, or if the raw data are not

Synopsis 3 Synopsis of Effect Sizes

Study	n	ES (Accuracy)		ES (Comprehension)	
Parent Tutored					
Heath 81	16	1.75		0.74	
Bush 82	18	2.20		1.14	
Carrick-Smith 82	28	0.75		0.42	
Morgan & Gavin 88	7	2.72		1.63	
Peer Tutored		tutees	tutors	tutees	tutors
Limbrick et al 85	3,3	10.00	4.30	4.01	1.84
Crombie & Low 86	12,12	2.65	1.38	1.53	0.63
Low et al 87	13,13	0.68	1.04		
Low & Davies 88	12,12		0.10		
Comparative Studies (almost all parent tutored)					
Jungnitz 84	11	1.96		4.56	
Dening 85	30	1.00		0.72	
Burdett 85	8	1.42	(parent & teacher tutoring)		
	8	1.15	(volunteer tutoring only)		
Sweetlove 87	24	0.76		0.69	

Mean Effect Size: Parent Tutoring	1.57 (σ 0.67)	1.41 (σ 1.33)
Mean Effect Size: Peer Tutors	1.71 (σ 1.57)	1.24 (σ 0.61)
Mean Effect Size: Peer Tutees	4.44 (σ 4.01)	2.77 (σ 1.24)
Mean Effect Size: Overall	2.12 (σ 2.26)	1.63 (σ 1.33)
Median Effect Size: Overall	1.40	1.14
Effect Sizes: Range	0.10-10.00	0.42-4.56

given in a way that enables computation of this. Six out of 18 studies incorporating control groups did not include information about variance in control group gains (and indeed one did not include information about size of mean gain in the control group). Synopsis 3 therefore relates to the 12 studies including non-participant control or comparison groups where the necessary data were available (number after author name is date of study). The mean effect sizes summarised at the foot of Synopsis 3 are not weighted according to the numbers of subjects in each study and are therefore skewed by the large effect sizes in the very small Limbrick et al. (1985) study. It should also be remembered that different reading tests were used in different studies.

The mean effect size for parent tutees, taken from pre-post parent-tutored projects and from those studies comparing techniques which incorporated parent-tutoring, averaged 1.57 for accuracy and 1.41 for comprehension, the latter showing considerably more variability. These indicators are drawn from 8 projects of which 7 had comprehension data, incorporating 142 Paired Readers. The effect sizes for peer tutees and peer tutors are drawn from a much smaller number of projects involving a much smaller number of subjects, particularly for reading comprehension. Peer tutees show a very large effect size (4.44) for reading accuracy, but with very great variability. Mean effect size for reading accuracy for peer tutors is much smaller, but again variability is considerable.

The overall effect size for reading accuracy in these 12 control group projects irrespective of nature of tutor and including results from peer tutors and tutees was 2.12, and for reading comprehension was 1.63. How meaningful this is, given the great variability in effect sizes in

general, is debatable. A further issue concerns how representative these 12 studies are of the 18 control group studies in the literature, and how representative the gains in control group studies are of the gains in all projects including those without control groups. However, the mean effect sizes evident here are large when compared to those cited in other meta-analytic reports (e.g. Cohen et al., 1982). This would be true even of the median effect size of 1.40 for accuracy and 1.14 for comprehension, computed to give an alternative perspective on arithmetic mean effect sizes which are easily inflated by a few abnormally high results from individual projects. The issue of meaningful comparison of effect sizes will be discussed in greater detail later, with respect to the outcome data from the current study.

Summary

Mean pre-post ratio gains for all 55 projects yielding norm-referenced data were 4.22 for accuracy (n = 934) and 5.39 for comprehension (n = 624), including outcomes for peer tutors and tutees. In 18 control group studies, mean experimental accuracy ratio gain was 3.84 (n = 298), the equivalent control figure being 1.56 (n = 278). In comprehension, mean experimental ratio gain was 4.57 (n = 195) and control ratio gain 2.29 (n = 181). In the 12 control group projects for which requisite data were available, mean effect size for reading accuracy was 2.12 (n = 218) and 1.63 for comprehension (n = 164), (median = 1.40 and 1.14 respectively).

On 16 studies comparing Paired Reading with other techniques, only 4 found Paired Reading markedly superior (to "Listening" in 3 cases, 2 of which were peer tutor projects). No study found Paired Reading

inferior. The general finding was that Paired Reading had similar effects to other methods. In many studies sample size was so small that finding statistical significance was unlikely. However, in the 11 studies yielding adequate norm-referenced data, the mean pre-post ratio gain for Paired Readers was 3.74 and that for other techniques 2.25 (peer tutor results included). Some studies report contamination between groups supposedly using different techniques, but some process studies have suggested that quality of Paired Reading technique is not necessarily related to reading test gains of individual subjects. Process studies have shown very variable degrees of conformity to the Paired Reading technique in different projects and different schools. However, in most studies there is evidence that training does result in some changes in tutor behaviour in the required direction. In larger field studies of parent tutoring of Paired Reading, conformity to good practice has varied from 75% to 43% of participants.

Studies of reading style have consistently shown that involvement in Paired Reading is associated with a reduction in error rate and refusal rate and an increase in the use of context and self-correction. There is evidence that phonic skills also increase. Rate or speed of reading has usually been found to increase, although studies using the Neale measure of rate have also found decreases on occasion. There is a relative paucity of follow-up data, that is available relating to follow-up periods ranging from 4 weeks to 1 year. Again, results are various and seem to depend upon the individual project or school, but overall there is little evidence of Paired Reading gains being "washed-out" by subsequent deceleration to below "normal" rates. In some projects, Paired Readers continued to make accelerated rates of gain

after the end of the intensive period for as long as 6 months, especially where tutoring continued. Even where the follow-up was over a period as long as a year and rates of gain during the follow-up period were below "normal", participant rates of gain were nevertheless better than those of control subjects. There is also some evidence that changes in reading style resulting from Paired Reading involvement can endure for as long as one year.

METHOD - SERVICE DELIVERY

The Kirklees local education authority services an area in the southern part of West Yorkshire which has two large centres of dense urban population, a number of smaller towns and also considerable upland area of low population density. Socio-economically, the area is correspondingly various, but there are pockets of severe deprivation within the urban concentrations. The total population is almost 400,000, making Kirklees the seventh largest of the Metropolitan districts. The school child population (including nurseries and sixth forms) was almost 62,500 as of January 1989.

The Department of Education and Science (1982) compared LEAs on 6 socio-economic indicators drawn from the National Dwelling and Housing Survey of 1977/78, and noted that Kirklees was one of the 16 LEAs (16.7% of the total) who were "above average on all indicators". Subsequently, data from the 1981 Census, analysed by the Department of the Environment, confirmed the picture of socio-economic deprivation: in West Yorkshire only the Bradford LEA had greater problems. Indicators relating to overcrowding and proportion of New Commonwealth Households were particularly high in Kirklees, and mortality was the highest in West Yorkshire. Over 18% of Kirklees' 758 Enumeration Districts fell within the worse 10% of Enumeration Districts in England and Wales. Of Kirklees Enumeration Districts, 7.8% fell within the worst 2% of all Enumeration Districts and were categorised as "extremely deprived" by the Department of the Environment (Vane, 1984).

In January 1984, at the inception of the project upon which this study is based, the Kirklees LEA directly maintained 236 schools, varying greatly in nature. In the Secondary sector, there were nineteen 11-18 comprehensive high schools, four 13-18 comprehensive high schools, one 11-18 grammar school, two 11-18 secondary schools, and three 16-18 sixth form colleges.

In the Primary sector there were 8 separate nursery schools, 39 first schools for pupils aged 5 to 9 or 10 years, 3 of which included nurseries, 43 infant schools for children aged 5 to 7 or 8 years, 3 of which included nurseries, 4 first and middle schools for children aged 5 to 12 years, 3 of which included nurseries, 29 junior schools for children aged 7 to 11 years, 18 middle schools for children aged 8, 9 or 10 to 12 or 13 years and 52 junior and infant schools for children aged 5 to 11, of which 3 included nurseries. In the Special Needs sector, there were 5 schools for children with moderate learning difficulties, 4 schools for children with severe learning difficulties, 2 schools for children with behavioural difficulties, 1 remedial centre, 1 peripatetic support service for pupils with behavioural difficulties and a school for physically handicapped children (which closed almost immediately as provison for physically and sensorially handicapped children moved into integrated units in ordinary schools). There were thus 29 schools in the secondary sector, 193 in the primary sector and 14 in the Special Needs sector.

In 1981 and 1982 interest in parental involvement in reading was developing in Kirklees, as was the case nationally. One junior school in particular (Deighton Junior School) had, in co-operation with the Psychological Service, developed a programme for Paired Reading Instruction by Non-Teachers (P.R.I.N.T.) (Arora and Sheppard, 1982). Encouraged by the success of this project, the Psychological Service proposed that the LEA seek Urban Aid funding to establish a support service on a larger project basis to encourage and enable other schools to develop similar kinds of work. This was effected and the work on which the present research is based commenced in November 1983 with the appointment as Project Leader of the current author. The original proposal was to target only children in junior and middle schools with reading difficulties, who would receive home visits from centrally based teaching staff, allowing the inclusion of 120 families per year in the project.

In the event, the project was to include many more families per year, incorporate home visits only from teachers in neighbourhood schools and rapidly extend to include children of all levels of reading competence of a much wider age range in a wider variety of types of school. The project was established with a budget of £16,908 (inflation proofed for subsequent years) which covered 20 hours of an educational psychologist's time as Project Leader, funding for up to 20 hours a week teacher home-visiting time and monies for travelling expenses, equipment and materials. A subsequent rearrangement of financing enabled the appointment of a clerical assistant for 20 hours per week. For one year, a part-time assistant was attached to the project, funded by the Manpower Services Commission Community Programme, who, after induction and training, helped schools with training meetings, assessment and testing and occasionally with home visiting. No other funding or resources were utilised.

Of the 236 schools in the LEA, it was thought that a programme of parental involvement in reading was unlikely to be relevant to high schools catering for the 13-18 age range, grammar schools or sixth form colleges, or to nursery schools, special schools for children with severe learning difficulties and the peripatetic support service for behaviour problem children. The special school for physically handicapped children was in any event on the verge of closure. The project's 'target' schools were therefore the 21 high and secondary schools for children aged 11-18, primary schools of all types except for the separate nursery schools (185), the 5 moderate learning difficulties schools, the 2 behaviour problem schools and the remedial centre (8 "special" establishments). Thus 214 of the 236 schools in the Authority were targeted as potential clients for service delivery.

In November and December 1983, a questionnaire survey of existing practices in parental involvement in reading in the LEA was carried out, partially to make an appraisal of the "baseline" situation and partly as a first step in raising teacher awareness of the existence of the new project. The schools were not asked to submit nil returns, so the response may actually have underestimated the true situation. Of 214 schools, 51 (29%) reported having a parental involvement in reading arrangement of some sort. Of the 51, 17 (33%) reported a Paired Reading scheme and the rest a scheme of another kind (a precise and detailed definition of the Paired Reading technique was included with the questionnaire, but it is not possible to be certain that all schools reporting a Paired Reading scheme had a clear understanding of the technique.) Five schools reported only using volunteer parents in school, 6 schemes not using Paired Reading appeared to have a fair degree of structure, while the remaining 23 appeared to be very loose and informal.

The reported Paired Reading projects were spread across the primary school age range and one was in a special school for moderate learning difficulty children. Schools with a Paired Reading scheme were much more likely to provide advisory leaflets and supportive home visits from teachers than was the case with other kinds of scheme. A large number of schools sent reading books home. Home-school recording systems were used in all kinds of scheme, although they were more frequently a feature in Paired Reading schemes. Many of the responding schools also had a system for volunteers (parents, teachers, Non-Teaching Assistants or other children) to help with children's reading in school. Helpers in school seemed much less likely to be given (particularly written) advice about their task than was the case with parents helping at home (Topping, 1984).

Of the schools reporting using Paired Reading, a number had learnt the approach from Deighton Junior School, but were tending to use Paired Reading only with a very few children with reading difficulty. Of the schools who

reported using Paired Reading, a number did not subsequently make contact with the central project or use any support services from it. Of the schools reporting other approaches to parental involvement in reading, a number subsequently "converted" to Paired Reading and utilised the central support services, while an equivalent number preferred to continue with their existing form of organisation.

Information about the central support services available via the Project Leader was circulated to all target schools in writing in December 1983, January 1984 and subsequently at frequent intervals thereafter. The services were to be available on an equal basis to all schools in the Authority (in line with the LEA's Equal Opportunities Policy). The schools who were interested then invited the Project Leader to contact them. The data reported here therefore by definition have been generated by relatively well-motivated schools who chose to develop parental involvement in reading using the services offered. However, a very large proportion of the target schools operated at least one Paired Reading Project within the period under study (39%).

In order to raise the awareness of teachers in the LEA regarding parental involvement in reading in general and the existence of the project and its support services in particular, other public relations activities were initiated. In the November of each year, a substantial conference was held, attracting an equal balance of Kirklees and national delegates. Presentations and workshops were offered by Kirklees teachers and national figures. The proceedings of the conference were then written up together with other feature articles on non-professional involvement in children's reading to form the "Paired Reading Bulletin", which was distributed to all schools in Kirklees the following Spring, as well as being available nationally (and indeed internationally). Articles on the topic were written by the project leader and published in a variety of national magazines,

journals and newspapers. The intention was to enable Kirklees teachers to feel that they were part of a venture which was attracting national attention, with a view to increasing their motivation to participate. Reports were made annually to the Kirklees Education Committee, and press releases sent out simultaneously, with the same purpose in mind.

The dissemination strategy was of a number of very broad awareness-raising initiatives linking with national and international perspectives, coupled with very detailed individual consultation and support services for individual schools. No attempt was made to mount in-service training programmes for groups of schools offering information at any intermediate level of detail, since it was felt that this was likely to result in a large number of schools establishing projects without the support, and therefore the supervision and monitoring, of the project leader - a situation which would not have facilitated either homogeneity of projects between schools or the consistent collection of research data. At the request of the Education Committee, additional promotional material was sent to schools in the most disadvantaged areas on one occasion, but to little extra effect.

The initial information about central support services available from the project conformed largely to the rather narrow specification contained in the Urban Aid Grant submission. However, by early 1984 it had become evident that a wider range of services to meet the variety of needs expressed by schools was necessary. The range of services offered was therefore extended, then remaining largely similar for the duration of the period under study.

Schools were offered an initial "briefing" service, involving the project leader visiting the school, to talk about the Paired Reading technique, demonstrate it on video, leave a variety of written handouts, urge the

participant teachers to practise the technique as soon as possible, briefly discuss the organisational parameters relevant to establishing the project and answer any questions. These sessions typically took place during after-school staff meetings or during a half-day in-service session ("Baker days"), and occasionally at lunch-times. The vast majority of schools receiving these briefing sessions went on to establish at least one in-school project. The second service offered was a "planning consultation service", which involved the project leader meeting with the teachers who had elected to be deeply involved with the project, to discuss all the organisational issues involved, make decisions about organisational format and note these, ensure that no organisational issues were forgotten and ascertain that tasks were clearly allocated to individuals, together with deadlines for their completion.

In-school projects were typically launched at a group training meeting for children and parents (or other helpers). For a school's first project, the project leader would be available to make the main presentation during the training meeting, describing the Paired Reading technique to participants and demonstrating the technique, most usually on video. The project leader also was available to cover any other organisational hiatus during launch meetings, based on the assumption that the launch meeting for a school's first project was particularly important in terms of community credibility, so that its smooth operation was essential. For a school's second and third project, a lesser role was negotiated for the project leader, the "fading" of central support leaving the school hopefully able to operate its fourth project completely independently with the same degree of success.

The project also made available to schools a range of practical materials and resources, including leaflets for parents and teachers, home-school

record sheets or cards, promotional badges, pens and certificates, at no charge to the school. Training videos were loaned to schools operating subsequent projects independently, and materials continued to be available to these schools. About a quarter of the schools involved chose to incorporate home visiting as part of project monitoring and follow-up, and the central support services included the possibility of paying teachers "overtime" for home visits carried out in the evening or otherwise outside of "Directed Time". Travelling expenses could also be paid for the same purpose. In the event, by no means all of the teachers who carried out home visits actually claimed any payment in connection with these. There was little evidence that the availability of these monies acted as an inducement to carry out home visiting for a significant number of teachers. Occasionally this funding was made available to enable teachers to work extra hours to carry out testing as part of the evaluation of a project.

The project leader also undertook a liaison function with other agencies to facilitate the operation of projects in some schools. Thus, for instance where schools were concerned about the inadequacy of their book stocks, it was possible to liaise with the Children's Services section of the local authority's Libraries Department, who were often able to provide an additional loan collection of books of an attractive nature and in good condition, which were especially selected to be relevant to the children involved in the project. Liaison with members of the LEA Inspectorate was undertaken in some instances, often in relation to finance for purchase of additional book stocks or for the acquisition of better display shelving etc. Once a school-based project had been "launched", continuing support and consultation was offered by the project leader, typically via a "review" meeting with the relevant staff in school after the first two to three weeks of a project, often held at a lunchtime. Teachers would raise problems

encountered with particular pairs or wider organisational problems with the project as a whole, and decisions would be made about the best way of solving them.

Of the schools involved in Paired Reading during the period under study, a very high proportion built in some form of structured evaluation, which the project leader encouraged. Evaluation activity of any reliability and validity tends to be extremely time-consuming, and the project leader offered advice about methods of evaluation, provided evaluative questionnaires to those schools requesting them and reading test materials to those schools requesting them, and was also able on occasion to help with professional time for evaluative testing, either by deploying the project assistant who was with the project for one year, or by deploying the project's clerical assistant (who was a qualified and experienced teacher) or by paying for schools to buy in extra teaching time to enable testing to be carried out. However, many schools did undertake norm-referenced testing without recourse to any support service other than advice.

A variety of other support services were requested by individual schools, and in some cases requests were made which extended well beyond the brief of the Project. These were however too varied to be worth detailing here. The overall intention was to deliver support services in a way which ensured, so far as possible, that a school's first project would be successful, establish community credibility and raise teacher confidence, thus giving the school a firm foundation from which to consolidate its expertise on an independent basis, and subsequently to develop its approach in whatever way the staff felt appropriate, with a view to moving towards a "whole- school policy" on parental involvement.

As part of the consultation service in connection with planning the organisation of the school's first project, the teachers involved were encouraged to consider their own context. A minimum of two enthusiastic participating teachers was considered desirable, with at least minimal support from their head teacher. Schools were encouraged to consider the socio-economic status of their catchment area and the existing relationships between school and community, together with the implications that might have for recruitment rates, nature of parent training and degree of follow-up and monitoring. If ethnic groups were to be targeted, particularly those speaking English as a second language, additional planning considerations would come into play.

Encouragement was given to consider the organisational context of the school, particularly with reference to existing reading standards, since the project leader did not wish schools to launch parental involvement projects if one effect might be to mask fundamental flaws in the organisation of reading teaching in school. Likewise, division of opinion about the merits of parental involvement among the teaching staff needed taking into account, since sabotage through non co-operation might be a possibility needing consideration. The project leader attempted to ensure that the teachers involved had a realistic view of the time and energy commitment necessary, and had accepted that the project would take priority above other distractions.

As a next step, the project leader would attempt to help the teachers involved clarify their feelings about a suitable target group of children, in terms of ages, reading level, class membership and numbers to be involved. Initially, the inclination of many teachers was to select a few of the worst readers from a fairly large number of classes for inclusion

in the project. This sort of approach rapidly became less common however, partly because it was more desirable to in-build success for a first project by including at least some children with less marked reading difficulty, partly to avoid the stigmatisation of a project with the notion that Paired Reading was only for poor readers, and partly to ensure that the Paired Readers were part of a coherent social group who had regular contact with each other, a group in which modelling of the required behaviours and mutual encouragement was more likely to take place.

Some schools were inclined to launch a first project with only a very few children (so the participants lacked a coherent peer group for mutual support), while other schools were inclined to include very large numbers of target children (running the risk of over-stretching staffing resources and rendering monitoring and follow-up sketchy and inadequate). The project leader advised schools to identify a target group of a size which, given the school's predicted take-up rate from the parents, would result in each participating teacher needing to follow up and monitor no more than ten pupils, especially if home visiting was to be incorporated.

In fact, over the years 1984-1987, an increasing number of schools moved towards offering Paired Reading involvement on a mixed ability basis for all members of existing class groups. Where predicted take-up rates were greater than 50%, schools often divided the class into halves, randomly or according to some selection criteria which would be considered equitable by the parent community, and operated a project for one half of the class in one term and a project for the other half in the following term. In mixed ability projects, this often left individual class teachers monitoring as

many as 15 children in one project, but as a number of the participants were already very able readers, they were considered to need considerably less monitoring and follow-up.

Paired Reading projects operated in almost all kinds of schools for children of all ages from 5 years through to 13 years. A majority of projects (approximately 60%) took place in Junior, Junior and Infant or Middle Schools. Proportionally fewer projects operated in Infant and First Schools, and in these schools it was typical for the oldest children to be considered most suitable for the Paired Reading approach. Although some infant schools did mount projects with middle infants (aged 5-6), and one or two schools mounted projects with reception infants (ages 4-5), there was a general assumption that participant children would benefit more from Paired Reading if they possessed at least a modicum of reliable sight vocabulary prior to commencement. Thus, those projects involving very young infant children tended to be in neighbourhoods of higher socio-economic status, where there was a greater probability of children arriving in school with pre-reading skills already relatively well developed. Projects in infant schools and involving children in the first two years of the junior department were particularly likely to operate on a mixed ability basis. The high school projects were much more likely to target children with some degree of reading difficulty, although there were nevertheless a number of instances of schools offering participation to very large numbers of pupils aged 11 and 12 on a mixed ability basis.

Although the majority of participant children were tutored by their natural parents, there were some families where because of lack of parental ability to read competently in English or because of disinclination, other members of the extended family were involved as tutors, most usually grand*parents*,

older brothers and sisters and aunts and uncles. Occasionally neighbours volunteered to act as tutors. Invitations to training meetings were couched in terms designed to attract such surrogate tutors to the meetings. Nevertheless, participating teachers still often felt that some of the children most in need of extra reading practice were unlikely to be supported by their own family. In some of these cases, teachers arranged alternative experiences for the most needy children in school, whether via volunteer (or "donor") parents, volunteer teacher tutors who did Paired Reading in their own free time, or via the deployment of other children as tutors, the peer tutors being of the same age as the tutees or older (cross-age tutoring).

However, the project leader generally advised schools to avoid expending disproportionate amounts of time and energy on complicated arrangements for a few individual pupils, since especially with the first project there was a danger that this would detract from the main exercise. It was usual to find that take-up rates from those families considered by the school to be "most in need" were better, even for a first project, than the teachers had expected. Furthermore, once an initial project had run successfully, those families who were initially rather doubtful about such projects became more confident and were more likely to involve themselves. Nevertheless, in many schools even after a second and third project there was still a residue of children who were weak readers for whom some form of surrogate tutoring still seemed to be desirable.

An organisational issue always considered in great detail during the course of planning consultation was that of child access to reading material within school. As Paired Reading project participants, children were free to choose to read from public library books, books at home, newspapers and

magazines and sundry other sources of text in the natural environment. However, many children, particularly in socio-economically deprived neighbourhoods lacking adequate facilities, largely tended to rely on the school for the provision of reading material. Existing locations for book stocks, and existing arrangements for child access and loan of books for out of school use were reviewed in detail. Some schools had a central school library while others had individual class libraries, and some schools had both, in some cases with one or the other collection biased towards fiction or non-fiction. Where schools employed a reading scheme, whether a single core scheme or a collection of reading schemes or other books banded into readability levels, these were often kept in yet another location.

Schools had very different existing policies as to which of these books could go home and very different arrangements for keeping records of which books had been borrowed by which children. In some schools, child access to book stocks was restricted, perhaps to a single "library period" once a week. The project leader encouraged schools to make as many books of as great a variety as possible available to the Paired Readers on the basis of daily access. Some schools were able to do this, while others restricted the children to a portion of the books available in school, although ensuring child access on a regular and frequent basis. The project leader generally advised that the children needed to be able to change their books every day from school based resources.

Most schools chose not to encourage children to take core reading scheme books home for Paired Reading purposes, and although this rarely became a definite prohibition, relatively few children chose such books in any event. Where wider collections of books were banded in readability, it was likely that schools would allow children to choose from this source, while

encouraging the children to explore other sources available within school. In order to give a project an air of being "special", some schools established a special "Paired Reading collection", either by selecting particularly relevant and attractive books from amongst existing school book stocks, or by arranging for a special loan collection from the Libraries Department. The special collections proved popular with participating children in the first weeks of the project, no doubt partially because of their exclusivity, with most children exploring a much wider range of book sources in the later weeks of a project.

Some schools would link in other literature based events with a Paired Reading project, such as a programme of visits to a local public library, a promotional event in relation to the school bookshop (of new or secondhand books), by arranging a visit to the school of the Library Services' mobile exhibition, or by arranging a visit from a children's author. Systems for recording child possession of particular books varied enormously, but rates of book loss during Paired Reading projects were invariably extremely low, and rates of book loss seemed to bear no relationship to the complexity of the recording procedure in operation. Few schools appeared to have evaluated their recording procedure carefully, and some schools appeared to operate an unnecessarily elaborate procedure which may well have deterred some children from borrowing books. Except in the case of the youngest children, many schools sought to place the administrative burden of recording upon the children, in some cases with a degree of light supervision from older children acting as "librarians", and this seemed to work adequately. Other schools used parent volunteers or non-teaching assistants as librarians, but many schools left self-recording entirely to the children.

Although Paired Reading operates on the principle of free child choice of reading material, particularly in the case of the youngest and the least able children, discriminatory skills of selecting high motivation books within a relevant range of readability do not develop spontaneously. In most projects the majority of the children would be choosing sensibly and carefully after the first two or three weeks, but in many projects there would be one or two children who needed additional instruction or individual guidance in appropriate selection. For the very youngest children in the infant school, a degree of guidance was more likely to be necessary for a greater number from the outset, and in these cases encouragement was often given for parents to come into school to help children choose the books that they were going to use for Paired Reading at home.

All projects incorporated an initial "launch" or training meeting, to which both children and parents were invited. Where possible, schools established communication with target parents prior to a specific invitation to a launch meeting. Preparatory information could be communicated verbally to individual parents, by informal contact or at open evenings or other school events, and in writing as part of other standard communications. Some parents needed reassurance as to why the children had been chosen to be invited. Where the school considered the child to have a reading difficulty, but had never made this clear to the parents, this parental reaction was common. Likewise, schools were advised to prepare a form of words to explain to parents the basis on which target groups had been selected, to ensure this was seen as "fair". Also in preparation, many teachers discussed the impending project with the relevant children, partially to help generate a degree of group enthusiasm, and partially to ensure that the children transmitted the right sort of information home to their parents.

All schools then issued a specific invitation to the launch meeting to the target families, usually with a reply slip for families to indicate likely attendance or otherwise. Most schools opted for a brief simple letter on school headed paper, but one or two schools produced much more striking promotional leaflets. A few schools involved the children in writing their own letters of invitation to their parents, or developed the format of the letter as a group project. Where teachers particularly wanted certain families to participate, an additional handwritten note could be attached to the standard letter, and these were sometimes followed up with a telephone call as well as other informal contact. A few schools made introductory home visits to selected families even before commencement of the project. Some schools put up advertising posters or mounted other displays concerned with the forthcoming project. Almost all schools issued a brief reminder note to all target parents the day before the launch meeting.

The structure of the training meeting was always carefully delineated in advance and a vast majority of them followed a very similar format. Most schools held just one launch or training meeting, typically in the evening at about 7 pm, to which parents and children were invited. A few schools chose to operate training meetings during school hours, and even fewer chose to have meetings immediately after school finished. Some schools offered parallel meetings, one in the evening and one during school hours, so that parents could attend whichever was most convenient, or one parent could attend one and the other parent the other. Local patterns of employment or unemployment, office hours and shift work, and alternative distractions needed to be considered when fixing a time, a day and a date for these meetings. For schools with very large catchment areas, most notably the special schools, arrangements to provide transport were sometimes necessary, although extremely complex and time-consuming to organise.

The location of the meeting needed careful consideration, since each family would be represented by at least two people, and adequate physical space, seating and ventilation was essential. Access to a television monitor and video tape-recorder had to be pre-planned and the location of electricity supplies ascertained. The practical element of the training meeting involved each pair having practice space with a degree of privacy and limited distraction from the noise of other pairs, so practice spaces in neighbouring rooms needed to be identified. The availability of books for practice purposes needed to be pre-planned, and most schools preferred to have children select their practice book during the course of the school day before the meeting, to avoid an organisational hiatus at a later point. All schools encouraged children to choose at least one book which was above their independent readability level, in order to make practice of supportive Reading Together a relevant activity. A minority of schools chose to make available some form of child minding facility within the school for other children in the family for whom parents were not able to make alternative arrangements. Many schools offered refreshments to parents at the end of the training meeting, in a way conducive to further discussion in a more relaxed social atmosphere. Usually parent volunteers arranged the refreshments.

Each meeting commenced with a very brief welcome and introduction from the Head Teacher or other co-ordinating teacher. In the majority of projects, this was followed by a humorous version of "how NOT to do it", usually on video, but occasionally by live role play from members of staff. This was intended to relax participants through humour, while making some serious points about undesirable parental practice, and allow latecomers to arrive without missing or disturbing anything essential. Then followed a talk on the essential components of the Paired Reading technique, closely conforming

to the "How To Do It" parent leaflet which was distributed at the end of the meeting (see Appendix 1). The two main aspects of the main Paired Reading procedure (Reading Together and Reading Alone) were demonstrated at the relevant point in the presentation, most usually on video for maximum visibility and audibility to a large group (both the leaflet and the training video will be found in Topping, 1988a). The video also had the advantage that it could be paused at any moment while a teaching point was made to the group. Some projects incorporated use of the video for demonstration, and then followed this with a brief live demonstration from parent and child "graduates" from previous projects, which may occasionally have incorporated less than perfect technique and been more difficult to see and hear in a large group, but which was undoubtedly well received by the parents and children present, constituting a convincing demonstration of local relevance. However, especially with young children, care was necessary that the meeting was not over long.

Pairs then dispersed into the practice space with the practice books in order to try out the technique, preferably starting with Reading Together on a more difficult section of text. While this was going on, the project leader and the participating teachers would circulate among the pairs, checking technique and encouraging as necessary. Praise was given to pairs who had mastered the technique. In pairs where difficulties were present, initially further verbal advice was given. If this was unsuccessful and Reading Together was proving problematic, the teacher would often join in Reading Together as a triad, and if this also failed the teacher would re-model the appropriate procedure using the child and the chosen book before asking the parent to continue. Staffing of the launch meetings needed to take into account that during a practice period of some 15 to 20 minutes each monitoring professional could hope to see no more than 5 or 6 pairs.

After practice, the pairs re-convened in the large group while one of the teaching staff briefly reviewed points of organisation of the project, such as availability of books, length of project period, dates of any follow-up meetings, use of home-school record cards and other support and problem-solving arrangements. It was usual to ask pairs only to decide whether they wanted to participate at the very end of the meeting, but in fact in almost every project all of the pairs present agreed to participate at the end of the meeting. Some schools formalised the agreement to participate by having parents sign a list and occasionally, especially with reluctant children, a more formal contract was signed by all parties. Participating pairs collected their "How To Do It" leaflet and home school record card, together with any other "equipment" provided such as "Advantages" handout (see Appendix 1A), plastic bag, badge, etc, then stayed for refreshments if they so wished.

The amount of follow-up support and monitoring provided was more variable between schools. This variability was partially a function of the time and energy the participating teachers felt able to commit to the project and partially a function of the needs of the target group as estimated by the teachers. In a mixed ability project including a large number of already able readers, follow-up was sometimes implemented on a discriminatory basis, with the least able readers offered much more intensive ongoing support.

All projects utilised a simple home-school diary card or sheet (see Appendix 2). Each night Paired Reading was undertaken, parent or child were to write down title of item read, for how long and who with, and the parent was to add (preferably positive) comments about child performance (rather than about the book itself). The practice session at the training meeting constituted the first diary entry. After a week or two some parents found

thinking of positive comments increasingly difficult, and in the later years of the project a "Dictionary of Praise" was produced (in Topping, 1988a), which was issued to pairs on a general basis at the launch meeting or a selective basis/on request two or three weeks into a school based project. Children were required to bring completed cards to show to their class teacher or other co-ordinating teacher in the school once a week, which gave the teacher the opportunity to discuss their reading with the child and enquire how things were going. The teacher was required to add a (hopefully positive) comment on the bottom of each week's entries, and then return the diary via the child to the parents. In this way, parent, teacher and child were kept in touch, and there was a cycle of mutual accountability wherein each participant could be seen to be fulfulling their contracted commitment.

Beyond this form of self-recording, amount and type of follow-up was much more various. Some schools offered individual in-school consultation with relevant teachers to all parents on request, but in very many cases it was clear that parents did not tend to take up such an open-ended invitation, even when it later transpired that they had actually been having some kind of difficulty. Some schools incorporated follow-up or "booster" meetings three to four weeks into a project, to gather together all the pairs once again for a group discussion on the positives and negatives of experience so far, for another practice session with its associated opportunity for remediation of "drifted" technique and for general encouragement to both parents and children. However, attendance at such meetings was always lower than that at launch meetings, and not all parents having difficulties attended them.

About a quarter of schools incorporated home visits as part of the follow-up. Of the schools incorporating home visits, about half made home visits

available to all participating children, while the other half made them available only to children considered to be likely to be having difficulties. The maximum frequency of home visits was approximately three visits during an eight or nine week project period. In most cases the visit was undertaken by the teacher in closest daily contact with the child, but in some cases other teachers volunteered to visit in order to spread the burden of so doing. A few visits took place in school-time, but the majority took place immediately after school or in the evening, since it was considered desirable to see the parents and the child at home. During home visits, every attempt was made to observe parent and child doing Paired Reading, in order to check whether the technique remained "correct", or at least that any variations adopted were functional and beneficial for the child concerned. Although a checklist of Paired Reading technique was made available to teachers (in Topping 1988a), in very few cases was this actually completed with respect to individual visits, so process data of any consistency was not available.

All school-based projects were offered, and the majority accepted, a review visit from the project leader three to four weeks into a school-based project, to meet with the staff concerned and discuss any organisational problems arising. These meetings were usually preoccupied with the discussion of individual cases where there were reasons to suspect that all was not going well, either because home-school diary cards were not being returned, or were being returned incomplete, or were being returned with many negative comments, or the pattern of reading was very erratic and incorporating material of doubtful suitability, and (very occasionally) because the diary card appeared to be forged. Home visits were usually well

received by the parents and children concerned, and appeared to serve as a general boost to participant motivation, but in by no means every case was technique considered "perfect".

In all projects, parents and children were asked to commit themselves to an initial contractual period of Paired Reading of between six and ten weeks, reading for a minimum of five minutes on five days each week. The intention was that this initial contractual period of intensive use of the technique would help to ensure that practice was sufficiently regular and frequent to enable pairs to become fluent and automatic in technique, and also so that parents were likely to be able to observe a discernible difference in their child's reading performance which was plausibly attributable to their Paired Reading efforts and which was likely to motivate them to continue using the technique in the longer term, albeit on a less frequent basis.

Towards the end of the "intensive period" of commitment, in almost all school-based projects parents (and often also children) were invited back to school to a group "feedback" meeting. At these meetings, the participating teachers reported to the group on the process of the project from the school's perspective and commented on positive and negative aspects. If norm-referenced evaluation results had been gathered and analysed in time for the meeting, a summary of these was often presented to the meeting, without identifying individual scores for any children. Parents were then asked for their observations on their experiences, both positive and negative, and for any suggestions for the improvement of the organisation of future projects. Where children were present, they were invited to make similar contributions. Children did not always feel able to make some verbal comment in a large group meeting with adults, particularly if they were young, but the amount of child response was very variable.

Parents were much more likely to express their views, but even so a substantial minority of parents would usually remain silent throughout the meeting, leaving the more vocal and confident parents to make many comments. Parental suggestions for improvements in project organisation were always recorded in writing for the reference of the school, but other more generalised parental feedback tended to be so variable and idiosyncratic, not to mention unrepresentative of the whole group, that recording and transcribing such feedback meetings would merely have yielded a mass of data which was not susceptible to analysis. Where children were present at a feedback meeting they were sometimes presented with a token of appreciation for their participation (a badge, an additional adornment to an existing badge, a certificate of merit or a Paired Reading pen), although some schools and some children preferred to have these presented during school hours in a full school assembly, which of course was useful in promotional terms. Occasionally parents requested or were spontaneously given similar tokens of appreciation, and some of the more imaginative schools produced Paired Reading displays, Paired Reading cakes, Paired Reading Easter eggs, and so on.

Every effort was made to ensure that parents and children did not see the feedback meeting as "the end" of the Paired Reading project. Teachers were advised to avoid using vocabulary with terminal implications. An important final component of feedback meetings was a review by individual pairs in a group setting of their decisions about continuation with Paired Reading. Some pairs wished to carry on Paired Reading five days a week, some pairs wished to carry on but only two or three days a week, some pairs wished to carry on reading at home but use a different method and some pairs wished to stop Paired Reading and have a "rest". In the context of the feedback

meeting, wherever possible individual pairs were encouraged to make a public commitment to their decision. It was usual for a majority of pairs to assert that they intended to continue doing Paired Reading with some degree of frequency and regularity. At many feedback meetings parents were given a "Beyond Paired Reading" handout (see Appendix 3), suggesting other continuation options.

As part of the organisational planning for each project, the project leader discussed with the participating teachers whether or not evaluation should be built into the project by means other than informal verbal feedback, and what forms of evaluative measure might best suit the context of the individual school and be maximally cost-effective within the limited time and resources each school was prepared to allocate to this aspect of project operation. In the event, a large proportion (84%) of school-based projects incorporated some form of norm-referenced evaluative data gathering. In some projects (including some where norm-referenced data was not gathered) subjective evaluative information was gathered by structured questionnaire from the participant parents or children or teachers or some combination thereof. Evaluation measures will be considered in greater detail later.

Peer Tutoring Projects

The organisation of peer tutoring projects followed a very similar structure. All the support services from the central project were made available in the same way, except of course the funding for payment for home visiting by teachers was not relevant. During the consultative planning process, schools decided whether to deploy same-age or cross-age peer tutors. The majority of schools opted for same-age peer tutoring, which is considerably less complex to organise. Target groups were identified, sometimes including whole class groups but sometimes (less frequently)

consisting of those members of class groups who volunteered or (occasionally) a few children from a variety of classes who were handpicked as being likely to be particularly competent tutors or representing needy tutees. As with parent tutoring, schools were encouraged to create a socially cohesive and self-supporting group and avoid stigmatisation of the exercise by the children.

Three criteria were employed when matching the tutors and tutees into pairs. The main criterion was the maintenance of a roughly similar tutor/tutee differential in reading ability. As a general principle, the tutors were ranked in order of reading ability, as were the tutees, and the procedure of "parallel matching" proceeded, with the most able tutor matched with the most able tutee and so on down the two ranked lists. The secondary selection criterion concerned pre-existing relationships between children, which were much more likely to present problems in same-age peer tutoring. Pairs where there was a known pre-existing poor relationship were considered, and the members re-allocated elsewhere. Some teachers were also disinclined to pair children who were already very close friends, particularly if it was felt likely that the pair would indulge more in social discourse than in Paired Reading.

The third selection criterion which was considered was any expression of child preference, although co-ordinating teachers virtually always reserved the right to over-ride the child's preference. Some tutees (especially boys) expressed disinclination to work with some tutors (especially girls), but most co-ordinating teachers declined to be influenced, regarding the existing pairing as a useful social learning experience for the children. In fact, many pairings which were the subject of initial complaint settled down to function satisfactorily, and problems with child preferences rarely

constituted a significant problem. During the course of tutoring problems did arise in a small minority of pairings ("personality clashes"), and sometimes teachers were forced to reallocate pairings within the very few children so affected. Some teachers appointed spare tutors to cover for the absence of regular tutors. Parental agreement to participation of tutors and tutees in the project was usually sought, or assumed, in an informative letter from the school. Written information from the school sometimes included a one-page summary of information about the effectiveness of peer tutoring (in Topping, 1988b).

In principle, tutees were allowed free choice of high motivation reading material, so long as it was within the independent readability level of the tutor. Some element of readability control was operated by guiding children to appropriate levels of a banded collection of reading books, teaching pairs simple readability self-checking procedures, encouraging tutors to utilise a prerogative of rejecting the tutee's choice on the grounds that it was too difficult, and by generalised teacher observation during tutoring. As with parent tutored projects, locations of different kinds of books within the school were reviewed, as were access arrangements. Tutees were often allowed to bring in material from the public library or home in addition. Occasionally problems arose where a tutee consistently self-selected books the content of which was of no interest whatsoever to the tutor, and in such cases teachers often allowed the tutor to choose the reading material on a regular but infrequent basis. In cross-age tutoring where the differential in reading ability between tutor and tutee was very great, more teacher supervision and adjustment of the readability of the books chosen were sometimes necessary.

Peer tutoring projects usually also ran for intensive periods of between six and ten weeks, with tutorial contact usually occurring five times weekly, but minimally three times weekly, usually for a period of 15 to 20 minutes, but occasionally for as long as 30 minutes. Same-age peer tutoring usually took place wholly during class time, while cross-age peer tutoring was more likely to be scheduled partly for class time and partly for children's free time. In both cases however, many schools made it possible for pairs to do extra Paired Reading during their own free time. Same-age peer tutoring usually took place in the host classroom or in a neighbouring shared or utility area. Cross-age peer tutoring was more likely to take place in some other room or area designated for the purpose elsewhere in the school, although where a whole class of tutors and a whole class of tutees were involved, half the tutorial pairs functioned in one classroom and the other half in the other.

Training was carried out in much the same way as for parent tutored projects. Monitoring of the process of tutoring was of course a great deal easier for peer tutored projects than for parent tutored projects. While tutoring went on, the co-ordinating teacher constantly circulated to monitor technique, other relevant behaviour and suitability of book choice. Occasionally spare tutors would be deployed to do the same. Members of tutorial pairs were encouraged to self-refer to the supervising teacher with problems and questions of all sorts, ranging from a query about the meaning of an individual word through to a complaint about relationship difficulties with their partner. As with parent-tutored projects, diary cards were kept by tutorial pairs, with tutors writing in positive comments and using the Dictionary of Praise. During the course of a project, tutors and tutees were occasionally withdrawn for group or individual discussion of the

positive and negative aspects of their experience so far. As with parent tutored projects, a group feedback meeting at the end of the intensive period was incorporated into project organisation by many teachers. Participant preferences about continuation or otherwise were solicited. Similar forms of evaluation were built in.

There was thus a high degree of homogeneity in structure and method among the school-based projects carried out during the period under study, 1984-1987. During this time, 83 schools were known to have operated 185 projects. However, only 155 of these projects (from 71 schools) yielded norm-referenced data. Of the 185 projects recorded, only 14 were not carried out under the close supervision of the project leader, some representing the subsequent independent efforts of schools who had been supported through their first projects by the project leader. Of the 14 projects whose process was not monitored by the project leader, 7 generated data which were nevertheless incorporated in the current analysis. Four such data-yielding projects were based in a special school and incorporated a total of 29 participants and one was based in a junior school and incorporated 14 participants. All 5 of these projects were in schools which had been operating Paired Reading prior to the arrival of the project leader in November 1983, and who saw themselves as in little need of external consultation. Thus only 2 projects not closely supervised by the project leader in schools not definitely known to be competent to deliver Paired Reading yielded data which were incorporated in the current analysis (n = 12 and 23). This obviously represents a very small proportion of the total data pool. The other 7 projects not closely monitored by the project leader yielded no data, either norm-referenced or subjective.

CHAPTER 6

METHOD - EVALUATION

Evaluation of the Kirklees project was essentially action research, data collection being shaped opportunistically around the exigencies of operating community-based projects wherein the prime objective was service delivery and not research. The evaluation research "design" was thus very imperfect.

Evaluative measures employed related almost wholly to the outcomes of Paired Reading, rather than to the detailed process of school-based projects. Previous research suggests the degree of tutor conformity to Paired Reading technique is very various from project to project, but detailed process research is very costly in time, and the many process research replications needed to permit validly the drawing of generalised conclusions were beyond the resources available in Kirklees.

Generalised self-recording by tutors and tutees was a feature of virtually all the Kirklees participating family and peer tutorial pairings, but this gave little indication of how precisely the tutors were conforming to the required technique. During 1985 and 1986, sets of home-school diary cards were collected from almost 600 participating families, and analysed with respect to the frequency of Paired Reading and the total time spent on Paired Reading. However, this sample of self-recordings was far from random, in that sets of cards were only available from those schools who had decided to collect these in at the end of the intensive period of the project, this not being an obviously logical thing to do except for research purposes. The reliability and validity of this kind of self-recording is very doubtful, although

attempts were made to relate the process parameters emerging from analysis of the diary cards to outcome parameters such as gains on reading tests.

In approximately 25% of parent tutored projects, home visits were carried out by co-ordinating teachers, partially to ensure that technique was not varying from the required model in an unproductive way, but data from these home visits were not recorded in a structured and systematic way permitting subsequent analysis. In peer tutor projects, activity was always monitored and supervised by co-ordinating teachers, but again detailed process data were not gathered. Although every attempt was made at training meetings to ensure that no participant left the meeting without having demonstrated mastery of the required technique, subsequent maintenance of the required behaviour could not be assumed. It is clear from anecdotal verbal information offered by some parents at feedback meetings that by no means all parent tutors conformed to the prescribed technique subsequent to the training meeting. Thus the subjective feedback from participants may be commenting on the effectiveness of what they actually did rather than on the effectiveness of what they were supposed to be doing. Much the same applies to the norm-referenced data.

Elliott (1989) has reported that the proportion of parents failing to utilise the Paired Reading technique "properly" varies from project to project. It seems likely that conformity to required technique varied considerably among the 155 projects yielding norm-referenced data, but insofar as detailed planning procedures and external monitoring were available via the project leader it is reasonable to assume that in the current study the proportion of parents deviating in technique was kept

to a minimum. Elliott's (1989) study was retrospective and based more on participant self-report in interview than on direct observation, and it is possible that some of her subjects were reporting deviations in technique subsequent to the conclusion of the intensive period of the project rather than during it.

At feedback meetings in the Kirklees project, parents were encouraged to discuss variations in technique which they might wish to utilise with particular children, but at initial launch meetings it was stressed that it was desirable for parents to adhere to "pure" Paired Reading technique for the duration of the intensive period. The impression gained from anecdotal verbal information from parents at feedback meetings was certainly that the majority of parents in school-based Paired Reading projects in Kirklees did conform to the prescribed technique, although it was not uncommon for parents to report that after the first few weeks their child tended largely to use the Reading Alone aspect. A smaller proportion of parents reported that their children over-dwelled on the Reading Together aspect, but most parents who reported this felt that it was beneficial and relatively few considered it an indication of "laziness".

The outcome measures utilised were of two main types: norm-referenced reading tests and structured subjective feedback questionnaires completed by the participants. A few schools utilised other approaches, such as Informal Reading Inventories, criterion-referenced cloze tests, tests of high frequency words known or sight vocabulary recalled, error frequency counts or error type analyses based on the administration of

passages of prose before and after intensive periods of projects, indications of progress in reading competence on levels of a core reading scheme which had not been available for Paired Reading purposes, paper and pencil indications of attitudes to reading on a before and after basis, and tape recordings of participants' verbal comments in individual or group settings. However, none of these measures was used by more than a few schools, and they were virtually all deployed in a manner idiosyncratic to the particular school, which rendered comparison of results virtually impossible.

"Objective" Measures

Where norm-referenced reading tests were utilised in a school-based project, a number of factors determined which test was selected. Where schools already carried out a programme of norm-referenced reading testing at regular intervals, many preferred to utilise the same test for evaluating the Paired Reading project in order to render results comparable with the period before and the period after the project. Additionally, schools felt more comfortable with a test with which they were already familiar; they were used to administering it and often had the materials readily available in school. However, a minority of schools did deliberately choose to use a different test for Paired Reading evaluation purposes, sometimes in order to try out a test which was new to them in order to determine whether it was more satisfactory for their purposes than the one which was already in use. Where it was specifically intended to compare the progress of the Paired Readers during the intensive period of the school-based project with their reading progress during the previous ("baseline") period, it was obviously essential that the same test be used, or a parallel form of the same test if such were available.

Both individual and group reading tests were utilised. Schools usually felt it necessary to utilise a group test where large numbers of children were involved in a project (particularly when a comparison or control group also was to be tested), and when no external help with staff time for individual testing was available. Teachers tended to express preference to carry out individual testing, but many were unable to find the time to do this. In cases where the project leader was able to make available an assistant to carry out testing, an individual reading test was always used, almost always either the Neale Analysis or the Daniels and Diack Test 1. Where the project leader made funding available to pay for extra teacher hours to carry out the administration and scoring of reading tests in connection with a school-based project, most schools preferred to utilise one of a wider range of individual tests.

Another very significant factor in selection of reading tests was the age of the target group of children. Many schools who could afford the time favoured the Neale Analysis for children in the junior age range, but this test does not register reading ages below 6.0 years and thus had too high a "floor" to be sufficiently discriminatory for many beginning readers in the infant school and/or very weak readers in the junior department. The author of the test also acknowledges in the manual that it is less reliably discriminatory above reading ages of 10 years (Neale, 1966). For younger and weaker readers many schools favoured the individually administered Daniels and Diack Test 1, which is briefer and less demanding for a young child and yields reading ages down to 5.2 years. Virtually all teachers were disinclined to use group reading tests with children aged 7 years and below, feeling that the validity of such an exercise would be severely limited.

When selecting what they considered an appropriate test for evaluation purposes, notwithstanding advice given by the project leader, teachers were rarely influenced by technical considerations of reliability and validity, as indicated by adequate statistical information. Many teachers were however concerned about the face validity of test content, and tended to express a preference for individual tests on the grounds that motivation to search for meaning in an interactive interpersonal situation was more likely to reflect the child's real potential. Many teachers expressed a preference for tests incorporating continuous meaningful prose, in contrast to those orientated towards word recognition.

Overall, 10 different reading tests were used on at least a pre-post basis with at least 30 children. These were, in order of frequency of use (number of children tested in brackets): Neale Analysis - Accuracy (638) (Neale, 1966), Daniels and Diack Test 1 (546) (Daniels & Diack, 1979), the Primary Reading Test (352) (France, 1981), the Schonell Word Recognition Test (233) (Schonell and Schonell, 1955) (schools often failing to specify to which set of norms the raw scores had been referred), the Daniels and Diack Test 12 (otherwise known as the Graded Test of Reading Experience) (134) (Daniels and Diack, 1979), the Holborn Reading Scale (128) (Watts, 1948), the New MacMillan Analysis (118) (Vincent and de la Mare, 1985), the Widespan Reading Test (71) (Brimer, 1972), the Salford Reading Test (63) (Bookbinder, 1976) and the Burt Word Recognition Test (31) (Burt and Vernon, 1938) (again, specification of standardisation tables utilised was often impossible to elicit from teachers scoring the test).

The Neale Analysis and New MacMillan Analysis tests both also yielded separate scores for reading comprehension, based on asking children questions at the conclusion of each passage of continuous prose orally read. Comprehension scores were available for 568 children on the Neale Analysis and 122 children on the New MacMillan. Of the tests cited above, the Widespan, Daniels and Diack Test 12 and the Primary Reading Test are all group-administered paper and pencil tests. A number of other reading tests were utilised on less than 30 participants in total. These included the Southgate Reading Test, Young's Cloze Test, the GAP Test, the GAPADOL Reading Comprehension Test and Young's Group Reading Test.

Additionally, some schools chose to utilise reading tests such as the Edinburgh Reading Tests which yielded standardised scores or quotients but not reading ages. As the majority of tests in use yielded reading ages but not standardised scores, and consistent analysis of the data was only available through either reading ages or standardised scores but not both, it was decided to conduct subsequent analyses based on reading ages alone and discard the very small amount of data from tests yielding only standardised scores. With the New MacMillan Test a further complexity arises, in that raw scores are related in the standardisation tables not to specific reading ages but to "Age Equivalent Ranges". In order to be able to assimilate data on this test into the larger analysis, raw scores were allocated a specific reading age constituting the arithmetic mean of the Age Equivalent Range, despite the inherent unsatisfactoriness of this procedure.

All of the 10 main tests used in the analysis can be criticised on a number of counts, irrespective of the purpose for and the manner in

which they are used. The Neale Analysis of Reading Ability tended to be preferred by teachers because it offered children continuous passages of meaningful prose which rendered the use of contextual cues possible and related quite closely to "real reading" as experienced in Paired Reading. However, the textual content and illustrations are now extremely dated, and some parts of the test can also be criticised on the grounds of cultural inappropriateness and sex stereotyping. Although the author asserts that the standardisation sample of 2,000 pupils was controlled or otherwise stratified with relation to geographical area, social background, age and sex, no details of how this was done are available. Furthermore, of the 3 supposedly parallel forms of the test, form A was standardised on 1,221 children but form B standardised on only 552 children and form C on only 489. Although it is possible to calculate reading ages up to 13.0 years for both reading accuracy and reading comprehension, reading age equivalents are extrapolated beyond 11 years 11 months for reading accuracy and beyond 11 years 8 months for reading comprehension, although it is not quite clear how this extrapolation was achieved, and all the children in the standardisation sample appear to have been primary school pupils.

The manual claims "high" parallel form reliability for accuracy (0.96 being the lowest co-efficient reported) and for comprehension (0.92), but a number of teachers in the current study considered these reliability co-efficients to be implausible in the light of their own experience, particularly the comprehension, at least in a short term pre-post test application. Given this perception, and the small numbers in the standardisation sample for forms B and C, some schools opted to test and re-test using only form A, while in an approximately

equivalent number of cases schools chose to use the parallel forms in an attempt to control for practice effects. The assessment of Reading Rate in the Neale Analysis has been criticised, but very few schools in the current study utilised this aspect of the test and no such data were incorporated in the analysis.

The Neale Analysis manual claims high validity coefficients (0.95) in relation to 2 factor-analytic studies carried out with 200 9-year-olds and 200 11-year-olds, who undertook a variety of other reading tests, but the supporting information is very sketchy and the pooling of scores on accuracy, comprehension and rate further clouds the issue. No information is given about inter-correlation between the accuracy and comprehension scores for tested children. It is unclear what the author's precise definition of reading comprehension might be, since clearly an element of comprehension enters into using contextual clues to read accurately, while asking children questions at the end of a text places considerable demands on memory as well as understanding, a compounding of different skills which is heightened by the absence in the administration instructions of anything to encourage testees to develop an orientation towards memorisation. This is particularly pertinent on the longer passages.

Practitioners utilising the Neale frequently comment on the stepwise effect of the criteria specified for administration of the comprehension questions. If on pre-test reading accuracy performance a child falls below the criterion level for administration of the comprehension questions, the child's comprehension score may be artifically depressed, while on post-test the child may make only 1 fewer errors, be able to

proceed to the comprehension questions and appear to make a large increase in comprehension reading age. To express it another way, especially on the longer passages, the accuracy test effectively sets a ceiling on the comprehension test so that children cannot achieve a higher level in comprehension than in accuracy. A study by Netley et al. (1965) suggests that the comprehension questions on form B are easier than those on form A. Although the samples used by Netley et al. were small, their findings suggested that form A had a higher reliability than the other forms in routine usage and tended to be less distorted by practice effects as a result of short inter-test periods and repeated exposure. They recommended that form A be used exclusively.

Winter (1985) makes further comment on "problems of non-equivalence between supposedly equivalent parallel forms" in the Neale Analysis. He points out that this is readily apparent in the reliability data cited in the test manual itself, where raw score differences between forms A and B of up to 2.65 on the accuracy scale and 2.37 on the comprehension scale may be found, the norms implying that raw score differences of that size would lead to differences of up to 3 and 4 months in accuracy and comprehension reading ages respectively. On a more positive note, Yule (1967) found a test/re-test reliability of 0.95 for reading accuracy and 0.93 for reading comprehension on a "slightly abnormal" sample of 140 children aged 9-11 years on the Isle of Wight, re-tested after an interval of 1 year, much longer than the Netley et al. (1965) inter-test period of approximately 15 weeks. However it is unclear whether Yule followed the Netley et al. recommendation to use form A exclusively.

Many of the other tests using the study are even more open to criticism. The content of the Daniels and Diack test 1 has dated rather less than some other aged tests, but there has never been adequate information available about its standardisation, reliability and validity, all of which must be suspect. Teachers in this study tended to like the test because it allowed the tester to generate a game-like, purposive and meaningful aura to its administration. Teachers considered this added to the test's relevance for evaluating Paired Reading projects as much as its heavy phonic bias detracted from this. Although having a relatively low "floor", the test also has a low "ceiling" and terminates halfway through the junior school age range.

Slee (1985) reports data on the relationship between Daniels and Diack Test 1 scores and scores on the Word Reading Test from the British Ability Scales, yielded by 120 backward readers, although the span of the tests overlapped for only 80 children. Correlation between the two sets of reading ages was 0.92. However, the relationship between the two sets of reading ages was not linear. DD Test 1 tended to underestimate BAS reading ages up to a reading age of 7.5 years, and overestimate them thereafter, and it was clear that DD1 did not discriminate adequately beyond a reading age of 8 years.

Daniels and Diack Test 12 is a multiple-choice group-administered sentence completion test giving reading ages from 6 to 14 years, although norms above the 10 year old level are unreliable. Slee (1981) has also criticized this test. The content of Test 12 has dated rather more than the content of Test 1. Both tests are now over 30 years old, and the norms may well have dated, even if they were reliable in the first place.

The Primary Reading Test was the most frequently used group reading test in the study. Parallel forms are available and were always used. Compared to other available tests, this test is well-structured, easily administered, has a clear manual, and appears to have satisfactory levels of reliability and validity. However, teachers often felt the results were very erratic when the test was used with top infant children, which was rare.

The Schonell Word Recognition (or Graded Word Reading) Test has been in use for many years, but information on the standardisation, reliability and validity of the tests in their original form is not available. The original norms were superseded by the "Salford" norms in 1972 and the "Cheshire" norms more recently, but there are differences between these and different teachers still refer to different sets of norms. Curr and Gourlay (1960) pointed out almost 30 years ago that this test suffers particularly from practice effects when children are repeatedly exposed to it over short inter-test periods. Thus, in the current study, despite the doubtful relevance of a word recognition test to a reading practice method emphasising continuity of oral reading from continuous prose, it might be expected that practice effects alone would produce high pre-post gain scores. This hypothesis is in fact checked empirically later.

The Holborn Reading Scale was constructed in 1948 and its norms are now very dated. There is no information about reliability or validity in the manual. Although its brevity has made it popular with some teachers, the same feature means that a small improvement in performance on the test can result in a large shift in reading age.

A later test of similar type, the Salford Sentence Reading Test, has the advantage of parallel forms but is even briefer. The test cannot be norm-referenced with children aged less than 6 years 10 months. Although the standardisation is more recent than that of the Holborn, the standardisation sample for Form A was small. Inter-form reliability is stated to be high, but again this is based on small samples. Although the sentence format nominally gives continuous prose, many of the sentences are semantically unpredictable and some users view the test as a word recognition test which happens to have the test words grouped into sentences. The sentences include place names and personal names which are unlikely to be familiar to young readers. Perhaps significantly, an inter-test correlation of 0.95 with the Schonell Graded Word Reading Test is reported. The author seems to expect the test to suffer from practice effects, since there is warning in the manual against its use more frequently than every 12 months. Parallel forms B and C were standardised on much larger numbers of children, but only within one small geographical area (Salford).

The New MacMillan Reading Analysis was the most recently published of the tests used, and was only utilised in the second half of the period under study. It follows the format of the Neale Test in offering 3 parallel forms, each incorporating a series of passages in continuous prose of increasing length, accompanied by illustrations, and followed by comprehension questions. The textual and illustrative content is much more modern, attractive and culturally relevant than is the case with the old Neale Test. The standardisation was carried out on children of average reading ability, rather than on children of a range

of reading ability. One effect of this may have been to lead the authors to over-estimate the reliability of the test when used with children of low (or indeed high) reading ability in relation to their age.

No empirical evidence of validity is offered in the manual. The authors note that some passages include "contrived prose" (presumably of low predictability) in order "to steepen the gradient of difficulty" in some passages. The test was standardised on a sample of only 600 children in one geographical area (London), who had obtained a standardised score of between 99 and 101 on the Primary Reading Test during the previous 12 months. In the standardisation sample of 600 children, each participant was administered two forms of the test, so that each form was standardised on only approximately 400 children, implying that the numbers of children tested on each form in each age group to create the standardisation table was very small indeed. The authors note that the results obtained do not conform to an expectation for group scores to increase with age in a linear fashion, and in particular the rate of increase in score tended to tail off in the upper age groups.

Test re-test reliabilities of 0.91 to 0.94 are reported for the 3 forms in reading accuracy, while the reliabilities for comprehension range from 0.76 to 0.83 and would be considered unsatisfactory by most workers in the field. Internal consistency reliability coefficients for comprehension are as low as 0.58 in some cases. A more detailed review will be found in Topping and Whiteley (1986), in which comment is made upon a "variously high floor" and "doubtful comparability between accuracy and comprehension components", together with the problem (also

suffered by the Neale Test) that the stepwise structure of the reading passages can produce inflated post-test scores in reading comprehension. In practice, individual teachers often reported very erratic scores on the MacMillan Test and this will be the subject of further comment later.

The Widespan has an unusual format, incorporating 80 items in each of which the child is asked to complete a sentence having one word missing by selecting the requisite word from another sentence by the side of the incomplete one. The second (cloze) sentence tends to be unpredictable and contrived, presumably to test understanding of the word supplied by the testee, but it is unclear how this relates to "normal" purposive reading. To add further complexity, answers are recorded on separate sheets. The test covers a wide chronological age range (7-15 years) and tended to be used by schools catering for a wide age or ability range of pupils, particularly Middle schools dealing with the age range 8-13 years. Instructions for determining the point at which to start testing were also complex, and in practice some teachers and children became confused during the completion of the test. The tests were standardised over a wide geographical area on a large sample of children including a minimum of 450 children on each of the two parallel forms in each year group.

Evidence of reliability is presented solely in terms of inter-form correlation, which ranges from 0.89 to 0.95 and can be considered satisfactory. No data is presented concerning validity, although the author asserts that the test should be construed as one of "reading comprehension". In practice, many teachers argued that the test was measuring factors other than this, including the ability to follow

complex instructions. The manual gives separate mean scores for each form for the first and second occasion of testing, "second occasion" being higher than "first occasion" means in the majority of cases. This implies that the test suffers from practice effects, and even when different forms of the test are administered consecutively, familiarity with the format may be a factor in the size of gain scores. Further comment on this point will be made later.

The Burt Word Reading Test suffers many of the disadvantages of the Schonell Word Recognition Test, but is of even more doubtful value at its extremes, particularly in the words selected to represent the "reading age" of 4-5 years. One year of reading age is covered by only 10 words, so small increments in performance can yield large changes in reading age. In addition to Burt's original norms (1921), there exists Vernon's 1938 rearrangement and restandardisation with Scottish children, a rearrangement and restandardisation in Cheshire in 1972 and a revision of the Vernon restandardisation in Scotland in 1974. Teachers using this test rarely specified to which set of norms they had referred raw scores.

The 10 most used reading tests were thus extremely various in structure, content and mode of administration, obviously measuring dissimilar aspects of the reading process. Two tests measured single word recognition by oral response in the one to one situation (Burt and Schonell), two measured word recognition skills in the context of isolated sentences by oral response in the one to one situation (Salford and Holborn), one required accurate oral reading of sentences comprising questions in the one to one situation with a view to

providing a semantically appropriate answer (Daniels & Diack Test 1), two presented passages of increasing length and difficulty accompanied by illustrations for oral reading in the one to one situation, followed by comprehension questions on each passage (Neale and New MacMillan), two were group-administered paper and pencil tests of a multiple-choice sentence completion nature (Daniels and Diack Test 12 and Primary Reading Test) and one was a group-administered paper and pencil sentence completion test in which the requisite word was to be selected from a neighbouring sentence (Widespan).

Some of these tests have parallel forms to help reduce practice effects (Widespan, Neale, Primary Reading Test, Salford, New MacMillan Analysis), but in the case of the Neale these were not always used, and the remaining tests offered no such facility. As will be demonstrated in more detail, correlations between different tests when used with the same subjects at the same time in the context of Paired Reading projects tended to be very low, implying low inter-test validity in this application. It can thus be argued that pooling the results from different reading tests is an exercise of very low validity, and this would be particularly true where only small numbers of subjects taking each type of test were involved. However, it is extremely difficult to identify any one of the outcome measures as likely to be the most valid and reliable on a theoretical basis.

Many teachers involved in projects considered the Neale Test to have the highest face validity for the purpose of evaluating progress during a Paired Reading Project, but the Neale Test produced very different gain scores to the apparently similar Macmillan Test, suffers its own serious

technical difficulties, and did not yield the highest gain scores among the tests utilised. In the absence of a single obvious "best" test for the purpose, the use of a variety of reading tests may actually be seen as a strength of the research programme rather than a weakness, since it enabled empirical comparison of results in this particular application. This created the possibility that Paired Reading could be demonstrated to result in improvements in certain aspects of the reading process sampled by different tests in a way that was not predictable on a theoretical basis. These data will be reported later.

In 23 school-based projects, data were collected on the primary test on the reading progress of the participants in the period prior to involvement in the project, in order that their reading progress could be compared before and during Paired Reading. These "baseline" periods varied greatly in length (from 6-51 weeks). In some cases, the data had already been collected by the school for another purpose, perhaps at a very different time of year, and the baseline tester was in some cases different from subsequent testers. Where schools deliberately took baseline measures as part of their preparation for mounting a Paired Reading project, it was more likely that the testers were the same throughout, especially where assistance with time for testing was provided via the central Project Leader. However, in all baselined projects the same test was used for baseline, pre-test and post-test.

In 37 projects, data were also collected on a "comparison group" of non-participant children. The comparison children were always of the same chronological age as the participant children and were drawn from the same class or classes as the participant children. During the course of

the project, the non-participant children were exposed to the normal teaching of reading carried on in class for all children. In some cases, some of the "non-participant children" were undoubtedly already reading at home to their parents on some informal basis, whether encouraged by the class teacher or not. Some teachers co-ordinating projects reported their informal observations of "contamination" of the comparison group during the project by "contagion of enthusiasm". Thus non-participant children, exposed to modelling of enthusiasm for reading at home by Paired Readers in their class, were sometimes noted to begin to take more books home themselves, although there was no way of determining what was done with those books at home or how much interaction with parents there might be.

In very few cases was the allocation of children to participant or comparison groups truly random. Even in projects where the co-ordinating teachers had agreed with the project leader to do this, subsequent monitoring by the project leader usually showed that the teacher had defaulted on the agreement and switched one or two children from comparison to participant groups at a late stage as "special cases". Quasi-random allocation to participant or comparison groups was more frequent in mixed ability projects for whole classes operated by a single class teacher, where half of the children in the class were invited to participate in a project in one term and the other half formed a "waiting group" to be invited to a repeat of the project the following term. However, this arrangement actually generated four different groups: those invited to project 1 and participating, those invited to project 1 and declining to participate, those invited to project 2 and participating and those invited to project 2 and not

participating. There was also the confusing factor of children who were invited to project 1 but declined to participate who then changed their mind and wished to participate in project 2. For practical purposes, however, the usual comparison made was between the progress of those children participating in project 1 and of the progress at the same time of those children not participating because they were waiting to be invited to project 2.

In other projects different organisational constellations arose. In a school which scheduled only one project but offered it on a mixed ability basis to the whole class, some families would self-select not to participate. Teachers were inclined to assume beforehand that non-participating families would tend to have low interest in literacy and incorporate some of the weaker readers. Experience indicated however that non-participants were equally likely to be the families of high-achieving readers, for whom involvement in Paired Reading was not seen as a priority by the parents. It is therefore not thought that the "self-selected non-participant" nature of some comparison groups necessarily introduced a bias towards lower gains in the comparison groups. On the contrary, a significant number (albeit a minority) of projects deliberately targeted below average or weak readers, who in the normal course of events would be expected to continue to make reading progress at the same rate as previously, that is, at a below average rate. This could be expected to bias the experimental group towards yielding lower gain scores than comparison groups (although this is complicated by issues of regression to the mean, which will be deal with in more detail later). Participant and comparison groups were always tested at the same time. Although comparison groups were quite various

in nature it was considered that they were numerous enough to permit pooling of data without introducing a consistent bias in any one direction.

In 17 projects follow-up data were gathered some time after the conclusion of the intensive period of the project. As with baseline periods, follow-up periods were very various (ranging from 6-82 weeks and averaging 26 weeks). As with the baseline data, some of these were gathered automatically by the school for other purposes, and with the longer term follow-up data in particular the tester was not always the same. Short term follow-up was more likely to involve the same tester, especially when help with time was provided by the project leader. In all cases, the test used was the same throughout. In a very few projects it was possible to gather follow-up data on both participant and comparison groups (in a number of schools, non-participants in project 1 subsequently became participants in later projects, and no longer useful for comparison purposes).

Despite doubts about the reliability and validity of norm-referenced reading test data, especially over a short test/re-test period, and concern about the use of different tests in different projects in different schools with different procedures for selection of target participant group, non-participant comparison group, different baseline and follow-up periods and sometimes different testers, the norm-referenced data were nevertheless subsequently pooled in order to detect overall emerging trends.

Subjective Measures

Four questionnaires were devised to enable the collection of the subjective observations and perceptions of the major participants in Paired Reading projects in a consistent and standardised way which were amenable to subsequent analysis.

A questionnaire for class teachers (see Appendix 5, "Teacher Evaluation Checklist") was designed to solicit the views of class teachers on the extent to which participation in a project might have produced effects in children which generalised into other reading activities in the ordinary classroom. Questions were asked about attitude to reading (regarding confidence, willingness, interest and pleasure in reading), oral reading skills (accuracy, fluency, expressiveness and pacing), amount and variety of reading, comprehension of reading, general motivation and concentration and general behaviour. For each item, respondents could choose one of four options:- more, less, same, and "not seen". The order of response options was randomised to allow for any tendency towards primacy or recency effects in responses.

The questionnaire for parents (see Appendix 6, "What Do You Think?") was similar, except only 3 response options were provided, the "not seen" option being eliminated, and the questions expressed in lower readability prose and less technical language. The item about pacing was omitted, as was the question about general concentration and motivation. Items were included about whether the child was behaving differently at home or was demonstrating a different general demeanour. Additionally parents were asked to indicate whether they intended to stop Paired Reading, go on doing Paired Reading twice a week, go on

doing Paired Reading five times a week or go on reading at home in a different way.

A largely similar questionnaire was used by peer tutors to report on the progress of their tutees (see Appendix 7, "What Do You Think?"), except that items about behaviour at home and demeanour at home were omitted and peer tutors were asked to indicate up to 3 options for continuation from the following: carry on peer tutoring as often as now, carry on tutoring but not so often, carry on tutoring but with a different tutee, be tutored yourself by someone better, tutor reading in a different way, or tutor some other curriculum area such as maths or spelling.

The tutees themselves, whether parent tutored or peer tutored, completed a quite different and simpler questionnaire which enquired into practical organisational details of the project as well as their own attitudes to it and their views on changes in their attainments (see Appendix 8, "What Was It Like?"). On this questionnaire, tutees were asked to tick which of 2 simple sentences was true of their own experiences for each item. Thus, "no change" and "not seen" options were not available. The order of positive and negative sentences was randomised through the items. Tutees were asked to indicate if they had any difficulties in obtaining books, finding time or finding a good place to read. They were asked whether they found Paired Reading easy to learn to do, whether they liked doing it, and whether the record sheet was helpful. They were asked to indicate whether their Paired Reading had resulted in their liking reading better, becoming a more skilful reader, or developing an improved relationship with their tutor.

Finally, they were asked to indicate whether they wished to carry on doing Paired Reading and whether they would tell other people about it.

The structures of these questionnaires were deemed the most appropriate in the search for simplicity with adequate meaning content. Not all projects utilised feedback questionnaires, and those projects that did utilise such devices rarely deployed all of them. All questionnaires were completed at the end of the "intensive period" of a project, at around the time of the feedback meeting. Some schools gave parents the questionnaires just prior to the feedback meeting, asking parents to bring the completed questionnaire to the meeting or to send it back to school if they were unable to attend the feedback meeting. Other schools distributed questionnaires to parents at the feedback meeting, for completion immediately or during the next day or two, and sent feedback questionnaires to those parents who were unable to attend the feedback meeting.

Tutee questionnaires were completed in school by participants in peer tutor projects, but those completed by participants in parent tutored projects were variously completed at home or at school. Whether either of these situations was likely to produce any consistent bias owing to the presence of the adult is debatable. As might be expected, return rates were lowest for parental questionnaires and tutee questionnaires in parent tutored projects, higher for tutors and tutees in peer tutor projects, and highest for class teachers. Parent questionnaires were used in 85 almost wholly primary school projects, teacher questionnaires in 29 primary school projects, child questionnaires in 57 mostly junior age projects and peer tutor and tutee questionnaires in 13 mostly junior age projects.

The questionnaires were piloted in 2 or 3 very early projects before the version used for subsequent data gathering was finalised. However, test/re-test reliability was not determined for any questionnaire. Although the questionnaires have an inherent face validity, other forms of validity in terms of relationships to other measures were not established until all the study data had been gathered. The relationship of questionnaire outcomes to outcomes on other questionnaires and performance on norm-referenced tests will be the subject of report later, but as no two outcome measures could be claimed to measure essentially similar variables, relationships between them might be expected to be low in magnitude. Certainly no one outcome measure could be considered sufficiently inherently valid or reliable to act as a criterion by which the validity of other outcome measures could be judged.

DESCRIPTIVE STATISTICS AND DATA ANALYSIS FORMAT

Data Coding

The data were coded according to the categories described below. Data were not available in all categories in all school-based projects. Each school was assigned a code name, and information entered about school size in terms of number of pupils on roll in 1984 (projects 1984-5) or 1986 (for projects 1986-7), school type (in terms of the LEA's designations - Infant, Junior, Middle, etc) and the age range of pupil served by the school.

An index of socio-economic status for the catchment area of each school was also entered. This was based on a breakdown of information from the 1981 Census by Enumeration District, related to LEA mapping of school catchment areas. Eight factors in the 1981 Census data considered relevant to an index of disadvantage with educational implications were included:- proportion of non-owner-occupied houses, proportion of households without exclusive use of amenities, over-crowded households, households with more than three children, households of single parents, households headed by a person born in the New Commonwealth or Pakistan, number of persons unemployed and a number of households without a car. This information was supplemented by local data on the incidence of free school meals by school gathered in 1985.

All nine indices of disadvantage were given the same weight. The numbers occurring in each category in each Enumeration District were expressed as a rate per thousand and then expressed as a standardised z score by subtracting the mean for all Enumeration Districts and dividing by the standard deviation of the latter. These standardised scores on

each index were then summed to form a composite index of disadvantage. Given the nature of the standardisation procedure, the index for any school's catchment area could be either positive or negative, positive indicating above average social disadvantage and negative below average social disadvantage.

Problems arose owing to lack of precision in delination of some school catchment areas, and this situation was worsened in the latter stages of the period under study as the effects of the 1980 Education Act resulted in an increasing number of pupils attending schools other than their neighbourhood school. This presented particular problems in densely populated urban areas, where many schools were located within a small geographical area. Further problems occurred with relation to denominational schools, particularly Roman Catholic schools, where the nominal catchment area could be very large and data about the relative socio-economic status of pupils attending denominational schools rather than alternative neighbourhood non-denominational schools were not available. Church of England denominational schools were much more likely to be found as the only school in a neighbourhood, but even here there were a small number of cases where Church of England schools had a catchment area which overlapped either wholly or partially with other, non-denominational, schools. Where total overlap occurred, both kinds of school were necessarily assumed to serve the catchment area equally.

Furthermore, the boundary lines of Enumeration Districts and nominal or official catchment areas were not always contiguous. The smallest unit of analysis in the data was set at half an Enumeration District, as a smaller unit of analysis would have generated excessive complexity

without materially improving accuracy and validity in the majority of cases. Thus there was not a perfect relationship between Enumeration Districts and catchment areas. The final major issue concerning reliability and validity of the socio-economic disadvantage index concerns possible demographic changes in the intervening period since the 1981 census data was collected. Nevertheless, the index was based on the best data available at the time (Topping et al., 1987). A similar exercise has been reported by Budgell (1985), except in this case an index of socio-economic advantage was also constructed.

Each project operated by an individual school was allocated a number and the incorporation or otherwise of home visits coded and frequency of home visits recorded in terms of the mean number of home visits carried out per project participant. Also recorded were date of project commencement, date of end of "intensive period", length of "intensive period", date of pre-test, date of post-test, inter-test period between pre- and post-test, main reading test used and secondary reading test used. Date of baseline testing was recorded together with the length of baseline period, and date of follow-up testing together with the length of follow-up period.

The project leader assigned to each project closely monitored by him an "Organisational Rating Code", a global rating of his perceptions of the smoothness and consistency of organisation and operation of the project. Coding categories were:

1. Very Good 2. Good 3. Average 4. Below Average 5. Poor

No other single person had oversight of all projects, so the establishment of inter-rater reliability was not possible. Furthermore, in many cases the Organisational Rating Code was not assigned "blind",

i.e. the project leader had had at least some evaluative feedback on the project, if only in terms of informal verbal responses from teachers, parents and children at feedback meetings, prior to the rating of each project. The O.R. Codes were thus highly subjective and susceptible to bias.

For each participant, initial and surname were recorded, together with sex and chronological age in years and months where available. Ethnic origin was coded as one of: Asian, Afro-Caribbean, White or Other. Tutor type was coded as Natural Parent, Adult Volunteer, Teacher Volunteer, Same-Age Peer Tutor or Cross-Age Peer Tutor. Baseline, pre-test, post-test and follow-up reading ages in accuracy (and comprehension where available) were recorded. Similar data were recorded on any secondary test employed. Similar data were also recorded for comparison group children and for peer tutors, except in the former case nature of tutor was of course omitted and in the latter home visits were not relevant. For each participant, the frequency of reading in terms of mean number of times reading occurred per week as indicated by self recording on home-school diary cards was coded, as was total reading time during the intensive period of the project from the same source, where available.

Regarding the subjective feedback, for parent, teacher and child feedback questionnaires "Positivity Indices" were constructed and coded, constituting the number of questionnaire responses favourable to Paired Reading, from which the number of responses unfavourable to Paired Reading were subtracted, with "not seen" or "same" options disregarded. Individual patterns of responses to each item on each

questionnaire were recorded but not submitted to computer analysis, and data cited below based on individual questionnaire items were all extracted by hand. The data were then analysed using the most recent version of the Statistical Package for the Social Sciences (SPSS Inc., 1986).

Descriptive Statistics

During the period under study, 1984-1987, 83 schools operated 185 projects and norm referenced data yielded by 155 projects in 71 schools. Norm referenced data were most likely to be collected from a school's first project, since once effectiveness of the approach had been demonstrated, the allocation of time to the evaluation of subsequent projects was often not considered cost-effective by the school (See Table 1). Thus for many schools, norm referenced data were only collected when experience of operating projects was at its lowest, but perhaps the novelty value of the approach was at its highest. A very small number of schools operating a large number of consecutive projects remained motivated to continue collecting data, in some cases because the nature of the projects changed over time, and this is reflected in the non-linearity of the frequency distribution, with 7 schools sustaining norm referenced data collection through to their fifth project. Additionally, a number of schools were thought to have operated second and subsequent projects independently without informing the project leader, but they yielded no data included in the current analysis. A number of schools also operated projects utilising approaches other than Paired Reading, but again these are disregarded for the current analysis.

TABLE 1

Number of Schools, Projects and Data Yielders 1984-7

PROJECTS OPERATED

number of schools = 83

number of projects = 185

number of projects per school:

	1	2	3	4	5	6	7	8
frequency	35	24	11	4	4	3	1	1

PROJECTS YIELDING NORM-REFERENCED DATA

number of schools = 71

number of projects = 155

number of projects per school:

	1	2	3	4	5	6	7	8
frequency	32	20	7	3	7	0	1	1

The very large number of different types of school present in the LEA renders the relating of numbers of data yielding projects to school type somewhat complex (See Table 2). Data yielding projects in First and Infants Schools appear to be few in relation to the number of such schools in the LEA. This may reflect greater doubt in such schools about the relevance of the Paired Reading technique for children of this age, together with a disinclination to utilise norm referenced evaluation with such a population. By contrast, projects in Middle schools appear over-represented in relation to the frequency of such schools in the LEA, but in fact a relatively small number of Middle schools operated a large number of projects each. Special schools and centres operated a proportionately large number of projects, despite the special organisational difficulties involved with large catchment areas and other factors, but of course the pupils of such schools may be seen as most in need of intervention.

The given figures for "Proportion of School Type in the LEA" are related to the total number of all types of school in the LEA, irrespective of the very various numbers of pupils attending such schools. Thus, although High Schools operating Paired Reading projects appear to be proportional to the number of such schools in the LEA, the proportion of High school pupils involved in Paired Reading projects is undoubtedly much smaller than the proportion of pupils involved in projects in smaller schools for younger children.

The length of intensive period of projects mostly ranged from 5 to 10 weeks, although 6.5% of projects were known to have longer intensive periods (See Table 3). The arithmetic mean length of intensive period

TABLE 2

Frequency of Data Yielding Projects by School Type, Compared to Incidence of School Type in the LEA

School Type	Number of Projects	Proportion Of Whole Number of Projects	Proportion of School Type in LEA	Expected Number of Projects by Proportion of School Type in LEA
Infant (and Nursery)	15	9.7%	20%	31
First (and Nursery)	19	12.3%	18%	28
Junior	25	16.1%	13%	20
Junior and Infant	37	23.9%	24%	37
First and Middle	2	1.3%	2%	3
Middle	29	18.7%	8%	12
High	18	11.6%	10%	16
Special+	9	5.8%	4%	6
Not Classified	1	0.6%	1%	2
TOTAL	155	100%	100%	155

+ Including 4 projects based on one day Moderate Learning Difficulties school (n = 6-9) and one project from each of two similar schools (n = 10,14), one project from a Remedial Centre (n = 7), and two projects from a residential Emotional and Behavioural Difficulties school where subejcts were tutored by volunteer adults from adults from residential and care staff (n = 5, 6).

TABLE 3

Frequency Distribution of Length of Intensive Period of Projects

Length of Project (weeks)	Frequency	Percent
5	9	5.8
6	12	7.7
7	11	7.1
8	64	41.3
9	26	16.8
10	22	14.2
11	2	1.3
12	1	.6
13	2	1.3
14	1	.6
15	2	1.3
17	1	.6
47	1	.6
Not known	1	.6
Total	155	100.0

Mean = 8.62 weeks
Standard Deviation = 3.63
Median = 7.71 weeks

was 8.62 weeks but the median was lower than this at 7.71 weeks, with a standard deviation of 3.63. Table 4 shows that the length of time between pre- and post-test with relation to these projects was even more variable, in some cases the inter-test period being much longer than the intensive period of participation in Paired Reading. Twenty four per cent of projects were known to have inter-test periods of longer than 10 weeks, and the arithmetic mean length of inter-test period was 10.73 weeks with a standard deviation of 6.47, while the median was considerably lower at 8.83 weeks. Where inter-test periods were considerably longer than intensive participant periods, this was usually because a school had decided to take as its pre-test measure reading test results which had been collected on participant children for other purposes some considerable time before the start of the Paired Reading project. Post-testing was almost always carried out immediately after the intensive period of the project.

The frequency distribution of length of baseline period in those projects where baseline data were collected is very flat (See Table 5), reflecting even more variability. The arithmetic mean baseline period was 19.23 weeks with a standard deviation of 10.78 and a median of 16.25 weeks, but these parameters are largely meaningless given the nature of the frequency distribution. Baseline data were available in 23 projects on the primary test which was also used for pre-post testing. In a further 8 projects baseline data were only available on secondary tests and not included in the current analysis. In baselined projects, baseline data were not always available for all subjects for whom pre-post data were available.

TABLE 4

Frequency Distribution of Length of Inter-Test Period of Projects

Length of Inter-Test Period (weeks)	Frequency	Percent
5	4	2.6
6	6	3.9
7	11	7.1
8	39	25.2
9	21	13.5
10	26	16.8
11	6	3.9
12	7	4.5
13	6	3.9
14	1	.6
15	2	1.3
16	1	.6
17	2	1.3
18	1	.6
20	2	1.3
22	5	3.2
27	1	.6
28	1	.6
51	1	.6
55	1	.6
Not known	11	7.1
Total	155	100.0

Mean = 10.73 weeks
Standard Deviation = 6.47
Median = 8.83 weeks

TABLE 5

Frequency Distribution of Length of Baseline Period of Projects

Length of Baseline Period (weeks)	Frequency	Percent
6	1	.6
8	3	1.9
9	1	.6
10	1	.6
12	1	.6
13	4	2.6
14	1	.6
15	2	1.3
16	1	.6
17	2	1.3
18	3	1.9
19	1	.6
22	1	.6
23	1	.6
25	1	.6
26	3	1.9
35	1	.6
40	1	.6
43	1	.6
51	1	.6
No baseline	124	80.0
Total	155	100.0

Mean = 19.23 weeks
Standard Deviation = 10.78
Median = 16.25 weeks

The frequency distribution of length of follow-up periods in those projects yielding follow-up data shows yet further variability (See Table 6). The distribution has an arithmetic mean of 25.85 weeks with a standard deviation of 19.41 and a median of 18.00 weeks, but these parameters are of limited relevance. For purposes of further analysis, projects yielding follow-up data were divided into those with follow-up data gathered at equal to or less than 17 weeks (Short-term Follow-up) and those with data gathered at greater than 17 weeks (Long-term Follow-up). In relatively few of these projects were follow-up data collected on both participant and comparison groups, not least because many comparison children did not remain non-participant during the follow-up period.

In subsequent analyses, the difficulties inherent in comparing gains over baseline and follow-up periods of such varying length with pre-post gains over intensive periods of participation which show much less variation necessitated the making of comparisons in terms of "Ratio Gains". Ratio Gain is defined for this purpose as the ratio of gain in reading age over the specified period to the chronological time passed during that period, based on the assumption that a "normal" rate of gain can be construed to be one month of reading age gain in one calendar "chronological" month. This latter assumption is of course highly suspect, not least because it assumes linearity in the individual child's development of reading skills.

Table 7 shows the distribution of project participants by type of tutor across all projects (i.e. whether data yielding or not). The vast majority of projects were parent tutored, with a significant minority

TABLE 6

Frequency Distribution of Length of Follow-Up Periods of Projects

Length of Follow-up Period (weeks)	Frequency	Percent
6	1	.6
7	2	1.3
8	2	1.3
12	1	.6
13	1	.6
15	1	.6
16	1	.6
18	1	.6
20	1	.6
28	1	.6
33	1	.6
34	1	.6
36	1	.6
38	1	.6
39	1	.6
45	1	.6
52	1	.6
82	1	.6
No follow-up	135	87.1
Total	155	100.0

Mean = 25.85 weeks
Standard Deviation = 19.41
Median = 18.00 weeks

TABLE 7

Distribution of Project Participants by Type of Tutor, all Projects

Type of Tutor	Frequency of Use	Proportional Usage
Natural Parent (NP)	2066	74.8%
Age-Peer Tutor (PT)	214	7.8%
Cross-Age Peer Tutor (XT)	174	6.3%
Adult Volunteer (AV)	29	1.1%
Teacher Volunteer (TV)	17	0.6%
Information Unrecorded	259	9.4%
Total	2759	100%

peer tutored. Projects deploying adult or teacher volunteers were too few to draw many valid conclusions about their relative efficacy. Returning to consideration only of projects yielding norm referenced data, Table 8 summarises the frequency of incorporation of supportive home visiting. Almost three quarters of data yielding projects did not incorporate home visits. Of those that did, about a third made less than one home visit per child during the project (i.e. made home visits only to a few participating children who were considered to especially need them, and then rarely more frequently than once during the course of the project), another third made between one and two home visits per child (i.e. usually visited all participant children once and those considered in greater need more frequently) and the final third made more than two home visits per child during the project (i.e. usually projects where several home visits were scheduled for all children during the course of the intensive period).

Home-school diary cards for the duration of the intensive period of projects were collected from approximately 600 families, although certainly not on a random basis, as previously mentioned. The mean frequency of reading per week (sessions of Paired Reading) was almost five, and the total reading time during the intensive period of the project averaged approximately seven and a quarter hours (see Table 9). Informal discussion with parents and children suggests that the number of reading sessions recorded is likely to be considerably more accurate and reliable than the total reading time recorded, since parents often guessed at the latter. Informal discussion with participants tends to suggest that total reading time recorded was more likely to be an under-estimate than an over-estimate. On average, for those project participants from whom data were collected, the frequency of Paired

TABLE 8

Frequency of Supportive Home Visiting in Data-Yielding Projects

	Number of Projects	Proportion of Projects
No Home Visits	114	73.5%
Less Than 1 Home Visit Per Child During Project	14	9.0%
1 - 2 Home Visits Per Child During Project	14	9.0%
More Than 2 Home Visits Per Child During Project	13	8.5%
TOTALS	155	100%

TABLE 9

Mean Reading Frequency and Total Reading Time

	Mean	Standard Deviation	Minimum	Maximum	Sample Size (n)
Mean Frequency of Reading Per Week	4.97	1.24	0.71	7.33	597
Total Reading Time During Project	7hrs 13mins	4hrs 7mins	20mins	36hrs 25mins	628

Note

These data are based on reading frequencies and times self-reported by participants in home-school diaries, completed daily by parents and monitored weekly by teachers.

Reading requested by project co-ordinators and informally contracted into by families (five times weekly) was met. However, there was considerable variation, and at one extreme some children barely participated at all. As previously mentioned, whether the reading recorded was actually "Paired Reading" as prescribed during the training meetings is another question.

Format of Data Analysis

Determining the most appropriate, valid and power-efficient form of data analysis is difficult in any study, given the very contradictory views on these matters expressed in the literature, and these problems are magnified in the current study, characterised as it is by great size and complexity as well as lack of a rigorous pre-determined research design. However, as Riecken and Boruch (1974) point out: "researchers are constantly faced with the necessity to keep a balance between practical and scientific necessities. By holding rigidly to an inflexible design, the experiment can be so over-controlled as to severely limit its utility, that is, it can become so artificial as to have no external validity". Campbell and Stanley (1963) make the same point, that the greater internal validity of a randomised experiment often comes at the cost of lower external validity.

A powerful critique of the traditional use of statistical significance testing has been offered by Carver (1978), who reiterates earlier doubts about the utility of the null hypothesis and is dismissive about the real world significance for practical purposes of the notion of statistical significance, especially in relation to the relatively small samples which are frequently found in research with human subjects. Carver concludes that data should be examined with scientific intent

from a whole variety of perspectives, and much more emphasis should be placed on the replication of results, rather than relying on statistical significance testing in small scale (albeit carefully designed) experiments in the hope that the latter will provide equivalent information. Carver thus equates reliability not with the internal reliability of a single experiment as reflected by statistical significance, but with reference to the replicability of the results. Statistical significance is now widely accepted to be more a function of sample size than of magnitude of effect, and as Hayes (1963) puts it: "virtually any study can be made to show significant results if one uses enough subjects, regardless of how nonsensical the content may be". Nevertheless, traditional statistical significance forms a large part of the data analysis which follows, but the interpretation of the statistical results will also be subject to the more important considerations of educational significance, in the light of the threats to validity which can be identified.

In the current study, analysis with reference to statistical significance of individual school based projects is of restricted value. The finding of statistical significance would be unlikely and of little practical importance. Therefore the data have been aggregated on the assumption that the individual school based projects were sufficiently homogenous to permit this with an adequate degree of validity. Nevertheless, school based projects were very various in terms of age and ability of children, length of intensive period of project, length of inter-test, baseline and follow-up periods, usage of different norm referenced tests and so on. Pillemer and Light (1980) have pro - vided a useful review of problems and methods in synthesising research outcomes from many studies.

These authors note that data aggregation or synthesis can mask discontinuity in data or conflicting outcomes, which can themselves offer important insights into programme effectiveness. They argue that different studies of the effectiveness of putatively the same programme in fact usually incorporate a multiplicity of differences in implementation, so that the computation of a single aggregate main effect indicator may well hide a number of setting x treatment interactions, which are themselves of major scientific and policy-making import.

As statistical significance is largely a function of sample size, it is argued that pooled data can appear spuriously significant in a situation where closer scrutiny would reveal a good deal of variation within programme type. An alternative increasingly proposed in recent years is the use of "effect size" as an indicator. A classic usage of this latter is the meta-analytic work on studies of the relationship between class size and pupil achievement (Glass and Smith 1979, Glass et al. 1982), wherein care was taken to tease out different effect sizes associated with data subsets of known homogeneity on important organisational variables ("blocking"). Glass asserts that there is no problem in mixing apples and oranges if we wish to generalise to fruit, but if it is also useful to make statements about apples, about oranges, and about the difference between them, blocked meta-analysis should yield this information.

In the class size meta-analysis, the basic unit of analysis was the comparison of achievement outcomes in classes of 2 different sizes (relatively "small" and "large"); 77 studies yielded 725 such comparisons. The differences between studies in measurement scale properties was dealt with by standardising all mean differences by the within-group standard deviation. Among the 725 comparisons, 435 or 60% favoured the smaller class. However,

results were also grouped by actual class size (1/2/3/4-5/6-10/11-16/17-23/24-34/>35) and other properties of individual studies were coded: when and where published, curriculum area (only 5.4% concerned reading), duration of instruction, estimate of subject overall ability, chronological age, type of experimental control and measurement, and actual achievement levels.

More detailed analyses then showed that in comparisons of class sizes of 2 and 28, in 90% of cases the smaller class showed higher achievement. However, this was true for only 50% of comparisons of class size 30 to size 60. Results favouring the small class were more likely if the small group had been so taught for over 100 hours. Where there was random assignment of pupils to different class sizes, results were more likely to favour the smaller class (85% of 109 random assignment studies favoured the smaller class.) Graphs were then produced showing mathematically smoothed curves of relationship between these variables, generated by least squares regression of logarithmic transformation of the differences between small and large class results.

However, the procedures of Glass and his co-workers involved assumptions of normality of distributions at various junctures, and these and other meta-analytic procedures have been criticised on this and other counts. A (partisan) review of the criticisms of meta-analytic procedures is offered by Rosenthal (1984) (see Chapter 7 - The Evaluation of Meta-analytic Procedures).

The Glass class size meta-analysis nevertheless remains a good example of what Pillemer and Light (1980) term "harnessing the benefits of contradictions". These latter authors note that different results may accrue not only from programme differences despite purported similarity and from setting x treatment interactions, but also from differences in research design (particularly procedures for assignation to treatment and control groups),

differing length of programme intervention, and differences in the unit or level of analysis (at pupil, class or school level). On the last point, Pillemer and Light (1980) observe that the more highly pre-aggregated the unit of analysis, the stronger the relationships found are likely to be. (In the Kirklees study, the vast majority of analyses were conducted at the level of the individual child, thereby incorporating conservatism).

Pillemer and Light (1980) note that there are meta-analytic techniques for combining statistical significance in addition to various methods for computing average and blocked effect sizes (listed in Rosenthal, 1984). These authors cite with approval the Rosenthal method of comparing blocked outcomes by casting the results into an overall analysis of variance with studies as a blocking variable, to give a measure of differential effectiveness. If the size of effects in the separate studies differs sharply, the studies x treatment interaction team in the ANOVA will be significant.

Other blocking variables meriting investigation are stated to include: other types of variation in students, amount of training of teachers, kind of project administrative organisation and socio-economic status of participants. These are explored or controlled for by homogeneity of service delivery in the Kirklees study. However, control or comparison group data were available for only a small sub-set (23) of the 155 Kirklees projects studied, so computation of effect size or statistical significance of experimental/control mean difference was only possible in these cases. Some form of analysis was required which was applicable to the whole of the data.

Simple forms of statistical analysis were chosen in the Kirklees study. For some variables, correlation analysis was carried out, both parametric (Pearson Product Moment Coefficient) and non-parametric (Spearman Rank Order Coefficient). The statistical significance was calculated for all correlations. For other variables, comparative analyses were carried out with respect to the means of sub-samples, statistical significance being determined by the application of the (parametric) t test and the (non-parametric) Mann-Whitney U test or the Wilcoxon test as appropriate. In a small number of cases the Chi-Squared test was used.

Regression analysis was also applied in a number of cases, and indeed some researchers may have considered this to be the analysis of choice for more widespread use. However, it was considered that although regression analysis may be interesting from a researcher's point of view, particularly insofar as partialling can indicate what proportion of the total variance is accounted for by the variable in question, it was felt that information of this kind would be considered less important by practising teachers than data which appeared more obviously to answer the usual question of whether the intervention had a significant effect, (notwithstanding the difficulties in answering this question, referred to above). It was felt that a much larger number of practitioners would be familiar with simple correlation and significance testing of means than with more complex forms of analysis, and ease of dissemination was considered a relevant factor in choosing the format of data analysis. In any event, where regression analysis was undertaken in this study in parallel with simpler forms of comparative analysis, relatively little additional information was yielded by the regression

analysis in comparison to the other, the significance of the relationship indicated by regression analysis often being finally tested by traditional t-test comparisons, yielding results very little different from the more traditional approach.

Another possible approach would have been the use of the analysis of co-variance, but this is even less familiar to practising teachers. Furthermore ANCOVA is only validly usable where clients are randomly assigned to groups, clearly not the case in the current study. ANCOVA conducted with pre-test scores as a blocking variable would be an obvious usage of the procedure were the conditions for its application satisfied, but such is the complexity of the current data that a very large number of other blocking variables could easily be identified, generating unwieldy complexity. ANCOVA is noted by Linn (1986) to be a "delicate" form of analysis, and Hargreaves and Attkisson (1978) discuss the several reasons why it should not be considered a robust form of analysis.

For many of the correlational and comparative analyses detailed below, statistical significance of findings is given in the Tables in parametric as well as non-parametric terms. Insofar as the comparative analyses are concerned, the aggregated data clearly do not meet the requirements for the valid application of the parametric t-test (Siegel, 1956). The observations cannot be assumed to be drawn from normally distributed populations, nor can these populations be assumed to have the same variance or known ratio of variances. Therefore, although the results of parametric tests of statistical significance are listed in the Tables, the results of non-parametric testing of significance were

considered of greater statistical validity and are therefore attributed a great deal more weight in the formulation of conclusions from the data. Statistical significances given in the text are almost always from non-parametric analyses.

The absence of normally distributed background populations of the same variance equally casts doubt on the application in this context of statistical analysis by repeated analysis of variance. However, where sample sizes are large, the failure of the aggregated data to meet the necessary assumptions for the valid application of parametric analyses can be considered to become less important, provided that the sample size for both groups in the comparison is large. In some of the analyses below, comparisons are made between one group represented by a large sample and another group represented by a small sample, and in this circumstance again much greater weight is accorded to the non-parametric statistical significance.

Where numbers in a group for any analysis are small, there must clearly be doubts about the possible effect of biasing factors such as the disproportionate use of certain reading tests with one sample as compared to the other sample. This could particularly occur in a situation where two groups of aggregated data of apparently similar n are compared, when in fact the members of one group are representatives of a very small number of school based projects each incorporating a large number of participants, while the other group represents a larger number of projects each incorporating a small number of participants and therefore automatically incorporating greater variability and a lesser likelihood of bias in any particular direction. For this reason,

levels of statistical significance from parametric tests will only be referred to in the text for analyses where the n for both groups is large and the number of projects in both groups is large. Significance levels cited in the text will have been generated from non-parametric tests, unless there is specific comment to the contrary.

In the event, when the statistical significance of parametric and non-parametric tests from the analyses detailed below were compared (see Table 10), in only 11% of cases did parametric and non-parametric tests differ in outcome, when referred to the usual criterion level of 0.05 (5%). In this small proportion of cases, findings of parametric significance with non-parametric non-significance outnumbered cases of parametric non-significance with non-parametric significance, but this tendency was not so strong as might be expected. As Hargreaves and Attkisson (1978) point out, where a large number of outcome measures have been employed, and only a few show a significant effect, this result may be due to chance, since 5% of any set of independent statistical tests would be erroneously "significant" at the .05 level of confidence even if there was no true effect. In the current study however, more than half of all analyses yielded results which were statistically significant.

In this study, many of the main outcomes are reported in terms of gain or difference scores. This seems superficially reasonable in the evaluation of an intervention, the whole purpose of which is to produce change. However, the use of gain scores has been widely criticised, principally on the grounds that they tend to be (a) unreliable and (b) correlated with the initial status of the subjects. When conducting an

TABLE 10

Frequency of Finding of Statistical Significance, All Analyses

Type of Finding re Statistical Significance	Number of occurrences	
Parametric and Non-Parametric Analyses both significant	136	(52%)
Parametric Analysis Significant, Non-Parametric not	17	(7%)
Non-Parametric Analysis Significant, Parametric not	11	(4%)
Neither Parametric nor Non-Parametric Analysis Significant	97	(37%)
Total	261	(100%)

Statistical significances of correlation coefficients have been included. Analyses from Tables 27 and 64 have been omitted, as in both cases there were very large numbers of significance tests relating to one major factor, and results significant parametrically only had already been discounted. Also, 3 tables involving only Chi-squared analyses were omitted.

analysis in terms of gain scores, test error can be compounded by further computation of the scores. Thus, a low pre-test score and a high post-test score owing to test error would effect the results, but if the two are combined as a gain score the error of measurement is also combined, and any apparent gain would be doubly exaggerated.

Gain scores are also generally highly correlated with pre-test scores, and low scorers at pre-test may be intrinsically likely to demonstrate higher gains, i.e. size of gain score or change and size of pre-test score or initial status are often negatively correlated. This problem may be exaggerated if higher scoring subjects are near the ceiling of the particular measuring instrument used. Nevertheless, gain score comparisons between groups have the advantage of ease of communication - they are readily understood by practising teachers and others who may be more concerned with the action implications of an intervention than with the statistical sophistication employed in the data analysis. Furthermore, as Linn (1986) argues, unreliability is of concern when difference scores are used to make decisions about individuals, but may not be a major concern when aggregating data from large numbers of subjects in order to draw general conclusions: "between-group differences in average change may be detected with considerable power despite the very low reliability of within-group change scores".

However, "a non-zero correlation between change and initial status in comparisons of groups is a potential problem because individuals who start at different levels will have different expected amounts of change in the absence of a treatment effect" (Linn,1986). Correlation of change with initial status is not a problem in itself since the observed

change is an unbiased estimate of an individual's true change, but if the groups are not equivalent initially, the differences in expected amounts of change irrespective of treatment effects will bias the estimate of the effect of the treatment.

Hays (1963) makes similar comments in a discussion of the "phenomenon" of "regression to the mean", defined thus: "where linear rules for prediction are used, it is always probable that an individual will fall relatively closer to the group mean on the thing predicted than he or she does on the thing already known", i.e. whatever the pre-test score of a subject may be, the prediction of the post-test score will always tend to be nearer the mean. This does not imply that an individual subject must or will fall relatively closer to the mean on post-test than he or she does on pre-test, but only that the best prediction is that he or she will do so. "Regression towards the mean is built into the statistical assumptions and methods we use for prediction and is not necessarily a feature of the natural world" (Hays, 1963).

In the current analysis, the implication is that high pre-test scores would be associated with a prediction of small gain scores, while low pre-test scores would be associated with a prediction of large gain scores. That is, if regression to the mean is operating, pre-test scores should be negatively correlated with size of gain. Table 11 indicates that there is a strong positive correlation between pre-test and post-test reading ages, as might be expected. Table 12 shows that for reading accuracy, with very large numbers of participants and very large numbers of projects in both groups, the correlation between pre-test scores and pre-post reading gains is indeed negative, but on non-

TABLE 11

Relationship Between Pre-test and Post-test Reading Ages

Parametric Correlation	Probability	Non-Parametric++ Correlation	Probability
	R E A D I N G A C C U R A C Y		
0.94 (2372)	0.000	0.95 (2372)	0.000
	R E A D I N G C O M P R E H E N S I O N		
0.86 (690)	0.000	0.88 (690)	0.000

( ) contains number of subjects in computation (n)
\+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

TABLE 12

Relationship between Pre-Post Reading Gains and Pre-Test Scores

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probabili-
	R E A D I N G A C C U R A C Y			
Pre-Test Score in Reading Accuracy ($\bar{x}$ = 8 yrs 3.3 mo σ = 22.7 mo)	-0.05 (2372)	0.01	-0.03 (2372)	0.10
Pre-Test Score in Reading Comp. ($\bar{x}$ = 8 yrs 10.3 mo σ = 18.5 mo)	0.12 (686)	0.00	0.08 (686)	0.01
	R E A D I N G C O M P R E H E N S I O N			
Pre-Test Score in Reading Accuracy	-0.05 (686)	0.10	-0.05 (686)	0.12
Pre-Test Score in Reading Comp.	-0.19 (690)	0.00	-0.19 (690)	0.00

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

parametric testing is so small as to fail to reach statistical significance, while on parametric testing the correlation coefficient is very small even though it does reach statistical significance. This implies that although there are detectable signs of "regression to the mean" in the reading accuracy data, these signs are very small, and this phenomenon could account for only a minute proportion of the total variance in the data.

In reading comprehension, again involving large numbers of participants in large numbers of projects, a negative correlation of - 0.19 is evident from both parametric and non-parametric analyses, both reaching statistical significance. Thus, in reading comprehension, the data are consistent with a small but significant tendency towards "regression to the mean", and the reading comprehension data will therefore be interpreted with a greater degree of caution. This finding accords with other evidence in the data, which will be cited later, that comprehension scores are innately more erratic than accuracy scores, i.e. more susceptible to random fluctuation. It should be noted that although the correlations cited are small when based on the aggregated data, this disguises the fact that between individual projects there was great variation. Within individual projects, correlations ranging from - 0.72 to + 0.96 were found.

In Table 13 the "regression to the mean" effect is pursued by exploring the relationship between pre-test scores and size of gains in both short-term and long-term follow-up periods. Numbers of participants involved were small, numbers of projects involved were small, and lengths of follow-up period were actually very variable. In Table 14, attempts

TABLE 13

Relationship Between Pre-Test Scores and Short-Term and Long-Term Follow-Up Gains

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probabilit
	R E A D I N G A C C U R A C Y			
Short-Term ($\leq$17 weeks) Follow-Up Gains	0.16 (100)	0.053	0.17 (100)	0.04
Long-Term ($>$17 weeks) Follow-Up Gains	-0.40 (170)	0.00	-0.40 (170)	0.00
	R E A D I N G C O M P R E H E N S I O N			
Short-Term ($\leq$17 weeks) Follow-Up Gains	-0.26 (75)	0.01	-0.24 (75)	0.02
Long-Term ($>$17 weeks) Follow-Up Gains	-0.05 (20)	0.41	0.14 (20)	0.48

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

TABLE 14

Relationship Between Pre-Test Scores and Short-Term and Long Term Follow-up Ratio Gains

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probability
		R E A D I N G A C C U R A C Y		
Short-Term (≤ 17 weeks)	0.09 (100)	0.20	0.10 (100)	0.16
Long-Term (> 17 weeks)	-0.28 (170)	0.00	-0.29 (170)	0.00
		R E A D I N G C O M P R E H E N S I O N		
Short-Term (≤17 weeks)	-0.21 (75)	0.04	-0.24 (75)	0.02
Long-Term (>17 weeks)	-0.05 (20)	0.41	0.01 (20)	0.48

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant, one-tailed

are made to compensate for this latter complication by expressing the relationships in terms of ratio gains. Pre-test scores in reading accuracy show no statistically significant correlation with follow-up ratio gains for short-term follow-up, but do show a statistically significant ($p = 0.00$) negative correlation (- 0.29) for long-term follow-up. In reading comprehension, there is a negative correlation (- 0.24) between pre-test scores and ratio gains in the follow-up period for short-term follow-up ($p = 0.02$), but the sample size is small (75). There was no statistically significant correlation for long-term follow-up in reading comprehension. Thus only two of the four correlations were negative and statistically significant, and it seems unlikely that regression artefacts are a strong influence in the long-term follow-up data. These trends could equally be hypothetically attributed to weaker readers tending to persist longer with Paired Reading, thus gaining longer term benefits and making more prolonged gains, while for more able readers Paired Reading is less of a priority and their lesser persistence with it results in the rate of acceleration falling off more rapidly after the intensive period of the project.

In recent years, as statistical significance testing has come under increased criticism and more concern has been expressed about the replicability and generalisability of findings, the use of various measures of "Effect Size" has become increasingly popular. A number of these are summarised in Rosenthal (1984) and Rosenthal and Rubin (1986). The effect size indicator used in the current analysis is that proposed by Glass (Glass, McGaw and Smith, 1981), which is probably the most widely used. This provides a measure of the difference between experimental and control group means by standardising the experimental

group mean with reference to the control group distribution. The formula is expressed as:-

$$\text{Glass's } \Delta \text{ (effect size)} = \frac{\text{experimental group mean - control group mean}}{\text{control group standard deviation}}$$

Thus an effect size of 1.0 would indicate that the mean of the experimental group was one standard deviation away from the mean of the control group. Glass's estimate of effect size is known to have a small sample bias, but this did not prove to be a problem in the current context.

A further point with reference to the format of data analysis is to note that most analyses are undertaken on the basis of data aggregated at the level of the individual child, while others are on the basis of data aggregated in terms of means of many individual school based projects. It will be mentioned in the ensuing text when analyses have been carried out at the level of aggregation of means of many individual projects. In all other cases the analyses have been undertaken at the level of aggregation of data on individual children.

Finally, it should be noted that analyses of reading test scores will be separated into "Reading Accuracy" and "Reading Comprehension" results. Data for "Reading Comprehension" are taken only from two tests which yield reading ages for "Comprehension" in addition to and separate from reading ages for "Accuracy", (viz Neale Analysis and New Macmillan Analysis). Results from tests which only yield a single reading age are all categorised under "Accuracy", although some of these were tests of decoding single words while others were of a cloze or sentence

completion nature and clearly placed demands on comprehension as well as decoding skills. Thus although all norm-referenced data-yielding subjects have at least pre-post reading "Accuracy" scores, and a proportion have reading "Comprehension" scores too, the distinction in the ensuing analysis is by no means clear or consistent. There is in any event considerable disagreement on the definition of "Reading Comprehension" (e.g. Wade and Dewhirst, 1983).

RESULTS - MAIN OUTCOMES AND EFFECT SIZES

Mean pre-post gains and baseline gains in reading accuracy and comprehension are given in Table 15. As pre-post inter-test periods were not the same, and baseline inter-test periods even more various, ratio gains are also given in both cases. In reading accuracy, the mean pre-post gain of 6.97 months of reading age was more than three times what might "normally" be expected, if an approximately "normal" expectation may be assumed to be a gain of one month of reading age in one chronological month. In reading comprehension, the mean pre-post gain of 9.23 months could correspondingly be considered to represent an increase of more than four times "normal" gains. However, given the great doubts previously expressed about the reliability and validity of comprehension scores, and the evidence for the existence of a degree of regression to the mean in these scores, it is doubtful whether any importance should be attached to the difference between reading accuracy pre-post gain and reading comprehension pre-post gain. A conservative interpretation would be that comprehension gains are no less than accuracy gains.

Pre-post gains are also given in Table 15 for subjects in the 23 projects for which some baseline data were available. However, baseline data were not available for all subjects for whom pre-post data were available, and the pre-post gain given includes pre-post scores for all participants in baselined projects, not just for those participants for whom baseline data was available. Although transient children are less likely to yield complete data sets, there is nothing in the data to suggest the introduction of any consistent bias. Time was not available for the lengthy selection exercise necessary to make an exact comparison.

TABLE 15

Mean Pre-Post Gains and Ratio Gains* for all projects, Baseline Gains and Ratio Gains, Pre-Post Gains and Ratio Gai for Projects with Baseline Data.

Baseline Gain	Baseline Ratio Gain	Pre-Post Gain	Pre-Post Ratio Gain	Pre-Post Gain for all Baselined Projects	Pre-Post Ratio Gain for all Baselined Projects
R E A D I N G A C C U R A C Y					
5.29 [6.63] (288)+	1.37 [1.77] (288)+	6.97 [8.02] (2372)	3.27 [3.99] (2372)	5.88 [7.14] (374)+	2.56 [3.14] (374)+
R E A D I N G C O M P R E H E N S I O N					
6.28 [9.11] (64)+	2.24 [3.24] (64)+	9.23 [10.13] (690)	4.39 [5.15] (690)	8.25 [8.64] (83)+	2.94 [3.21] (83)+

in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
* Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests.
\+ baseline data on secondary tests excluded. Number of baselined projects = 23. Baseline data not available for all subjects.

Given the extremely various length of baseline periods, comparisons can only be made in terms of ratio gains. In baselined projects, pre-post rates of gain in reading accuracy were almost twice pre-project rates of gain. In reading comprehension, pre-post rates of gain in baseline projects were high (2.94), but the rate of gain during the baseline period was also unusually high (2.24), although sample numbers are much lower for this comparison. It is also evident from Table 15 that baselined projects tended to produce lower pre-post gains than non-baselined projects, although it is not clear why this should be the case. This difference expressed in ratio gains is even more marked, but this is because the pre-post inter-test period for baselined projects tended to be longer than that for non-baselined projects, relatively depressing the pre-post ratio gain for baselined projects.

Comparison of baseline ratio gains with pre-post ratio gains in all baselined projects (see Table 16) indicates that in reading accuracy, pre-post gains are statistically significantly greater than baseline ratio gains ($p = 0.000$). In reading comprehension, the difference is not significant. A similar analysis was conducted within each individual project for which baseline data were available, and in 9 of the 23 projects pre-post ratio gains in reading accuracy were statistically significantly greater than baseline ratio gains. These 9 projects tended to have larger numbers of participants than the other 14 projects. In reading comprehension, only 1 project out of 7 yielded statistically significant differences.

Comparison group (Quasi-"Control" Group) data was available from 37 individual school-based projects (see Table 17). Within individual

TABLE 16

Comparison of Baseline Ratio Gains with Pre-Post Ratio Gains, All Baselined Projects

Baseline Ratio Gain+	Pre-Post Ratio Gain+	Difference	Significance of Difference: Non-Parametric Probability	Significance of Difference: Parametric Probability
		R E A D I N G A C C U R A C Y		
1.37 [1.77] (286)	2.74 [3.00] (286)	1.37	0.000	0.000
		R E A D I N G C O M P R E H E N S I O N		
2.24 [3.24] (64)	2.82 [3.27] (64)	*0.58*	*0.219*	*0.200*

\+ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests.
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Wilcoxon
Parametric Analysis: t-test
Data from 23 projects
Bold figures are statistically significant, probabilities one-tailed

TABLE 17

Comparison of Pre-Post Gains in Participant and Control Groups, Across All Projects

Participant Pre-Post Mean Gain+	Control Pre-Post Mean Gain+	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
R E A D I N G A C C U R A C Y				
7.42 [7.26] (580)	4.69 [6.56] (446)	2.73	0.000	0.000
R E A D I N G C O M P R E H E N S I O N				
9.90 [8.86] (170)	6.08 [7.40] (159)	3.82	0.000	0.000

\+ in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t -test
Data from 37 Projects
Bold figures are statistically significant, probabilities one-tailed.

projects, numbers of subjects in participant and comparison groups were often different. Participant pre-post mean reading age gain in control group projects was somewhat higher for both reading accuracy and reading comprehension than was the case for all projects combined. For both reading accuracy and reading comprehension, the difference in mean reading age gain between participant and comparison subjects is highly statistically significant ($p = 0.000$) in favour of the participant group. Peer tutored projects are included in this comparison. A similar analysis conducted within individual school-based projects showed that in reading accuracy, 10 projects had a participant/comparison group difference which was statistically significant on both parametric and non-parametric tests, and 24 not. In reading comprehension, one project yielded a statistically significant difference and 11 not. Of course, numbers in groups were often small. Although within each comparison group project, participant and comparison subjects were always tested at the same time, given the varying numbers of participant and comparison subjects within projects, it was possible that when the data were aggregated the mean inter-test periods of participant and comparison subjects would not be the same. To allow for this, comparison between participant and "control" groups in terms of ratio gains was also carried out (See Table 18), and again the differences were highly statistically significant ($p = 0.000$) in favour of the participant group, in both accuracy and comprehension.

Follow-up data collected some time after the end of the "intensive" pre-post period of individual school-based projects were available for 17 projects. Given the highly various length of follow-up period, it is necessary to analyse these data largely in terms of ratio gains. Short-term (less than or equal to 17 weeks) follow-up data is available from 7 projects and long-term (greater than 17 weeks) follow-up data is available

TABLE 18

Comparison of Pre-Post Ratio Gains in Participant and Control Groups, Across All Projects

Participant Pre-Post Mean Ratio Gain	Control Pre-Post Mean Ratio Gain	Difference	Significance of Difference Non-Parametric Probability	Significance of Difference Parametric Probability
READING ACCURACY				
3.35 [3.51] (580)	1.99 [2.63] (446)	1.36	0.000	0.000
READING COMPREHENSION				
4.55 [4.34] (170)	2.51 [3.15] (159)	2.04	0.000	0.000

[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis = Mann-Whitney
Parametric Analysis = t-test
Data from 37 projects
Bold figures are statistically significant, probabilities one-tailed
Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests.

from 10 projects. As for the baseline data, follow-up data were not available for all subjects for whom pre-post data were available, and the pre-post gains of participants for whom follow-up data were not available are included in the analysis. As before, it is difficult to estimate whether this form of sample attrition has introduced any consistent bias into the data. Mean pre-post ratio gains in accuracy and comprehension for projects with follow-up data were higher than the mean for all projects combined, and it is possible that the former were particularly successful projects. Table 19 shows that during short-term follow-up periods, participants were still gaining more than 2 months of reading age for each chronological month elapsed in both reading accuracy and comprehension. Table 20 shows that during long-term follow-up periods, participants were still gaining more than one month of reading age for each calendar month elapsed in both reading accuracy and comprehension. However, sample numbers are relatively small for the comprehension follow-up data, particularly at long term follow-up, and the comprehension results must be interpreted with caution, although they are very similar to the accuracy results.

Given the importance of follow-up data in evaluating the cost effectiveness of an intervention, more detailed data for those follow-up projects for which adequate information is available are given in Table 21, not least to take into account the very various follow-up periods. The results from the 3 projects involving peer rather than parent tutors were markedly the worst, and these were therefore separated out to be dealt with in isolation (see the foot of Table 21). Considering the 6 short-term follow-up projects and 8 long-term follow-up projects remaining, all involving natural parent tutors, it is evident that ratio

TABLE 19

Pre-Post and Short-Term Follow-Up (≤17 weeks) Gains and Ratio Gains for Projects with Short-Term Follow-Up Data

Pre-Post Gain+	Pre-Post Ratio Gain++	Short-Term Follow-Up Gain+	Short-Term Follow-Up Ratio Gain++
	R E A D I N G A C C U R A C Y		
11.27 [8.17] (126)*	5.93 [4.39] (126)*	4.38 [4.39] (102)*	2.01 [2.09] (102)*
	R E A D I N G C O M P R E H E N S I O N		
12.55 [10.66] (101)*	6.28 [5.53] (101)*	5.56 [6.14] (77)*	2.32 [2.59] (77)*

\+ In months of Reading Age gain
++ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests
Number of Projects = 7
[ ] contains standard deviation [σ]
( ) contains sample number (n)
* Short-term follow-up data not available for all subjects

TABLE 20

Pre-Post and Long-Term Follow-Up (>17 weeks) Gains and Ratio Gains for Projects with Long-Term Follow-Up Data

Pre-Post Gain+	Pre-Post Ratio Gain++	Long-Term Follow-Up Gain+	Long-Term Follow-Up Ratio Gain++
READING ACCURACY			
8.35 [6.70] (207)*	3.73 [3.01] (207)*	11.31 [9.25] (170)*	1.20 [0.92] (170)*
READING COMPREHENSION			
7.28 [10.01] (39)*	4.11 [5.99] (39)*	8.80 [7.52] (20)*	1.36 [1.16] (20)*

\+ In months of Reading Age gain
++ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests
Number of projects = 10
[ ] contains standard deviation [σ]
( ) contains sample number (n)
* Long-term follow-up data not available for all subjects

TABLE 21 Pre-Post and Short- and Long-Term Follow-up Ratio Gains For Projects with Follow-up Data, By Project According to Tutor Type

PROJECT NO. PROJNO	(n)	FOLLOW-UP RATIO GAIN (Accuracy) FPRGACC	FOLLOW-UP RATIO GAIN (Comprehension) FPRGCOMP	FOLLOW-UP PERIOD (weeks) FOLLPER	HOME VISITS (Yes or No)	ORGAN-ISATION CODE* ORCODE	SES**	RETARD-ATION (Accuracy) AGEGAP1	RETARD-ATION (Comp'n) AGEGAP2	PRE-POST RATIO GAIN (Accuracy) PPRGACC	PRE-POST RATIO GAIN (Comp'n) PPRGCOMP
PARENT TUTORED PROJECTS AT "SHORT TERM" (≤ 17 weeks) FOLLOW-UP:-											
BATØ2	12	2.69	1.67	13	N	4	4.9620	NA	NA	10.14	7.38
BISØ1	13	1.56	2.87	15	N	4	-3.8125	9.31	14.23	2.44	3.92
CROØ1	11	1.07	NA	7	Y	3	5.6240	11.36	NA	3.94	NA
MANØ1	14	1.19	NA	6	N	4	4.1401	-6.93	NA	4.08	NA
RAVØ1	19	4.48	4.23	8	Y	4	9.3897	6.89	9.17	7.80	6.23
SHEØ1	10	2.82	2.49	8	N	2	-7.6038	NA	NA	8.13	7.67
Σ =	79	$\bar{x}$ = 2.46	$\bar{x}$ = 3.01	$\bar{x}$ = 9.5	Y=2,N=4	$\bar{x}$ = 3.5	$\bar{x}$ = 2.9389	$\bar{x}$ = 4.91		$\bar{x}$ = 6.12	$\bar{x}$ = 6.20
PARENT TUTORED PROJECTS AT "LONG-TERM" (> 17 weeks) FOLLOW-UP:-											
CHIØ1	9	1.50	NA	33	Y	4	9.0325	18.22	NA	5.60	NA
COWØ2	11	1.62	NA	38	N	3	1.5421	-14.27	NA	4.28	NA
CROØ2	15	1.19	NA	39	Y	3	5.6240	11.04	NA	4.80	NA
LEPØ1	10	0.99	NA	18	N	3	-5.8959	NA	NA	2.77	NA
LEPØ2	21	1.21	NA	34	N	3	-5.8959	NA	NA	1.83	NA
NORØ1	15	1.01	NA	20	Y	2	-4.8636	13.69	NA	3.69	NA
TILØ2	27	1.12	NA	82	N	3	7.4172	NA	NA	4.30	NA
TILØ4	32	1.76	NA	36	Y	3	7.4172	0.09	NA	2.57	NA
Σ =	140	$\bar{x}$ = 1.33		$\bar{x}$ = 37.5	Y=4,N=4	$\bar{x}$ = 3.0	$\bar{x}$ = 2.6036	$\bar{x}$ = 0.04		$\bar{x}$ = 3.50	
PEER TUTORED PROJECTS AT "SHORT TERM" (≤ 17 weeks) FOLLOW-UP:-											
ROYØ4	23	0.46	0.71	12		5	2.6841	47.44	46.74	6.48	6.69
PEER TUTORED PROJECTS AT "LONG TERM"(> 17 weeks) FOLLOW-UP:-											
COLØ1	10	0.73	NA	52		2	-4.7954	NA	NA	2.85	NA
HIGØ3	20	0.50	1.36	28		3	10.2926	4.69	7.54	6.37	6.42

*1 = very good, 5 = poor all means weighted by n NA = not available

** high = low ses, low = high ses n = number of participants with follow-up data

gains for short-term follow-up are markedly higher than those for long-term follow-up. It might be assumed that one factor in this could be a greater practice effect over shorter inter-test periods, but the correlation between size of follow-up ratio gain and length of follow-up period is not statistically significant (coefficient = -0.36, p = 0.20, analysis at the level of project means rather than pooling data on individual participants). Regression analysis and analysis of variance on this relationship also yielded insignificant results. The follow-up projects involving home visits showed better follow-up ratio gains than those without visits (mean with visits = 1.84, without = 1.65), but this difference is far from being statistically significant (t = 0.34, p = 0.74, parametric test used on the grounds that means of individual projects may more readily be assumed to be likely to be normally distributed than the aggregated gains of individual subjects). The size of follow-up gain is not significantly correlated with the rating of quality of organisation of a project (OR Code) (coefficient = 0.27, p = 0.65).

Follow-up projects were in schools whose catchment area was very various in terms of socio-economic status. The mean index of socio-economic status for follow-up project schools was slightly below the average for the LEA as a whole. Pre-post gains in all projects were related to socio-economic status, with low-s.e.s. areas doing well (see below), but no such statistically significant effect appears with these follow-up gains (correlation coefficient = 0.23, p = 0.57). In reading accuracy, subjects in short-term follow-up projects tended to be more reading retarded in relation to their chronological age at pre-test than long-term follow-up

project subjects. However, the correlation between follow-up ratio gains and degree of retardation is not significant (coefficient = -0.04, p = 0.91).

Subjects in short-term follow-up projects made bigger pre-post gains in reading accuracy and comprehension than subjects in long-term follow-up projects and than all subjects in all projects (including those without follow-up) taken together. As previously mentioned, it is possible that the high short-term follow-up gains reflect follow-up of particularly successful projects (c.f. Tables 19 and 20, but note that these show data at the individual level not mean of the project level and include peer tutor projects). Follow-up ratio gains in reading accuracy correlated with pre-post ratio gains strongly and statistically significantly (coefficient = 0.73, p = 0.003, regression equation: fprgacc = 0.33 + 0.29 X pprgacc). It is not clear whether this implies that participants who are successful with Paired Reading in the intensive period also tend to be successful with it in the follow-up period, perhaps because they are motivated by the initial high level of success to continue, or whether those schools who are successful in producing high pre-post gains (irrespective of the Organisational Rating code) are also good at sustaining pupil interest subsequently. Whichever interpretation is placed upon it, this finding is not surprising.

Of the 14 parent tutored projects, 7 used the Neale Test, 6 the Daniels and Diack 1 Test, and 1 project used the Schonell Test. While the Neale and Daniels and Diack 1 Test do show some tendency to produce above average results (this is discussed in more detail later), 2 other tests of the 10 used in the whole study on more than 30 participants yielded better

average results, so any tendency for pre-post and follow-up gains to be higher in the follow-up projects as a result of the specific test used should not be too pronounced. The short-term follow-up projects showed no better organisation (as indicated by mean Organisational Rating code) than long-term follow-up projects and were less likely to include home visits. Nevertheless, the short-term follow-up gains might be construed to be higher than those likely to be found were all projects followed up, but this is hypothetical. The empirical data available certainly suggest that although there is a tendency to deceleration in reading progress after the initial intensive project period, reading progress nevertheless continues at above "normal" levels. Thus one may expect participants to remain relatively advantaged compared to non-participants. Accelerated participant gain during the intensive period is not followed by progress at below normal rates resulting in overall "normal" gain over the total period, ie Paired Reading gains do not "wash out" at follow-up.

The follow-up data for peer-tutored projects were much less encouraging, but it remains difficult to draw definite conclusions. All short-term peer tutor follow-up data were drawn from one project (ROY 04), where the twelve-week follow-up period included the six weeks summer holiday and which had the worst Organisational Rating code. Projects COL 01 and HIG 03 have been written up separately in the literature (see Cawood and Lee 1985, Lee 1986, Townsend 1987) and are perhaps best viewed in isolation. Table 21 suggests that peer tutor projects may yield worse follow-up gains than parent tutored projects, and this is what might be expected, since parent tutors are arguably more likely to continue tutoring beyond the end of the intensive period of the project. Whether it is reasonable on the basis of the data cited here to assert that peer tutor projects show signs

of "wash-out" is more debatable. In project COL 01 the follow-up gains of the participant group were markedly better than the follow-up gains of the comparison group, and in project HIG 03 children who were parent tutored and then peer tutored made lesser gains than children who were peer tutors throughout and similar gains to children who were peer tutees throughout. Follow-up of peer tutored Paired Reading projects is clearly an area needing further research.

Returning to analysis at the level of aggregated data on individual subjects, Table 22 details short-term and long-term follow-up gains in control group projects. Only for long-term follow-up data in reading accuracy was sample size large enough to permit the drawing of conclusions. Long-term gains in reading accuracy for participants were greater than those for comparison children, but this difference did not reach statistical significance. This suggests that the pre-post difference between participant and comparison groups (which was highly statistically significant) had not "washed-out" at follow-up, i.e. relative gains were sustained although relative acceleration was not. Of 6 individual projects with long-term follow-up data for comparison groups, 2 showed statistically significant differences in long-term follow-up gains between participant and comparison groups on within project analysis (LEP 02, TIL 02). Short-term follow-up data were available for only 1 control group project. A similar analysis was conducted with reference to short-term and long-term follow-up ratio gains in control group projects (see Table 23), to compensate for the very various lengths of follow-up period, with very similar results.

TABLE 22

Short-term and Long-Term Follow-up Gains in Control Group Projects

Period of Follow-up	Mean Gain+ Participant Group	Mean Gain+ Control Group	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y			
Short Term (≤17 weeks)	3.44 [4.08] (27)	3.17 [2.89] (12)	0.27	0.415	0.405
Long Term (> 17 weeks)	12.05 [9.97] (100)	9.81 [13.28] (89)	2.24	0.147	0.100
		R E A D I N G C O M P R E H E N S I O N			
Short Term (≤17 weeks)	9.92 [6.81] (13)		no data		
Long Term (>17 weeks)	8.67 [6.81] (12)		no data		

\+ in months of Reading Age gain.
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
Data from 7 projects
"Bold" figures are statistically significant, probabilities one-tailed.

TABLE 23

Short-Term and Long-Term Follow-up Ratio Gains in Control Group Projects

Period of Follow-up	Mean Ratio Gain Participant Group	Mean Ratio Gain Control Group	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y			
Short Term	1.36 [1.71] (27)	2.29 [2.09] (12)	-0.93	0.208	0.195
Long Term	1.06 [0.85] (100)	1.03 [1.56] (89)	0.03	0.894	0.905
		R E A D I N G C O M P R E H E N S I O N			
Short Term	2.87 [1.97] (13)		no data		
Long Term	1.36 [1.16] (20)		no data		

[ ] contains standard deviation (σ)
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
Data from 7 projects
Bold figures are statistically significant, probabilities one-tailed.
Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests.
Short Term = $\leq$ 17 weeks. Long Term = $>$ 17 weeks.

Effect Sizes

Using Glass's formula, pre-post effect sizes were computed for each of the 34 control group projects. Reading comprehension data were available for 12 of these projects (see Table 24). In reading accuracy, the mean effect size was +0.87, although the range of variability was substantial. In reading comprehension, the mean effect size was +0.77. The smaller average effect size for reading comprehension is a reflection of the lesser differential between participant and comparison groups in reading comprehension than in reading accuracy, despite the tendency for pre-post gains in comprehension to be higher than pre-post gains in accuracy. Glass's effect size formula is known to have a tendency towards over estimation when used with small samples, and in some of these control group projects, sample size was small. Hedges (1982) has proposed a correction factor to compensate for small sample size. The magnitude of correction suggested by Hedge's correction factor applied to the data from control group projects is very small, however. The average effect size would be corrected downwards by only 3%, i.e. to +0.84 for accuracy and +0.75 for comprehension.

The mean effect sizes displayed in Table 24 are of the order described by Cohen (1977) as "large". However, Glass et al. (1981) point out that the interpretation of effect sizes does not depend on any set levels of significance and that the use of descriptive adjectives is potentially misleading. They assert that there is no inherent value in any particular effect size dissociated from a context of comparative value in decision making, since depending on what benefits can be achieved at what cost, an effect size of 2 might be construed as 'poor' and an effect size of 0.1 might be construed as 'good'. However, the Binomial Effect Size Display

TABLE 24

Pre-Post Effect Sizes, Control Group Projects (EScg)

Mean Effect Size+	Number of Projects (n)	Standard Deviation (σ)	Highest ES	Lowest ES
	R E A D I N G A C C U R A C Y			
0.87	34	1.04	+5.82	-0.13
	R E A D I N G C O M P R E H E N S I O N			
0.77	12	0.72	+2.94	+0.04

+Effect sizes were computed for each control group project, using the formula for Glass's △ , ie

$$\triangle \text{ or } ES_{cg} = \frac{\text{mean pre-post gain in reading age for experimental group} - \text{mean pre-post gain in reading age for control group}}{\text{standard deviation of control group}}$$

[See Glass, G.V, McGaw, B. & Smith, M.L. (1981) Meta-Analysis in Social Research, Beverly Hills California: Sage]

The statistics in the table summarise the distribution of resulting effect sizes.

(Rosenthal and Rubin, 1982) indicates that effect sizes of this order are equivalent to an "increase in success rate" from 10% to 90%. These results may also be compared with the average attainment gain effect sizes reported in Cohen et al. (1982) in a meta-analysis of a variety of other approaches for peer tutoring - average effect sizes of 0.33 for tutors (38 studies) and 0.40 for tutees (52 studies) are cited. The mean effect sizes in Table 24 are clearly substantially larger than this.

To give a crude and approximate indication of comparability of effect size in control group projects to size of effect in the other projects, for each individual project the mean pre-post gain in reading age was expressed as a deviation from the "normal" expectation of increase in reading age (equivalent to the inter-test period), standardised by the standard deviation of the pre-post gains (see Table 25). The use of inter-test period instead of control group gains was expected to inflate the resulting effect sizes, while the use of the standard deviation of pre-post gains instead of the standard deviation of control group gains was expected to tend to reduce them. Although this is an approximation of extremely doubtful statistical validity, the approximated effect size estimate for reading accuracy for all projects was very similar to the valid one for control group projects computed using the Glass formula. The reading comprehension estimate is substantially higher and obviously should be interpreted with extreme caution. Even where effect sizes are encouragingly large, experimental effects are still usually only accounting for a very small proportion of the total variance.

TABLE 25

Pre-Post "Effect Sizes," All Projects (ESpp)

Mean Effect Size+	Number of Projects (n)	Standard Deviation (σ)	Highest ES	Lowest ES
	R E A D I N G A C C U R A C Y			
+0.81	141	0.61	+2.45	-2.72
	R E A D I N G C O M P R E H E N S I O N			
+0.92	55	0.59	+2.65	-0.49

+Effect sizes were computed for each data-yielding project, using (in the absence of control group data for all projects) the approximation:

$$ESpp = \frac{\text{mean pre post gain in reading age - inter-test period}}{\text{s.d. of pre-post gains}}$$

The statistics in the table summarise the distribution of resulting effect sizes.

RESULTS - EFFECT OF ORGANISATIONAL FACTORS

There is little evidence of a relationship of any strength between the size of pre-post gains in reading and the chronological age of the subjects, as will be reported in greater detail later. This of course begs the question of whether a gain in reading age of a specified size means the same when indicating progression from different basal reading ages (Pumfrey 1986, 1987), given the non-linearity of individual development in reading skill and the increase in the range of reading test scores found as reading age increases. Tables 26 and 27 give details of a related analysis, looking at differences in size of pre-post gains in reading accuracy and comprehension according to the type of school in which projects operated. The classification allows only a gross measure of an age-related variable, not age per se. The great number and variety of types of school in the LEA rather confuse the picture, and for some kinds of school the group size is so small that little confidence can be placed in the reliability of the finding, given the variation in results from different reading tests and other background factors at play in the data.

In Table 26, school types are ranked in relation to size of pre-post gains in reading, type of school producing the largest mean pre-post gain being ranked first. In Table 27, those differences between schools which reach statistical significance are listed. In reading accuracy, it is difficult to see any obvious relationship between size of pre-post gains and the age of child served by the different types of school. However, from visual inspection there seems to be some tendency for Middle schools and High schools (serving older pupils) to have produced somewhat better results than Junior and Junior and Infant schools

TABLE 26

Ranking of School Types in Relation to Size of Pre-Post Gains in Reading

Rank	School Type	Mean Gain Pre-Post +	Standard Deviation (σ)	Group Size (n)
	READING ACCURACY			
1	First and Nursery	8.41	6.43	58
2	Middle	8.12	10.44	502
3	Infant	6.97	5.70	130
4	High	6.83	7.48	248
5	Junior	6.56	6.58	241
6	Junior and Infant	6.55	7.70	450
7	First	6.12	7.63	137
8	Infant and Nursery	5.79	6.22	190
9	First and Middle	4.61	3.78	23
10	Special	4.37	4.40	65
	READING COMPREHENSION			
1	First	14.31	17.41	29
2	First and Nursery	12.00	8.46	34
3	Junior and Infant	10.90	8.63	191
4	Junior	10.03	8.68	39
5	Special	9.82	11.42	49
6	First and Middle	9.09	7.23	11
7	Middle	8.09	10.73	138
8	High	7.08	9.78	126
9	Infant and Nursery	2.73	8.12	22
10	Infant		No data	

+ in months of reading age.

TABLE 27

Pre-Post Gains as a Function of Type of School

COMPARISON +			Significance of Difference	
			Parametric	Non-Parametric
READING ACCURACY				
First and Middle	vs	Infant	0.008	0.037
First and Middle	vs	Middle	0.000	0.020
First and Nursery	vs	Infant	0.004	0.001
First and Nursery	vs	Junior and Infant	0.023	0.022
First and Nursery	vs	First	0.017	0.036
First and Nursery	vs	High	0.053	0.029
First and Nursery	vs	Junior	0.027	0.042
First and Nursery	vs	Special	0.000	0.000
Infant and Nursery	vs	Junior and Infant	0.094	0.026
Infant and Nursery	vs	High	0.056	0.039
Infant and Nursery	vs	Infant	0.041	0.007
Infant and Nursery	vs	Junior	0.106	0.009
Infant and Nursery	vs	Middle	0.000	0.000
Junior and Infant	vs	Middle	0.004	0.001
Junior and Infant	vs	Special	0.001	0.004
First	vs	Middle	0.007	0.015
First	vs	Special	0.021	0.010
High	vs	Middle	0.027	0.008
High	vs	Special	0.001	0.005
Infant	vs	Special	0.001	0.001
Junior	vs	Middle	0.007	0.012
Junior	vs	Special	0.001	0.001
Middle	vs	Special	0.000	0.000
READING COMPREHENSION				
First and Middle	vs	Infant and Nursery	0.016	0.032
First and Nursery	vs	Infant and Nursery	0.000	0.000
First and Nursery	vs	High	0.003	0.002
First and Nursery	vs	Middle	0.013	0.008
Infant and Nursery	vs	Junior and Infant	0.000	0.000
Infant and Nursery	vs	First	0.002	0.001
Infant and Nursery	vs	Junior	0.001	0.004
Infant and Nursery	vs	Middle	0.005	0.040
Infant and Nursery	vs	Special	0.002	0.006
Junior and Infant	vs	High	0.000	0.000
Junior and Infant	vs	Middle	0.006	0.001
First	vs	High	0.019	0.005
First	vs	Middle	0.037	0.011
High	vs	Junior	0.039	0.034

+ Comparison of mean pre-post reading age gain for two types of school specified. Differences significant only parametrically omitted. All other differences non-significant.

(serving younger pupils), if groups with small n and very variable results from schools catering for similar age ranges are disregarded. In reading comprehension, however, Junior and Junior and Infant schools seem to have done rather better than Middle and High schools, i.e. the situation is reversed, although numbers in groups are very small for many types of school. Special schools performed relatively better in reading comprehension than in reading accuracy, although they might have been expected to be bottom of the list in both areas of reading skill gain. As with younger children, the relative increase in reading age of special school pupils might be more impressive than the absolute increase.

Considering only statistical significance, for reading accuracy Middle schools produced significantly better pre-post gains than Junior and Infant schools, Junior schools and High schools. In reading comprehension, Junior and Infant schools yielded statistically significantly higher pre-post gains than both High and Middle schools. These results are consistent with the small but positive and statistically significant correlations between pre-post gains in reading accuracy and chronological age of tutee which will be reported in more detail later. However, interpretation of all these results should be cautious in the light of Pumfrey's (1986, 1987) assertion that the same reading age gains from higher basal reading ages are less educationally significant than those from lower basal reading ages.

Effect of Time of Year

The second (Spring) term of the academic year was by far the most popular term among schools for the operation of data yielding projects

(see Table 28). The Autumn term was much less popular and the Summer term the least popular. However, projects operating in the Autumn term (Term 1) produced statistically significantly better pre-post gains in reading accuracy and reading comprehension than projects in the Spring term (Term 2). For outcome comparisons in reading accuracy between other terms, the probabilities yielded by parametric and non-parametric tests are very different despite the large sample sizes, presumably indicating some gross discontinuities in the data. Adoption of the more stringent 1% level of statistical significance for these comparisons would seem appropriate, making the conservative assumption of non-significance.

It is of interest that by far the most popular term for operating projects yields lower mean pre-post gains in reading accuracy than the other two terms. For reading comprehension, mean pre-post gains are lowest for the summer term, and the difference between this term and the Autumn term reaches statistical significance although the sample size for the Summer term was only 119. This finding is more in line with the expectations of practising teachers, who tend to regard the Summer term as unfavourable for the operation of projects because of its fragmentation by holidays and out of school trips and because warmer weather and more hours of daylight are assumed to reduce children's motivation to stay indoors after school and read. However, the difference between pre-post gains for the Spring and Summer terms was small and did not reach statistical significance.

Effect of General Organisational Quality

The statistically significant differences in mean pre-post gains between projects placed in the five different rating categories of general

TABLE 28

Pre-Post Gains as a Function of Time of Year Project Operated

School Terms Compared	Mean Gain+ Pre-Post			Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G	A C C U R A C Y			
Term 1 Term 2	7.92 6.47	[7.22] [7.53]	(560) (1422)	1.45	0.000	0.000
Term 1 Term 3	7.92 7.51	[7.22] [10.46]	(560) (380)	0.41	0.030*	0.255*
Term 2 Term 3	6.47 7.51	[7.53] [10.46]	(1422) (380)	-1.04	0.236*	0.035*
		R E A D I N G	C O M P R E H E N S I O N			
Term 1 Term 2	10.62 9.13	[8.45] [9.99]	(147) (424)	1.49	0.020	0.040
Term 1 Term 3	10.62 8.21	[8.45] [12.23]	(147) (119)	2.41	0.007	0.035
Term 2 Term 3	9.13 8.21	[9.99] [12.23]	(424) (119)	0.92	0.128	0.227

\+ in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
Data from 147 Projects
Bold figures are statistically significant, probabilities one-tailed.

Term 1 = Autumn Term (August-December); Term 2 = Spring Term (January-March); Term 3 = Summer Term (April-July).

*in view of the discrepancy between parametric and non-parametric probabilities, the conservative assumption of non-significance has been made.

organisational quality are mapped in Figure 1 (reading accuracy) and Figure 2 (reading comprehension). There was clearly a tendency for projects rated as having better overall organisational quality to produce significantly better pre-post gains in reading accuracy, although the sample size in category 1 was not very large. Projects with "very good" organisation produced statistically significantly better results than all other types of project, and projects with "good" and "average" organisation produced better results than projects with "poor" organisation.

The data for reading comprehension show more statistically significant differences (see Figure 2), despite smaller sample sizes (and very small sample size in category 1), but the inter-relationships between organisational categories are not what might be expected. Projects with "very good" organisation produce significantly better results than all other types of project, but the sample size is so small as to render this finding of doubtful reliability and validity. However, "average" projects produce significantly worse results than "poor" projects. The pattern of outcome for organisational categories 3 and 5 does not conform to expectations. This may reflect the greater variability and suspected lesser reliability of the reading comprehension tests used, compared to the reading accuracy tests. However, as the rating exercise was not carried out "blind", the reliability and validity of this whole exercise is open to grave doubt, and results must be interpreted with extreme caution.

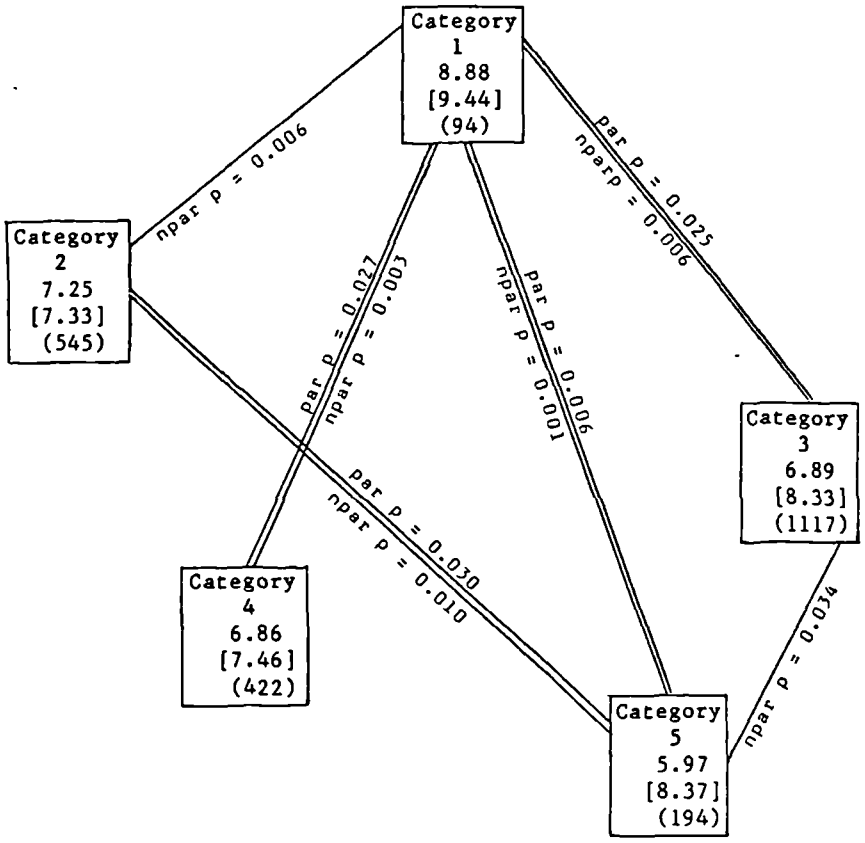
Effect of Experience in Project Operation

Mean pre-post gains were calculated separately for projects which represented a school's first project, the same school's second project,

FIGURE 1

Pre-Post Gains in Reading Accuracy as a Function of Organisational Factors in Projects

Categories of Organisational Quality:		No. Projects	Total Participan
1	Very Good	4	100
2	Good	35	610
3	Average	67	1325
4	Below Average	33	490
5	Poor	16	234
	Total	155	2759



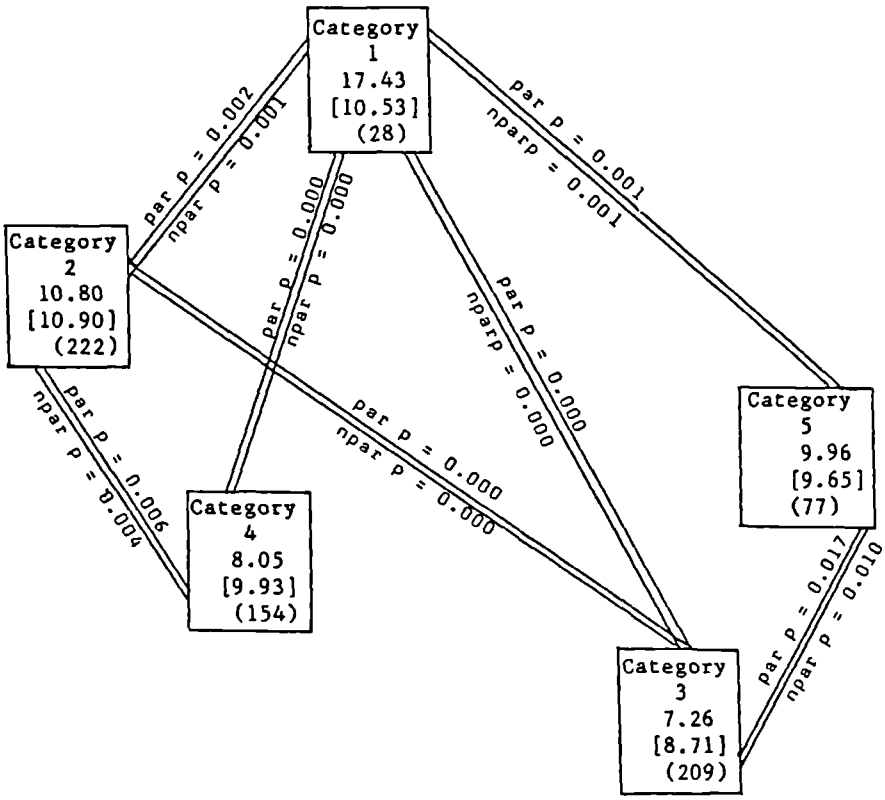
For each category of project, mean pre-post gain in reading accuracy age in months is indicated, together with standard deviation [σ] and group size (n). Par = parametric probability, t-test; Npar = non-parametric probability, Mann-Whitney. Comparisons significant parametrically only were omitted; all other comparisons non-significant.

FIGURE 2

Pre-post Gains in Reading Comprehension as a Function of Organisational Factors in Projects

Categories of Organisational Quality:

1. Very Good
2. Good
3. Average
4. Below Average
5. Poor



For each category of project, mean pre-post gain in reading comprehension age in months is indicated, together with standard deviation [σ] and group size (n). Par = parametric probability, t-test; Npar = non-parametric probability, Mann-Whitney. Comparisons significant parametrically only were omitted; all other comparisons non-significant.

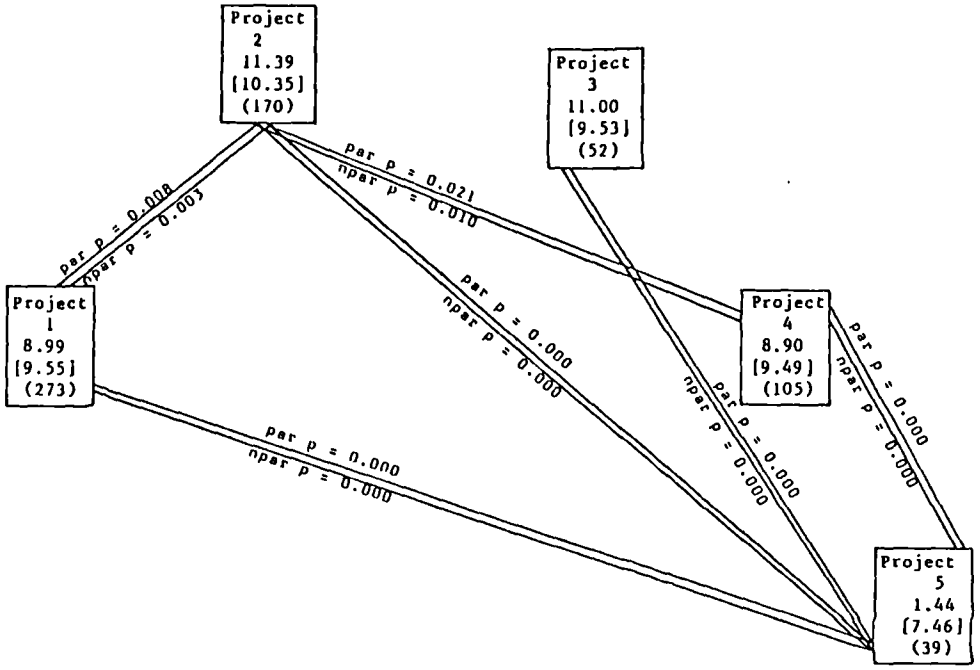
and so on, to give an indication of any relationship between size of gains and a school's experience in operating projects. For reading accuracy, only one difference reached statistical significance, first projects yielding significantly higher gains than third projects ($p = 0.03$).

In reading comprehension, a much stronger pattern of statistically significant relationships emerged, despite smaller sample sizes, and these are mapped in Figure 3. There is some evidence that for reading comprehension, a school's second project is likely to yield higher pre-post gains than either a first or a fourth project. Results were particularly poor for fifth projects, but the group size here was very small. It may be hypothesised that this pattern reflects an improvement in project outcomes as a function of increased experience, which is followed by a tailing off, perhaps as teacher enthusiasm and novelty wears off. However, if this were the case, one would expect a similar finding for reading accuracy, and the reading comprehension data have been treated with greater caution throughout the data analysis.

Effect of School Size, Project Length and Socio-Economic Status

There is no evidence of a statistically significant relationship between the number of pupils on roll in a school and the mean pre-post gains in Paired Reading projects in either reading accuracy or reading comprehension (see Table 29). Further correlational analysis of school type as a factor in the relationship between school size and reading gains yielded no significant results. The length of intensive period of a project showed a small but statistically significant correlation with the size of pre-post gains in reading accuracy (coefficient = 0.22, p =

Comparison of Pre-Post Gains in Reading Comprehension as a Function of a School's Experience in Operating Projects



For each project, mean pre-post gain in Reading Comprehension age in months is indicated, together with standard deviation [σ] and group size (n). Par = parametric probability, t-test; Npar = non-parametric probability, Mann-Whitney. Comparisons significant parametrically only were omitted, all other comparisons non-significant.

Note: For projects beyond the 5th, numbers in groups were too small to justify inclusion.

FIGURE 3

TABLE 29

Relationship of School Size, Project Length and Socio-Economic Status of Catchment Area to Gains in Reading Accuracy and Comprehension

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probability
	R E A D I N G A C C U R A C Y			
School Size	0.06 (141)	0.22	0.11 (141)	0.09
Length of Intensive Period of Project	**0.34** (141)	**0.00**	**0.22** (141)	**0.01**
Socio-Economic Status of Catchment Area - Index	0.07 (132)	0.22	**0.17** (132)	**0.02**
	R E A D I N G C O M P R E H E N S I O N			
School Size	-0.11 (53)	0.21	-0.08 (53)	0.29
Length of Intensive Period of Project	0.07 (54)	0.32	0.07 (54)	0.31
Socio-Economic Status of Catchment Area - Index	0.10 (45)	0.27	0.07 (45)	0.33

( ) contains number of projects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant
nb: in S.E.S. index, higher figures = greater disadvantage
all analyses at project level

0.01), but this was not the case with reading comprehension. There is thus some indication that children who are involved in Paired Reading projects with longer intensive periods tend to make larger pre-post gains in reading accuracy, but this finding must be set in the context that longer project intensive periods were usually specified by co-ordinating teachers for older pupils, and the educational significance of larger pre-post gains for pupils with higher basal reading ages is a questionable issue, as aforementioned. Leaving this aside, it could be construed that this positive correlation lends some support to the idea that Paired Reading works by virtue of a "practice effect", i.e. the longer you do it the more reading practice you get and therefore the better you get. However, other evidence will be cited below which suggests that this would not be a valid interpretation of this positive correlation.

There is also evidence in Table 29 of positive correlation which is statistically significant (0.17, $p = 0.02$) between the index of socio-economic disadvantage of a school's catchment area and the size of pre-post gains in reading accuracy. Given the structure of the index, this means there is evidence of a tendency for socio-economically disadvantaged areas to yield <u>higher</u> reading accuracy gains, and advantaged areas lower gains. The relationship between socio-economic status and pre-post gains is explored in further detail in Tables 30 - 32. Table 29 includes analyses all of which were conducted at <u>project</u> level, but Tables 30-32 incorporate aggregated data from individual subjects.

In order to explore the interaction between socio-economic disadvantage of catchment area and frequency of home visiting in terms of the effect

TABLE 30

Relationship Between Pre-Post Reading Gains, Socio-Economic Status of Catchment Area, and Frequency of Home Visiting

Indices of socio-economic status for school catchment areas were categorised thus:- >5 (low s.e.s.) = category 0, 0.001 to 5 = category 3, 0 to -4.9 = category 6, ⩽-5 (high s.e.s.) = category 9. Frequency of home visiting within projects was expressed in terms of mean visits per participant and categorised thus: frequency 0 = category 0, < 1 = category 1, ⩾1 = category 2. Further categories of interaction between home visiting frequency and socio-economic status were then created (labelled NNV). The table shows under each NNV category the sample size (n), mean pre-post gain ($\bar{x}$) and standard deviation (σ), first for Reading Accuracy and secondly for Reading Comprehension.

		Home Visit Frequency →	0	<1	⩾1	Total n
		HVF Category →	0 = ↓	1 = ↑	2 = ↑↑	
SES Index	>5	SES Category 0 = ↓↓	NNV 0 322 7.47 8.23 5 10.40 2.97	NNV1 82 9.80 8.45 25 13.88 12.16	NNV2 84 8.92 6.32 26 12.38 8.53	488 (29%) 56 (11%)
	0.001 to 5	SES Category 3 = ↓	NNV 3 296 6.83 7.38 79 10.61 9.04	NNV 4 37 8.73 7.08 0 0 0	NNV 5 61 7.02 5.84 35 5.34 9.67	394 (23%) 114 (22%)
	0 to - 4.9	SES Category 6 = ↑	NNV6 395 5.47 5.76 220 7.36 9.18	NNV 7 37 10.70 9.18 21 20.33 10.39	NNV 8 81 5.36 4.78 31 9.06 6.31	513 (31%) 272 (53%)
	⩽-5	SES Category 9 = ↑↑	NNV9 208 6.46 5.21 68 10.54 13.62	NNV10 29 4.03 9.22 0 0 0	NNV11 42 12.43 12.12 7 17.57 8.12	279 (17%) 75 (14%)
Total n		Accuracy Comprehension	1221 (73%) 372 (72%)	185 (11%) 46 (9%)	268 (16%) 99 (19%)	1674 517

Note: only gains made by parent-tutored tutees included.

TABLE 31

Ranking of NNV Categories According to Size of Pre-Post Gains in Reading Accuracy and Comprehension in Months

RANKING	Reading Accuracy				Reading Comprehension			
	NNV	$\bar{x}$	n	σ	NNV	$\bar{x}$	n	σ
1	ses↑↑ 11 hv ↑↑	12.43	42	12.12	7	20.33	21	10.39
2	ses↑ 7 hv ↑	10.70	37	9.18	11	17.57	7	8.12
3	ses↓↓ 1 hv ↑	9.80	82	8.45	1	13.88	25	12.16
4	ses↓↓ 2 hv ↑↑	8.92	84	6.32	2	12.38	26	8.53
5	ses↓ 4 hv ↑	8.73	37	7.08	3	10.61	79	9.04
6	ses↓↓ 0 hv ↓	7.47	322	8.23	9	10.54	68	13.62
7	ses↓ 5 hv ↑↑	7.02	61	5.84	0	10.40	5	2.97
8	ses↓ 3 hv ↓	6.83	296	7.38	8	9.06	31	6.31
9	ses↑↑ 9 hv ↓	6.46	208	5.21	6	7.36	220	9.18
10	ses↑ 6 hv ↓	5.47	395	5.76	5	5.34	35	9.67
11	ses↑ 8 hv ↑↑	5.36	81	4.78	No data in categories NNV4, NNV10			
12	ses↑↑ 10 hv ↑	4.03	29	9.22				

TABLE 32 Statistical Significance of Significant Differences in Pre-Post Reading Gains Between Categories of Socio-economic Status/Home Visit Frequency Interaction

COMPARISON		ACCURACY				COMPREHENSION			
		PAR		NON-PAR		PAR		NON-PAR	
NNV0	NNV1	0.014	s	0.043	s	i.d.		i.d.	
0	2	0.042	s	0.087		i.d.		i.d.	
0	3	0.155		0.027	s	i.d.		i.d.	
0	6	0.000	s	0.000	s	i.d.		i.d.	
0	7	0.024	s	0.020	s	i.d.		i.d.	
0	8	0.002	s	0.006	s	i.d.		i.d.	
0	9	0.043	s	0.031	s	i.d.		i.d.	
0	10	0.031	s	0.017	s	i.d.		i.d.	
0	11	0.007	s	0.022	s	i.d.		i.d.	
1	3	0.003	s	0.002	s	0.122		0.136	
1	5	0.011	s	0.045	s	0.003	s	0.003	s
1	6	0.000	s	0.000	s	0.008	s	0.004	s
1	7	0.308		0.210		0.030	s	0.035	s
1	8	0.000	s	0.000	s	0.041	s	0.091	
1	9	0.001	s	0.003	s	0.131		0.064	
1	10	0.003	s	0.002	s	i.d.		i.d.	
2	3	0.006	s	0.001	s	0.190		0.135	
2	5	0.032	s	0.061		0.002	s	0.004	s
2	6	0.000	s	0.000	s	0.004	s	0.004	s
2	7	0.144		0.128		0.004	s	0.004	s
2	8	0.000	s	0.001	s	0.054		0.093	
2	9	0.001	s	0.001	s	0.218		0.075	
2	10	0.006	s	0.003	s	i.d.		i.d.	
2	11	0.042	s	0.099		i.d.		i.d.	
3	5	0.415		0.156		0.004	s	0.004	s
3	6	0.005	s	0.049	s	0.004	s	0.004	s
3	7	0.009	s	0.003	s	0.001	s	0.000	s
3	8	0.016	s	0.163		0.157		0.304	
3	11	0.003	s	0.001	s	i.d.		i.d.	
4	6	0.005	s	0.007	s	i.d.		i.d.	
4	8	0.006	s	0.012	s	i.d.		i.d.	
4	9	0.035	s	0.063		i.d.		i.d.	
4	10	0.014	s	0.018	s	i.d.		i.d.	
4	11	0.049	s	0.114		i.d.		i.d.	
5	6	0.029	s	0.021	s	0.127		0.096	
5	7	0.017	s	0.015	s	0.000	s	0.000	s
5	8	0.037	s	0.051		0.033	s	0.014	s
5	9	0.253		0.194		0.014	s	0.012	s
5	10	0.060		0.041	s	i.d.		i.d.	
5	11	0.005	s	0.014	s	i.d.		i.d.	
6	7	0.001	s	0.000	s	0.000	s	0.000	s
6	9	0.017	s	0.020	s	0.038	s	0.041	s
6	11	0.001	s	0.000	s	i.d.		i.d.	
7	8	0.001	s	0.000	s	0.000	s	0.000	s
7	9	0.005	s	0.002	s	0.001	s	0.000	s
7	10	0.003	s	0.003	s	i.d.		i.d.	
8	9	0.044	s	0.083		0.231		0.431	
8	11	0.001	s	0.000	s	i.d.		i.d.	
9	10	0.088		0.047	s	i.d.		i.d.	
9	11	0.002	s	0.001	s	i.d.		i.d.	
10	11	0.001	s	0.002	s	i.d.		i.d.	

PAR = Parametric Test NON-PAR = Non-parametric Test
i.d. = insufficient data
($n < 20$)
s = statistically significant

on pre-post reading gains, indices of socio-economic disadvantage for school catchment areas were categorised thus: - more than 5 (low s.e.s.) = category 0, 0.001 to 5 = category 3, 0 to - 4.9 = category 6, less than or equal to - 5 (high s.e.s.) = category 9. Frequency of home visiting within projects was expressed in terms of mean visits per participant and categorised thus:- frequency 0 = category 0, less than 1 = category 1, greater than or equal to 1 = category 2. Further categories of interaction between home visiting frequency and socio-economic disadvantage were then created (labelled NNV). Table 30 shows under each NNV category the sample size (n), the mean pre-post gain ($\bar{x}$) and standard deviation (σ), first for reading accuracy and secondly for reading comprehension. Only data from parent tutored subjects are included.

The final column of Table 30 indicates that children in projects evaluated by reading accuracy tests came from all social classes, although the highest quartile of socio-economic status was markedly under-represented (containing 17% of the total n), while the lowest quartile of s.e.s. and the second quartile of s.e.s. (above average) were somewhat over-represented (29% and 31% respectively). Comprehension data were distributed much less evenly, and children in projects evaluated by reading comprehension tests were most likely to be from the second s.e.s. quartile (above average - 53%), and this may have skewed the comprehension results, the highest and lowest quartiles being markedly under-represented. However, it should not be assumed that any bias present would necessarily be in any particular direction. Most participants were in projects which did not incorporate home visits (73%). Of those in projects incorporating home visits (27%), the majority (16%) were in projects incorporating at least one visit to

every participant, while the minority (11%) participated in projects incorporating visits only for a few participants.

Despite the disparity in distribution of data yielders across socio-economic disadvantage categories in reading comprehension, when mean pre-post gains according to category of interaction between socio-economic disadvantage and home visiting frequency are ranked according to size, there is a substantial degree of rank correlation between results for reading accuracy and reading comprehension (see Table 31, Spearman coefficient = 0.79, $p = 0.006$, disregarding NNV4 and NNV10 for which no comprehension data existed). From Table 31 it is apparent that for reading accuracy, the highest and lowest rankings are inconsistent, above average s.e.s. participants receiving some home visits doing both very well and very badly - the influence of background variables such as different reading tests and small group sizes probably accounts for some of this. Apart from this anomaly, the second highest and subsequent five rankings in reading accuracy are occupied by participants of below average socio-economic status, with the lowest s.e.s. participants doing particularly well. Of the participants receiving no home visits, those of lowest s.e.s. are ranked highest, and the very large group size in this cell suggests the finding is likely to be reliable. A similar picture emerges with respect to reading comprehension, although the trend is less strong, and of course the group size in some of the cells is very small indeed.

Considering the statistical significances of significant differences in pre-post reading gains between categories of socio-economic disadvantage/home visit frequency interaction displayed in Table 32, for participants in the lowest s.e.s. category (high disadvantage), home

visits at some frequency were associated with larger pre-post accuracy gains to a statistically significant degree. Although there was a tendency for participants in the "below average" s.e.s. category to do less well than participants in the lowest s.e.s. category, this difference was statistically significant only for those receiving high frequency home visits. There was a tendency for home visited participants in the "below average" s.e.s. category to produce higher gains than those not home visited, but this effect did not reach statistical significance. For participants in the highest s.e.s. group, receiving home visits was associated with higher gains to a statistically significant degree only when the home visits were frequent, occasional home visits being inexplicably associated with low gains for this group. Although participants of above average s.e.s. tended to do the converse, i.e. do well with occasional home visits but badly with frequent home visits, this trend did not reach statistical significance. It seems clear that above average and high s.e.s. groups tend to produce below average gains in the absence of any home visits.

In summary, although there was a tendency for participants receiving home visits to yield bigger gains than those receiving none, this tendency was not so strong as the tendency for low s.e.s. (high disadvantage) participants to yield bigger gains than high s.e.s. (low disadvantage) participants. However, even in the lowest s.e.s. category, home visits at some frequency were associated with pre-post gains in reading accuracy larger to a statistically significant degree than for participants in projects with no visiting. Home visits appeared to be less strongly associated with larger gains for "below average" (third quartile) s.e.s. participants. The inconsistency in results for "high" and "above average" s.e.s. groups with "selective"

and "universal" visiting makes it difficult to draw definite conclusions. However, there was some evidence that above average and high s.e.s. groups yielded below average gains in the absence of any home visits, so it cannot be assumed that home visits are not a relevant variable with the upper two quartiles of socio-economic status.

EFFECT OF WITHIN-PAIR FACTORS

Effect of Age and Retardation

The chronological age of subjects at the point of pre-test was known for 1,560 subjects. Reading accuracy pre-post gains showed a small but statistically significant correlation with chronological age (coefficient = 0.08, p = 0.00), i.e. older subjects tended to yield larger gains in reading accuracy (see Table 33). As a specific reading age gain is less educationally significant for a more able reader than the same gain for a less able reader, this result may itself have little educational significance. For reading comprehension, the same relationship is not statistically significant.

Defining "retardation in reading" as the subject's reading age at pre-test subtracted from the subject's chronological age at pre-test, a small but statistically significant correlation existed between retardation in reading accuracy and size of pre-post gain in reading accuracy, i.e. more retarded children tended to make bigger gains (coefficient = 0.12, p = 0.00). A similar relationship was found with reference to reading comprehension (coefficient = 0.18, p = 0.00, n = 414). Although small positive correlations resulted when data on all individual subjects were aggregated, this disguised the fact that within individual projects there was great variation. For individual projects on individual variables, correlations ranging from -0.72 to + 0.96 were evident. The correlation coefficients from the aggregated data were so small that the factors involved could only account for a very small proportion of the total variance in reading age gains. However, the finding that more retarded children show a small tendency to make bigger pre-post gains does not concur with the expectations of most practising teachers, and is of interest in this regard, although "a year's retardation" has different implications at different basal reading ages.

Effect of Gender of Tutee

The mean pre- and post-test reading ages for male and female participant and comparison ("control") subjects are detailed in Table 34. Over all projects, on average girls started off (pre-test) better than boys in reading accuracy, and remained so by the end of the intensive period of projects, although the gap had closed. In reading comprehension, boys started off better, and by post-test had slightly increased the gap.

However, as Table 35 indicates, the difference in pre-post gains between male and female subjects did not reach statistical significance for either reading accuracy or reading comprehension. Of all data-yielding project participants, 54.2% were male, 42.0% were female, and in 3.8% of cases gender was unrecorded. The data thus show some tendency for male participants to do better than female participants during intensive periods of Paired Reading projects, contrasting with the common finding of greater reading progress for girls in the primary school age range (e.g. Thompson, 1975). In comparison ("control") group projects, Table 34 shows that mean pre-test scores in reading accuracy were lower for participants than controls for both sexes, but the opposite was true for reading comprehension. There is thus little evidence here that the "self-selected non-participant" nature of some comparison groups consistently biased comparison groups towards disproportionate inclusion of low attaining children.

The effect of sex of tutee on the difference between baseline and pre-post ratio gains is shown in Table 36, the use of ratio gains again being necessitated by the very variable length of baseline period. In reading accuracy, both male and female tutees made very statistically significantly greater ratio gains pre-post than during baseline ($p \leq 0.001$). Male tutees showed a greater improvement from baseline than did

TABLE 34

Mean Pre- and Post - Test Reading Ages (In Months) for Male and Female Participants and Controls

Type of Test (Pre- or Post-; Accuracy or Comprehension)	Type of Subject (Male or Female; Participant or Control)	Mean Reading Age (months)	Sample Size (n)
	ALL PROJECTS		
Pre Acc	Female Participant	100.60	1019
Post Acc	Female Paricipant	106.92	995
Pre Comp	Female Participant	105.95	243
Post Comp	Female Participant	114.81	247
Pre Acc	Male Participant	98.53	1364
Post Acc	Male Participant	105.70	1344
Pre Comp	Male Participant	107.57	434
Post Comp	Male Participant	117.23	433
	CONTROL GROUP PROJECTS		
Pre Acc	Female Control	98.82	180
Post Acc	Female Control	102.60	183
Pre Comp	Female Control	98.82	80
Post Comp	Female Control	103.47	76
Pre Acc	Male Control	96.69	208
Post Acc	Male Control	102.15	202
Pre Comp	Male Control	97.68	75
Post Comp	Male Control	104.23	77
Pre Acc	Female Participant	91.74	261
Post Acc	Female Participant	98.47	257
Pre Comp	Female Participant	103.93	72
Post Comp	Female Participant	111.72	72
Pre Acc	Male Participant	93.72	309
Post Acc	Male Participant	100.91	304
Pre Comp	Male Participant	98.60	101
Post Comp	Male Participant	110.14	99

n.b. this table gives data for all subjects for whom results were available on either occasion of testing. Later tables citing gains only include subjects for whom pre and post data were available.

TABLE 35

Pre-Post Gains as Function of Sex of Tutee

Sex of Tutee	Mean Gain+ Pre-Post	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y		
Male	7.29 [7.86] (1327)	0.069	0.088	0.020
Female	6.60 [8.24] (976)			
		R E A D I N G C O M P R E H E N S I O N		
Male	9.52 [10.63] (427)	0.76	0.286	0.175
Female	8.76 [9.51] (238)			

\+ in months of Reading Age gain.
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
"Bold" figures are statistically significant, probabilities one-tailed.

TABLE 36

Effect of Sex of Tutee on the Difference Between Baseline and Pre-Post Ratio Gains, All Projects

Sex of Tutee	Baseline Ratio Gain+	Pre-Post Ratio Gain+	Difference	Significance of Difference: Non-Parametric Probability	Significance of Difference: Parametric Probability
		READING ACCURACY			
Male	1.36 [1.49] (165)	2.93 [3.20] (165)	1.57	0.000	0.000
Female	1.45 [2.19] (108)	2.63 [2.78] (108)	1.18	0.001	0.001
		READING COMPREHENSION			
Male	1.90 [3.22] (40)	2.90 [3.75] (40)	1.00	0.175	0.146
Female	2.94 [3.27] (23)	2.69 [2.37] (23)	-0.25	0.470	0.398

+ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis = Wilcoxon
Parametric Analysis = t-test
Data from 23 Projects
Bold figures are statistically significant, probabilities one-tailed

female tutees in reading accuracy, but as Table 37 indicates, this difference did not reach statistical significance. In reading comprehension, unusually high rates of gain during baseline were evident for both male and female tutees, but especially for female tutees. For both male and female tutees, the difference between baseline and pre-post ratio gains did not reach statistical significance.

The difference in pre-post gains between participant and control groups according to the gender of the tutee are detailed in Table 38. In reading accuracy, the difference between participant and control groups was statistically significant for both male and female tutees (p equal to or less than 0.002), although the difference was larger for female tutees. In reading comprehension in control group studies, the pre-post gain of male participants was considerably greater than the mean gain of comparison group children, and this difference reached statistical significance ($p = 0.000$). Female participants showed much smaller mean pre-post gains however, and the difference between participant and control group gains for female participants for reading comprehension did not reach statistical significance. The follow-up data were similarly analysed according to gender of tutee, but resulting group sizes were too small to permit the drawing of valid conclusions, there being a preponderance of male subjects (66.7%) yielding follow-up data.

There was a consistent tendency for male participants to do better than female participants, although in no case did the difference between male and female participants reach statistical significance.

Effect of Ethnic Origin

Analysis by ethnic origin was conducted solely by comparison between "white" participants and "Asian" participants (i.e. those with family origins in

TABLE 37

Difference Between Baseline and Project Rates of Gain* in Reading Accuracy and Comprehension for Male and Female Participants

Sex of Participant and Type of Test	Sample Size (n)	Mean Difference ($\bar{x}$)	Standard Deviation (σ)	Parametric + Probability	Non-Parametric Probability +
Male Accuracy	165	1.57	3.06	0.20	0.27
Female Accuracy	108	1.18	3.70		
Male Comprehension	40	1.00	5.93	0.18	0.21
Female Comprehension	23	-0.25	4.61		

* Ratio Gains = Ratio of Reading Age Gain to Chronological Time Passed between tests.

+ Parametric Analysis = t-test Non-Parametric Analysis = Mann-Whitney. Bold figures are statistically significant, probabilities one-tailed.

NB: Other comparisons discarded as sample sizes too small

TABLE 38

Differences in Pre-Post Gains Between Participant and Control Groups According to Sex of Tutee

Sex of Tutee	Mean Gain Participant Group+	Mean Gain Control Group+	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y			
Male	7.60 [7.46] (303)	5.70 [6.51] (196)	1.90	0.002	0.002
Female	7.07 [6.94] (256)	3.86 [7.18] (178)	3.21	0.000	0.000
		R E A D I N G C O M P R E H E N S I O N			
Male	11.40 [9.13] (99)	6.57 [7.46] (75)	4.83	0.000	0.000
Female	7.80 [8.06] (71)	6.18 [7.75] (71)	1.62	0.094	0.112

\+ in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test

Bold figures are statistically significant, probabilities one-tailed.

Note

Peer Tutor gains are included

South Asia, the Indian subcontinent). Numbers in other ethnic categories coded (C Caribbean, O Other) were too few to justify valid inclusion. Only a very small proportion of Caribbean participants had been ethnically coded during data collection. It is thus possible that the "White" category includes a small number of mis-coded Afro-Caribbean or mixed-race subjects.

Fifty projects in 30 schools included Asian subjects (mean number of Asian subjects in these projects = 5.08, standard deviation = 7.42, maximum = 47, minimum = 1). The largest proportion of Asian subjects in any one project was 74% (17 out of 23). Across all projects, 9.4% of subjects were Asian, compared to an incidence of 15.4% in the background school child population (1977 figures). This indicated a disproportionately low participation rate by the Asian population, but this may of course reflect a low rate of offering participation to such populations, as well as other factors.

The mean pre-test and post-test reading ages for Asian and white participant and control subjects are detailed in Table 39. At pre-test, Asian participants were attaining less well than white participants in both reading accuracy and comprehension, lagging behind to a similar degree in each area. This finding should be interpreted in relation to the disproportionately lower take up rate into projects of the Asian population, which may have skewed the pre-test attainments of the Asian participants. By post-test, in reading accuracy the Asian subjects had reduced the gap between themselves and the white participants, but in comprehension had fallen further behind white participants. Table 40 indicates that although the mean pre-post gain for Asian participants was larger than the mean pre-post gain for white participants, this difference did not reach statistical significance. However, the reading comprehension gain for Asian participants was considerably less than the gain for white participants, and

TABLE 39

Mean Pre- and Post-Test Reading Ages (in Months) for Asian and White Participants and Controls

Type of Test (Pre- or Post-; Accuracy or Comprehension)	Type of Subject (Asian or White; Participant or Control)	Mean Reading Age (months)	Sample Size (n)
	A L L P R O J E C T S		
Pre Acc	Asian Participant	93.15	248
Post Acc	Asian Participant	100.76	247
Pre Acc	White Participant	99.89	2171
Post Acc	White Participant	106.61	2126
Pre Comp	Asian Participant	101.60	47
Post Comp	Asian Participant	107.13	48
Pre Comp	White Participant	106.68	654
Post Comp	White Participant	116.44	654
	C O N T R O L G R O U P P R O J E C T S		
Pre Acc	Asian Control	90.03	59
Post Acc	Asian Control	94.24	58
Pre Comp	Asian Control	88.50	8
Post Comp	Asian Control	95.38	8
Pre Acc	White Control	99.51	381
Post Acc	White Control	104.15	380
Pre Comp	White Control	101.03	141
Post Comp	White Control	106.84	147
Pre Acc	Asian Participant	92.23	61
Post Acc	Asian Participant	99.30	60
Pre Comp	Asian Participant	93.57	9
Post Comp	Asian Participant	98.56	9
Pre Acc	White Participant	92.88	519
Post Acc	White Participant	99.95	511
Pre Comp	White Participant	101.27	162
Post Comp	White Participant	111.61	160

n.b. this table gives data for all subjects for whom results were available on either occasion of testing. Later tables citing gains only include subjects for whom pre and post data were available.

TABLE 40

Pre-Post Gains as a Function of Ethnic Origin of Tutee

Comparison	Mean Gain+ Pre-Post	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
	R E A D I N G A C C U R A C Y			
Asian	7.38 [7.60] (241)	0.43	0.146	0.207
White	6.95 [8.05] (2096)			
	R E A D I N G C O M P R E H E N S I O N			
Asian	6.04 [7.95] (47)	**-3.52**	0.025	0.003
White	9.56 [10.25] (640)			

\+ in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
Bold figures are statistically significant, probabilities one-tailed.

this difference did reach statistical significance ($p = 0.003$), although it is based on a group size of only 47 Asian participants.

Remembering that the Asian subjects are drawn from many projects utilising many different reading tests, it may be expected that the data for the relatively small numbers of Asian participants are likely to be less biased by the preponderant use of particular reading tests in a small number of projects than is the case with other forms of data cited in this analysis. However, the issue of cultural relevance of reading test content must also be considered. Some of the tests used included material of doubtful cultural relevance for white indigenous children, which was likely to be even more alien to Asian participants. However, the finding of lesser progress by Asian participants relative to white participants in reading comprehension than in reading accuracy accords with the experience of many practising teachers, who tend to report a preoccupation with accuracy in English reading amongst South Asian families, wherein a broad English language base for the exploration of the semantic aspects of reading may be lacking in at least some of the members of family. Additionally, as pre-test scores by Asian participants were less than those of white participants, the validity of direct comparison of size of gain is again a relevant issue.

As noted, at post-test in reading accuracy the Asian subjects had reduced the lead of the white participants, but in comprehension had fallen further behind white participants. However, Asian participants in Paired Reading projects still made rates of gain in reading comprehension similar to those of white comparison children, at well above the rates of gain "normally" expected in the white population at large. In control group projects, in reading accuracy Asian participants appeared to start at the same level as white participants at pre-test and gain as much as white participants, while

Asian controls started well behind both participant groups at pre-test (and may therefore be unrepresentative) and gained much less, as did the white control group (which started ahead of white participants). Comprehension data in control group projects were available on too few Asian subjects to permit valid conclusions to be drawn.

The difference in pre-post gain between participant and control groups reached statistical significance for white participants, but just failed to reach statistical significance for Asian participants ($p = 0.057$), although this comparison is based on small numbers of Asian participants in control group projects (See Table 41). The lack of statistical significance in the reading accuracy comparison is associated with a number of factors: Asian participants gained a little less than white participants in control group projects (which was not the case in all projects combined), the Asian control group gained more than the white control group, and the numbers in the Asian groups were much smaller than in the white groups.

Analyses were conducted on the effect of ethnic origin of tutee on short-term and long-term follow-up gains, but short-term follow-up data were only available for 20 Asian tutees and long-term follow-up data for only 24 Asian tutees, in reading accuracy. Reading comprehension data for Asian tutees were only available for 15 participants at short-term follow-up. As group sizes for this comparison were so small, the results were disregarded, but in any event no statistically significant differences were found.

Effect of Tutor Type

Mean pre-test and post-test reading ages for all participants according to their type of tutor are given in Table 42. In reading accuracy, the numbers of participants tutored by adult volunteers and teacher volunteers were very

TABLE 41

Differences in Pre-Post Gains Between Participant and Control Groups According to Ethnic Origin of Tutee

Ethnic Origin of Tutee	Mean Gain+ Participant Group	Mean Gain+ Control Group	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y			
Asian	7.18 [8.37] (60)	5.14 [5.06] (56)	2.04	0.051	0.057
White	7.47 [7.14] (509)	4.66 [6.87] (371)	2.81	0.000	0.000
		R E A D I N G C O M P R E H E N S I O N			
Asian	*				
White	10.25 [8.94] (159)	6.25 [7.46] (140)	4.00	0.000	0.000

\+ in months of Reading Age Gain
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test

Bold figures are statistically significant, probabilities one-tailed.
* numbers in group too small to justify inclusion

TABLE 42

Mean Pre- and Post - Test Reading Ages (in Months) for Participants According to Type of Tutor

Type of Test (Pre- or Post; Accuracy or Comprehension)	Tutor Type	Mean Reading Age (months)	Sample Size (n)
Pre Acc	Adult Volunteer	95.35	29
Post Acc	Adult Volunteer	101.35	29
Pre Comp	Adult Volunteer	101.19	27
Post Comp	Adult Volunteer	109.11	27
Pre Acc	Natural Parent	94.63	1790
Post Acc	Natural Parent	101.47	1757
Pre Comp	Natural Parent	106.16	567
Post Comp	Natural Parent	116.00	568
Pre Acc	True-Age Peer Tutor	111.14	206
Post Acc	True-Age Peer Tutor	117.83	196
Pre Comp	True-Age Peer Tutor	98.00	19
Post Comp	True-Age Peer Tutor	107.94	17
Pre Acc	Cross-Age Peer Tutor	99.39	170
Post Acc	Cross-Age Peer Tutor	105.47	171
Pre Comp	Cross-Age Peer Tutor	110.69	70
Post Comp	Cross-Age Peer Tutor	117.03	71
Pre Acc	Teacher Volunteer	75.63	16
Post Acc	Teacher Volunteer	79.93	15
Pre Comp	Teacher Volunteer	no data	0
Post Comp	Teacher Volunteer	no data	0

n.b. this table gives data for all subjects for whom results were available on either occasion of testing. Later tables citing gains only include subjects for whom pre and post data were available.

small, and in reading comprehension this was true of these two groups and also same-age (true-age) peer tutor participants. In all these cases, results should be interpreted with great caution.

At project commencement (pre-test), all participant tutor groups had higher reading comprehension scores than reading accuracy scores, except for the same-age peer tutor group where the opposite was true. Peer tutors (especially same-age) tended to work with more able readers than was the case in parent and adult volunteer tutored groups, but this may merely reflect the fact that peer tutoring is usually deployed by practising teachers with older children, while parent tutoring is more widely used with younger children. The teacher volunteer tutored group had very much lower pre-test scores in reading accuracy than all other groups, and produced the smallest gains.

The statistical significance of differences in pre-post gain between participant groups tutored by different kinds of tutor are given in Table 43. Pre-post gains in reading accuracy for the subjects who were natural parent, same-age peer, cross-age peer or adult volunteer tutored were all statistically significantly better than those for subjects who were tutored by teacher volunteers, but the usefulness of this finding is negligible in view of the latter group's much lower mean pre-test score. In reading comprehension, natural parent tutors produced statistically significantly higher gains than cross-age peer tutors, but not significantly higher than same-age peer tutors. The main finding from this analysis is thus that, with the exception of teacher volunteers who constitute a special case, all types of tutor produce pre-post gains which do not differ to a statistically significantly degree. However, if parents tend to work with younger children and peer tutors tend to work with older children, the fact that

TABLE 43

Pre-Post Gains as a Function of Type of Tutor

Comparison	Mean Gain+ Pre-Post			Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G	A C C U R A C Y			
Natural Parent	6.90	[7.17]	(1727)	3.63	0.010	0.001
Teacher Volunteer	3.27	[3.69]	(15)			
Age-Peer Tutor	6.92	[8.73]	(195)	3.65	0.011	0.002
Teacher Volunteer	3.27	[3.69]	(15)			
X-Age Peer Tutor	6.51	[8.12]	(168)	3.24	0.052	0.004
Teacher Volunteer	3.27	[3.69]	(15)			
Adult Volunteer	6.00	[4.94]	(29)	2.73	0.026	0.023
Teacher Volunteer	3.27	[3.69]	(15)			
		R E A D I N G	C O M P R E H E N S I O N			
Natural Parent	9.58	[10.39]	(555)	2.75	0.035	0.011
X-Age Peer Tutor	6.83	[9.14]	(70)			

\+ in months of Reading Age Gain, Accuracy only.
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test

Bold figures are statistically significant, probabilities one-tailed.

Note

Statistically significant differences where not shown in any of the other tutor type comparisons, which included all permutations of natural parent, age-peer tutor, cross-age peer tutor, adult volunteer, teacher volunteer. No comprehension data available for teacher volunteers.

both tutors tend to produce the same size of pre-post gain may not necessarily imply that they have the same impact on the development of reading skills.

An analysis of the effect of type of tutor on the difference between baseline and pre-post ratio gains resulted in group sizes in all tutor type categories other than natural parent to be too few to permit the drawing of valid conclusions. The second largest group size was n = 37, (same-age peer tutored participants in reading accuracy), where the difference between baseline and pre-post ratio gains was statistically significant, and indeed was twice as large as the gain evident for natural parent tutored children, but little practical importance can be attached to this result (see Table 44).

Similar problems pertained in an analysis of the differences in pre-post gains between participant and control groups according to type of tutor, although the difference between mean pre-post gains for participant and control groups for natural parent tutored children was highly statistically significant in both reading accuracy ($p = 0.000$) and comprehension ($p = 0.000$). Participant children tutored by same-age peers demonstrated very large gains in relation to comparison groups, and these reached statistical significance ($p = 0.031$), but the group sizes were very small and these results should be discounted.

Likewise, short-term follow-up gains from natural parent tutored children were much larger than those from cross-age tutored children, and this difference reached statistical significance ($p = 0.000$ for reading accuracy and for reading comprehension), but these results are based on a cross-age tutor group size of 23 and should be discounted. Considering long-term follow-up gains, natural parent tutored children demonstrated much bigger

TABLE 44

Effect of Type of Tutor on the Difference Between Baseline and Pre-Post Ratio Gains, All Projects

Type of Tutor	Baseline Ratio Gain+	Pre-Post Ratio Gain+	Difference	Significance of Difference Non-Parametric Probability	Parametric Probability
		R E A D I N G A C C U R A C Y			
Natural Parent	1.59 [0.13] (208)	2.47 [0.18] (208)	0.88	0.000	0.000
True Age-Peer Tutor	0.90 [0.18] (37)	2.83 [2.50] (37)	1.93	0.000	0.000

+ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis = Wilcoxon
Parametric Analysis = t-test
Data from 23 Projects
Bold figures are statistically significant, probabilities one-tailed

gains in reading accuracy than age-peer tutored children, and this difference is statistically significant ($p = 0.009$), but this result is based on a same-age peer tutor group size of 15 and should be discounted.

The long-term follow-up data may be more validly analysed by visual inspection of Table 21.

Effects of Peer Tutoring

The norm referenced outcomes of peer tutoring using the Paired Reading technique will now be considered in greater detail. Pre-test and post-test data were collected on the peer tutors as well as peer tutees in 15 peer tutor projects, mostly same-age peer tutored. Table 45 shows that the mean pre-post gains for tutors were greater than those for tutees, although this difference did not quite reach statistical significance (data only available in sufficient numbers for reading accuracy). For tutees only, a comparison of baseline ratio gains with pre-post ratio gains in reading accuracy only in all peer tutor projects is detailed in Table 46. Pre-post ratio gain was substantially higher than baseline ratio gain, the difference being highly statistically significant ($p = 0.000$). As mentioned above, group sizes for control group data in peer tutored projects were too small to permit valid analysis.

When considering the organisation of peer tutor projects in Paired Reading, practising teachers often raise the question of whether children should be paired with a child of the same gender, or not. An attempt to explore this question empirically, albeit only in terms of norm referenced data, is detailed in Table 47. Data for all peer tutor projects, whether same-age or cross-age, were aggregated, pre-post outcome data for male and female tutees separated, and then these groups sub-divided according to whether pairings were same-sex or different-sex. It is evident from Table

TABLE 45

Pre-Post Gains of Tutors and Tutees in Peer Tutor Projects

Tutor Gains+	Tutee Gains+	Difference	Significance of Difference: Non-Parametric Probability	Significance of Difference: Parametric Probability
8.08 [13.03] (207)	6.39 [9.10] (204)	1.69	0.17	0.07

\+ in months of Reading Age Gain, for Reading Accuracy only
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Mann-Whitney
Parametric Analysis: t-test
Data from 15 Projects
Bold figures are statistically significant, probabilities one-tailed.

Results for Reading Comprehension available for only 12, 10 cases. Results for Long-Term Follow-Up available for only 8,6 cases. No results available for Short-Term Follow-Up.

TABLE 46

Comparison of Baseline Ratio Gains in Reading Accuracy with Pre-Post Ratio Gains for Tutees in All Peer Tutor Projects

Baseline Ratio Gain+	Pre-Post Ratio Gain+	Difference	Significance of Difference	
			Non-Parametric Probability	Parametric Probability
0.78 [1.09] (69)	3.70 [4.08] (69)	2.92	0.000	0.000

\+ Ratio Gain = Ratio of Reading Age Gain to Chronological Time Passed between tests, Reading Accuracy only.
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Non-Parametric Analysis: Wilcoxon
Parametric Analysis: t-test
Data from 23 projects (secondary test data excluded)
Bold figures are statistically significant, probabilities one-tailed
No data for Reading comprehension.

TABLE 47

Mean Pre-Post Gains for Differing Sex Combinations of Peer Tutoring Pairings

		TUTEE OUTCOME	TUTOR OUTCOME	TOTAL
		MMee	MMor	
MALE TUTORS	MALE TUTEES	7.58 [8.80] (71) R.G. = 3.78	8.07 [9.62] (71) R.G. = 4.02	142
		MFee	MFor	
	FEMALE TUTEES	4.95 [5.70] (21) R.G. = 2.49	7.57 [9.08] (21) R.G. = 3.78	42
		FMee	FMor	
FEMALE TUTORS	MALE TUTEES	3.00 [11.78] (23) R.G. = 1.50	11.87 [19.21] (23) R.G. = 5.92	46
		FFee	FFor	
	FEMALE TUTEES	7.42 [8.30] (71) R.G. = 3.70	4.35 [12.06] (71) R.G. = 2.17	142
	TOTAL	6.66 (186) R.G. = 3.32	7.06 (186) R.G. = 3.52	372

\+ In months of Reading Age Gain, for Reading Accuracy only
RG = Ratio Gain = Ratio of Reading Age to Chronological Time Passed between tests
[ ] contains standard deviation [σ]
( ) contains sample number (n)
Data from 15 projects

MMee = tutee outcome, male tutor, male tutee
MMor = tutor outcome, male tutor, male tutee
MFee = tutee outcome, male tutor, female tutee
MFor = tutor outcome, male tutor, female tutee
FMee = tutee outcome, female tutor, male tutee
FMor = tutor outcome, female tutor, male tutee
FFee = tutee outcome, female tutor, female tutee
FFor = tutor outcome, female tutor, female tutee

47 that same-sex pairings were much more common, and for these projects the project leaders' estimation is that this is much more likely to have reflected the choice of the teacher than the choice of the children. Group size numbers for male tutors with female tutees and for female tutors with male tutees are thus very small, and data in these groups must be interpreted with caution. The statistical significance of comparison of mean pre-post gains for differing sex combinations of peer tutoring pairings is given in Table 48. Both male and female same-sex tutorial pairings yield very similar pre-post gains in reading accuracy for tutees, the small difference not reaching statistical significance. However, the picture concerning outcomes for tutors is very different, with male same-sex pairings producing much higher pre-post reading gains in reading accuracy for tutors than do female same-sex pairings, and this difference is statistically significant ($p = 0.02$). For the cross-sex pairing comparisons, group sizes were much smaller, and results must be interpreted with great caution. However, there appears to be a tendency for mixed-sex combinations to be good for the tutors but poor for the tutees, particularly for female tutors with male tutees. The practical implications of these findings are discussed further in Topping and Whiteley (1988).

TABLE 48

Significance of Comparisons in Mean Pre-Post Gains+ for Differing Sex Combinations of Peer Tutoring Pairings

Comparison	Tutee Outcome	Tutor Outcome	Difference	Probability
Male Tutor with Male Tutee	7.58	8.07	-0.49	0.47
Male Tutor with Female Tutee	4.95	7.57	-2.62	0.05*
Female Tutor with Male Tutee	3.00	11.87	-8.87	0.03
Female Tutor with Female Tutee	7.42	4.35	3.07	0.02
Male Tutors with Male Tutees vs Female Tutors with Male Tutees	7.58 3.00		4.58	0.04
Female Tutors with Female Tutees vs Male Tutors with Female Tutees	7.42 4.95		2.47	0.07
Female Tutors with Female Tutees vs Female Tutors with Male Tutees	7.42 3.00		4.42	0.03
Male Tutors with Male Tutees vs Female Tutors with Female Tutees		8.07 4.35	3.72	0.02
Female Tutors with Male Tutees vs Female Tutors with Female Tutees		11.87 4.35	7.52	0.02

+ in months of Reading Age Gain, for Reading Accuracy only
Bold figures are statistically significant, Mann-Whitney nonparametric test except for * (Kolmogorov-Smirnoff test more appropriate here for small n), probabilities one-tailed. All other comparisons non-significant.

RESULTS - SUBJECTIVE FEEDBACK

Parent Feedback

Feedback questionnaires for parent tutors were used in 85 projects, but not all participating parents (1466) returned questionnaires, and not all returned questionnaires (1068) had responses to all items (questionnaire return rate 73%). As aforementioned, the parent questionnaire format offered three response options: negative/no change/positive. A positivity index was recorded for each returned questionnaire, constituting the number of positive responses less number of negative responses with other responses disregarded. The mean positivity index (PI) for the 1068 returned parental questionnaires was 6.2, (the maximum possible positivity index on this questionnaire being 12, with continuation options not included in the index). The mean percentage of positive responses in relation to the total number of responses of all kinds on the questionnaire was 61%.

Further details of parental subjective feedback are reported in Table 49. Seventy eight per cent of parents reported that since Paired Reading their child was more confident in reading. Between 71% and 73% of parents reported that their child was reading more various kinds of book, was enjoying reading more, was reading more fluently and/or was reading more accurately. Sixty eight per cent of parents reported that their child was showing better comprehension of books and 67% of children were reading more in total volume. Sixty five per cent of children were more willing to read and 62% were more interested in reading. Sixty two per cent of children were also reading more

TABLE 49

Aggregate Responses to Items in Parent Questionnaire

At Home, Child Reported To Be:+	Positive Responses	Total Responses	Proportion Positive
1. Reading More	661	983	67%
2. Reading Different Kinds of Book	721	986	73%
3. Understanding Books More	649	955	68%
4. More Confident in Reading	761	971	78%
5. More Willing to Read	634	974	65%
6. More Interested in Reading	596	960	62%
7. Enjoying Reading More	723	994	73%
8. Making Less Mistakes	695	976	71%
9. Keepinga Steadier Flow	707	985	72%
10. Reading with More Expression	589	954	62%
11. Behaving Better at Home	138	936	15%
12. Happier at Home	168	874	19%
Continuation Options:			
13a Stop Paired Reading?	83	1032	8%
13b Continue Twice Weekly?	387	1032	38%
13c Continue Five Times Weekly?	339	1032	33%
13d Continue in a Different Way	223	1032	22%

+ Questionnaire format offered 3 options: negative/no change/positive

expressively. Nineteen per cent of participant children were reported to seem happier at home and 15% of children were behaving better at home. These latter percentages are small but certainly significant, justifying the inclusion of these items in the questionnaire, which was sometimes a source of puzzlement to parents who had not responded positively to those items.

Seventy one per cent of parents said they intended to carry on with Paired Reading, 33% five times weekly and 38% twice weekly. Twenty two per cent of parents intended to continue reading at home in a different way, and only 8% intended to stop reading activities with their children at home. To what extent parents subsequently conformed to these declared intentions is of course another matter.

Teacher Feedback

The teacher feedback questionnaire offered 4 options; negative/no change/no observations made/positive. In the 29 projects using teacher questionnaires, teachers did not return questionnaires for all participating children (475), and not all returned questionnaires (430) had responses to all items (questionnaire return rate = 91%). On the whole, teacher feedback was less positive than parent feedback (see Table 50). This may reflect a higher return rate eliminating the positive bias associated with lower return rates, or the effect of offering an additional response option which was not positive, or may simply indicate that improved reading at home did not necessarily generalise back into the classroom.

TABLE 50

Aggregate Responses to Items in Teacher Questionnaire

In Class, Child Reported To Be:+	Positive Responses	Total Responses	Proportion Positive
1. Reading a Greater Amount	266	420	63%
2. Reading More Widely and Variously	221	415	53%
3. Showing Better Reading Comprehension	241	409	59%
4. Showing More Confidence in Reading	289	415	70%
5. Showing More Willingness to Read	200	414	48%
6. Showing More Interest in Reading	237	416	57%
7. Showing More Pleasure in Reading	238	414	57%
8. Showing Greater Accuracy in Reading	279	416	67%
9. Showing Greater Fluency in Reading	265	415	64%
10. Showing Greater Expressiveness in Reading	156	408	38%
11. Showing Better Pacing in Reading	186	409	45%
12. Showing Better Concentration and Motivation	151	410	37%
13. Showing Better Behaviour Generally	57	412	14%

+Questionnaire format offered 4 options:
negative/no change/no observations/ positive

The mean positivity index for teacher questionnaires was 6.5, higher than that for parent questionnaires, while the mean percentage positive responses per questionnaire for teacher feedback was 53%, lower than that for parent questionnaires, and this suggests that teacher feedback differed from parent feedback not because teachers gave more negative feedback, but that they gave more feedback which was neither negative nor positive. A minor complication here is that item 13 was not the same in teacher and parent questionnaires, and was included in the positivity index in the former but not in the latter, but any effect of this is unlikely to be significant. Teacher feedback was available from relatively few projects, and it is difficult to know to what extent these are representative of all projects. An analysis of norm referenced data with regard to the absence or presence of teacher feedback could not have helped answer this question, as teacher feedback correlated very poorly with norm referenced data (as discussed in more detail later).

Table 50 indicates that 70% of teachers reported that Paired Reading children were showing more confidence in reading in class, 67% were showing greater accuracy, 64% greater fluency, 63% were reading a greater amount, 59% were showing better comprehension, 57% were showing more interest in reading and more pleasure in reading, 53% were reading more widely and variously, 48% were showing more willingness to read, 45% were showing better pacing in reading and 38% greater expressiveness. Thirty seven per cent were showing better concentration and motivation in class and 14% better behaviour generally in school.

Child Feedback (Parent Tutored)

Tutee feedback questionnaires were completely different to the other feedback questionnaires, and offered only two response options; negative or positive. It is thus unsurprising that the mean positivity index for parent tutee feedback is 7.0 and the mean percentage positive responses per questionnaire 83%, both higher than for parent or teacher feedback questionnaires, but also relating to a smaller number of questionnaire items. In the 57 parent tutored projects using child questionnaires, not all participating children (964) returned questionnaires and not all returned questionnaires (692) had responses to all items (questionnaire return rate = 72%, very similar to the return rate of parent questionnaires).

Some child questionnaires in parent tutored projects were completed at school, but many were completed at home. Ninety five per cent of responding children felt that they were now better at all kinds of reading as a result of their Paired Reading experiences (see Table 51). Ninety two per cent reported now liking all reading better and a similar proportion a better relationship with their tutoring parent(s). Ninety per cent said they would tell others about Paired Reading and 70% said they wished to go on doing Paired Reading, a very similar proportion to that reported by parents. Regarding the actual organisational process of Paired Reading, 87% of children reported it was easy to learn to do, 86% that it was easy to find a good place to do it, 83% liked doing it, 83% found the record sheet was helpful, 77% reported it was easy to obtain books, and the lowest proportion positive (59%) was with respect to the item concerning the ease of finding time to do Paired Reading.

TABLE 51

Aggregate Responses to Items in Child Questionnaire

Child Self-Reports:+	Positive Responses	Total Responses	Proportion Positive
1. It Was Easy to Get Books	528	682	77%
2. It Was Easy to Find Time	400	674	59%
3. It Was Easy to Find a Good Place	590	684	86%
4. It Was Easy to Learn to Do	586	672	87%
5. I Liked Doing It	560	674	83%
6. The Record Sheet Was a Help	563	677	83%
P.R. has led to:			
7. Liking All Reading Better	607	662	92%
8. Getting Better at All Kinds of Reading	638	674	95%
9. Getting on Better with Each Other	588	637	92%
10. I Want to Go On Doing P.R.	475	678	70%
11. I Will Tell Other People About P.R.	601	670	90%

+Questionnaire format offered 2 options: negative/positive

Peer Tutor Feedback

The peer tutor feedback questionnaire was identical to the parent feedback questionnaire with respect to the first ten items, but the items about behaving better at home and general demeanour at home were omitted and the continuation options available were more numerous and quite different. Like the parent feedback questionnaire, the peer tutor feedback questionnaire offered three options: negative/no change/positive. In the 13 projects using peer tutor feedback questionnaires, not all participating tutors (185) returned questionnaires and not all returned questionnaires (158) had responses to all items (questionnaire return rate = 85%). Peer tutor feedback questionnaires were used in a relatively high proportion of peer tutor projects, and the questionnaire return rate was higher than for parent and parent tutee feedback, but lower than that for teacher feedback. The mean positivity index was 5.2, although the mean percentage positive responses per questionnaire was 64%, indicating that the peer tutors were more likely than parents or teachers to give negative responses on the feedback questionnaire, even allowing for the availability of fewer items on the peer tutor questionnaire.

In the peer tutoring situation, peer tutors reported their tutees to be reading more variously in 75% of cases, reading more in total volume in 73% of cases and more confident in reading in 72% of cases (see Table 52). Sixty eight per cent of peer tutees were felt to be understanding books more and sixty five per cent were more willing to read, enjoying reading more and reading more accurately. Fifty eight per cent were reading more fluently, 56% showing more interest in reading and 50% reading with more expression.

TABLE 52

Aggregate Responses to Items in Peer Tutor Questionnaire

In Tutoring Situation, Tutee Reported To Be:+		Positive Responses	Total Responses	Proportion Positive
1.	Reading More	114	156	73%
2.	Reading Different Kinds of Book	114	151	75%
3.	Understanding Books More	100	148	68%
4.	More Confident in Reading	107	149	72%
5.	More Willing to Read	97	150	65%
6.	More Interested in Reading	82	146	56%
7.	Enjoying Reading More	98	150	65%
8.	Making Less Mistakes	97	150	65%
9.	Keeping a Steadier Flow	87	150	58%
10.	Reading with More Expression	76	153	50%
	Continuation Options:++			
11.	Go On as Often as Now?	26	87	30%
12.	Go On, But Not So Often?	41	87	47%
13.	Go On, With a Different Tutee?	52	87	60%
14.	Be Tutored Yourself, by someone better?	32	87	37%
15.	Tutor Reading in a Different Way?	30	87	34%
16.	Tutor Something Else, eg Maths or Spelling	60	87	69%

+ Questionnaire format offered 3 options: negative/no change/positive
++Respondents could choose up to 3 options

For the continuation options (items 11 to 16 on the peer tutor questionnaire, of which respondents could choose up to 3), data were available from only 6 projects incorporating 87 children responding to these items. Continuation with peer tutored Paired Reading in the existing format with the existing frequency was the least chosen option (30%). However, 47% of respondents indicated that one of their 3 optional choices would be to carry on with peer tutored Paired Reading with a lesser frequency, while 34% opted to tutor reading but in a different way. Thirty seven per cent indicated an interest in being tutored themselves by somebody better. However the most chosen options were to carry on with peer tutored Paired Reading but after swapping partners (60%) or to tutor some other curriculum area, for example maths or spelling (69%), which latter was the most chosen option.

Peer Tutee Feedback

Peer tutees completed the same questionnaire as children who were tutored by their parents, the questionnaire format offering only negative and positive options. In the 13 projects using tutee questionnaires (a high proportion of peer tutor projects operated), not all participating tutees (185) returned questionnaires and not all returned questionnaires (173) had responses to all items (questionnaire return rate = 94%, the highest return rate of all feedback questionnaires). The mean positivity index was 4.9, while the mean percentage positive responses per questionnaire was 74%, implying that while there were many positive responses, there was also a substantial number of negative responses, i.e the opinions of the peer tutees appeared more divided than was the case with any other group providing feedback.

Nevertheless, the overall picture is still very positive (see Table 53). Ninety three per cent of peer tutees reported improving at all kinds of reading as a result of their Paired Reading experiences, 82% now liked all reading better and 79% felt that they had a better relationship with their peer tutor. Sixty eight per cent wished to carry on doing Paired Reading and 69% said they would tell other people about it. Seventy nine per cent found it easy to learn to do and 78% reported it was easy to find time. Seventy four per cent found it easy to find a good place, 66% liked doing it and 66% said the record sheet was a help. A minority (49%) said it was easy to get books, 51% reporting that it was hard to obtain suitable books.

Comparisons Between Subjective Feedback

The subjective feedback from parent tutors and class teachers is compared in Table 54. On every comparable item, the responses of class teachers were less positive than those of parents. This may be partially attributable to teachers having an additional multiple choice, a lesser positive bias in the teacher feedback owing to the higher return rate, or a failure of Paired Reading effects to generalise from the home into the classroom in some cases. Conversely, it could be hypothesised that the much lower proportion of projects using teacher questionnaires, compared to the proportion of projects using parent questionnaires, might have introduced some positive bias into the teacher feedback, which would otherwise have been even less positive in relation to the parent feedback.

In the event, the difference in positive responses between the two kinds of feedback is so great for certain items than it can not be readily

TABLE 53

Aggregate Responses to Items in Peer Tutee Questionnaire

Tutee Self-Reports:+		Positive Responses	Total Responses	Proportion Positive
1.	It Was Easy to Get Books	85	172	49%
2.	It Was Easy to Find Time	131	169	78%
3.	It Was Easy to Find a Good Place	125	170	74%
4.	It Was Easy to Learn to Do	134	169	79%
5.	I Liked Doing It	113	172	66%
6.	The Record Sheet Was a Help	111	169	66%
	P.R. has led to:			
7.	Liking All Reading Better	138	168	82%
8.	Getting Better at All Kinds of Reading	157	169	93%
9.	Getting on Better with Each Other	129	164	79%
10.	I Want to Go On Doing P.R.	117	172	68%
11.	I Will Tell Other People About P.R.	118	171	69%

+Questionnaire format offered 2 options: negative/positive.

TABLE 54 Subjective Feedback from Parent Tutors and Classteachers.

Concerning Reading, Child Reported to Be:- +	At Home, By Parents			At School, By Teachers			Difference in Proportions Positive	Significance of Difference++
	Positive Responses	Total Responses	Proportion Positive	Positive Responses	Total Responses	Proportion Positive		
1 Reading More	661	983	67%	266	420	63%	+ 4	0.172
2 Reading More Widely	721	986	73%	221	415	53%	+20	0.000
3 Understanding More	649	955	68%	241	409	59%	+ 9	0.002
4 More Confident	761	971	78%	289	415	70%	+ 8	0.001
5 More Willing	634	974	65%	200	414	48%	+17	0.000
6 More Interested	569	960	62%	237	416	57%	+ 5	0.081
7 Enjoying More	723	994	73%	238	414	57%	+16	0.000
8 More Accurate	695	976	71%	279	416	67%	+ 4	0.135
9 More Fluent	707	985	72%	265	415	64%	+ 8	0.004
10 More Expressive	589	954	62%	156	408	38%	+24	0.000
11 Better Behaved	138	936	15%	57	412	14%	+ 1	0.787

+ Parent questionnaire format offered 3 options, negative/no change/positive
Teacher questionnaire format offered 4 options, negative/no change/not observed/positive

++ X^2 ,d.f. = 1, probabilities two-tailed, Bold figures statistically significant

explained in terms of artifacts within data gathering. Mean percentage positive responses per questionnaire was 61% for parents and 53% for teachers. Parents were much more likely than teachers to report that their children were now reading more widely, were more willing to read, were enjoying reading more and were reading more expressively ($p = 0.000$ in all cases). Other statistically significant differences were that parents were more likely than teachers to report that their children were more confident in reading ($p = 0.001$), understanding reading more (0.002) and more fluent in reading ($p = 0.004$).

The subjective feedback from parent tutors and peer tutors is compared in Table 55. Considerable similarity is evident, although there is a tendency for parent feedback to be more positive. However, on only two items were there statistically significant differences between parent tutor and peer tutor feedback - parents were more likely than peer tutors to consider that their tutees had become more fluent and more expressive in reading ($p = 0.001$ and $p = 0.006$ respectively). The higher response rate for peer tutor questionnaires may imply that the peer tutor feedback is less likely to be positively biased than the parent tutor feedback.

Subjective feedback from children tutored by their parents and children tutored by peers is compared in Table 56, although again the substantially higher return rate for peer tutee questionnaires may mean than the peer tutee subjective feedback is less likely to be positively biased than the parent tutee feedback. On the whole, the subjective feedback from parent tutees is considerably more positive than that from peer tutees, when both same-age and cross-age peer tutee data are

TABLE 55 Subjective Feedback from Parent Tutors and Peer Tutors

Concerning Reading, Tutee Reported to Be:- +	By Parent Tutors			By Peer Tutors			Difference in Proportions Positive	Significance of Difference++
	Positive Responses	Total Responses	Proportion Positive	Positive Responses	Total Responses	Proportion Positive		
1 Reading More	661	983	67%	114	156	73%	- 6	0.170
2 Reading More Widely	721	986	73%	114	151	75%	- 2	0.612
3 Understanding More	649	955	68%	100	148	68%	0	0.996
4 More Confident	761	971	78%	107	149	72%	+ 6	0.089
5 More Willing	634	974	65%	97	150	65%	0	0.989
6 More Interested	569	960	62%	82	146	56%	+ 6	0.199
7 Enjoying More	723	994	73%	98	150	65%	+ 8	0.071
8 More Accurate	695	976	71%	97	150	65%	+ 6	0.120
9 More Fluent	707	985	72%	87	150	58%	+14	0.001
10 More Expressive	589	954	62%	76	153	50%	+12	0.006

+ For both questionnaires, format offered 3 options, negative/no change/positive
++ χ^2, d.f. = 1, probabilities two-tailed, Bold figures statistically significant

TABLE 56 Subjective Feedback from Parent Tutees and Peer Tutees

Concerning P.R., Tutee Self-Reports:+	Parent Tutees			Peer Tutees			Difference in Proportions Positive	Significance of Difference++
	Positive Responses	Total Responses	Proportion Positive	Positive Responses	Total Responses	Proportion Positive		
1 Easy to Get Books	528	682	77%	85	172	49%	+28	0.000
2 Easy to Find Time	400	674	59%	131	169	78%	−19	0.000
3 Easy to Find Good Place	590	684	86%	125	170	74%	+12	0.000
4 Easy to Learn to Do	586	672	87%	134	169	79%	+ 8	0.012
5 Liked Doing It	560	674	83%	113	172	66%	+17	0.000
6 Record Sheet Helpful	563	677	83%	111	169	66%	+17	0.000
P.R. has led to:								
7 Liking Reading Better	607	662	92%	138	168	82%	+10	0.001
8 Better at Reading	638	674	95%	157	169	93%	+ 2	0.507
9 Getting on Better with Each Other	588	637	92%	129	164	79%	+13	0.000
10 Will Go On with P.R.	475	678	70%	117	172	68%	+ 2	0.674
11 Will Tell Others re P.R.	601	670	90%	118	171	69%	+21	0.000

+ Questionnaire format offered 2 options: negative/positive
++ X^2, d.f. = 1, probabilities two-tailed, Bold figures statistically significant

aggregated, except for one item. The mean percentage positive responses per questionnaire was 83% for parent tutees and 74% for peer tutees. Parent tutees were much more likely to report that it was easy to get books, that they would tell others about Paired Reading, that they liked doing Paired Reading, that they found the record sheet helpful, that they were now getting on better with their tutor, and that it was easy to find a good place to do Paired Reading (all $p = 0.000$).

Additionally, parent tutees were more likely to report that they now liked reading better ($p = 0.001$) and that they found Paired Reading easy to learn to do ($p = 0.012$). However, on two crucial items, feedback between parent and peer tutees was not significantly different - the groups were equally likely to report that Paired Reading had led to being better at reading and that they would carry on doing Paired Reading. Peer tutees were much more likely than parent tutees to report that it was easy to find time to do Paired Reading ($p = 0.000$), as might be expected, since peer tutoring was usually scheduled to take place at least partly during class time.

Many of the differences reported for these comparisons are too large to be explicable purely in terms of bias introduced by the difference in questionnaire return rates between the two groups. The largest difference, parent tutees being much more likely than peer tutees to report it being easy to obtain books, is likely to be particularly surprising to those practising teachers who assume that easy access to a wide variety of interesting books is more likely to be a feature of the average school than of the average home. It should also be noted that although the general attitudinal indicators to reading were better for

parent tutees, they were still very positive and favourable for peer tutees. In both groups a high proportion felt that they were now more proficient at reading and wished to carry on with Paired Reading.

The subjective feedback from peer tutees and from peer tutors was then further analysed according to whether the peer tutoring had been on a same-age or cross-age basis. Table 57 compares the subjective feedback from same-age peer tutees and cross-age peer tutees. In the ten projects using same-age peer tutee questionnaires, not all participating tutees (99) returned questionnaires, and not all returned questionnaires (91) had responses to all items (questionnaire return rate = .92%). In the three projects using cross-age peer tutee questionnaires, not all participating tutees (86) returned questionnaires and not all returned questionnaires (82) had responses to all items (questionnaire return rate = 95%). Although the questionnaire return rate was similar for both groups, the fact that the cross-age feedback data were generated from only three projects might have resulted in some element of bias in the data.

The mean percentage positive responses per questionnaire for same-age peer feedback was 78%, and the mean percentage of positive reponses per questionnaire for cross-age feedback was 69%. The feedback from the same-age peer tutees was on the whole considerably more positive than that from the cross-age peer tutees. Despite the small numbers involved, the greater positivity of same-age peer tutee feedback reached statistical significance on a number of items:- ease of finding a good place ($p = 0.000$), finding the record sheet helpful (0.004), liking

TABLE 57 Subjective Feedback from Age-Peer Tutees and Cross-Age Peer Tutees

Concerning P.R., Tutee Self-Reports:+	Age-Peer Tutees			Cross-Age Peer Tutees			Difference in Proportions Positive	Significance of Difference++
	Positive Responses	Total Responses	Proportion Positive	Positive Responses	Total Responses	Proportion Positive		
1 Easy to Get Books	52	91	57%	33	81	41%	+16	0.044
2 Easy to Find Time	72	89	81%	59	80	74%	+ 7	0.643
3 Easy to Find Good Place	78	89	88%	47	81	58%	+30	0.000
4 Easy to Learn to Do	74	91	81%	60	78	77%	+ 4	0.615
5 Liked Doing It	63	90	70%	50	82	61%	+ 9	0.278
6 Record Sheet Helpful	68	89	76%	43	80	54%	+22	0.004
P.R. has led to:								
7 Liking Reading Better	81	90	90%	57	78	73%	+17	0.008
8 Better at Reading	87	89	98%	70	80	88%	+10	0.021
9 Getting on Better with Each Other	69	86	80%	60	78	77%	+ 3	0.744
10 Will Go On with P.R.	61	90	68%	56	82	68%	0	0.924
11 Will Tell Others re P.R.	62	89	70%	56	82	68%	+ 2	0.976

+ Questionnaire format offered 2 options: negative/positive
++ X^2, d.f. = 1, probabilities two-tailed, Bold figures statistically significant

reading better as a result of Paired Reading ($p = 0.008$), feeling more proficient at reading ($p = 0.021$) and finding it easy to get books ($p = 0.044$).

Same-age peer tutorial pairs usually operated in their own classroom, resulting in fewer problems of finding a good place to do the reading, while in cross-age tutorial pairs movement was necessary for at least one member of the pair. Difficulty with obtaining books was reported by quite a high proportion of tutees, especially cross-age tutees, and the organisational aspects of this clearly require closer scrutiny when organising peer tutor projects. Likewise, the usage of record sheets in cross-age peer tutorial projects needs further scrutiny. Although same-age peer tutees were more likely than cross-age peer tutees to report improved attitudes to and skill in reading, again it should be noted that even the responses of the cross-age tutees were very positive in this respect.

The differences in results between same-age and cross-age peer tutored groups has implications for the results given in Table 56, which compared feedback from parent tutees with feedback from all peer tutees taken together. There is greater similarity between the feedback from parent tutees and the feedback from same-age peer tutees than between that from parent tutees and cross-age peer tutees. However, parent tutee feedback remains markedly more positive than same-age peer tutee feedback on 3 items (Liked Doing It, Getting On Better, Will Tell Others).

The subjective feedback from peer tutors according to whether tutoring occurred on a same-age or cross-age basis is detailed in Table 58. In the 10 same-age peer tutor projects using tutor questionnaires, not all participating tutors (99) returned questionnaires, and not all returned questionnaires (89) had responses to all items (questionnaire return rate = 90%). In the 3 cross-age projects using peer tutor questionnaires, not all participating tutors (86) returned questionnaires, and not all returned questionnaires (69) had responses to all items (questionnaire return rate = 80%). The mean percentage of positive responses per questionnaire from same-age peer tutors was 64%, and that from cross-age tutors was 62%. However, the differences in feedback between the two groups were very small and even the largest difference failed to reach statistical significance.

The action implications of the analysis of subjective feedback from the different groups of participants in Paired Reading projects are discussed in greater detail in Topping and Whiteley (1989).

TABLE 58 Subjective Feedback from Age-Peer Tutors and Cross-Age Peer Tutors

Concerning P.R., Tutee Reported to Be:- +	Age-Peer Tutors			Cross-Age Peer Tutors			Difference in Proportions Positive ++
	Positive Responses	Total Responses	Proportion Positive	Positive Responses	Total Responses	Proportion Positive	
1 Reading More	62	88	70%	52	68	76%	- 6
2 Reading More Widely	64	85	75%	50	66	76%	- 1
3 Understanding More	56	81	69%	44	67	66%	+ 3
4 More Confident	60	84	71%	47	65	72%	- 1
5 More Willing	54	85	64%	43	65	66%	- 2
6 More Interested	43	82	52%	39	64	61%	- 9
7 Enjoying More	58	84	69%	40	66	61%	+ 8
8 More Accurate	52	83	63%	45	67	67%	- 4
9 More Fluent	49	85	58%	38	65	58%	0
10 More Expressive	48	87	55%	28	66	42%	+13

+ Questionnaire format offered 2 options: negative/positive

++ None of the differences between groups reached statistical significance X^2, d.f. = 1, probabilities two-tailed. E.g. item 10, p = 0.158

RESULTS - INTER-RELATIONSHIPS BETWEEN OUTCOME MEASURES

It has already been noted that the reliability and validity of the 5 different kinds of feedback questionnaire were not separately established prior to their use in this study. Furthermore, the reliability and validity of the majority of the norm referenced reading tests used was in doubt on one count or another, considering each test independently even before datawere aggregated. Since it was debatable whether the norm referenced or the subjective data could be considered the more valid and reliable, attempts to establish the validity of one with reference to its relationship with the other could be construed as circular and essentially paradoxical. However, attempts were made to explore the inter-relationships between the different kinds of outcome measure, since the detection of consistency across two disparate kinds of measure might lead one to place greater confidence in both of them. Additionally, the relationship between different outcome measures and the frequency of reading and total time spent reading reported by participants was also analysed.

Effect Of Practice

"Time on task" is acknowledged as an important variable in the effectiveness of educational interventions. Children involved in Paired Reading projects almost certainly spent more time reading during the intensive period of projects than was previously the case. Parents often reported that increased oral reading in the context of a Paired Reading "session" was accompanied by an increase in spontaneous silent reading by the child, and teachers made similar comments. It is thus important to explore whether Paired Reading works merely by increasing

the amount of reading practice. If this is in fact the case, one might expect extra reading practice using any parental involvement technique to yield largely similar results.

Over 500 sets of home/school reading record sheets or diary cards for the intensive period of projects were analysed, and the mean number of reading sessions per week calculated, together with the total time recorded as spent reading during the intensive period of the project. Data on reading comprehension gains were available in relation to 200 sets of these diary sheets or cards. The relationship between pre-post reading gains and the frequency of reading and total time spent reading is expressed in Table 59 in terms of correlation coefficients. The correlation between frequency of reading and total time spent reading and pre-post gains in reading accuracy was extremely small (coefficient = 0.03, $p = 0.27$ or 0.28), far from achieving statistical significance. For reading comprehension, frequency of reading per week correlated positively with pre-post reading comprehension gains to a degree that achieved statistical significance (coefficient = 0.13, $p = 0.04$). The relationship between time spent reading and reading comprehension pre-post gains yielded a smaller coefficient (0.11) which failed to reach statistical significance ($p = 0.053$).

Even the one correlation coefficient achieving statistical significance was very small, and the factor involved can only account for a minute proportion of the underlying variance. Although self-recorded number of sessions and time spent are very imperfect indicators of amount of reading practice, it may be concluded that there is little evidence here that sheer volume of reading practice contributes significantly to the effectiveness of Paired Reading as measured by gains on norm referenced reading

TABLE 59

Relationship Between Reading Gains and Frequency of Reading and Total Time Spent Reading

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probability
	R E A D I N G A C C U R A C Y			
Frequency of Reading Per Week	-0.01 (503)	0.41	0.03 (503)	0.27
Total Time Spent Reading During Intensive Period of Project	0.02 (533)	0.29	0.03 (533)	0.28
	R E A D I N G C O M P R E H E N S I O N			
Frequency of Reading Per Week	0.09 (195)	0.10	0.13 (195)	0.04
Total Time Spent Reading During Intensive Period of Project	0.09 (200)	0.10	0.11 (200)	0.053

( ) Contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

tests. A similar correlation analysis with reference to short-term and long-term follow-up gains yielded no significant coefficients, numbers in cells being small (n = 9-41).

Subjective Feedback

A correlation analysis of the relationship between positivity of questionnaire feedback from the 3 main groups of participants (parents, teachers and parent tutored children) is summarised in Table 60. The overall positivity index of feedback questionnaires gives only a very general indication of overall subject response, but it was felt that detailed comparison between the 3 groups at the level of individual items would prove unmanageably complex, as well as difficult on account of the slightly different structure of the feedback questionnaires for the different participant groups.

From Table 60, it is clear that a moderately substantial positive correlation exists between parent positivity and child positivity in the 668 cases where this comparison could be made (0.39, p = 0.00). However, neither parent positivity nor child positivity correlated non-parametrically at a statistically significant level with teacher positivity, and if the parametric correlations can be accorded any weight on an aggregated variable in these smaller samples (274, 222 respectively), there was some suggestion of a tendency for parents and children to agree but for teachers to think the opposite (Parent cf Teacher coefficient = - 0.19, p = 0.00; Teacher cf Child coefficient = - 0.22, p = 0.00).

TABLE 60

Relationship Between Positivity of Questionnaire Feedback from Parents, Teachers and Children

Comparison	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probability
Parent Positivity Index cf. Child Positivity Index	0.42 (668)	0.00	0.39 (668)	0.00
Parent Positivity Index cf. Teacher Positivity Index	-0.19 (274)	0.00	0.03 (274)	0.30
Teacher Positivity Index cf. Child Positivity Index	-0.22 (222)	0.00	0.03 (222)	0.34

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

The relationships between the positivity of parent, teacher and child questionnaire feedback and pre-post gains on reading tests are detailed in Table 61. The positivity of parent feedback showed a small but statistically significant positive correlation with pre-post gains in reading accuracy (coefficient = 0.15, p = 0.00). The positivity of child feedback likewise showed a small but statistically significant correlation with pre-post gains in reading accuracy (coefficient = 0.14, p = 0.00). However, the inter-relationship between reading gains and positivity of teacher feedback was more equivocal, a statistically significant (p = 0.00) positive non-parametric correlation (0.21) being balanced by a parametric correlation close to zero which did not reach statistical significance.

In reading comprehension, a similar picture emerged, parent positivity correlating modestly but statistically significantly with reading gains (coefficient = 0.22, p = 0.00) and child feedback likewise (coefficient = 0.23, p = 0.00). The feedback from teachers was however again diverse with reference to reading comprehension, a non-parametric co-efficient of virtually zero being accompanied by a negative parametric correlation, although neither reached statistical significance.

As some variation in outcome results according to the type of reading test used was evident from observation, the relationship between reading test gains and reading frequency, total reading time and positivity of questionnaire feedback from parents, teachers and children was analysed further with reference to the reading test used (see Table 62). Inevitably, group sizes for some comparisons were very small, and results must be interpreted with extreme caution. Generally, comparisons involving group sizes of less than 50 have been disregarded.

TABLE 61

Relationship Between Reading Gains and Positivity of Parent, Teacher and Child Questionnaire Feedback

Factor	Parametric+ Correlation	Probability	Non-Parametric++ Correlation	Probability
	R E A D I N G A C C U R A C Y			
Positivity Index of Parent Feedback Questionnaire	0.11 (831)	0.00	0.15 (831)	0.00
Positivity Index of Teacher Feedback Questionnaire	-0.02 (332)	0.37	0.21	0.00
Positivity Index of Child Feedback Questionnaire	0.10 (608)	0.01	0.14 (608)	0.00
	R E A D I N G C O M P R E H E N S I O N			
Positivity Index of Parent Feedback Questionnaire	0.21 (263)	0.00	0.22 (263)	0.00
Positivity Index of Teacher Feedback Questionnaire	-0.11 (187)	0.07	0.01 (187)	0.42
Positivity Index of Child Feedback Questionnaire	0.23 (244)	0.00	0.23 (244)	0.00

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

TABLE 62

Relationship Between Reading Gain and Reading Frequency, Total Reading Time and Positivity of Questionnaire Feedback from Parents, Teachers and Children as a Function of Reading Test Used

Factor	Parametric + Correlation	Probability	Non-Parametric++ Correlation	Probability
R E A D I N G A C C U R A C Y				
DANIELS & DIACK TEST ONE (individual)				
Reading Frequency	0.21(111)	0.02	0.17(111)	0.03
Total Reading Time	0.13(112)	0.09	0.11(112)	0.12
Parent Feedback	0.15(210)	0.01	0.17(210)	0.01
Teacher Feedback	0.41 (57)	0.00	0.46 (57)	0.00
Child Feedback	0.17 (41)	0.15	0.23 (41)	0.07
HOLBORN READING TEST (individual)				
Parent Feedback	-0.33 (42)	0.02	-0.20 (42)	0.10
Teacher Feedback	0.51 (24)	0.01	0.66 (24)	0.00
Child Feedback	-0.32 (32)	0.04	-0.29 (32)	0.052
NEALE ANALYSIS OF READING ABILITY (individual)				
Reading Frequency	0.09(154)	0.13	0.10(154)	0.11
Total Reading Time	0.27(159)	0.00	0.31(159)	0.00
Parent Feedback	0.26(224)	0.00	0.28(224)	0.00
Teacher Feedback	0.25(121)	0.00	0.26(121)	0.00
Child Feedback	0.05(187)	0.26	-0.02(187)	0.38
NEW MACMILLAN READING ANALYSIS (individual)				
Reading Frequency	-0.59 (52)	0.00	-0.42 (52)	0.00
Total Reading Time	-0.43 (52)	0.00	-0.50 (52)	0.00
Parent Feedback	0.06 (71)	0.31	0.12 (71)	0.16
Teacher Feedback	-0.05 (68)	0.37	0.11 (68)	0.18
Child Feedback	0.12 (87)	0.14	0.14 (87)	0.10
PRIMARY READING TEST (group)				
Reading Frequency	-0.02 (83)	0.44	-0.03 (83)	0.39
Total Reading Time	-0.04 (83)	0.36	-0.06 (83)	0.30
Parent Feedback	-0.02(119)	0.40	0.02(119)	0.43
Teacher Feedback	0.38 (57)	0.00	0.35 (57)	0.00
Child Feedback	0.10(127)	0.13	0.10(127)	0.14
SCHONELL WORD READING TEST (individual)				
Reading Frequency	0.02 (56)	0.44	0.04 (56)	0.37
Total Reading Time	-0.10 (80)	0.19	-0.03 (80)	0.40
Parent Feedback	0.14 (69)	0.12	0.20 (69)	0.05
Child Feedback	0.06 (79)	0.29	0.14 (79)	0.11

Table 62 continued

Factor	Parametric + Correlation	Probability	Non-Parametric++ Correlation	Probability
	R E A D I N G C O M P R E H E N S I O N			
	NEALE ANALYSIS OF READING ABILITY (individual)			
Reading Frequency	0.14(143)	0.054	0.15(143)	0.04
Total Reading Time	0.16(148)	0.03	0.22(148)	0.00
Parent Feedback	0.13(188)	0.03	0.14(188)	0.03
Teacher Feedback	0.24(119)	0.01	0.25(119)	0.00
Child Feedback	0.05(153)	0.29	0.02(153)	0.38
	NEW MACMILLAN READING ANALYSIS (individual)			
Reading Frequency	-0.48 (52)	0.00	-0.32 (52)	0.01
Total Reading Time	-0.36 (52)	0.00	-0.36 (52)	0.00
Parent Feedback	0.09 (75)	0.22	0.10 (75)	0.20
Teacher Feedback	-0.04 (68)	0.38	0.01 (68)	0.48
Child Feedback	0.11 (91)	0.14	0.08 (91)	0.23

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

Note

Other tests used included the Burt Word Reading Test, Daniels and Diack Test Twelve, the Salford Reading Test, the Southgate Reading Test and the Widespan Reading Test. On many of these, data was only available for a small number of subjects (n) and no significant correlations emerged on the factors considered here.

An exception is the data given for the Holborn Reading Test, where negative non-parametric correlations (which did not however reach statistical significance) were found between reading gains and both parent and child feedback, whereas there was a large positive correlation between reading gains and teacher feedback. This latter finding may of course be because the teachers giving the feedback also gave the reading test, not necessarily in that order, but certainly the results for the Holborn Test are completely the opposite to those for all reading tests combined.

The Daniels and Diack Test 1 also showed a positive correlation with teacher feedback (coefficient = 0.46, p = 0.00), but correlated positively with parent feedback as well (coefficient = 0.17, p = 0.01) and showed a positive relationship with child feedback which did not however reach statistical significance. This test also correlated positively with reading frequency (coefficient = 0.17, p = 0.03), although not significantly with total reading time.

The inter-relationships for reading accuracy on the Neale Analysis of Reading Ability also showed a degree of coherence. Pre-post reading accuracy gains correlated positively with total reading time (coefficient = 0.31, p = 0.00), parent feedback positivity (coefficient = 0.28, p = 0.00) and also teacher feedback (coefficient = 0.26, p = 0.00). Much the same applied to the inter-relationship of these variables with pre-post scores in reading comprehension on the same test, despite reservations about the reliability and validity of this aspect of the test expressed elsewhere.

The Schonell Word Reading test/gains barely showed any statistically significant correlation with anyother/variable in this study, and the Primary Reading Test only showed a positive correlation with teacher feedback (coefficient = 0.35, p = 0.00). The New MacMillan Reading Analysis, like the Holborn Reading Test, was associated with unusual patterns of inter-relationship data. On the accuracy component of this test, substantial negative correlations were found between pre-post gains and reading frequency (coefficient = - 0.42, p = 0.00) and total reading time (coefficient = - 0.50, p = 0.00). A similar picture emerged on the reading comprehension component of the MacMillan Test (coefficient for reading frequency = - 0.32, p = 0.01), and for total reading time (coefficient = - 0.36, p = 0.00). No statistically significant correlations with the 3 kinds of subjective feedback were found for this test.

As some teachers were aware of test results before completing questionnaires, there may have been some contamination of teacher subjective perceptions, although it is not clear how this could have occurred differentially with reference to some tests but not with reference to others. The Daniels and Diack Test 1 and the Neale Analysis of Reading Ability appeared to yield pre-post gains which correlated best with other relevant outcome measures and process variables in Paired Reading projects.

Inter-relationship Between Reading Tests

The mean pre-post reading gain was computed for each of the 10 main reading tests used in the project, taken separately. Tests used on less than 30 subjects were not included in the analysis. Table 63 gives a ranking of the 10 reading tests in relation to the size of the mean pre-

TABLE 63

Ranking of Reading Tests in Relation to Size of Mean Pre-Post Gains in Reading

Rank	Reading Test	Mean Gain Pre-Post+	Standard Deviation [σ]	Group Size (n)
	R E A D I N G A C C U R A C Y			
1	Widespan*	9.18	18.32	71
2	Burt Word Recognition	8.26	4.02	31
3	Neale Analysis - Accuracy	7.68	6.74	638
4	Daniels & Diack Test 1	7.64	6.32	546
5	Holborn Reading Scale	7.04	7.22	128
6	Daniels & Diack Test 12* (GTRE)	6.85	9.63	134
7	Primary Reading Test*	6.38	10.47	352
8	Schonell Word Recognition	5.17	5.38	233
9	Salford	4.25	3.45	63
10	New MacMillan Analysis	4.17	6.34	118
	R E A D I N G C O M P R E H E N S I O N			
1	Neale Analysis - Comprehension	10.61	10.11	568
2	New MacMillan Analysis - Comprehension	3.11	7.67	122

\+ in months of Reading Age. Tests where $n < 30$ discarded.
* = group reading test.

post gain in reading they produced, with the largest mean pre-post gain at the top. From this it is evident that the Widespan group test has tended to produce very high mean gain scores but with very great variability, perhaps casting doubt on its reliability in this sort of application. The other two group reading tests in the list (Daniels and Diack Test 12 and the Primary Reading Test) also showed higher than average levels of variability. However, there is no evidence that individual reading tests as a type tended consistently to produce higher pre-post gains than group reading tests as a type, or vice-versa.

Even within the different kinds of individual reading test, little similarity in outcomes was found between those of similar construction, the Burt Word Recognition Test being ranked second while the Schonell Word Recognition Test was ranked eighth, and the Holborn Sentence Reading Test being ranked fifth while the Salford Sentence Reading Test was ranked ninth. Perhaps most striking is the discrepancy between the mean outcomes on the Neale and New MacMillan Tests, of very similar construction, with the Neale ranking third and the MacMillan tenth. For reading comprehension the disparity between the Neale and the MacMillan was even greater, the MacMillan Test producing very low scores.

As noted earlier, some of these tests are more likely to show practice effects in short-term repeated measures applications than others, but the Schonell Test which is known to suffer from practice effects actually ranked lower in Table 63 than did the Neale Test which is known to be less likely to suffer practice effects. Nor was there any obvious relationship between ranking by size of mean pre-post gain and whether or not a test was available and used in parallel forms. The

Widespan, Neale, Primary, Salford and New MacMillan Tests have parallel forms, but were very variously ranked at 1, 3, 7, 9 and 10. A similar analysis of baseline reading age gains ranked these tests in a very similar order, but numbers in many groups were very small.

The statistical significance of differences in mean pre-post gains between the different reading tests used is given in Table 64, and the significant differences for reading accuracy are summarised in Table 65. This latter indicates that the Burt Test produced gains significantly higher than 6 of 9 other reading tests. The Neale Test produced gains significantly less than the Burt but significantly greater than 5 other tests. The Daniels and Diack Test 1 produced gains significantly greater than 5 other tests. The Holborn Test produced gains significantly higher than 3 tests but significantly lower than another 3, and the Primary reading test followed a similar pattern. The Daniels and Diack Test 12 produced gains significantly greater than 3 tests, as did the Widespan Test (albeit erratically). The Schonell, Salford and New MacMillan Tests produced gains significantly less than 6, 7 and 8 other tests respectively.

The need to choose a test carefully which has proved to be reasonably stable in this particular application is underlined by the data in Table 66, which reviews the correlations between pre-post reading gains on the main reading test and on a secondary reading test used concurrently, in individual projects. As this analysis was at the level of individual

TABLE 64

Comparison of Mean Pre-Post Gains as a Function of Reading Test Used

COMPARISON +			Significance of Difference	
			Parametric Probability	Non-Parametric Probability
R E A D I N G A C C U R A C Y				
Daniels & Diack Test 1	vs	Holborne	0.196	0.032
Daniels & Diack Test 1	vs	New Macmillan	0.000	0.000
Daniels & Diack Test 1	vs	Primary Reading Test	0.022	0.003
Daniels & Diack Test 1	vs	Salford	0.000	0.000
Daniels & Diack Test 1	vs	Schonell	0.000	0.000
Daniels & Diack Test 12	vs	New Macmillan	0.005	0.002
Daniels & Diack Test 12	vs	Salford	0.003	0.002
Daniels & Diack Test 12	vs	Schonell	0.033	0.013
Holborn	vs	Neale	0.178	0.030
Holborn	vs	New Macmillan	0.001	0.002
Holborn	vs	Salford	0.000	0.000
Holborn	vs	Schonell	0.011	0.041
Holborn	vs	Burt	0.105	0.013
Neale	vs	New Macmillan	0.000	0.000
Neale	vs	Primary Reading Test	0.018	0.004
Neale	vs	Salford	0.000	0.001
Neale	vs	Schonell	0.000	0.000
Neale	vs	Burt	0.228	0.042
New Macmillan	vs	Primary Reading Test	0.004	0.016
New Macmillan	vs	Schonell	0.072	0.045
New Macmillan	vs	Widespan	0.015	0.003
New Macmillan	vs	Burt	0.000	0.000
Primary Reading Test	vs	Salford	0.002	0.003
Primary Reading Test	vs	Burt	0.021	0.047
Salford	vs	Widespan	0.015	0.021
Salford	vs	Burt	0.000	0.000
Schonell	vs	Widespan	0.037	0.008
Schonell	vs	Burt	0.000	0.000
R E A D I N G C O M P R E H E N S I O N				
Neale	vs	New MacMillan	0.000	0.000

+ Comparison of mean pre-post reading age gain on two types of reading test specified. Tests where $n < 30$ discarded. Differences significant parametrically only omitted. All other differences non-significant.

TABLE 65

Significant Differences in Mean Pre-Post Gains in Reading Accuracy between Reading Tests

Note The tests are ranked 1-10 in order of size of mean pre-post gain yielded, i.e. no. 1 gives the largest pre-post gains, no. 10 the smallest.

Reading Test	1 Wide Span	2 Burt	3 Neale	4 DD1	5 H'born	6 DD12	7 PRT	8 Schon-ell	9 Sal-ford	10 NMR
1 W/Span										
2 Burt										
3 Neale		X								
4 DD1										
5 Holborn		X	X	X						
6 DD12										
7 PRT		X	X	X						
8 Schonell	X	X	X	X	X	X				
9 Salford	X	X	X	X	X	X	X			
10 NMR	X	X	X	X	X	X	X	X		

TABLE 66

Relationship Between Reading Gains on Main and Secondary Test, By Project

Project	Main Test	Secondary Test	Parametric Correlation+	Prob.	Non-Parametric Correlation++	Prob.
			R E A D I N G A C C U R A C Y			
ALM01	Neale	D Diack 12	0.40 (7)	0.19	0.25 (7)	0.29
EAS01	D Diack 1	D Diack 12	0.41 (5)	0.25	0.34 (5)	0.28
FIE01	Neale	Primary RT	-0.05 (11)	0.45	0.19 (11)	0.29
GOM01	Schonell	D Diack 12	0.05 (18)	0.43	0.04 (18)	0.44
GOM02	Schonell	D Diack 12	-0.06 (20)	0.40	0.05 (20)	0.41
GOM03	Schonell	D Diack 12	0.07 (18)	0.39	-0.04 (18)	0.44
HEC02	Primary RT	Holborn	0.18 (24)	0.20	0.22 (24)	0.15
HOL02	D Diack 12	Burt	0.06 (37)	0.36	0.05 (37)	0.38
LEP01	Neale	Young Cloze	0.33 (10)	0.18	0.26 (10)	0.23
LEP02	Neale	Young Cloze	0.53 (14)	0.03	0.48 (14)	0.04
LON01	Neale	Salford	0.10 (8)	0.41	0.18 (8)	0.34
ROY01	Neale	Schonell	-0.30 (14)	0.15	-0.26 (14)	0.18
WEL05	Widespan	GAP	0.51 (13)	0.04	0.51 (13)	0.04
			R E A D I N G C O M P R E H E N S I O N			
ALM01	Neale	D Diack 12	0.63 (7)	0.06	0.58 (7)	0.09
LON01	Neale	Salford	0.69 (8)	0.03	0.50 (8)	0.10
ROY01	Neale	Schonell	-0.28 (14)	0.17	-0.33 (14)	0.12

( ) contains number of subjects in computation (n)
+ = Pearson Product-Moment
++ = Spearman Rank Order
Bold figures are statistically significant

projects, group sizes were inevitably small, and the finding of statistically significant relationships correspondingly unlikely.

However, in 13 projects only 2 positive correlations were found which reached statistical significance (both $p = 0.04$). Although in many projects the main reading test and the secondary reading test might be considered to be measuring very different aspects of the reading process, this very low level of positive correlation has disturbing implications for the reliability of individual results on some, or possibly all, reading tests in the current application. Over-reliance on results from a small number of subjects on one particular test, no matter how well structured the research design, is contra-indicated by these findings. The data in Table 66 could be interpreted as support for the meta-analytic approach synthesising many replications. Equally, there is some support here for the contention that the inclusion of many different reading tests in the current study proved to be in some respects a strength as well as a weakness.

DISCUSSION

Paradigmatic shifts in methodology have seemed more frequent in the social sciences than the natural sciences. Given the higher reflexivity of the former compared to the latter, this is unsurprising.

The classical format of tightly controlled experimental design, random allocation of (usually small numbers of) subjects to experimental/control conditions and evaluation of significance of differences by (usually parametric) statistical testing has become increasingly unfashionable in recent years as doubts have grown about content validity and generalisability of findings. A variety of qualitative or ethnomethodological modes of enquiry have been increasingly utilised, each possibly illuminating in its own right, but tending to generate studies which were difficult to summarise, meta-analyse or otherwise relate to each other.

Another strand of methodological development has been reflected in growing interest in replication and generalisability of findings, moving away from the notion of the single perfect experiment to methods of detecting general trends from many studies with the same focus, albeit incorporating sundry imperfections. Those favouring this meta-analytic approach appear to assume that it gives a "truer" view of the real world, i.e. a truer estimate of the probability of a given outcome for subsequent replications of the experimental intervention. Of course, it also highlights wide variance in outcomes between studies where this exists, a valuable feature in the context of academic debates in which individual studies are often cited highly selectively.

The current study does not fit neatly into any single methodological paradigm, as prior "experimental design" proved possible to only a small degree in this multi-faceted action research project in which data

gathering was essentially opportunistic. It is hoped that adequate descriptive and qualitative information has been given to enable further replication with a reasonable degree of equivalence. In terms of outcomes, this study most closely approximates to a meta-analytic methodological approach. It is felt that this will be of maximal interest to practising teachers, who tend to be more concerned with the probability of an innovation being effective if replicated in their own school than with the outcomes reported from a single study in a "laboratory" or other special setting. In other words, teachers are more interested in the replicative durability of innovation effectiveness in a highly imperfect world than with the scientific propriety of a single study under artificially controlled conditions.

This raises issues concerning the effectiveness of effectiveness research, the evaluation of evaluation. Should the objective of action research be the generation of reliable and valid conclusions about effectiveness (as in summative evaluation), or should the objective of action research be the generation of more action? (as in formative evaluation). These two are not mutually exclusive, but carrying out any research has resource implications. The question then becomes: assuming the objectives of a research effort can be clearly specified, which methodological paradigm is likely to be most cost-effective in meeting those objectives?

Insofar as the current study has generated a very large quantity of replicatory data of a type likely to be encouraging to practising teachers, the results have been and will be widely disseminated and the data have been gathered at very low cost as a tangential exercise to a service delivery project which itself showed high cost-effectiveness, it can be argued that the cost-effectiveness of the current study in terms

of generating further action is likely to be fairly high in comparison to other educational research.

Given that practitioners are motivated to act by many factors other than exposure to scientifically unimpeachable information, the action-generation objective may however prove partially independent from the conclusion-validation aspect in action research. The current study is riddled with methodological weaknesses which threaten to invalidate putative conclusions, and the inescapability of these weaknesses within the resources available to conduct the research is no defence. Extreme caution in the interpretation of the results is thus indicated, and detailed consideration of the results with reference to the conclusion - validation objective follows.

Specific Threats To Validity

Many specific threats to validity inherent in the way the research reported here was conducted have already been mentioned in the previous text, and will be summarised here.

1. PARTICIPANT SELF-SELECTION

Parents, children and teachers involved in projects operating in individual schools self-selected so to do. Furthermore, schools operating projects and choosing to avail themselves of the central support services offered likewise self-selected so to do. Thus all participants were arguably highly motivated, and the results reported here are not necessarily generalisable to other parents, children, teachers and schools. In one sense this is not problematic, since none of these could be "forced" to participate against their will. However, care must be taken not to assume a high probability of success where participants are reluctant and/or "persuaded". Equally, it should not be assumed that

similar success would be obtained in other Local Education Authorities where central support services were not as readily available and/or where the public relations aspect of the work was less well developed. It should be noted that the results cited here emanate mostly from projects operated by schools with relatively little experience of such exercises, but in circumstances where the novelty of the approach was presumably at a maximum. Although there is little evidence in this study that a school's projects subsequent to the first resulted in significantly different gains in reading accuracy on reading tests, to some degree the Hawthorne Effect may be postulated to have influenced the results as a whole.

2. COMPREHENSION

In the tables of results "Comprehension" has been taken to refer to scores on the separate comprehension scales of two reading tests. In fact, "accuracy" scores also reflect the deployment of some degree of skill in comprehension. This raises the question of the definition of "reading comprehension", itself an area of considerable debate. The "comprehension" results cited here proved very erratic, including those for baseline and control group comparisons. Furthermore, the comprehension data are drawn from a sub-sample skewed towards schools of high socio-economic status, and therefore less representative than the total sample. For all these reasons, little credence can be given to the comprehension results, and much more emphasis will be given to the results for "reading accuracy", even though these emanate from a variety of reading tests sampling different ranges of reading skills.

3. DATA POOLING

For purposes of meta-analysis in the current study, data from many school-based projects were pooled. Although these projects had many common features and a substantial degree of homogeneity in some respects, there was considerable non-homogeneity in other respects. Individual projects were different in terms of host school, target groups selected, reading tests used, testers, baseline and comparison groups, chronological age of participants, reading ability of participants, socio-economic status of catchment and presence or otherwise of home visits or other extended follow-up.

However, the impact of some of these subsidiary variables has been teased out in separate analyses. This is also true regarding the pooling of outcome data from projects deploying different kinds of tutors (parent, peer, etc), although here there is the additional complexity that peer tutoring was usually deployed with older children, and that at higher chronological ages given gains in reading age are less educationally significant.

Pre-post inter-test periods were very various, and some were very short, but pre-post gains were nevertheless pooled as they stood, rather than being expressed in ratio gains. Furthermore, the pooling of the data has served in some cases to mask great variability in results between different individual school-based projects. Consequently, for any analysis cited, where the n in <u>any</u> sub-sample is small, the comparison made is in grave doubt. This also applies even with moderately large n in both sub-samples for any comparison if the number of <u>projects</u> involved is small.

However, the aggregation of very various results may mask the existence of a minority of children for whom project participation was actually counter-productive.

4. GAIN SCORES

Gain scores are notoriously unreliable and tend to be correlated with the initial status of the subject, being associated with a large (combined) error of measurement which in turn is likely to increase any tendency towards regression to the mean in the data. Again, the different significance of a similar reading age gain from different basal reading ages is relevant. However, there is little evidence in the current study that regression to the mean was a significant factor in the reading accuracy data, although this was true to a somewhat greater extent in the comprehension data, which were inherently more erratic for a host of reasons.

5. "NORMAL" COMPARISON

Test results from Paired Reading projects have sometimes been cited in terms of multiples of "normal" rates of gain, on the assumption that a gain of one month of reading age in one chronological month is "normal". This assumption is extremely doubtful, however. The baseline and control group data cited in the current study demonstrate rates of gain during non-participation in Paired Reading which is greater than "normal" in the majority of cases. This is particularly true for reading "comprehension". Although there may be doubts concerning the relationship of the baseline and control group data in this study to the standardisation data originally gathered on the tests in question, an assumption that a ratio gain of 1.0 is "normal"

cannot be justified. A further complication here is that in some individual school-based projects the pre-post inter-test period was not contiguous with the intensive period of activity of the project, sometimes being considerably longer. Thus the pre-post period incorporated a period of presumably "normal" progress prior to the project and a period of accelerated progress during the intensive period of activity.

6. BASELINE AND COMPARISON GROUPS

The gathering of comparison group and baseline data was essentially opportunistic, and projects featuring these comparisons are by no means a random sample of total projects. Indeed, the data show that pre-post gains for participants in projects with baseline and comparison group data differ from the average pre-post gains for all projects combined - i.e. baseline and comparison group projects are not typical. Baseline data were gathered at times by a different tester than were pre-post data, and some reverse attrition in baseline projects may have produced effects which were compounded by the non-elimination of subjects lacking full data sets.

Comparison (quasi-control) groups were extremely various in nature and allocation to experimental and control groups was never purely random, although this was approached in one or two projects. The majority of comparison group subjects were self-selected non-participants. However, there is little evidence that their pre-test attainment in reading differed substantially from participants on average. A possibility of contamination between experimental and comparison subjects has already been

acknowledged, and this may partially account for the "above normal" gains of control subjects. Whether contaminated by spin-off from a school-based Paired Reading project or not, it is of course perfectly possible that non-participant control children were in any event still doing some reading at home with their parents, as many children spontaneously do (especially in the lower reaches of the primary school).

7. RATIO GAINS

The use of ratio gains to compare gains in test scores over different chronological periods is of extremely doubtful validity. The reading development of individual children is not linear, and direct comparison of a one month gain in reading age made by children of different chronological ages from different basal reading ages is not valid. The range and variance of reading test scores expressed in reading age increases with chronological age, another reason for non-equivalence of given reading age gains - a gain of 3 months at a reading age of 7 represents a very different standard deviate than a similar gain at a reading age of 10 years. "Retardation" with reference to different basal reading ages cannot be validly compared for exactly the same reason. Thus rate scores inherently tend to compare like with unlike. However, given the very various baseline and follow-up periods inherent in the data collected in this study, it is extremely difficult to discern any more satisfactory alternative approach. Attempts to separate out baseline and follow-up data into groups of similar period length result in the sample size in each group being so low as to itself vitiate any possibility of drawing conclusions.

8. PROCESS

The current study is severely lacking in process data, and there is little to indicate whether participants actually did Paired Reading during the intensive period of a project. It is not known whether Paired Reading was done for all, part or any of the project period, nor what relative emphasis was given to Reading Alone and Reading Together. Even in peer tutor projects, amenable to closer monitoring, no detailed process data are available. The background literature is not encouraging in this respect, between 43% and 75% of participants doing Paired Reading "properly" subsequent to training in large field studies. If participants are not doing Paired Reading, it is not clear what is being measured. The subjective feedback from participants may be commenting on what they have actually done, rather than on the effectiveness of Paired Reading, although some of them may have been convinced that what they were doing was Paired Reading even when it bore little relationship to what they were told to do. Likewise, the norm referenced data reflect the impact of what the Pair actually did, not necessarily Paired Reading.

The current study was based on a crude input/output model, and the outcome data cited may say as much if not more about the structure of service delivery in these projects than about the effectiveness of the Paired Reading technique itself. It is certainly desirable to conduct detailed process research, but this is extemely time-consuming and was not possible within the resources available in this study. The identification of crucial process aspects leading to both positive and the occasional negative outcome is needed to determine what aspects of a Paired Reading project are crucial to

success. Nevertheless, the whole may prove to be more than the sum of its parts, and attempts to separate out critical variables might not be successful.

9. FOLLOW UP

In the current study, although a substantial addition has been made to the follow-up data on Paired Reading, inadequacies are still evident. More follow-up data are needed, over longer follow-up periods. Long-term follow-up testers were not always the same as pre-post testers. Inevitably, follow-up samples showed some attrition, although not to an extraordinary degree. There is evidence that follow-up projects were not necessarily typical of all projects, and may have tended to be particularly successful projects. Peer tutor follow-up remains extremely sparse. An issue is the question: "How long a follow-up period is long enough?" The reasonableness of expecting the impact of a relatively brief and lightweight intervention to remain visible to unsophisticated measuring instruments at very long-term follow-up is open to debate. Professionals may differ on the follow-up objectives that should be set for parental involvement projects - is the maintenance of relative gain compared to non-participant children satisfactory, or should the maintenance of relative acceleration be sought? Whether one opts for the more ambitious or the less ambitious of these objectives will doubtless be influenced by reference to the long-term effectiveness of other educational interventions.

10. TESTS

The reading tests used in this study were numerous, very various in type, and even more various in terms of the adequacy of background information on norms, reliability and validity. Some were very dated in content and some incorporated very elderly and possibly dated normative information. The tests covered different age ranges, exacerbating the aforementioned problem of the valid comparison of reading age gains at different ages. It is highly likely that even the same test was administered in a slightly different way by different testers in different projects. In the case of one or two tests reference was made to different tables of norms.

Information about reliability and validity, where present in the test manual, was of doubtful relevance to the usage of the test in the current application, as is indicated by the data comparing results on different tests in the current study. In some cases the structure of the test tended to inflate post-test scores (e.g. Neale Comprehension) and in other cases the structure of the test probably resulted in substantial practice effects. Some supposedly parallel forms proved in practice to be of doubtful comparability, and the adequacy of standardisation sometimes differed for parallel forms of the same test. Some tests had no parallel forms and were thus arguably more susceptible to practice effects in general. The low correlation between individual test scores where two tests were used in parallel suggest low inter-test validity in the current application. Given this, the use of a variety of tests in the current study may not represent a major

methodologicalweakness, however. Campbell (1969) comments that the imperfect validity of measures often leads the researcher to "the use of multiple measures of independent imperfection" (page 415).

The reading skills sampled by the reading tests were extremely various, and no test could pretend to give a complete window on the subtleties of an individual child's deployment of multiple reading strategies. Some of the tests involved silent reading rather than oral reading, and (paradoxically) aside from the Neale and Macmillan, group tests were more likely than individual tests to incorporate a degree of comprehension. However, assessment of higher order reading skills was not a significant feature of any test. In practice, test choice was severely constrained by test availability, the age of the children participating in the project and the time available to the school to carry out testing. The gathering of baseline and/or comparison group and/or follow-up data introduced further restrictions. Schools would not claim to have done other than made the best of a bad job. All reading test usage in the current study can be criticised on the ground of the very short pre-post inter-test period, this being usage for which none of the tests were specifically designed even when incorporating parallel forms, and which was advised against by the authors of more than one.

11. SUBJECTIVE FEEDBACK

Subjective feedback of all kinds is subject to what Campbell (1969) calls the "grateful testimonials" effect. Campbell proposes that the feedback will be more favourable as: (a) the

clearer is the evaluative implication of the response measure, (b) the more directly the responder is identified by name, (c) the more the responder gives the answer directly to the intervention agent, (d) the more that agent will continue to be influential in the responder's life in the future, (e) the more the responses deal with feelings rather than observables and (f) the more the responders are a small, self-selected or agent-selected subset of all service recipients.

In the current study, parent feedback focused on child behaviour rather than parental feeling, child feedback focused on organisational factors and their own behaviour and teacher feedback on child behaviour. Although strong attitudinal indicators were included by implication, feelings were rarely solicited directly. Questionnaires had a space for the name of the responder, but no-one was directly asked to complete this, although many chose to do so spontaneously. Responders certainly knew that feedback would be returned to teachers in school who continued to be influential in the life of their children. Response rates were on the whole adequate, although any response rate less that 100% can be assumed to introduce some positive bias into the results. It can be argued that paper and pencil feedback is not reliable, yet there is little evidence to suggest that oral face to face feedback is any more reliable.

The difference in the structures of the questionnaires of the various participant groups, together with the differences in response rates from these groups, renders their feedback not

directly comparable, although attempts have been made to do this so far as possible. In the child questionnaire, with only two response choices per item, the subjects may have felt forced into "yea-saying". Completion of child questionnaires could have been influenced by where the questionnaire was completed and under whose supervision. Some teachers may have been aware of reading test results before completing the teacher questionnaire, although in some cases the teacher completing the questionnaire was not the teacher co-ordinating the project or carrying out the testing.

Perhaps of greatest concern is that the reliability of these feedback instruments had not been established in independent studies prior to their usage in the current study, and any validity beyond simple face validity was likewise unestablished, although attempts to relate feedback questionnaire outcomes to other outcomes were made subsequently. The validity of the Positivity Index is naturally in even greater doubt.

12. STATISTICAL ANALYSIS

The descriptive, correlational and comparative statistical analyses applied in the current study can be criticised on the grounds of excess simplicity and crudity. It was felt that the application of complex procedures of statistical analysis could create confusion, by dignifying data which were inherently unsatisfactory and potentially masking their inadequacy. In any event, the application of techniques such as ANOVA and ANCOVA would have been of doubtful value as many of the assumptions for their valid use were violated by the current data. Complex

statistical analysis has the further disadvantage of restricted comprehensibility, and it was desirable for the current research programme to be open to scrutiny by lay practitioners so far as possible.

13. COMPOUNDING OF VARIABLES

Given the lack of process data in the current study it is extremely difficult to disentangle the critical variables in the total service delivery package. What proportion of effects are due to the Paired Reading technique itself, what proportion to the format and organisation of support service delivery and what proportion to the effect of group "ethos" and motivation in this particular local education authority with its attendant publicity cannot be determined. Nor can any estimate be made of the proportion of effectiveness attributable to Hawthorne Effect, although certainly the in-service training and public relations aspects of the exercise were in a sense deliberately calculated to produce such an effect. Given the inadequacies of the data, it is not even possible to be definitive on the issue of whether Paired Reading works merely by increasing time on task at reading, although the evidence from the current study and elsewhere gives no support for this proposition.

14. MISCELLANEOUS

Other threats to the validity of conclusions can be identified. The current study deals with the impact of home visiting in a very crude way, by identifying mean number of home visits per person and assigning that statistic to the whole project. Exact numbers of visits per individual subject are not correlated with

individual outcome measures. Additionally there is no process data concerning what actually occurred on home visits, which may have been extremely various in nature. No adequate measures of time on task at reading were available, data from self-recording in home-school diaries being collected from a decidedly non-random sample and being of very doubtful accuracy, particularly concerning total time spent.

The index of socio-economic status can only be regarded as approximate, as much of the census data were considerably out of date and it was extremely difficult to determine the socio-economic status of schools with large catchment areas such as denominational and high schools and those schools in dense urban areas. The Organisational Rating coding was not carried out "blind", and must be considered highly subjective and unreliable. Conclusions regarding differences between white and Asian participants must be extremely tentative, since not only were the reading tests culturally inappropriate for many Asian children, it was not possible to partial out the effect of socio-economic status between the Asian and white groups. There might also be some bias in the outcome data for Asian participants resulting from the disproportionately low take-up rate in the Asian population. Finally, it is likely that at some point in the manipulation of this large volume of data invalidity owing to computational or clerical error has entered.

General Threats To Validity

Useful lists of "Factors Jeopardising Internal and External Validity" are provided by Campbell and Stanley (1963) and Campbell (1969). Nine threats to internal validity are listed in Campbell (1969):-

1. History - other causative events between pre- and post test
2. Maturation - within-subject spontaneous change, e.g. growth, fatigue.
3. Instability - unreliability of measures (both single and repeated) and sampling fluctuations.
4. Testing - practice effects and awareness/expectation effects.
5. Instrumentation - changes in measures or measurers.
6. Regression - pseudo-gains of extreme scorers at pre-test.
7. Selection - differential recruitment of comparison groups.
8. Attrition - differential loss of subjects from groups.
9. Selection-Maturation Interaction-selection bias yielding differential rates of maturation.

Campbell (1969) also lists six threats to external validity, i.e. threats to valid generalisation of the results to other settings, other versions of the treatment and other measures of the effect:-

1. Interaction Effects of Testing - sensitising effects of pre-test on subjects inhibits generalisation of treatment effects to non-tested subjects.
2. Interaction of Selection and Treatment - unrepresentative responsiveness of the treated population.
3. Reactive Effect of Experimentation - "artificiality" of experimental setting inhibits generalisation to more "regular" settings.
4. Multiple Treatment Interference - composite effect of multiple treatment greater than the sum of its parts.

5. Irrelevant Responsiveness of Measures - some components of complex measures irrelevant yet may produce apparent effects.
6. Irrelevant Replicability - replications of complex treatments may fail to include the causative components.

Campbell (1969) notes that "statistical tests are relevant to at best one of fifteen threats to validity" (page 410) - the one being Instability in sampling and measures. However, Campbell defends the use of statistical tests in non-randomised studies, in order to exclude findings which are probabilistically trivial. For Campbell, this is the point at which self-criticism commences, via detailed consideration of "rival explanations" which could be used to "explain away" an effect. However, "the mere possibility of some alternative explanation is not enough - it is only plausible rival hypotheses that are invalidating" (page 411). Similarly, Linn (1986) comments: "by emphasising falsification and the search for and testing of plausible rival hypotheses much can be learned from non-randomised studies" (page 96, emphasis not original).

It will be salutory to consider the data from the current study with respect to these factors. With such a large data pool, it seems implausible that other causative events consistently occurring between pre- and post-test could account for the changes evident. Regarding the maturation factor, subjects in the current study were expected to show spontaneous change in growth but participant progress was expected to be greater than this - problems of interpretation only arise if the spontaneous growth is not steady and linear. Indeed, there is evidence from these data of seasonal variations in impact of Paired Reading participation. Linear development of reading skill cannot be assumed.

However, in such a large data pool non-linear maturation should to an extent be equalised in the aggregation, except for seasonal fluctuations which are widespread and consistent. Additionally, the baseline and comparison group aspect of the current study enables a partialling out of the maturation factor.

The unreliability of single and repeated measures coupled with sampling fluctuation are a very real threat to validity in the current study. However, it has been argued that the incorporation of a multiplicity of imperfect measures is not necessarily a major weakness, while gain scores have not been shown to demonstrate the usual adverse side effects and sampling fluctuations are to some extent compensated for by the aggregation of many disparate samples. Insofar as testing is concerned, practice effects and awareness/expectation effects will certainly have applied in some projects, to the extent of some teachers informing participants at post-test of the purpose of the test. However, it seems implausible to attribute the whole of the gains demonstrated to practice, awareness/expectation and Hawthorne effects.

So far as instrumentation is concerned, changes in measures were very few and changes in measurers only affected some baseline and follow-up data. Pseudo-gains owing to regression were a small factor in comprehension data but barely a significant factor in reading accuracy data. Differential recruitment of comparison groups was a problem with some comparison groups, but there was no evidence that on average comparison groups were more or less able than participant groups. Attrition was a problem with some baseline and follow-up data, but the proportion of subjects involved was within normal limits for an exercise of its kind. There was no real evidence of the selection bias yielding differential rates of maturation.

Turning to threats to external validity, the "sensitising effect" of pre-test on subjects was certainly commented upon by some teachers, some of whom continued to test participants in subsequent projects long after the effectiveness of the project had been demonstrated in their school. Some parents referred to the "ritual" of a Paired Reading project, and reported greater child enthusiasm when part of an "official" school-based group project. However, it is implausible to attribute more than a small proportion of the total variance in outcome to the impact of this particular phenomenon. Some schools never tested participants yet were subjectively delighted with the observed response of participants.

In the current study, the "treated population" were by definition unrepresentatively responsive, since they had volunteered to participate and were highly motivated to do so. However, as there is no intention to attempt to generalise this approach to families who are not prepared to volunteer, this threat to external validity is virtually irrelevant. Concerning "artificiality" of experimental setting, the current study is arguably of much greater generalisable validity than single studies in individual schools which were previously reported in the literature, although it may be "artificial" by virtue of providing central support services in a co-ordinated way and maintaining a high public relations profile.

Concerning Campbell's fourth threat to external validity, it may indeed be true that Paired Reading can be construed as a "multiple treatment", the composite effect of which is greater than the sum of its parts. Indeed, attempts to isolate or partial out the crucial, causative factors in this multiple treatment may be vitiated by this very

feature. However, until further research is carried out, this issue lies in the realm of speculation. Irrelevant responsiveness of measures seems unlikely to have been a significant factor in the current study. The issue of irrelevant replicability is however an important one - the intervention evaluated by the current study was indeed complex, including not just a particular technique but many facets of a service delivery package. Failure to replicate both aspects may well result in failure to replicate the positive results.

Bearing these many threats to validity in mind, cautious conclusions will now be made.

CHAPTER 14

SUMMARY AND CONCLUSIONS

Conclusions deriving from the current study will be listed and enumerated. Subsequently, brief general comparison of the outcomes of the current study with outcomes reported in the previous literature will be made. Finally, areas needing further research will be delineated.

Conclusions

1. From 1984 to 1987, 83 schools of all types operated 185 projects. Norm referenced data were generated by 155 projects in 71 schools involving 2,372 children, 54% of which were known to be boys and 42% girls. "Comprehension" data were available for 690 children. Data from separate projects were pooled.
2. The average period of intensive activity in projects was 8.62 weeks, the average baseline period in 23 baselined projects was 19.23 weeks with great variability, and the average follow-up period in 17 follow-up projects was 25.85 weeks with great variability. Comparison ("control") groups were a feature of 37 projects.
3. The majority of participants (74.8%) were parent tutored, 7.8% were same-age peer tutored and 6.3% cross-age peer tutored. Other tutors were adult volunteers, teacher volunteers or unrecorded.
4. Home visits were incorporated in 26.5% of projects, equally divided between less than 1 per child per project, between 1 and 2 per child and greater than 2.
5. Home/school reading diary cards from approximately 600 non-random families were analysed, showing the mean frequency of reading per

week to be 4.97 and the total reading time during project intensive period to be 7 hours 13 minutes, the former statistic probably being more accurate.

6. Of 261 statistical analyses on the data, 56% achieved non-parametric statistical significance, which is not to assume educational significance.

7. Results were analysed in terms of gain in reading age, but there was little sign of regression to the mean operating in reading accuracy data, although this may have occurred to a limited extent with the more erratic comprehension data and to a degree in some of the follow-up data.

8. The mean pre-post gain in reading accuracy was 6.97 months of reading age, and in comprehension 9.23 months. Expressed as ratio gains, i.e. as a multiple of chronological time passed, RG = 3.27 for accuracy and 4.39 for comprehension. As comprehension scores are more erratic and comprehension not readily differentiable from accuracy, this difference is not meaningful.

9. In 23 baselined projects incorporating 374 participants, of whom 288 were baselined children, for accuracy baseline ratio gain was 1.37 while pre-post ratio gain was 2.56, a highly statistically significant difference. In comprehension, both baseline and pre-post gains were high and the difference not statistically significant.

10. In 37 comparison group studies incorporating 580 participant and 446 comparison children, the scores in both accuracy and comprehension for participant children were statistically significantly greater than those for comparison children.

11. At short-term follow-up (equal to or less than 17 weeks) 102 children in seven projects averaged ratio gains over the follow-up period of 2.01 for accuracy and 2.32 for comprehension.

12. At long-term follow-up (greater than 17 weeks) 170 children in ten projects averaged ratio gains over the follow-up period of 1.20 in accuracy and 1.36 in comprehension.

13. Follow-up ratio gain did not correlate to a statistically significant degree with the length of follow-up period (coefficient = -0.36, $p = 0.20$). Ratio gains cannot validly be interpreted as multiples of "normal" progress. Whether the follow-up result is "good" or not is thus debatable, but it compares favourably to the follow-up data on other brief educational interventions.

14. Where follow-up data are available in comparison group projects, follow-up gains in reading accuracy show no statistically significant difference between participant and comparison groups. Thus, one may expect participants to remain relatively advantaged compared to non-participants, since relative pre-post gains are sustained in the long run although relative acceleration is not.

15. Follow-up ratio gains show no significant correlation with presence or otherwise of home visits, socio-economic status of catchment area or degree of retardation of participants, but there is a significant correlation with pre-post gains, i.e. children doing best during the intensive period also do best at follow-up.

16. The follow-up data for peer tutoring are much less encouraging than those for parent tutoring, but as yet they are sparse and equivocal.

17. Overall Effect Sizes for reading accuracy were +0.87 and for comprehension + 0.77. These compare favourably with related reports in the literature based on meta-analysis of individual published studies and theses.

18. Type and size of school showed little relationship with size of gains. Older children tended to show slightly higher gains in reading accuracy, but this is not educationally significant given the structure of reading tests.

19. The Spring Term was the most popular for the operation of projects but the Autumn Term yielded statistically significantly better results. The least popular Summer Term yielded acceptable results in reading accuracy.

20. Projects rated as of better organisational quality yielded higher pre-post gains in reading accuracy, but the objectivity of the rating was in doubt. There was little evidence that a school's increasing experience in operating projects (associated with reducing novelty) had any significant influence on gains in reading accuracy.

21. Children from all social classes in the LEA were involved in projects, 60% of participants being of below average socio-economic status. There was a tendency for participants of lower socio-economic status to make larger gains in reading accuracy, even if not home visited, and vice versa. However, home visiting made an additional significant positive difference for participants in the lowest quartile of socio-economic status. There are implications here for the cost-effectiveness of differential inclusion of home visiting support in this kind of service delivery. Comprehension results were disproportionately derived from higher SES groups.

22. A tendency was evident for more retarded readers to make larger gains in accuracy and comprehension, although this was small and in any event retardation (and reduction therein) has different implications at different basal reading ages.

23. At pre-test, on average girls were ahead of boys in accuracy and boys ahead of girls in comprehension. Boys made bigger gains than girls in accuracy and comprehension, but not statistically significantly. In comparison group projects, male participants made gains statistically significantly larger than those of controls, but females did not.

24. Participants of South Asian origin were recorded in 50 projects yielding norm-referenced data, operated in 30 schools. The average number of such participants in these projects was 5 (range 1 - 47, maximum proportion 74%). Asian participants constituted 9.4% of total participants, compared to 15.4% in the total schoolchild population. At pre-test, Asians were behind whites on average in accuracy and comprehension. Asians made gains greater than white participants in accuracy (but not significantly) and significantly smaller gains in comprehension. However, there are doubts about the validity of direct comparison of gain size from different basal reading ages, the non-partialling of socio-economic status for Asian participants and the cultural relevance of the reading tests. Asian participant pre-post gains were greater than the gains of non-participant children of any ethnic origin.

25. Pre-post gains in reading accuracy were similar for parent tutored, same-age peer tutored and cross-age peer tutored participants. Adult volunteers and teacher volunteers tutored

very small numbers of participants and teacher volunteers tended to tutor much weaker children. However, peer tutoring tended to operate with older children, so this finding may underestimate the true impact of parent tutoring.

26. Pre-post gains of peer tutors were greater than those of peer tutees in reading accuracy, but the difference was not statistically significant.

27. In peer tutor projects, same sex pairings were much more common than mixed-sex pairings. Boy same-sex pairings yielded significantly higher gains for tutors than did girl same-sex pairings, although no difference was evident in outcomes for tutees. There was a tendency for mixed-sex combinations to be good for tutors but poor for tutees.

28. Parent feedback questionnaires were returned by 1,068 parents (response rate 73%). Parents reported that 78% of their children were more confident in reading, 71% more accurate and 68% showing better comprehension. Seventy one per cent of parents wished to continue with Paired Reading.

29. Teacher feedback (n = 430, RR = 91%) was somewhat more equivocal and less positive. Teachers reported 70% of children to be more confident, 67% more accurate and 59% showing better comprehension in class. Better concentration and motivation in class was evident in 37%. Parental responses were significantly more positive than teacher responses on several items but this could be partially attributed to different questionnaire structure and response rate.

30. Feedback from 692 parent-tutored children (RR = 72%) was very positive but their questionnaire had fewer choice options. As a

result of Paired Reading, 95% of children felt they were better at reading and 92% liked reading more, while 70% wished to carry on doing P.R.

31. Subjective feedback from 158 peer tutors (RR = 85%) was very similar to parent tutor feedback although not quite so positive, being significantly different on only 2 items. Sixty per cent wished to continue doing Paired Reading but with a different partner.

32. Feedback from 173 peer tutees (RR = 93%) showed opinions a little more divided than other respondent groups: however, 93% felt they were better at reading, 82% liked it better and 68% wished to carry on doing Paired Reading.

33. All five forms of subjective feedback yielded some indicators of a more positive attitude to reading, including greater enthusiasm and a much wider choice and greater volume of reading material.

34. Parent tutee feedback was more positive than peer tutee feedback, especially cross-age peer tutee feedback. Feedback from same-age peer tutees was more positive than from cross-age peer tutees. However, same-age and cross-age peer tutor feedback was very similar.

35. The positivity of parent feedback questionnaires showed a significant positive correlation with the positivity of child feedback questionnaires, but neither correlated significantly with teacher feedback positivity.

36. Parent feedback positivity and child feedback positivity correlated positively and significantly with pre-post gains in accuracy and comprehension, but the picture was more equivocal for teacher feedback.

37. Self-recorded frequency of reading and time spent reading showed no significant relationship to reading accuracy gains. Although the validity of informal self-recording as a measure of time on task is in doubt, there is no evidence here that Paired Reading works by increasing time spent on reading.

38. Of the reading tests used, the Neale and Daniels and Diack Test 1 showed coherent positive correlations with subjective feedback from parents, teachers and children and to a degree with self-reported frequency of reading. Other tests showed few positive correlations and the Macmillan and Holborn Tests showed very low coherence with other data.

39. As different reading tests are relevant to different chronological age ranges, direct comparison of gains on them is a dubious exercise. However, there was a tendency for group reading tests to show greater variability than individual reading tests, but not consistently higher or lower pre-post gains. Tests of similar type and construction produced very different results in some cases. Gains on particular tests showed no consistent relationship with the likelihood of practice effects from the structure of the test, the availability or otherwise of parallel forms or the reading age range relevance of the instrument. The Burt, Neale and Daniels and Diack Test 1 yielded stable results which tended to be average or above. The Holborn, Daniels and Diack Test 12 and Primary Reading Test yielded stable results which were average or below. The Neale and the Macmillan tests produced very different results in both accuracy and comprehension despite their similar structure. Very erratic results were

evident on the Widespan test and the MacMillan test, the former tending to average out very high and the latter very low. The Schonell and Salford tests also tended to produce low gain scores.

40. Where two reading tests were utilised on a pre-post basis concurrently, correlations between them tended to be very low. Even allowing for the different nature of the two tests, this is somewhat disturbing and over-reliance on results on a single test on a small number of subjects would seem contra-indicated. This finding gives support for the meta-analytic approach.

41. Overall, then, the study provides a number of pointers to the probable success of the Paired Reading approach. However, the present study, extensive as it is, has not been in a position to offer conclusive evidence, for which we must await the findings of properly controlled studies.

Previous Research

Brief comparison of outcomes from the Kirklees data and from previous research can now be made. The Kirklees data incorporate norm referenced results from 2,372 subjects in 155 projects, while all the previous literature taken together gave data on 934 subjects in 55 projects. The Kirklees research has thus greatly added to the volume of outcome data. As is evident from Synopsis 4, mean ratio gains cited in the literature tended to be somewhat higher than those found in Kirklees (4.22 cf 3.27 for accuracy, 5.39 cf 4.39 for comprehension). This is to be expected given the positive bias in published results stemming from submission and publication policies. It is nevertheless most encouraging that in the very large field study in Kirklees incorporating a large proportion of the schools in the LEA, average results were nevertheless so little behind those of published studies. The Paired Reading technique combined with organised and supported service delivery seems associated with a substantial degree of generalisability, replicability and durability.

Synopsis 4 Comparison of Mean Ratio Gains, All Projects

Data Source	Number of Projects	Accuracy		Comprehension	
		n	Mean Ratio Gain	n	Mean Ratio Gain
Literature	55	934	4.22	624	5.39
Kirklees	155	2372	3.27	690	4.39

Comparison of mean ratio gains in control/comparison group projects in the literature and in Kirklees is given in Synopsis 5. In the literature, there are results from 18 control group projects incorporating almost 300 subjects. In Kirklees, data are available from twice this number of "control" group projects incorporating almost twice as many subjects. Again, the results from the literature are somewhat more positive than those from Kirklees, Literature experimentals tending to do better than Kirklees experimentals and Literature controls tending to do less well than Kirklees controls. This difference is less marked for comprehension data.

The tendency for Literature outcomes to be better than Kirklees outcomes is more pronounced when considering Effect Sizes (see Synopsis 6). In reading accuracy, the Literature Mean Effect Size is more than twice that for Kirklees, and this undoubtedly reflects greater variability within Kirklees projects as well as smaller absolute gains. Again, this difference is less marked for reading comprehension.

The current study has added substantially to the research literature with respect to follow-up data. In the previous literature, follow-up gains were cited for only 5 projects and these were very various. Some follow-up ratio gains were less than 1 but greater than control groups. Full data were not always cited. By contrast, the Kirklees research gives follow-up data on 17 projects, indicating mean short-term follow-up ratio gains in reading accuracy and comprehension of greater than 2, and in the longer term of 1.20 in accuracy and 1.36 in comprehension. The Kirklees follow-up data is thus considerably more substantial and encouraging.

Synopsis 5 Comparison of Mean Ratio Gains in Control Group Projects

Data Source		Number of Projects	Accuracy		Comprehension	
			n	Mean Ratio Gain	n	Mean Ratio Gain
Literature	E	18	298	3.84	195	4.57
	C		278	1.56	181	2.29
Kirklees	E	37	580	3.35	170	4.55
	C		446	1.99	159	2.51

E = Participant Experimental Group C = Comparison (Control) Group

Synopsis 6 Comparison of Distribution of Effect Sizes in Control Group Projects

	No of Projects	Mean Effect Size	Standard Deviation	Highest ES	Lowest ES
		READING ACCURACY			
Literature	12	2.12	2.26	+ 10.00	+ 0.10
Kirklees	34	0.87	1.04	+ 5.82	- 0.13
		READING COMPREHENSION			
Literature	9	1.63	1.33	+ 4.56	+ 0.42
Kirklees	12	0.77	0.72	+ 2.94	+ 0.04

Further Research

The citation of crude pre-post gains as evidence of effectiveness is clearly unsatisfactory, particularly with relation to reading comprehension. The use of control or comparison groups and/or baseline measures are essential if sound conclusions are to be drawn. Control groups need not necessarily be randomised, as the artificiality of this can vitiate the effectiveness of service delivery. However, great care is needed to avoid contamination between control and experimental groups.

Further research should explore the cost-effectiveness of various approaches to the training and follow-up aspects of service delivery. It is important to begin to partial out what aspects of the total package are crucial in terms of effectiveness. For instance, the exploration of the use of a home/school reading diary alone would be a useful first step. However, this task may not be easy and the whole may prove to be more than the sum of its parts.

The collection of more follow-up data is clearly desirable. This is particularly true with respect to peer tutoring. Such follow-up research should relate to longer follow-up periods and concentrate on reading style and other outcomes in addition to norm referenced testing. Such follow-up also needs conducting with control or comparison groups, examining the sustaining of relative gain in comparison to the sustaining of relative acceleration. Differences between schools and projects in capability to sustain follow-up gains require close examination. The effect of involvement in Paired Reading for longer periods or involvement in consecutive projects, perhaps with different tutor types, needs scrutiny.

Although extremely time-consuming and expensive, further detailed research into the process of Paired Reading in the field setting is essential. Clarification as to what proportion of parents actually do Paired Reading properly is required, and whether this makes a significant difference to outcomes. It may be that other aspects of service delivery are in fact more influential on outcomes than the technique itself, but as yet this is a speculation. It is of course possible that even if parents do not do Paired Reading properly, it nevertheless has an effect on inhibiting previous unhelpful behaviour, and thus nevertheless results in improved outcomes. Further study of process on a follow-up basis is also necessary - if parents start doing Paired Reading properly, how soon if at all do they modify it, with or without further training and in what circumstances? When parents express the desire in their feedback to continue with Paired Reading, do they actually go on doing it and if so for how long? The study of process X outcome interactions will be essential to determining what aspects of the Paired Reading service delivery package are crucial to effectiveness. Audio and/or video recording are likely to be important vehicles for this kind of research.

Subsequent research could usefully deploy alternative reading tests, not necessarily yielding reading ages. One example would be the Edinburgh Reading Test, which samples a wide range of reading skills, although it is time-consuming and expensive for a group test. Perhaps more importantly, further attention should be paid to less tangible outcome variables, particularly of an attitudinal nature, such as enthusiasm, confidence, etc. The effect of participant belief in the likelihood of improving might merit further study in the context of the

almost evangelical nature of Paired Reading service delivery in some areas. These aspects are of course extremely difficult to measure with a satisfactory degree of reliability and validity, and efforts in this area have often been characterised by results failing to reach statistical significance. A more observable indicator might be the number and variety of books read spontaneously by children over a longer period, as recorded in some form of extended reading diary. The impact of Paired Reading on vocabulary, both orally and in creative writing, would be worth further exploration.

More detailed comparison of the similarities and differences between parent and peer tutoring using Paired Reading would be valuable. The differences in follow-up results to date and the need for further follow-up research on peer tutoring has already been mentioned. Again, this should focus on reading style and attitudinal variables as well as crude norm referenced reading outcomes. Attention could also be paid to increased vocabulary and to personal and social outcomes. The latter are often of considerable interest to teachers who co-ordinate peer tutoring programmes, who tend to see such exercises as a practical vehicle for personal and social education.

Attempts to control systematically for the impact of other ancillary factors in the Paired Reading service delivery package will be important. Factors which could be important in promoting the success of Paired Reading projects include the development of a group ethos, child free choice of reading material, greater access to books, and so forth. All of these merit study as important variables in their own right. However, such is the nature of home based Paired Reading

projects that they are by definition extremely difficult to monitor and it may be that some of these questions could be more economically explored in the peer tutoring format which allows closer monitoring. Other issues requiring investigation are whether Paired Reading is particularly suited to children of a certain chronological age, degree of reading ability and with or without specific learning difficulties. Further studies comparing the impact of different non-professional tutoring techniques with groups of children of different kinds may result in clarification about which techniques are most effective for which children with which difficulties at which ages.

Further research should explore the use of alternative units and formats of analysis. Reference has already been made to the undesirability of using gain scores as primary units of analysis. In a controlled study involving pairs matched on the pre-test measure, members of which are then randomly allocated to experimental or control groups, the unit of analysis could be differences between experimentals and controls in post-test score, whether expressed as raw score, reading age or standardised deviation score (deviation "reading quotient"), the latter having the additional advantage of automatic relativity to normal expectations of progress through passage of time. Outcomes could readily be expressed with reference to statistical significance, effect size or both and the results of many small studies using the same research design meta-analysed, utilising blocking on variables indicated by the Kirklees study as potentially influential. However, variable take-up rates in parent tutored projects could adversely affect this design.

Another alternative approach avoiding the use of gain scores might be analysis of covariance on post-test scores (of whatever type) with pre-test scores as a co-variate, but care that the conditions for the valid application of ANCOVA were fulfilled would be necessary. If some kind of gain score is to be used, blocking by test used and comparing shifts in raw score may be more valid, given that towards the ceiling of some reading tests a small increase in raw score can yield a large increase in reading age. However, the differential scalar properties of raw score shifts from different basal scores would remain problematic. Direct comparison of post-test scores to normative expectations is another possibility. For each subject, considering the deviation quotient associated with the pre-test raw score, a normative prediction of post-test raw score which is associated with the identical deviation quotient for the child's age at post-test at the end of the project period could be made. The unit of analysis would be the difference between predicted post-test raw score and actual post-test raw score. This type of analysis could be applied irrespective of the existence or otherwise of control group data. A roughly similar but less satisfactory procedure could be applied with tests yielding only reading ages. However, the validity of assuming such predictions based on the norming population are applicable to the experimental population must remain in doubt.

Avoiding the use of ratio gain scores presents further problems. This could most obviously be achieved by a meta-analysis blocked by pre-post test interval (this latter not necessarily being coterminous with project length in the activity duration sense). Although analysis by comparison of post-test raw scores to normative expectations superficially avoids this problem, such comparisons would remain influenced by the scalar properties of the normative data, and a blocked meta-analysis may still be expected to show considerable variation according to project length, as indeed is suggested by the Kirklees data.

Furthermore, interesting as these alternatives may be, direct comparison to the previous literature (the vast majority of which reports only reading age data) would prove difficult. Previous studies reporting control group data and/or statistical significance could of course be subjected to comparative meta-analysis, but such information is unavailable for many studies.

Alternative forms of statistical analysis should also be considered. In the Kirklees study, many comparisons employed the non-parametric Wilcoxon matched-pairs signed-ranks test. If it is considered that the use of gain scores seriously impairs comparison in terms of magnitude of differences, a statistical test utilising information simply about the direction of differences could be applied, namely the sign test. While this is less statistically powerful as a procedure, it makes no assumptions whatsoever about the form of distribution of differences, except that it is continuous (Siegel, 1956). The sign test could be applied readily to comparisons of post-test raw scores with normative expections as outlined above.

Another option would be the use of multiple regression techniques, although as previously reported the Kirklees data breached some of the conditions for the valid application of this procedure, and there must be concern that the application of linear or planar methods of analysis to aggregated data might mask fundamental discontinuities of practical importance. In the Kirklees study, many of the organisational variables studied in relation to outcomes clearly accounted for a very small portion of the total variance, albeit at times a portion of some potential practical significance. Any multiple regression analysis on post-test scores entering pre-test scores and any other organisational factor as variables would certainly result in the pre-test scores accounting for an overwhelmingly large portion of the total variance. Equally, entering two or more organisational factors alone would result in very large uninterpretable residuals.

Likewise, the application of two-way analysis of variance, while seeming of potential relevance in an occasions x treatment format, would require that samples be drawn at random and the variances of the sample populations be equal. Again, many of the comments regarding partitioning of variance made above would apply. Nevertheless, Tizard et al. (1982) used ANOVA on participant and control group post-test scores in the seminal Haringey project. Considerable dissatisfaction with the general linear model in statistical analysis has been expressed in the literature in recent years, and this has led workers such as Goldstein (1987) to propose wider use of newer techniques such as multi-level modelling. These offer a degree of flexibility facilitative of fitting the analysis to the data, rather than vice versa, and may prove useful for the analysis of new data or reanalysis of the Kirklees data.

Finally, research effort is necessary in the area of dissemination and replication. A most important step in further replicatory research is the recruitment of subjects, i.e. how do you get people to do it? A study of the comparative effectiveness of various means of dissemination and recruitment would itself be valuable.

Summary

The impact of services to help schools to guide and support parents (and peer tutors) in the use of the Paired Reading technique for improving children's reading was evaluated. Compared to all studies previously reported in the literature taken together, the Kirklees research yielded more than double the volume of pre-post norm referenced outcome data, double the amount of control or comparison group data and triple the amount of follow-up data. Additionally, in Kirklees baseline data were compared with pre-post data to give a time series comparison. Although outcomes on reading tests were slightly less favourable than those selectively reported in the previous literature, the research

nevertheless suggested that project effects were discernible even in a large field study incorporating many schools in one Local Education Authority, representing a significant test of the generalisability and replicability of the technique. The Kirklees study also examined the influence of a number of organisational, demographic and within-subject factors as they related to outcomes. In addition a very large volume of subjective feedback from teacher, parent and child participants was collected in a systematic way, and proved mostly positive. The research also examined the inter-relationship of the various outcome measures deployed with a view to assessing their relative reliability and validity for this purpose. As very few process data were gathered it was not possible to demonstrate what proportion of participants actually utilised the Paired Reading technique in the way they were trained. It is thus difficult to partial out to what extent the positive outcomes are due to the impact of the technique and/or the service delivery support package. However, the technique and service delivery package combined are suggested by the data to have contributed towards improving children's reading skill and attitude to reading. Given the low cost of the total programme, this is encouraging.

It has been argued that community interventions should be:- (1) simple, (2) inexpensive, (3) effective, (4) compatible with the existing values and needs structures of the population, (5) flexible, (6) decentralised and (7) sustainable. The Kirklees research appears to suggest that the LEA's Paired Reading service delivery package meets many of these criteria. Further research is however needed into the last of these, concerning whether programme effects endure for individual participants in the much longer term and whether schools persist in offering project

involvement to potential tutors even when impact has been demonstrated, in the face of other professional distractions.

REFERENCES

Allington, R.L. (1980)
Teacher Interruption Behaviours During Primary-Grade Oral Reading
Journal of Educational Psychology 72, 3, 371-7

Arora, C.M.J. and Sheppard, J. (1982)
P.R.I.N.T. Project in Kirklees (Paired Reading Instruction by Non-Teachers)
(unpublished document, Kirklees Psychological Service, Huddersfield)

Baker, C. (1980)
Reading Through Play: the easy way to teach your child
London Macdonald Educational

Barrett, J. (1986)
Reading with Mother: a Paired Reading programme described and evaluated.
Reading (U.K.R.A.) 20, 3, 173-8

Barrett, J. M. (1987)
Paired Reading: psycholinguistics in practice
Reading (U.K.R.A.) 21, 3, 152-8

Blatchford, P., Burke, P., Farquar, C., Plewis, I. & Tizard, B. (1985)
Educational Achievement in the Infant School: the influence of ethnic origin, gender and home on entry skills.
Educational Research 27, 1, 52-60

Bloom, W. (1987)
Partnership with Parents in Reading
London Hodder and Stoughton

Boehnlein, M. M. & Hager, B. H. (1985)
Children, Parents and Reading: an annotated bibliography
Newark, Delaware International Reading Association

Bookbinder, G. (1976)
Salford Sentence Reading Test
London Hodder and Stoughton

Branston, P. & Provis, M. (1986)
Children and Parents Enjoying Reading
London Hodder and Stoughton

Brimer, A. (1972)
Widespan Reading Test
London Nelson

Budgell, P. (1985)
The Allocation of Resources Using Indeces of Socio-Economic Advantage and Disadvantage
paper presented at Annual Conference of the Association of Educational Psychologists, University of York

Burdett, L. (1985)
Parental Involvement in Reading: a comparative study of Paired Reading and Individualised Reading with poor readers.
Unpublished M. Ed. thesis, University of Hong Kong

Burdett, L. (1986 a)
A Study of P. R. and Individualised Reading in Hong Kong.
Paired Reading Bulletin 2, 53-60.

Burdett, L. (1986 b)
Two Effective Approaches for Helping Poor Readers
British Journal of Special Education 13, 4, 151-4

Burt, C. & Vernon, P. E. (1938)
Burt (Rearranged) Word Reading Test
London Hodder and Stoughton

Bush, A.M. (1982)
A "Paired Reading" Experiment Involving Parents in a Multicultural Educational Priority Area.
Unpublished Advanced Diploma Dissertion, Leeds Polytechnic

Bush, A.M. (1983)
Can Pupils' Reading be Improved by Involving their Parents?
Remedial Education 18, 4, 167-70

Bush, A.M. (1985)
Paired Reading at Deighton Junior School
in: Topping and Wolfendale (op cit)

Bushell, R., Miller, A. & Robson, D. (1982)
Parents as Remedial Teachers
Journal of the Association of Educational Psychologists, 5, 9, 7-13

Butkowsky, I. S. & Willows, D. M. (1980)
Cognitive - Motivational Characteristics of Children Varying in Reading Ability: evidence of learned helplessness in poor readers.
Journal of Educational Psychology 72, 3, 408-22

Byron, D. (1987)
Parent/School Liaison: the Hartcliffe Comprehensive Paired Reading Project
Bristol Educational Psychology Service, Hartcliffe

Byron, D. & Brock, R. (1984)
Working with Parents: the Golden Valley Paired Reading Project
Bristol Educational Psychology Service, Hartcliffe

Campbell, D. T. (1969)
Reforms as Experiments
American Psychologist 24, 409-29

Campbell, D.T. & Stanley, J.C. (1963)
Experimental and Quasi-experimental Designs for Research in Teaching
in: N. L. Gage (ed.)
Handbook of Research on Teaching
Chicago Rand McNally

Carbo, M. (1978)
Teaching Reading with Talking Books
The Reading Teacher 32, 267-73

Carrick-Smith, L. (1982)
An Investigation into the Effectiveness of Paired Reading
Unpublished M.Sc. (Educational Psychology) thesis, University of Sheffield.

Carrick-Smith, L. (1985)
A Research Project in Paired Reading
in: Topping and Wolfendale (op cit)

Carver, R.P. (1978)
The Case Against Statistical Significance Testing
Harvard Educational Review 48, 3, 378-99

Cawood, S. & Lee, A. (1985)
Paired Reading at Colne Valley High School
Paired Reading Bulletin 1, 46-50

Cohen, J. (1977)
Statistical Power Analysis for the Behavioral Sciences
New York Academic Press

Cohen, P. A., Kulik, J.A. & Kulik, C-L.C. (1982)
Educational Outcomes of Tutoring: a meta-analysis of findings.
American Educational Research Journal 19, 2, 237-48

Cooknell, T. (1985)
An Inner-city Home-reading Project
in: Topping and Wolfendale (op cit)

Crombie, R. and Low, A. (1986)
Using a Paired Reading Technique in Cross Age Peer Tutoring
Paired Reading Bulletin 2, 10-15

Curr, W. & Gourlay, N. (1960)
The Effect of Practice on Performance in Scholastic Tests
British Journal of Educational Psychology 30, 156-67

Curtis, M.E. (1980)
Development of Components of Reading Skill
Journal of Educational Psychology 72, 5, 656-69

Cutting, B. (1982)
Reading Matters (helping your child with reading)
Leeds Arnold-Wheaton

Daniels, J.C. & Diack, H. (1979)
The Standard Reading Tests (fourteenth impression)
St Albans Hart-Davis Educational

Dening, T. (1985)
A Comparison of Three Parent Tutoring Methods and the Type of Post-Training Support Necessary to Improve Children's Reading Skills.
Unpublished M. App. Psych. thesis, Murdoch University, Perth, W. Australia.

Department of Education and Science (1982)
A Classification of Local Education Authorities by Additional Educational Needs (Cluster Analysis).
Statistical Bulletin 8/82 (ISSN 0142-5013)
London D.E.S.

Dowhower, S.L. (1989)
Repeated Reading: research into practice
The Reading Teacher 43, 502-7

Eldredge, J.L. & Butterfield, D. (1986)
Alternatives to Traditional Reading Instruction
The Reading Teacher 40, 32-7

Elliott, J. (1989)
Home Reading Styles: variations in technique and project organisation
Paired Learning 5, 4-13

ERIC (1984)
Resources for Involving Parents in Literacy Development: Digest
Urbana, Illinois ERIC Clearinghouse on Reading & Communication Skills

ERIC (1985)
Parent Involvement in Reading: Resources
Urbana, Illinois ERIC Clearinghouse on Reading & Communication Skills

Evans, A. (1984)
Paired Reading: a report on two projects
Unpublished Paper, Division and Institute of Education, University of Sheffield

France, N. (1981)
The Primary Reading Test (revised edition)
Windsor NFER-Nelson

Fredericks, A.D. & Taylor, D. (1985)
Parent Programs in Reading: guidelines for success
Newark, Delaware International Reading Association

Friedlander, J. (1981)
Parent Involvement
chapter in: Early Reading Development: a bibliography
London Harper and Row

Fry, L. (1973)

Token Reinforcement and the Reading Ability of Retarded Readers

New Zealand Journal of Educational Studies 1, 165-76

Gautrey, F. (1988)

Paired Reading in Reaside Middle School

Paired Reading Bulletin, 4, 12-15

(also in Reading [UKRA] 22, 3, 175-9)

Glass, G. V. & Smith, M. L. (1979)
Meta-analysis of the research on class size and achievement
Educational Evaluation and Policy Analysis 1, 2-16

Glass, G.V., McGaw, B. & Smith, M.L. (1981)

Meta-Analysis in Social Research

Beverley Hills, Ca. Sage

Glass, G. V., Cahen, L. S., Smith, M. L. & Filby, N. N. (1982)
School Class Size: research and policy
Beverly Hills and London: Sage

Goldstein, H. (1987)
Multi-level Models in Social and Educational Research
London: Griffin Press

Gollop, S. (1984)

A Paired Learning Project in a Junior School

Focus on Language 6, 10-11

(Essex Reading & Language Centre)

Greaney, V. (1986)

Parental Influences on Reading

The Reading Teacher 39, 8, 813-8

Green, F. (1987)

Analysis of Reading Tutorial Results with an Emphasis on Consideration

of the Effectiveness of the Paired Reading Procedures

unpublished paper, Lyndon State College, Lyndonville, Vermont, USA

Grigg, S. (1984)
Parental Involvement with Reading: an experimental comparison of Paired Reading and Listening to Children Read.
Unpublished M.Sc. (Educational Psychology) dissertation, University of Newcastle-upon-Tyne

Grundy, P. (1987)
Parent, Teenage Volunteer and Age Peer Tutors in the Primary School
Paired Reading Bulletin 3, 45-49

Hancock, R. (1988)
Parental Involvement in Children's Reading: results of a survey of Brent primary school headteachers.
Reading (U.K.R.A.) 22, 3, 168-74

Hannon, P. (1987)
A Study of the Effects of Parental Involvement in the Teaching of Reading on Children's Reading Test Performance.
British Journal of Educational Psychology 57, 1, 56-72

Hannon, P. W. & Cuckle, P. (1984)
Involving Parents in the Teaching of Reading: a study of current school practice.
Educational Research 26, 1, 7-13

Hannon, P. & Jackson, A. (1987)
The Belfield Reading Project: Final Report
London National Children's Bureau

Hannon, P. & Jackson, A. (1987)
Parents' and Teachers' Strategies in Hearing Young Children Read
Research Papers in Education 1, 1, 6-25

Hannon, P., Long, R., Weinberger, J. & Whitehurst, L. (1985)
Involving Parents in the Teaching of Reading: some key sources
Sheffield Division of Education, University of Sheffield

Harding, L.M., Beech, J.R. & Sneddon, W. (1985)
The Changing Pattern of Reading Errors and Reading Style from 5 to 11 Years of Age
British Journal of Educational Psychology 55, 45-52

Hargreaves, W.A. & Attkisson, C.C. (1978)
Evaluating Program Outcomes
in: C.C. Attkisson et al. (eds.)
Evaluation of Human Service Programs
New York Academic Press

Haskett, G.J. & Lenfestey, W. (1974)
Reading-Related Behavior in an Open Classroom: effects of novelty and modelling on preschoolers
Journal of Applied Behavior Analysis 7, 2, 233-41

Hays, W.L. (1963)

Statistics for Psychologists

New York Holt, Rinehart and Winston

Heald, C. & Eustice, V. (1988)

Ready to Read

Leamington Spa Scholastic

Heath, A. (1981)

A Paired Reading Programme

Edition 2, I.L.E.A. School Psychological Service, 22-32

Heath, A. (1985)

A Study of the Effectiveness of Paired Reading

in: Topping and Wolfendale (op cit)

Heckelman, R.G. (1966)

Using the Neurological Impress Technique

Academic Therapy, 1, 235-9

Heckelman, R. G. (1986)

N.I.M. Revisited

Academic Therapy 21, 411-20

Hedges, L.V. (1982)

Estimation of Effect Size from a Series of Independent Experiments

Psychological Bulletin 92, 490-9

Hewison, J. (1987)
The Haringey Reading Project: long term effects of parental involvement in children's reading
Paired Reading Bulletin 3, 50-53

Hewison, J. & Tizard J. (1980)
Parental Involvement and Reading Attainment
British Journal of Educational Psychology 50, 209-15

Holdaway, D. (1979)
The Foundations of Literacy
Sydney Ashton Scholastic

Holdsworth, P. (1986)
Direct Instruction, Precision Teaching and Shared Reading
Paired Reading Bulletin 2, 32-5

Holdsworth, P. (1985)
Parental Involvement at Mowbray School
in: Topping & Wolfendale (op cit)

Horowitz, R. & Samuels, S.J. (1985)
Reading and Listening to Expository Text
Journal of Reading Behavior 17, 3, 185-98

Hoskisson, K. (1975)
The Many Facets of Assisted Reading
Elementary English 52, 312-5

Hoskisson, K., Sherman, T.M. & Smith, L.L. (1974)
Assisted Reading and Parent Involvement
The Reading Teacher 26, 710-4

Hutson, B.A., Cowger, D.E. & Wallbrown, F.H. (1980)
Assessing Psycholinguistic Orientation and Decoding Strategies for Remedial and Non-remedial Readers
Journal of School Psychology 18, 3, 263-9

Jackson, A. & Hannon, P. (1981)
The Belfield Reading Project
Rochdale Belfield Community Council

Jeffs, A. (1984)
Children and Parents and Reading
Gravesend, Kent Home and School Council

Jenkins, J.R. & Larson, K. (1979)
Evaluating Error-Correction Procedures for Oral Reading
The Journal of Special Education 13, 2, 145-56

Johnson, S. (1982)
Listening and Reading: the recall of 7-9 year olds
British Journal of Educational Psychology 52, 24-32

Jones, J.M. (1987)
Hawarden Infants Parent Involvement Project
Paired Reading Bulletin 3, 54-61

Joscelyne, T. (1989)
Changes in Reading Style Following Peer Tutored Paired Reading and Listening to Reading
Paired Learning 5, 88-93

Jungnitz, G. (1984)
A Descriptive and Comparative Study of a Parental Involvement in Reading Project with Asian Families.
Unpublished M.A. Thesis, University of York

Jungnitz, G. (1985)
A Paired Reading Project with Asian Families
in: Topping and Wolfendale (op cit)

Jungnitz, G., Olive, S. & Topping, K. J. (1983)
The Development & Evaluation of a Paired Reading Project
Journal of Community Education 2, 4, 14-22

Kidd, A.E. & Levey, M. (1983)
Parental Involvement in Reading in a Comprehensive School
Unpublished paper, Birmingham Psychological Service

Koven, J.T. & LeBow, M.D. (1973)
Teaching Parents to Remediate the Academic Problems of their Children
Journal of Experimental Education 41, 4, 64-73

Lane, C. (1987)
Aiming ARROW at Learning Targets
British Journal of Special Education 14, 3, 99-101

Lane, C.H. & Chinn, S.J. (1986)
Learning by Self-Voice Echo
Academic Therapy 21, 4, 477-82

Lawrence, D. (1972)
Counselling of Retarded Readers by Non-Professionals
Educational Research 15, 48-51

Lawrence, D. (1973)
Improved Reading Through Counselling
London Ward Lock

Lawrence, D. (1975)
Improved Reading Through Counselling
in: A.F. Laing (ed.)
Trends in the Education of Children with Special Learning Needs
Swansea Faculty of Education University College of Swansea

Lawrence, D. (1985)
Improving Self-Esteem and Reading
Educational Research 27, 3, 194-200

Lee, A. (1986)
A Study of the Longer Term Effects of Paired Reading
Paired Reading Bulletin 2, 36-43

Lees, E. (1985)
An Account of a P.R. Project
Paired Reading Bulletin 1, 68-9

Lees, E.A. (1986)
A Cognitive Developmental Analysis of Reading Skills in Good and Poor Readers.
Paper presented at Annual Conference of Developmental Psychology Section, British Psychological Society, September 1986. (Submitted to the Journal of Reading Research).

Lees, E. (1987)
An Analysis of the Skills Involved in Reading Development Resulting from Paired Reading.
Paired Reading Bulletin 3, 110-112

Limbrick, E., McNaughton, S. & Glynn, T. (1985)
Reading Gains for Underachieving Tutors and Tutees in a Cross-Age Tutoring Programme
Journal of Child Psychology & Psychiatry 26, 6, 939-53

Lindsay, G., Evans, A. & Jones, B. (1985)
Paired Reading and Relaxed Reading: a comparison
British Journal of Educational Psychology 55, 3, 304-9

Linn, R.L. (1986)
Quantitative Methods in Research on Teaching
in: Wittrock, M.C. (ed.)
Handbook of Research on Teaching (3rd edition)
New York Macmillan

Long, R. (1986)
Developing Parental Involvement in Primary Schools
London Macmillan Education

Loveday, E. (1988)
Making Reading Child's Play!
Special Children 23 (September), 18-19

Loveday, E. & Simmons, K. (1988)
Reading at Home: does it matter what parents do?
Reading (U.K.R.A.) 22, 2, 84-8

Low, A. & Davies, M. (1988)
Infant/Infant Cross-Age Peer Group Tuition
Paired Reading Bulletin, 4, 7-10

Low, A., Madden, L. & Davies, M. (1987)
Infant/Junior Cross-Age Peer Group Tuition
Paired Reading Bulletin 3, 91-5

McKerracher, D. W. (1967)
Alleviation of Reading Difficulties by a Simple Operant Conditioning Procedure
Journal of Child Psychology and Psychiatry 8, 51-6

McMillan, G., Young, J., Johnson, J. & Noble, J. (1988)
Paired Reading in Edinburgh
in: Innovating Practice, volume III
Scottish Education Department/Regional Psychological Services

McNaughton, S.S., Glynn, T. & Robinson, V.R. (1980)
Parents as Remedial Reading Tutors: issues for home and school
Wellington, N.Z. New Zealand Council for Educational Research

Meyer, V. (1982)
Prime-O-Tec: a successful strategy for adult disabled readers
Journal of Reading 25, 512-5

Miller, A. (1986)
The Role of Paired Reading in Home-School Collaboration Schemes
Curriculum 7, 1, 33-8

Miller, A. (1987)
Is There Still a Place for Paired Reading?
Educational Psychology in Practice, 3, 1, 38-43

Miller, A., Robson, D. & Bushell, R. (1986)
Parental Participation in Paired Reading: a controlled study.
Educational Psychology 6, 3, 277-84

Moffatt, J. & Wagner, B.J. (1976)
Student-centred Language Arts and Reading, K-13
New York Houghton Mifflin

Moon, C. & Wells, G. (1979)
The Influence of Home on Learning to Read
Journal of Research in Reading 2, 1, 52-62

Moorhouse, A.J. & Yule, W. (1974)
A Comparison of the Neale and the Daniels and Diack Reading Tests
Reading 8, 3, 24-7

Morgan, R.T.T. (1976)
"Paired Reading" tuition: a preliminary report on a technique for cases of reading deficit
Child Care Health and Development 2, 13-28

Morgan, R.T.T. (1985)
Paired Reading: origins and future
in: Topping and Wolfendale (op cit)

Morgan, R. (1986)
Helping Children Read: the Paired Reading handbook
London Methuen

Morgan, R. & Gavin, P. (1988)
Paired Reading: evaluation & progress
Support for Learning 3, 4, 201-6

Morgan, R.T.T., & Lyon, E. (1979)
"Paired Reading" - a preliminary report on a technique for parental tuition of reading-retarded children
Journal of Child Psychology and Psychiatry 20, 151-60

Morris, A. (1984)
A Paired Reading Approach - Parental Involvement at a Comprehensive School
Unpublished paper. Division and Institute of Education, University of Sheffield.

Mudd, N. (1987)
Strategies Used in the Early Stages of Learning to Read: a comparison of children and adults.
Educational Research 29, 2, 83-94

Murfitt, J. & Thomas, J.B. (1983)
The Effects of Peer Counselling on the Self-Concept and Reading Attainment of Secondary Aged Slow Learning Pupils.
Remedial Education 18, 2, 73-4

Neale, M.D. (1966)
Neale Analysis of Reading Ability (second edition)
London Macmillan Education

Netley, C., Rachman, S. & Turner, R.K. (1965)
The Effect of Practice on Performance in a Reading Attainment Test
British Journal of Educational Psychology 35, 1-8

Neville, M.H. (1968)
Effects of Oral and Echoic Responses in Beginning Reading
Journal of Educational Psychology 59, 362-9

Neville, M.H. (1975)

Effectiveness of Rate of Aural Message on Reading and Listening

Educational Research 18, 1, 37-43

Newson, J. & Newson, E. (1977)

Perspectives on School at Seven Years Old

London Allen and Unwin

Niedermeyer, F.C. (1970)

Parents Teach Kindergarten Reading at Home

Elementary School Journal 70, 438-45

O'Hara, M. (1985)

Paired Reading in a School for the Physically Handicapped

Paired Reading Bulletin 1, 16-19

Pillemer, D. B. & Light, R. J. (1980)
Synthesizing Outcomes: how to use research evidence from many studies.
Harvard Educational Review 50, 2, 176-95

Pitchford, M. & Taylor, P. (1983)

Paired Reading, Previewing & Independent Reading

Unpublished paper, Leeds Psychological Service

Potter, F. (1982)

The Use of the Linguistic Context: do good and poor readers use different strategies?

British Journal of Educational Psychology 52, 16-23

Pumfrey, P.D. (1986)

Paired Reading: promise and pitfalls

Educational Research 28, 2, 89-94

Pumfrey, P.D. (1987)
A Critique of Paired Reading
Paired Reading Bulletin 3, 62-6

Quisenberry, N.L., Blakemore, C. & Warren, C.A. (1977)
Involving Parents in Reading: an annotated bibliography
The Reading Teacher 30, 34-9

Rachman, S. (1962)
Learning Theory and Child Psychology - Therapeutic Possibilities
Journal of Child Psychology and Psychiatry 3, 149-63

Reid, D. (ed.) (1987)
Working Together: parents, teachers, and children
Reading, Berks. Reading & Language Information Centre, University of Reading

Richards, I.A. (1942)
How to Read a Page
New York Norton

Richardson, J. and Brown, J. (1978)
A Study of Three Methods of Helping Children with Reading Difficulties
Reading (U.K.R.A.) 12, 2, 10-21

Richardson, J.R. (1986)
Paired Reading in Suffolk
Paired Reading Bulletin 2, 94

Richek, M.A. & McTague, B.K. (1988)
The "Curious George" Strategy for Students with Reading Problems
The Reading Teacher 42, 3, 220-26

Riecken, H.W. & Boruch, R.F. (1974)
Social Experimentation: a method for planning and evaluating social intervention
New York Academic Press

Ripon, C., Winn, B. and Ingleby, S. (1986)
Can Parents Teach their Own Children with Reading Problems?
Paired Reading Bulletin 2, 76-81

Robinson, A. (1985)
Reading: parents, children and teachers
London National Children's Bureau

Robinson, F., Prichard, D. & Rennie, J. (1984)
Reading: involving parents
Coventry Community Education Development Centre

Robson, D., Miller, A. & Bushell, R. (1984)
The Development of Paired Reading in High Peak & West Derbyshire
Remedial Education 19, 4, 177-83

Root, B. (1988)
Help Your Child Learn to Read
London Usborne

Rosenberg, M.S. (1986)
Error-correction During Oral Reading: a comparison of three techniques
Learning Disability Quarterly 9, 3, 182-92

Rosenthal, R. (1984)
Meta-Analytic Procedures for Social Research
Beverley Hills, Ca. Sage

Rosenthal, R. & Rubin, D.B. (1982)
A Simple General Purpose Display of Magnitude of Experimental Effect
Journal of Educational Psychology 74, 166-9

Rosenthal, R. & Rubin, D.B. (1986)
Meta-analytic Procedures for Combining Studies with Multiple Effect Sizes
Psychological Bulletin 99, 3, 400-6

Rowe, K.J. (1989)
Literacy Development in 5000 5-14 Year-Olds
personal communication
(paper in preparation for Journal of Child Psychology and Psychiatry)

Ryback, D. & Staats, A.W. (1970)
Parents as Behavior Therapy Technicians in Treating Reading Deficits (Dyslexia)
Journal of Behavior Therapy and Experimental Psychiatry 1, 109-19

Sartain, H.W. (ed.) (1981)
Mobilizing Family Forces for Worldwide Reading Success
(Selected Papers from the Seventh IRA World Congress on Reading)
Newark, Delaware International Reading Association

Schneeberg, H. (1977)
Listening While Reading: a four year study
The Reading Teacher 29, 629-35

Schonell, F.J. & Schonell, F.E. (1955)
Graded Word Reading Test
in: Schonell Reading Tests
London Oliver and Boyd

Scott, J. (1983)
Process Evaluation of Paired Reading
Occasional Papers of the Division of Educational & Child Psychology
British Psychological Society, 7, 1, 82

Shake, M.C. (1986)
Teacher Interruptions During Oral Reading Instruction
Remedial and Special Education 7, 5, 18-24

Sharpley, C.F. & Rowland, S.E. (1986)
Palliative vs. Direct Action Stress-Reduction Procedures as Treatments for Reading Disability
British Journal of Educational Psychology 56, 40-50

Siegel, S. (1956)
Non-parametric Statistics for the Behavioral Sciences
New York McGraw-Hill

Simpson, M. (1985)
PATCH - Parents and Teachers with Children at Home
Paired Reading Bulletin 1, 66-7

Singh, N.N. & Singh, J. (1986)
A Behavioural Remediation Programme for Oral Reading: effects on errors and comprehension
Educational Psychology 6, 2, 105-14

Slee, F. (1981)
Using the Daniels and Diack Test at Nine Plus
Remedial Education 16, 1, 18-20

Slee, F. (1985)
How Useful Is the Daniels and Diack Test One?
Remedial Education 20, 2, 64-7

Smith, D.D. (1979)
The Improvement of Children's Oral Reading Through the Use of Teacher Modelling
Journal of Learning Disabilities 12, 3, 172-5

Smith, F. (1978)
Understanding Reading: a psycholinguistic analysis of reading and learning to read (second edition)
New York Holt Rinehart and Winston

Spalding, B., Drew, R., Ellbeck, J., Livesey, J., Musset, M. & Wales, D. (1984)
"If You Want to Improve Your Reading, Ask Your Mum"
Remedial Education 19, 4, 157-61

Spiby, G. (1986)
A Comparative Study of the Effectiveness of Paired Reading and Reading Aloud among a Sample of First Year Pupils in the Remedial Department of a Comprehensive School.
unpublished M.Ed. Thesis, University College of Swansea

S.P.S.S. Inc. (1986)
Statistical Package for the Social Sciences (SPSSx)
New York McGraw-Hill

Staats, A.W. (1973)
Behavior Analysis and Token Reinforcement in Educational Behavior Modification and Curriculum Research
in: C E Thorensen (ed.)
Behavior Modification in Education
Chicago NSSE

Staats, A.W., Minke, K.A. & Butts, P. (1970)
A Token-Reinforcement Remedial Reading Program Administered by Black Therapy-Technicians to Problem Black Children.
Behavior Therapy 1, 331-53

Sweetlove, M. (1985)
Withins School Paired Reading Project - Assessment
Unpublished document, Withins School, Bolton

Sweetlove, M. (1987)
Paired Reading and Direct Instruction Corrective Reading - comparative and joint effectiveness
Paired Reading Bulletin 3, 74-82

Tannenbaum, A.J. (1968)
An Evaluation of STAR: a non-professional tutoring program
Teacher's College Record 69, 5, 433-48

Thirkell, B. (1989)
Peer Tutoring Using Paired Reading in a Multi-Cultural Setting
M.Sc. (Ed. Psych.) Thesis Polytechnic of East London

Thompson, G.B. (1975)
Sex Differences in Reading Attainments
Educational Research 18, 1, 16-23

Tizard, J., Schofield, W.N. & Hewison, J. (1982)
Collaboration Between Teachers and Parents in Assisting Children's Reading
British Journal of Educational Psychology 52, 1-15

Toepritz, P.I. (1982)
A Study of Parent - Child Interaction in a Paired Reading Project
Unpublished M.Sc. (Educational Psychology) Thesis, University of Sheffield

Topping, K.J. (1984)
Parental Involvement in Children's Reading Development - Survey Results
unpublished paper, Kirklees Psychological Service, Huddersfield

Topping, K.J. (1986)
WHICH Parental Involvement in Reading Scheme? a guide for practitioners.
Reading (U.K.R.A.) 20, 3, 148-56

Topping, K.J. (1987)
Correction Procedures in Parental Involvement in Reading Techniques
Paired Reading Bulletin 3, 102-108

Topping, K.J. (1988a)
Paired Reading Training Pack (3rd edition)
Huddersfield Kirklees Paired Learning Project

Topping, K.J. (1988b)
The Peer Tutoring Handbook: promoting co-operative learning
London Croom Helm: Cambridge, Ma. Brookline Books

Topping, K.J. (1989)
A Whole-School Policy on Parental Involvement in Reading
Reading (U.K.R.A.) 23, 2, 85-97

Topping, K.J. & Shaikh, R. (1989)
Paired Reading in English, Urdu and Gujerati (training video)
Huddersfield Kirklees Paired Learning Project

Topping, K. & Whiteley, M. (1988)
Sex Differences in the Effectiveness of Peer Tutoring
Paired Reading Bulletin 4, 16-23

Topping, K. & Whiteley, M. (1989)
Participant Evaluation of Parent-tutored and Peer-tutored Projects in Reading
Journal of Educational Research 31, 3

Topping, K.J. & Wolfendale, S.W. (eds.) (1985)
Parental Involvement in Children's Reading
London Croom Helm: New York Nichols

Topping, K.J., Davies, L., Holling, P., Atkinson, R. & Lawton, G. (1987)
Disadvantage in School Catchment Areas in Kirklees
Huddersfield Kirklees Directorate of Educational Services

Townsend, J. (1987)
Paired Reading with Parent and Peer Tutors at High Bank First School
- an update
Paired Reading Bulletin 3, 83-4

Townsend, J. & Topping, K.J. (1986)
An Experiment Using P.R. With Peer Tutors vs. Parent Tutors
Paired Reading Bulletin 2, 26-31

Tunnell, M.O. & Jacobs, J.S. (1989)
Using "Real" Books: research findings on literature based instruction
The Reading Teacher 43, 470-7

Tyre, C. and Young, P. (1985)
Parents as Coaches for Dyslexic and Severely Reading-Retarded Pupils
in: Topping and Wolfendale (op cit)

Ulmer, R. & Green, F. (1988)
Parent Tutored Paired Reading in a Summer Program for Migrant Children
Paired Reading Bulletin, 4, 33-5

Vane, D.C.W. (1984)
Multiple Deprivation in Kirklees
(Report to Resources and Planning Sub-Committee, 23.8.1984)
Huddersfield Kirklees Directorate of Technical Services

Vaughey, S. and MacDonald, J. (1986)
Paired Reading Projects with Asian Families
Paired Reading Bulletin 2, 6-9

Vellutino, F.R. & Scanlon, D.M. (1986)
Experimental Evidence for the Effects of Instructional Bias on Word Identification
Exceptional Children 53, 2, 145-55

Vincent, D. & de la Mare, M. (1985)
New Macmillan Reading Analysis
London Macmillan Education

Wade, B. & Dewhirst, W. (1983)
Reading Comprehension Revisited
Educational Research 25, 3, 171-6

Walberg, H.J. (1984)
Improving the Productivity of America's Schools
Educational Leadership 41, 8, 19-27

Wareing, L. (1983)
Comparison of the Relative Merits of Three Methods of Parental Involvement in Reading
Unpublished M.Sc. (Educational Psychology) thesis, North East London Polytechnic

Wareing, L. (1985)
A Comparative Study of Three Methods of Parental Involvement in Reading
in: Topping and Wolfendale (op cit)

Watts, A.F. (1948)
Holborn Reading Scale
London Harrap

Weinberger, J., Jackson, A. & Hannon, P. (1986)
Involving Parents in the Teaching of Reading: will they respond?
Reading (U.K.R.A.) 20, 3 179-83

Welch, S. (1985)
An Investigation into the Use of a P.R. Approach by Teachers
unpublished M.Sc. (Educational Psychology) thesis, North East London Polytechnic

Welsh, M. & Roffe, M. (1985)
Paired Reading Projects with Asian Families
Paired Reading Bulletin 1, 34-7

Whiteley, M. & Topping, K. (1986)
The Macmillan Reading Analysis: a preliminary evaluation
Paired Reading Bulletin 2, 85-9

Wilkinson, A.C. (1980)
Children's Understanding in Reading and Listening
Journal of Educational Psychology 72, 4, 561-74

Williams, T. (1987)
Paired Reading: the Shropshire approach
Paired Reading Bulletin 3, 85-90

Winter, S. (1985)
Giving Parents a Choice
in: Topping & Wolfendale (op cit)

Winter, S. (1987)
Parents and Pupils as Remedial Tutors
Bulletin of the Hong Kong Psychologists' Society, 18, 15-31

Winter, S. (1988)
Paired Reading: a study of process and outcome
Educational Psychology 8, 3, 135-51

Winter, S. & Low, A. (1984)
The Rossmere Peer Tutor Project
Behavioural Approaches with Children 8, 2, 62-5

Wisner, M. (1988)
Reading While Listening
Paired Reading Bulletin, 4, 73-7

Wolfendale, S. (1983)
Parental Participation in Children's Development and Education
London Gordon and Breach

Wolfendale, S. (1989)
A Whole School Policy on Parental Involvement in Reading
Paired Learning 5, 1-3

Wolfendale, S. & Gregory, E. (1985)
Involving Parents in Reading: a guide for in-service training
Northampton Reading & Language Development Centre

Young, P. & Tyre, C. (1983)
Dyslexia or Illiteracy? Realizing the Right to Read
Milton Keynes Open University Press

Young, P. & Tyre, C. (1985)
Teach Your Child to Read: a guide for parents
London Fontana

Yule, W. (1967)

Predicting Reading Ages on Neale's Analysis of Reading Ability

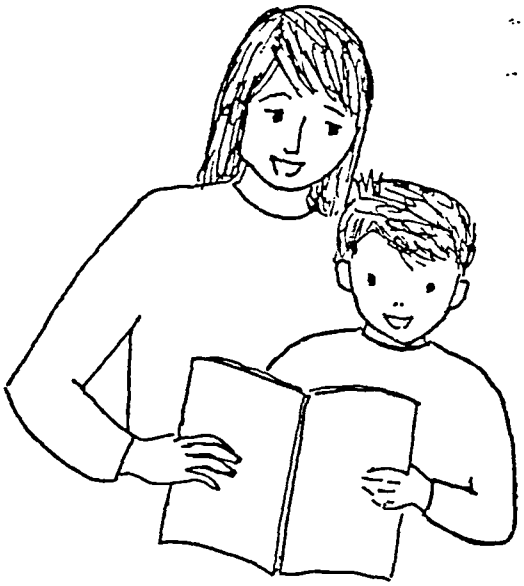
British Journal of Educational Psychology 37, 252-5

The Paired Reading Bulletins and Paired Learning are available on loan internationally from the Educational Resources Information Centre (ERIC) through any library (microfiche reference numbers: 1985 - ED285124; 1986 - ED285125; 1987 - ED285126; 1988 - ED 298429, 1989 - ED313656).

Appendix 1 : How To Do It Leaflet for Parents

PAIRED READING

HOW TO DO IT



how mums and dads
can help their kids
to read better

(reduced from A4 size, as are all subsequent Appendices)

PAIRED READING is a very good way for parents to help with their children's reading. It works really well with most children, and their reading gets a lot better. Also, Paired Reading fits in very well with the teaching at school, so children don't get mixed up. Most children really like it - it helps them want to read.

WHAT YOU NEED

Books

to choose from, at home or from school or the library. School will tell you about the libraries.

Your child should choose the book. Children learn to read better from books they like. Don't worry if it seems too hard. Your child will soon get used to picking books that aren't too hard.

If your child gets fed up with a book, and wants to change it, that's O.K. Only read a book again if the child wants to.

Time

to do Paired Reading. Try very hard to do some Paired Reading nearly every day. You only need to do 5 minutes each day, if you want. Don't do more than 15 minutes unless your child wants to carry on.

Don't make children do Paired Reading when they really want to do something else.

If mum or dad haven't got time to do Paired Reading for 5 minutes 6 days a week, grandma or grandad or older brother or sister can help. They must do Paired Reading in just the same way.

Place

to do Paired Reading.

Try to find a place that's quiet. Children can't read when it's noisy, or when there's lots going on. Get away from the T.V., or turn it off.

Try to find a place that's comfy. If you're not comfortable, you'll both be shifting about. Then you won't be able to look carefully and easily at the book together.

Get close - reading together can be very warm and snuggly.

continued . . .

New Ways

of helping.

It's often harder for parents to learn new ways than it is for children !

With Paired Reading, the hardest things for parents to get used to are:-

(1) When your child gets a word wrong, you just tell your child what the word says. Then your child says it after you. You DON'T make the child struggle and struggle, or 'break it up' or 'sound it out'.

(2) When your child gets words right, you smile and show you are pleased and say "good". You DON'T nag and fuss about the words your child gets wrong. Praise for: good reading of hard words, getting all the words in a sentence right, and putting wrong words right before you do (self-correction).

Talk

Show interest in the book your child has chosen. Talk about the pictures. Talk about what's in the book as your child goes through it. It's best if you talk at the end of a page or section, or your child might lose track of the story. Ask what your child thinks might happen next. Listen to your child - don't you do all the talking.

Notes

It is a help for both child and school teacher to keep a note each day of what has been read, and how your child is going on.

There is a diary on the last page that you can use for this. If your child has done well, write this on the paper.

At the end of the week, your child can take the paper to show the teacher at school, and get some extra fuss for doing well. This helps to keep them keen.

continued . . .

HOW TO DO IT

Paired Reading has 2 steps:-

Reading together

You and your child both read the words out loud together. You must not go too fast. Make your speed as fast or as slow as your child's.

Your child must read every word. If your child struggles and then gets it right, show you are pleased. But don't let your child struggle for more than 5 seconds.

If your child:-

(a) struggles too long, or (b) struggles and gets it wrong

then you:-

(1) just say the word right yourself, and

(2) make sure your child then says it right as well.

Make sure your child looks at the words. It can help if one of you points to the word you are both reading with a finger. It's best if your child will do the pointing.

Reading alone

When you are Reading Together and your child feels good enough, your child might want to read a bit alone. You should agree on a way for your child to ask you to be quiet.

This could be a knock, a sign or a squeeze. (You don't want your child to have to say "be quiet", or your child will lose track of the reading). You stop reading out loud straight away, and praise the child for making the sign.

When your child struggles for more than 5 seconds, or struggles and gets it wrong, you read the word out loud right for your child. Make sure your child then says it right as well.

Then you both go on reading out loud together, until your child again feels good enough to read alone, and again asks you to be quiet.

Appendix 2 : Paired Reading – What Are The Advantages ?

PAIRED READING – What are the Advantages?

1. Children are encouraged to pursue their own interests in reading material. They have more enthusiasm from reading about their own favourite things, and so try harder. Paired Reading gives them as much support as they need to read whatever book they choose.

2. Children are more in control of what's going on – instead of having reading crammed into them, they make decisions themselves in the light of their own purposes (eg about choice of books, going on longer than 10 minutes, and going onto Reading Alone.)

3. There is no failure – it is impossible not to get a word right within 5 seconds or so.

4. Paired Reading is very flexible – the child decides how much support is necessary according to the current level of interest, mood, degree of tiredness, amount of confidence, difficulty of the book, and so on.

5. The child gets lots of praise – its much nicer to be told when you're doing well, instead of just being moaned at when you go wrong.

6. There's lots of emphasis on understanding – getting the meaning out of the words – and that's what reading is all about. It's no use being able to say the words mechanically without following the meaning.

7. Paired Reading gives continuity – it eliminates stopping and starting to 'break up' hard words – which often left you having forgotten the beginning of the sentence by the time you got to the end. This means it's easier for children to make sensible guesses at new words, based on the meaning of the surrounding words.

continued . . .

8. During Reading Together, a child can learn (by example) to read with expression and the right pacing - for example, by copying how the adult pauses at punctuation, or gives emphasis to certain words.

9. Children are given a perfect example of how to pronounce difficult words, instead of being left to work it out themselves and then thinking their own half-right efforts are actually correct. In Paired Reading, children learn by what psychologists call modelling.

10. When doing Paired Reading, children get a bit of their own peaceful, private attention from their parents, which they might not have otherwise had. There is some evidence that just giving children more attention can actually improve their reading. Of course, this also applies to other schemes for non-teachers to help with children's reading.

11. Paired Reading increases the amount of sheer practice at reading children get. Because children are supported through books, they get through them faster. So the number of books read in a week goes up, and the number of words children look at in a week goes up - and more words must stick in the child's memory.

12. Paired Reading gives parents a clear, straightforward and enjoyable way of helping their children - so they don't get confused, worried or bad-tempered about reading.

So you can see how Paired Reading helps - children

A. WANT to read more,

B. have more CONFIDENCE, and

C. show more UNDERSTANDING.

Name: _____

HOME READING RECORD SHEET

DAY	BOOK CHOSEN	TIME SPENT	WITH WHOM?	COMMENTS
Monday				
Tuesday				
Wednesday				
Thursday				
Friday				
Saturday				
Sunday				

Teacher's Comment: Signed: Date:

Appendix 4 : Beyond Paired Reading

BEYOND PAIRED READING

WHERE do you want to go from here?

(1) Stop Paired Reading for a Rest (and perhaps start again later)?

(2) Go on Reading Together and Independently, but only twice a week?

(3) Go on Reading Together and Independently, just as often as nc

(4) Go on to Stage 3 - Reading Mini-help,

(5) Go on to Stage 4 - Reading silently

(6) Go on to Stage 5 - Reading So

→ for continuingly committed families →
→ for increasingly competent readers →

HOW to go on from here

Paired Reading and Beyond - The Five Stages

REMEMBER - LOTS OF PRAISE AT EVERY STAGE

CHILD (and signals to go on to the next stage when confident) chooses starting stage

1. READING TOGETHER

2. READING INDEPENDENTLY
(and together as necessary)

3. READING MINI-HELP
(Child attempts all words independently; parent gives correct example of error word only (including omissions, excluding additions); child repeats correctly and continues unaided).

4. READING SILENTLY
(Pair both read each page silently sitting together, then discuss and question each other about content at some natural break (picture, paragraph and page end)

5. READING SOLO
(After initial discussion with parent about book, child reads silently sitting alone, then comes to parent at natural break for discussion/ questions. If the child tends to continue too long alone, a suitable natural break is agreed in advance.

PARENT goes back to support of previous stage if many errors of accuracy or understanding.

Appendix 5 : Teacher Feedback Questionnaire

PAIRED READING TEACHER EVALUATION CHECK-LIST

Name of Child: _____

Please circle the answer indicated by your observations. Only indicate change if you feel it has occurred since Paired Reading started and is definite and significant.

A. General

Is the:-

(1) Amount of Reading done	less	more	same	not seen
(2) Width and Variety of Reading	more	less	not seen	same
(3) Comprehension of Reading	same	not seen	more	less

B. Attitude

Is there:-

(4) Confidence in Reading	not seen	same	less	more
(5) Willingness to Read	same	less	more	not seen
(6) Interest in Reading	more	not seen	less	same
(7) Pleasure in Reading	less	more	same	not seen

C. Oral Reading

Is:-

(8) Accuracy	worse	better	same	not seen
(9) Fluency	better	same	not seen	worse
(10) Expressiveness	same	worse	better	not seen
(11) Pacing	not seen	better	worse	same

D. Other

Is:-

(12) Concentration & Motivation generally	same	worse	not seen	better
(13) Behaviour generally	worse	not seen	better	same

(14) Any other significant changes (specify):-

Thank you for giving your observations.

Appendix 6 : Parent Feedback Questionnaire

PAIRED READING WHAT DO YOU THINK?

Name of Child _____

PLEASE TICK WHICH IS TRUE FOR YOU

A. Is your child:-

(1)	Reading more	: about the same:	Reading less
(2)	Sticking to the same kind of book	: about the same:	Reading different kinds of book
(3)	Understanding books more	: about the same:	Understanding books less

B. Is your child:-

(4)	Less confident in reading	: about the same:	More confident in reading
(5)	More willing to read	: about the same:	Less willing to read
(6)	Less interested in reading	: about the same:	More interested in reading
(7)	Enjoying reading more	: about the same:	Enjoying reading less

C. When reading out loud,is your child:-

(8)	Making more mistakes	: about the same:	Making less mistakes
(9)	Keeping a steadier flow	: about the same:	Stopping & starting more
(10)	Reading in a lifeless,boring way	: about the same:	Reading with more life and expression

D. Is your child:-

(11)	Behaving better at home	: about the same:	Behaving worse at home
(12)	Happier at home	: about the same:	Less happy at home

E. Are you going to:-

(13) Stop Paired Reading, and perhaps start again later? ☐

Go on doing Paired Reading, but only twice a week? ☐

Go on doing Paired Reading 5 times a week? ☐

Go on reading at home, but in a rather different way? ☐

Any other comments:-

Thank you for telling us what you think.

Appendix 7 : Peer Tutor Feedback Questionnaire

PAIRED READING WHAT DO YOU THINK?

Name of Tutor _____ Name of Tutee: _____

PLEASE TICK WHICH IS TRUE FOR YOU

A. Is your tutee:-

(1)	Reading more	: about the same:	Reading less
(2)	Sticking to the same kind of book	: about the same:	Reading different kinds of book
(3)	Understanding books more	: about the same:	Understanding books less

B. Is your tutee:-

(4)	Less confident in reading	: about the same:	More confident in reading
(5)	More willing to read	: about the same:	Less willing to read
(6)	Less interested in reading	: about the same:	More interested in reading
(7)	Enjoying reading more	: about the same:	Enjoying reading less

C. When reading out loud, is your tutee:-

(8)	Making more mistakes	: about the same:	Making less mistakes
(9)	Keeping a steadier flow	: about the same:	Stopping & starting more
(10)	Reading in a lifeless, boring way	: about the same:	Reading with more life and expression

D. Would you like to:- CHOOSE UP TO THREE:

(11)	Go on Peer Tutoring as often as now?	YES	☐
(12)	Go on Tutoring, but not so often?	YES	☐
(13)	Go on Tutoring, with a different tutee?	YES	☐
(14)	Be Tutored yourself, by someone better?	YES	☐
(15)	Tutor reading, but in a different way?	YES	☐
(16)	Tutor something else, like maths or spelling?	YES	☐

Any other comments:-

Thank you for telling us what you think.

PAIRED READING — WHAT WAS IT LIKE?

TICK WHICH IS TRUE FOR YOU

1.	a.	It was hard to get books	OR	b.	It was easy to get books
2.	a.	It was easy to find time	OR	b.	It was hard to find time
3.	a.	It was hard to find a good place to read	OR	b.	It was easy to find a good place to read
4.	a.	It was easy to learn to do	OR	b.	It was hard to learn to do
5.	a.	I soon got fed up with it	OR	b.	I liked doing it
6.	a.	The Record Sheet was a help	OR	b.	The Record Sheet was no use

PAIRED READING HAS LED TO:

7.	a.	Not liking all kinds of reading	OR	b.	Liking all reading better
8.	a.	Getting better at all kinds of reading	OR	b.	No better at all kinds of reading
9.	a.	Getting on worse with each other	OR	b.	Getting on better with each other
10.	a.	I want to go on doing Paired Reading	OR	b.	I want to stop Paired Reading for now
11.	a.	I won't tell anyone about Paired Reading	OR	b.	I will tell other people about Paired Reading

12. Can you tell us one thing we can do to make Paired Reading better? (or the way we tell you about it?) Write what you think here:-

Name: _____