The Effectiveness of Additional Interventions for Children with Literacy Difficulties in Years 3 & 4.

by

Timothy Kenneth Bunn

A thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Education

University of Warwick, Institute of Education

September 2006

Table of Contents

List of tables and charts

Acknowledgements

I should like to express my great respect and thanks to all the staff at the University of Warwick Institute of Education who have tutored and supported me throughout the six years of the EdD programme. I owe a particular debt of gratitude to my supervisor, Dr Jonathan Solity, who has guided and above all encouraged me when analysis and writing of this thesis seemed to me to have become almost impossible.

I should also like to thank Derek Lucas, former Principal Educational Psychologist for Northamptonshire County Council, for financial support in my first year of study, and subsequent Education Officers, especially Gary Sturgess, who continued to encourage me to continue the research in the face of the demands of the "day job".

During the data collection phase of the research I received substantial help and support from my then personal assistant, Mrs Jacqui Henry, whose thoroughness and positive support were always very much appreciated.

I could not have embarked on the EdD programme without the support of my wife, Carol Passingham, to whom I am more grateful than I usually admit, and whose direct and indirect help through listening, gentle encouragement and occasional reminders have been immeasurable.

Finally, I should like to thank the teachers and teaching assistants who gave me considerable time and whose data forms the basis for this research. I did not realize when I sought their help how long it would take to reach conclusions but I hope their contributions will ultimately lead to the development of better help for children with literacy difficulties.

Declaration of Authorship

The contents of this thesis include nothing which has previously been published, other than quotations and references to the work of other authors or researchers which are stated in the references section. These references include those to the author's previous published and unpublished work, but this work does not form part of the present thesis.

The research has not involved collaboration with others.

The research has not been submitted for a degree at another university.

Abstract

This study compared the progress in reading and spelling of 256 children in 11 classes in 9 English primary schools in Years 3 and 4, and a partially overlapping sample of 126 children who received additional help with literacy during one year. Teachers and teaching assistants used either Additional Literacy Support (ALS), a highly structured set of small group teaching materials devised by the English National Literacy Strategy, or a wide variety of other materials including other published intervention programmes, reading scheme based , computer based and individually designed interventions, or a combination of ALS and other interventions. The influence of a broad range of contextual factors were investigated, especially whether children's qualities, school factors such as Socio-Economic Status and class size, and delivery differences made significant differences to the outcomes of the different interventions. The study used a naturalistic quasi-experimental design, in which teachers were asked to record details of their children and interventions without altering their professional decisions, which has not been used before in investigating literacy difficulties in context.

ALS was marginally more effective than other interventions in the majority of classes, but was clearly superior in value for money terms. Children's qualities did not appear to affect outcomes. Although children receiving additional help made better than average progress, below average children receiving only class teaching made more progress. Overall catch-up was limited, especially in spelling. There appeared to be a larger influence of class teaching than expected. A tentative theory of how class teaching and additional interventions combine is suggested. The study considers how research of this type could be advanced, the need for further development of both class literacy teaching and additional interventions, and raises some questions about national policy towards literacy interventions.

List of abbreviations:

ALS	Additional Literacy Support
ANOVA	Analysis of Variance
BAS	British Ability Scales
BPS	British Psychological Society
COP	Code of Practice for Special Educational Needs
CVA	Contextual Value Added
DfE	Department for Education
DfEE	Department for Education and Employment
DfES	Department for Education and Skills
DISTAR	Direct Instructional System for Teaching & Remediation
EAL	English as an Additional Language
ELS	Early Literacy Support
ERR	Early Reading Research
FLS	Further Literacy Support
GCSE	General Certificate of Secondary Education
HI	Hearing Impairment
ICT	Information & Communications Technology
ILS	Integrated Learning System
IQ	Intelligence Quotient
KS1,2,3	Key Stage 1,2,3
LEA	Local Education Authority
NC	National Curriculum
NLS	National Literacy Strategy
NRC	National Research Council

NRP	National Reading Panel
OFSTED	Office for Standards in Education
OISE	Ontario Institute for the Study of Education
ORT	Oxford Reading Tree
PA	Phonological Awareness (occasionally also Phonemic Awareness)
PhAB	Phonological Abilities Battery
PAT	Phonological Abilities Test
PAT	Phonological Approaches to Teaching
PIRLS	Progress in International Reading and Literacy Study
QCA	Qualifications and Curriculum Authority
RAN	Rapid Automatic Naming
RITA	Reader's Intelligent Teaching Assistant
SAF	Social Advantage Factor
SAT	Standard Assessment Task/Test
SEN	Special Educational Needs
SES	Socio-Economic Status
SLI	Specific Language Impairment
SpLD	Specific Learning Difficulty
SPELLIT	Study Programme to Evaluate Literacy Learning through Individualised Teaching
SPSS	Statistics Package for the Social Sciences
TA	Teaching Assistant
THRASS	Teaching Handwriting, Reading and Spelling Skills
USA	United States of America

Chapter 1: Introduction

"In the area of learning disabilities, there are no empirically validated answers to the question of what intervention(s) work best for which children in what setting(s) for what duration and for what reason."

(Foorman et al 1997)

1.1 The origins of this study

The kernel of this study is an investigation into whether a particular intervention, Additional Literacy Support (ALS), is effective in helping children to catch-up in reading and spelling in their middle primary years in England. ALS was the first in a series of small group interventions delivered mainly by Teaching Assistants (TAs) designed to deal more efficiently than previous interventions with the very long-standing problem of children falling behind in reading and spelling at primary school. The others were Early Literacy Support (ELS) for younger primary children and Further Literacy Support (FLS) for early secondary children. It was part of the National Literacy Strategy (NLS), which was introduced in England & Wales in 1998 to try to raise educational standards, a very prominent part of the New Labour Government's programme. After decades of wrangling about what type of schools England and Wales should have, the New Labour slogan "standards not structures" seemed to many educationalists to be a very significant turning point. At last we would start to try to teach all children better rather than simply sorting them into different social classes.

The NLS itself began the Herculean task of reforming the teaching of reading and writing in English primary schools. Schools were not obliged by law to use the new methods of the NLS, but it was clear that only very brave schools would stand out against it. It was based on the National Literacy Project (Sainsbury 1998), which tried out a combination of new methods of teaching literacy in primary classrooms, most notably a daily Literacy Hour including whole-class instruction, small group practice and a plenary, without the previous emphasis on teachers spending much of their time on hearing children read. It concerned many teachers and researchers that the project only ran for a year, and that it seemed that NLS was being written before the project was completed. Although evidence based policy development was much talked about, the political necessity for results in time for the next election seemed to put its effectiveness at risk because there was not enough time to iron out all the problems.

ALS was the first of the second stage initiatives, part of a "Three Waves" model for dealing with literacy difficulties. The first wave was class teaching, reformed and improved through the NLS. The second was the series of small group interventions, ALS, ELS and FLS, intended to deal with the majority of catching up problems which were seen as a largely a result of socio-economic deprivation. The third would be a series of more individually delivered interventions intended to help the small minority of children who had not made sufficient progress through the first two waves, and who perhaps suffered from biologically based special needs. As this

research was being conducted, the Government of England & Wales introduced its approach to third wave interventions, "Targeting Support" (DfES 2003), which inevitably affected this study, by widening its scope and raising further questions about the 3 waves strategy and how interventions should be evaluated.

ALS itself was developed quickly by NLS on the basis of current research ideas and intervention practice, especially the work of Marie Clay on Reading Recovery (Clay 1985). It was introduced in 1999, but had not been properly validated through research. This study was intended to do that job.

Immediately a number of questions were raised: what sort of evidence was needed, what sort of research methods should be used, how broad should the scope of a study be? These questions were bound to be coloured by practical questions: what sort of study be conducted by a researcher working full-time as an education officer for a Local Education Authority, with no dedicated time and no possibility of research funding?

The epistemological and methodological stance of any researcher is probably determined (or should it be "chosen"?) more by background and training than by conscious analysis. My first degree was in a psychology of an exclusively empirical, experimental, quantitative type. My dissertation was a joint paper on ecological influences on the behaviour of baboons in Ethiopia (Aldrich-Blake et al 1971). I later trained as an educational

psychologist, and again the training was in a hard-science tradition. I spent two years researching literacy teaching and pupil outcomes in Year 2 in the 1990s in Leicester, but did not succeed in bringing substantial data gathering and analysis to publishable conclusions. In retrospect, I was trying in a very limited way to raise the kind of questions that the National Literacy Project put forward about class teaching of literacy. Currently as an education officer for a Local Education Authority (LEA) I am concerned with broad questions of how to achieve better results for children with special educational needs (SEN).

But balancing these influences, I had also studied philosophy. I worked as a teacher of primary and secondary children, and with children with severe learning difficulties. As an educational psychologist, I worked with many children with literacy difficulties. Their problems and often their suffering seemed to call for both much greater effort to collect good and systematic information about what might be causing their difficulties and a more personal engagement with their struggle. I had studied the sociology of special education in its early stages (Bunn 1987) at a time when special educational systems seemed to be becoming increasingly mechanical. And I had trained in family therapy and worked in Child Guidance within a loosely dynamic orientation. And as an education officer, I not infrequently received strong and distressing accounts from parents about how their children with dyslexia needed so much more help.

It may be impossible to be sure whether a decision to focus on a particular type of question is based on objective "real-world" conditions or upon subjective meaning constructions. The decision to focus on "product" rather than "process", the product being as large a scale evaluation as possible of the results of using ALS, rather than the processes by which TAs and teachers might support children of 8 and 9 to develop their understanding of how to read and spell, seemed to be made by objective conditions. The English education community seemed to need to know much more urgently how well ALS worked. It was not crying out to know what inter-subjective or intra-individual cognitive processes were involved in using ALS. Researchers, especially those in training, need to answer questions that other people beside themselves will find interesting. This does not mean that qualitative methods are without interest or value in this particular field. In the final analysis, there probably is a greater role for research on how special educational processes take place in mainstream schools, because of limitations which may, perhaps, be inherent in experimental and quasi-experimental research paradigms in literacy interventions. These questions will be taken up again in the final chapter.

Research on how children learn to read and spell and on what might be causing difficulties has made great progress in the last 30 years (Stanovich 2000). Researchers mainly using correlational and experimental methods based on cognitive psychology have been able to resolve divisive questions about how children learn to read (called "reading wars" in the USA) (Snow et al 1998). The very success of these

efforts has reinforced the influence of researchers who advocate a stronger experimental scientific influence upon policy, especially in the United States (see Reid-Lyon 2004, for example). It would have been odd not to base understanding and explanation on the most successful research tradition available.

The most important determinants of the ways in which this research was conducted may have been practical, however. It was almost impossible to find time during the working day to collect data directly about children, and nearly as difficult to make time to interview or observe teachers or TAs. Much of the work of an education officer is reactive. A school finds it is having major difficulties dealing with a child's behaviour. A parent fells that they have been let down by recent decisions by a school. Or helping a child to progress looks as though it is going to cost more – is there any alternative? It can become extremely difficult to set aside time to indulge an interest in research when there are immediate crises to attend to.

LEAs do in fact collect data on paper files about students with literacy difficulties, and have been doing so for years. Unfortunately, the data is collected unsystematically and in ways that do not allow evaluation of the effectiveness of interventions to be made, except in quite broad and probabilistic terms. They have been collecting data electronically for about 10 years, biographical and financial details, but not what interventions children receive. The results of SATs and GCSEs are not in a form that can be readily linked with information about how much help they have had

or the kinds of difficulties that they have been identified as having. I have always felt that a huge amount of effort is wasted in collecting and holding information about children which cannot be used to ask questions about what kinds of teaching really did help the children make progress. Would it be possible to obtain, in a form that could be analysed, information about children's literacy difficulties (the most common educational difficulty) without expecting schools to do unreasonable amounts of extra work? If it was possible, could LEAs begin to collect and distribute back to schools information that would help them make better decisions about helping children?

It might then be possible to use an obvious and serious limitation on what time I could give directly to data collection as a source of strength: only the most important data would be asked for, and its value would be tested in the utility of the conclusions which came out.

Thus this study took the form it did because of a combination of an objective demand for evidence about the effectiveness of ALS, a personal history, a powerful and successful international research effort to understand how children learn to read, and an opportunity to collect data in a way that might open doors from research onto the LEA shop-floor.

1.2 Brief outline of the 5 chapters

In the light of these questions and issues about how to tackle the key research questions, this report is organised in 5 further chapters, using a

standard empirical science sequence of literature review, research methods, results, discussion of the results and conclusions.

1.2.1:Chapter 2: Review of the Literature.

The review covers three main areas, the early development of literacy and the nature of the difficulties which some children experience, including within-child and context factors, national and local policy development in the last 10 years in dealing with literacy difficulties, and interventions in England and Wales, with a particular focus on a landmark synthetic study of interventions by Greg Brooks (Brooks 2002).

1.2.2: Chapter 3: Research Methodology

Two possible research designs were considered: experimental and quasi-experimental. Practicality, cost and validity issues were all relevant to the decision to use a naturalistic rather than a truly experimental design.

The key variables included two types of measures of reading and spelling gain, which it was hoped could be integrated, a specially designed measure of children's qualities, school factors such as size and SES, and how interventions are delivered in schools.

The sample of schools from a single English Local Education Authority (LEA) was very important in shaping the study, and its rationale is discussed, as well as the unexpected problems in data collection processes which led to the actual data set. This chapter also presents

research aims, and concludes with a detailed set of research questions, the answers to which structure the next chapter.

1.2.3: Chapter 4: Results

The study collected outcome data into two overlapping databases, "Classes" and "Additional Help", intended to complement each other. The chapter explains the differences in the databases. The larger but less detailed Classes is used to examine differences between children who received additional interventions and those who did not.

The intended integration of the two main types of gain measure, age scores and NC levels, seemed at first to be possible. But when the outcomes of interventions were compared, the results did not support integration. Answers to the two main questions, the effectiveness of ALS compared to other interventions, and the role of class teaching compared with additional interventions, nevertheless emerge, but are somewhat qualified.

The search then begins for explanations for the differences between interventions, looking at selection for different types of help, differences between children and between schools, and between classes. A different way of analysing types of interventions is used to further explore the relationship between interventions and outcomes. Delivery factors, including staffing, are also considered. Staff and children's views of the materials are taken into account.

Finally, multiple regression is used to examine how much delivery, school and pupil factors explain the outcomes of the interventions.

1.2.4: Chapter 5: Discussion

A comparison of class teaching and additional interventions was one of the key questions in this study. A theory is suggested based on the findings about the relationship between these two ways of helping children with literacy difficulties. Brooks' conclusion about the role of class teaching in helping children catch up is called into question. Comparison between interventions gave only qualified answers in terms of outcomes. But when "value for money" analysis is used, the answers were much clearer. The role of pupil qualities continues to tease and surprise by its elusiveness. Although the study was not designed to consider school processes in general, the ways in which interventions were delivered were significant.

As the implications of the results became clear, the methodological and technical limitations of the study also became clearer. Two complementary ways in which the study could be improved are considered: technical improvements on the design and sample, and new work on the school and class processes through which intervention decisions are made.

Finally, some policy implications are considered.

Chapter 2: Review of the Literature

2.1 Introduction:.

This study focuses on the most "mechanical" end of the literacy continuum, especially decoding print and encoding it in writing. This is mainly because a working assumption was made that until the possible value of a narrow focus is shown to be exhausted, more progress in research can be made here than by working with value, motivation and other harder to measure features of school and child performance. There is a substantial research tradition which questions approaches to literacy intervention that appear to involve only technical issues (see for example, Wearmouth 2002), but it is not considered here.

The agenda for this chapter is therefore to review what is most relevant to understanding interventions for children with literacy difficulties from

1) Literature about the early development of literacy and the nature of children's literacy difficulties, including dyslexia;
2) National and local policy options on literacy development and interventions to deal with difficulties, particularly in England and Wales, but with some comparisons with the United States of America (USA);
3) Interventions in England and Wales in the last 25 years, with a particular focus on a recent synthetic analysis of methodologically adequate research (Brooks 2002).

2.2 Literacy development and difficulties:

2.2.1 Child factors predicting literacy achievement

Research in Britain, the USA and Europe has been converging for some time towards a consensus about which within-child factors on school entry predict later reading achievement. By the late 1980s letter knowledge, phonological awareness, vocabulary knowledge and verbal reasoning, and knowledge about print were seen as the most important prerequisites, in roughly that order (Adams 1990, Ch4). Towards the end of the 1990s, the picture had altered only slightly, with more specific aspects of language skills, including receptive and expressive language, receptive and expressive vocabulary and verbal memory all having independent effects (Snow et al 1998). Most recently this evidence has been subjected to meta-analysis (Scarborough 1998) and review (Elbro & Scarborough 2004). The main factors have not changed but some subtleties in the language domain (for example about the role of rapid naming) have emerged, which raise questions about the links between different factors.

The best predictor is probably letter identification, with average correlations of between r=.52 and r=.57 (depending on how narrowly letter identification is defined) across 21 (letter sound knowledge) and 24 (naming letters) studies (Elbro & Scarborough 2004). But as Adams (Adams 1990, pp 61-64) pointed out, it has also proved unproductive to specifically train children in letter identification as a way of preventing reading difficulties.

Over the last 20 years, phonological awareness (PA) has been shown to be the most crucial prerequisite for literacy achievement. Scarborough found that the average correlation between phonological awareness and later reading achievement over 27 studies was r=.46, second only to letter identification (Elbro & Scarborough 2004). Unlike letter identification, training in phonological awareness does have a major effect on reading achievement. Phonemic Awareness training has been extensively investigated and was the subject of the first subgroup of the American National Reading Panel (NRP) investigation into how to help children with reading difficulties (NRP 2000). Clear benefits have been found for combining PA with work on letter-sound correspondences (Bradley & Bryant 1983). A similar result was obtained using Teaching Assistants with slightly older children (seven year olds) by Hatcher, Hulme and Ellis (Hatcher, Hulme & Ellis 1994). Their groups were given PA alone, reading practice alone (similar to the Marie Clay Reading Recovery programme, Clay 1985) or PA with reading practice and explicit letter-sound correspondence practice. The third group made the greatest progress.

Receptive vocabulary knowledge, measured using tests which require children to identify which of a set of pictures corresponds best to a spoken word, was not as strong a predictor as expressive vocabulary, measured by asking the child to name a series of pictures (correlations of r=0.33 and r=0.45 respectively, from Elbro & Scarborough 2004). While the former is

probably a better measure of the extent of a child's word knowledge, which in turn is important in reading because children almost certainly find it harder to read words whose meaning they do not know (see Adams 1990, p140), the latter is probably a better measure of accurate retrieval of phonological representations of words and of their production in speech

Children's familiarity with the mechanics of how meaning is encoded in print in books (especially the concept of "word") is an important predictor (7 studies with an average correlation of r=0.46 in Scarborough 1998). However, its predictive power declines with age.

General language abilities are important, with 7 studies in Scarborough's meta-analysis showing an average correlation of r=0.46 with reading achievement (Scarborough 1998). Children with Specific Language Impairment (SLI) have been studied extensively, because they have a particularly strong risk of reading difficulties. Interventions, particularly early speech therapy, while effective in overcoming speech difficulties, may not prevent later reading difficulties (Huntley et al 1988). It remains unclear precisely what aspects of SLI are most important in causing reading difficulties (Dockerell & Lindsay 2004). There are some indications that milder difficulties which do not affect the earlier stages of literacy acquisition might be linked to under achievement later, as comprehension becomes more important (Snowling & Nation 1997).

Grammatical knowledge and especially children's ability to spell morphological features have been studied quite recently by Bryant, Nunes and colleagues (Bryant, Nunes and Bindman 1998, Kemp & Bryant 2003, Bryant and Nunes 2004), who have suggested a stage model for the development of spelling in which morphological features such as the final "ed" past tense spelling in English is achieved correctly only after an earlier phonological stage (in which letter-sound correspondences are learned). There is evidence that alongside PA, grammatical awareness is a continuing predictor of reading achievement at the later primary stage. Muter and Snowling (Muter & Snowling 1998) found reading accuracy at 9 was best predicted by "grammatical and phonemic awareness, which, along with vocabulary, accounted for 70% of the total variance in reading accuracy skill" (p327). They argue (following Tunmer 1989) that grammatical awareness may have a significant role in reading as well as spelling, because it may function to improve decoding through the use of context cues, especially at the middle primary stage of education.

The role of verbal short-term memory in reading acquisition remains controversial. Digit span has for a considerable time been considered the best measure of short-term memory limitations in reading (for example, Miles 1983) but recently Scarborough (Scarborough 1998) suggested that sentence or story recall may be a better measure, and others (Muter & Snowling 1998) have suggested that non-word repetition may be a better measure at the later primary stage.

Finally, rapid automatic naming (RAN) has recently been suggested by Wolf and colleagues as a separate and important predictor of reading difficulties. Scarborough found an average correlation of r=0.38 across fourteen studies (Scarborough 1998). Stanovich and many other had argued that RAN was one aspect of phonological processing (Stanovich 1986) and RAN was included in tests of phonological abilities such as PhAB (Frederickson et al 1996). However Wolf and her colleagues have found evidence that phonology and RAN constitute independent difficulties, and that phonological and RAN difficulties constitute a "double deficit"(Wolf & O'Brien 2001). In particular there is considerable evidence that in languages with more regular letter sound correspondences than English such as German and Spanish that RAN is a more powerful predictor as phonology becomes less significant (Landerl & Wimmer 2000). Snowling and Griffiths (Snowling & Griffiths 2004) have recently speculated that a "speed of processing" factor, which they link to RAN, may be involved in impeding some children's ability to learn reliable mappings between orthography and phonology.

2.2.2 Specific disabilities associated with literacy difficulties, including dyslexia

Children with SLI are particularly at risk of later literacy difficulties, as we have discussed above. Children with hearing impairment (HI) or deafness are also particularly vulnerable to literacy difficulties. Population studies by Di Francesca in the US (Di Francesca 1972) and Conrad in England and Wales (Conrad 1977, 1979) have found that average reading ages at age

15 to 16 were 9:2 years and 9:0 years respectively. More recent research has not altered the pattern (Allegria 2004, p 460-461).

The largest group of children whose literacy difficulties are associated with a disability are dyslexics. Dyslexia itself remains a controversial, confusing but very extensively researched area (see the Fourth, Fifth and Sixth British Dyslexia Association International Conference Reports, Hulme & Snowling 1997, Fawcett 2001 and Reid & Fawcett 2004). The key questions for dyslexia are very similar to those for literacy difficulties in general: what are its causes, and what can or should be done to help children who experience it. The concept of dyslexia itself adds further questions: is there a distinct group of children with literacy difficulties who have significant differences from other children (that is, children with a developmental disorder rather than delay), and are there distinct methods of helping such children?

A turning point for research came when Stanovich argued (Stanovich 1991) that the identification of children as dyslexic had been based on unproven assumptions. He went on to show (Stanovich & Seigel 1994) using large samples from previous studies that there were no good grounds for distinguishing between children identified as reading disabled because of low achievement rather than a discrepancy between their IQ and achievement. Fletcher and colleagues (Fletcher et al 1994), using new data, confirmed the that, "although differences between children with impaired reading and children without impaired reading were large,

differences between those children with impaired reading who met IQ-based discrepancy definitions and those who met low reading achievement definitions were small or not significant." Stanovich labelled these two groups "garden variety" and "discrepancy" dyslexics. Both groups had difficulties which were predominantly associated with phonological processing difficulties.

Vellutino, Scanlon and Sipay supported this finding in a longitudinal intervention study, in which children were classified into 6 groups according to how well they responded to remedial interventions based on 1:1 tutoring (4 groups from "Very Good Growth" to "Very Low Growth", with two "normal" reader groups of average and above average IQ) (Vellutino et al 1997). They were able to compare the cognitive abilities of children who responded well to remediation and those who did not: "we found that phonologically based skills – such as phoneme segmentation; rapid naming; and memory for words, sentences and nonsense syllables – reliably and robustly differentiated not only the normal readers from the tutored children who were the most difficult to remediate, but, in many instances, the most difficult and the most readily remediated children as well" (p372). Children who do not respond to remediation are sometimes called "treament resistors" (Torgeson 2001, Blachman 1997). Torgeson has argued that the best response may be to increase the intensity of interventions, or better still to intervene earlier (Torgeson 2001).

However, Fletcher, Morris, Reid Lyon, Stuebing, Sally Shaywitz, Shankweiler, Katz and Bennett Shaywitz have conducted a series of methodologically sophisticated studies using different types of cluster analysis, in which stable subtypes do emerge (Fletcher at al 1997). Fletcher and colleagues raise the important possibility that subtypes may respond differently to different reading interventions. They suggest, for example, that subtypes with verbal short-term memory difficulties may require more drill and repetition than others. They point out that "it is possible that a treatment *on the average* looks ineffective, but actually works well with some subjects and not others." (Fletcher at al 1997, p 112).

2.2.3 Family and School influences on literacy.

Twin studies have shown substantial genetic influence on reading performance, especially on phonemic awareness, non-word-reading and rapid naming (Grigorenko 2001, p108). However, the relative importance of genetic and environmental influences remains unclear, because of the interpretation of sampling issues. Bishop, for example has recently argued the possibility that "the environmental cause is much more common than the genetic cause, but that the genetic cause leads to much more severe and extensive problems, which are likely to attract parental and clinical concern" (Bishop 2001, p 196). Both types of influence are likely in studies of familial effects upon dyslexia. Prospective studies in which parental abilities are identified first are probably the strongest evidence of the relationship: Elbro, Borstrom & Peterson (Elbro et al 1998) showed that

37% of children of dylexic parents had significant reading difficulty after one year of schooling, compared with 5-10% of the Danish population. Across a number of studies, parents' reading disabilities predict between 31% and 62% (Snow et al 1998, p120). Once parental abilities are accounted for, home experience factors are probably only modestly important. Two recent meta-analyses on studies of parents reading to their pre-school children showed average correlations of r=.28 (Bus et al 1995; Scarborough & Dobrich 1994).

Socio-economic status (SES) is measured from some combination of variables such as household income, parents' education and occupation. Typically, a proxy variable such as the percentage of children receiving free school meals in an area is used in educational studies, because direct measurement of variables such as parental income is difficult and time consuming. Generally, there has been agreement from research that reading achievement is correlated with SES. But multi-level modelling has shown that correlations between low SES and reading achievement are stronger at a school level (average correlation size across 93 samples of 0.68) and substantially weaker at an individual level (average correlation size across 174 samples of 0.23) (White 1982). The National Research Council team in the US (Snow et al 1998, p 126) argued that this difference is likely to be a result of the quality of school experiences: children from low SES families will tend to attend schools with similar children, whose schools are less likely to be able to deliver effective education.

In England & Wales, the role of schools in affecting achievement has been highly controversial and at times highly politicised. In the late 80's it had become clear that schools overall achievement levels were influenced by the SES of their pupils, but that there is a consistent "school effect", generally estimated at between 8% to 12% of the variance in pupil achievement (Reynolds 1997). There was also evidence that some schools were able to overcome the disadvantages of lower achieving intake. This led to the creation of a regime of school inspection, testing and school league tables which was intended to put pressure on teachers and Local Education Authorities to improve schools which were "failing". This focus on schools and teaching methods will be considered below in discussing national policy development, and the particular issue of additional interventions in England & Wales will be considered in the third section of this chapter. However, it is important to draw attention to a number of factors at the school level which may affect literacy achievement.

Class size is one of the most apparently straightforward factors that might affect achievement. The STAR project in Tennessee, USA, used an experimental methodology, in which children were randomly assigned to either small (average size 17) or "regular" (average size 23) classes, with and without extra classroom assistants (TAs) (Finn & Achilles 1999). They found clear advantages on attainments of small classes in young children. Peter Blatchford and his colleagues at the Institute of Education in London

built on the work of Finn & Achilles, but used a non-experimental design over three years to try and replicate the STAR project findings and to seek explanations for the classroom processes that might underlie class size effects (Blatchford 2003). They found that there was a clear class size effect on literacy and numeracy in the first (reception) year of English schooling, which was still evident in Year 1 but which had dissipated by the end of Year 2. In the Reception year, there was also a greater effect of class size on less able pupils in literacy, though not in maths.

The study also looked closely at the effects on attainment of TAs. As the STAR project had found, Blatchford concluded "There was no clear evidence for any year, for either literacy or maths, that additional staff or additional adults in the class had an effect on children's progress, and there was no apparent 'compensation effect' [in larger classes] of having extra adults in the class" (Blatchford 2003, p147).

There are of course continuing and broader issues about teacher quality and training and how these factors might affect achievement, and identifying key qualities remains problematic. Blatchford, for example, did not find that headteacher or self-ratings of teacher quality in his sample (of 330 classes) were related to pupil achievements. Teacher quality remains extremely difficult to capture or measure either qualitatively or quantitatively. Wragg et al, for example relied on nominations by Local Authority advisors for their small sample of expert literacy teachers (Wragg et al 1998) although they corroborated these nominations from

achievement data and classroom observation. Wragg and his team were unable to distil a small set of teacher qualities that were essential to effective literacy teaching in the primary school.

By contrast, researchers in the school improvement tradition have focused on whole-school changes, which might be expected to alter the skills and expectations about literacy of all teachers. Perhaps the most notable is the "Success for All" project (eg Slavin et al 1996), in which the school as a whole was taken as the unit within which a number of initiatives were taken with the aim of preventing academic deficits appearing and intervening intensively to deal with any deficits that do appear, with a joint emphasis on strong class teaching and individual tutoring of children by qualified additional teachers, who also take part in group teaching. The project achieved positive effect sizes at schools participating in the project compared with control schools over a number of years, and was taken as a model for the development of more effective literacy teaching by the English National Literacy Strategy, especially in its emphasis on a "literacy hour" and group reading activities within the literacy hour (Beard 1998).

2.2.4 Individual and social factors in the prediction of reading difficulties

The relative importance of individual children, families, and schools (including teacher and method variables) in the prediction of reading difficulties matters because it affects the type of intervention chosen to try to improve reading achievement. The National Research Council concluded that "no single risk factor, on its own, is sufficiently accurate to

be of practical value for predicting reading difficulties. In combination, however, measures of various kinds of risk – individual, familial and demographic – can provide useful estimates of future reading achievement levels" (Snow et al 1998, p 131).

Scarborough in 1998 found that 57% of variance could be accounted for by combining predictions using multiple regression across seven studies (Scarborough 1998). Elbro and Scarborough in 2004 suggest that background factors such as parental income, educational level and occupation are not strongly associated with children's later reading achievement, at an individual level. The recent focus has been on how parents influence their children's reading achievement, but its effects are no more than modest: parental reading ability, which may have a genetic link with children's achievement, is the strongest factor, but then ".. probably no more than 8% of the variance in early literacy skills is directly accounted for by differences in home literacy environments" (Elbro & Scarborough 2004, p343). Overall, differences between individual children on school entry, especially in language related abilities, are considered by current research to be the most important predictors of children's reading difficulties.

The importance of differences in children's qualities after some years of schooling are much less well understood. In his pivotal article on Matthew Effects, Stanovich argued the number of causes of reading difficulties

might be less than commonly supposed, because of the effects of success and failure in reading over time (Stanovich 1986): He argued that (p381):

> "The effect of reading volume on vocabulary growth, combined with large skill differences in reading volume, could mean that a 'rich-get richer' or cumulative advantage phenomenon is almost inextricably embedded within the developmental course of reading progress. The very children who are reading well and who have good vocabularies will read more, learn more word meanings, and hence read even better. Children with inadequate vocabularies – who read slowly and without enjoyment – read less, and as a result have slower development of vocabulary knowledge, which inhibits further growth in reading ability."

He also suggested that reading failure may itself have effects upon a broader range of cognitive and motivational processes, perhaps from around the middle primary stage (if not before), and that this process of "reciprocal causation" may help to explain the wide range of correlations between reading difficulty and cognitive processes.

This suggests that some important relationships at later stages of literacy development may not be linear (Elbro & Scarborough 2004) and that we should expect the analysis of within-child explanations for reading difficulties in the middle years of development to be complex, perhaps needing different kinds of explanation. This is perhaps where the effects of

interventions begin to be as or more important than qualities which children bring to their learning.

2.3 National & local policy development on literacy difficulties in England and Wales since the early 1990s

2.3.1 The recent policy context

Looking back to the early 90's, it is clear there were a number of policy options to deal with "literacy difficulties" (if the problems to be addressed were conceived at the individual or small group level) or "raising standards in literacy" if the problems were conceived at population level. These options were something like this, placing them in rough order of magnitude:

- Reorganize the English National Curriculum
- Change the ways in which mainstream teaching of literacy is practised
- Positively discriminate towards children from socially and linguistically disadvantaged backgrounds
- Intervene systematically at the Foundation Stage for children "at risk" of literacy difficulties
- Intervene systematically with children who have fallen behind in the initial year of teaching
- Intervene intensively with children with severe difficulties in literacy learning.

Although there remained some scope for LEAs to develop local teaching initiatives (such as the "Catch Up Project" in Norfolk, Cornwall and

Wakefield) and to set up local volunteer or parental involvement projects (such as "Better Reading Partnerships" in Durham, Redcar, Cleveland and Worcestershire), decisions about most options were largely at the national government level.

In the early 1990's there was much political enthusiasm for setting up "Reading Recovery" as a national "second wave" system. The "Three Waves" model was originally put forward by Marie Clay in New Zealand (Clay 1985), and was intended as a framework for intervention with all children, most of whom would make sufficient progress from ordinary class teaching, some from individual interventions using Reading Recovery, and a small minority, who did not benefit sufficiently from Reading Recovery, from longer term individual interventions. But the incoming Labour Government in 1997 chose instead to pursue a reform of mainstream teaching through the National Literacy Strategy (NLS). With the benefit of hindsight, it seems that the successive governments of England & Wales began with the most far-reaching options, and have gradually narrowed their focus.

Sir Ron Dearing's review of the National Curriculum in 1994 (Dearing 1994) produced a somewhat simpler and more manageable version of the English curriculum in 1995 (DfE 1995). A pilot National Literacy Project was set up in 1996 by the government to give direct advice on teaching literacy to mainstream teachers (Sainsbury 1998), and this was adopted in 1998 as The National Literacy Strategy (NLS). There was thus in England

a clear decision that the first option, curriculum revision, was insufficient and that the greatest priority was to alter mainstream teaching practice.

The National Literacy Strategy has occupied centre stage for several years, and has brought about significant improvements. Recent evaluations show (OFSTED 2002, OISE 2003) a plateau was reached in overall national results in 2001-2 on which we remain.

The government has introduced substantial positive discrimination towards children experiencing social disadvantage through the "Sure Start", "Excellence in Cities" and "Education Action Zones" initiatives. Such interventions are necessarily long-term and link with the government's broader strategies for the elimination of child poverty (OFSTED 2003)

In 1999 the government began to introduce new materials to help children who had not made a good start, beginning with Additional Literacy Support (ALS) at KS2 (DfES 1999), and extending this to Early Literacy Support (ELS) at KS1 and then Further Literacy Support (FLS) at KS3. These initiatives effectively took the place of intensive 1:1 interventions such as Reading Recovery, which was the original and perhaps the leading model for remedial intervention in the 80's and early 90's. Reading Recovery requires the use of individual teachers who have been trained for a year in its delivery, and is therefore expensive and slow to introduce. By the mid 90's, evaluations of its effectiveness had cast doubts on its ability to deliver sustained benefits. Shanahan and Barr (Shanahan & Barr 1995)

for example, had found that children successfully completing Reading Recovery programmes did not maintain their gains much better than similar children who had not followed the programme. At the same time evaluations of interventions involving trained and supported volunteers or teaching assistants (such as the American "Book Buddies"), which were heavily influenced in the type of curriculum and delivery methods by Reading Recovery, suggested that good progress could be made for much lower costs.

No thorough evaluations of the ALS/ELS/FLS approach have yet appeared. OFSTED's evaluation of the NLS (OFSTED 2002) has commented positively on these interventions, but it remains important to clarify what benefits they bring, and whether they are the best methods for "second wave" intervention.

Changes have been made at broader levels of intervention (the national curriculum, national teaching methods, social disadvantage, and initial interventions). Not all possible improvement has been exhausted at these broader levels, and some possible approaches have not been attempted, especially preventative intervention at the Foundation stage. Further improvement is expected in the development of methods of mainstream teaching, especially the use of phonics.

It is encouraging that improvement is now being sought on different fronts at the same time, but there is a danger that, as the single focus provided

by the NLS becomes only one of a number of policy initiatives, it will become harder to identify what has and has not worked.

2.3.2 The Three Waves model:

The "Three Waves" model is summarized in "Targeting Support" (DfES 2003):

> "Wave 1 The effective inclusion of all children in a daily and high quality literacy hour
>
> Wave 2 Additional small-group intervention for children who can be expected to catch up with their peers as a result of the intervention
>
> Wave 3 Specific targeted approaches for children as requiring SEN support (on *School Action, School Action Plus* or with a Statement of special educational needs)"

The three waves model derives from the work of Marie Clay, who argued that slow and reluctant identification of children with literacy difficulties as dyslexic or socially disadvantaged, as they progressed haltingly through the school system, would inevitably be ineffective and inadequate (Clay 1985). She argued that all children who had fallen behind in the first year of mainstream teaching should receive daily individually delivered teaching help from specially trained teachers using Reading Recovery methods, until they reached the attainments of the class average. The small minority who did not reach the average level during the second year would continue to need individual help. This effectively divided children into three groups or "waves": the first wave consists of children who are able to

benefit from regular mainstream teaching, the second of those who recover to average after additional help, and the third of those who need prolonged help. It has subsequently been suggested that perhaps the second wave consists mainly of children who are disadvantaged socially, while the third group includes those with more severe difficulties.

Although the government of England & Wales has chosen not to use Marie Clay's Reading Recovery (because of doubts about its longer-term cost effectiveness), a modified wave model has been adopted in its policy for dealing with children with literacy difficulties (Beard 1998, DfES 2003). The English and Welsh "Code of Practice for Special Educational Needs" (COP) (DfES 2001) includes a "staged approach to assessment" and is highly congruent with the three waves model.

The three waves model is based on assumptions which may be questionable, and which are relevant to this research.

First, some children, it assumes, will not learn to read and write adequately from mainstream class teaching alone. Roger Beard suggested there would be around 20% of children who needed second wave teaching (Beard 1998), and analysis of SATs results suggests that 20% needing additional help at the end of KS2 might not be far wrong. But if we assume that a proportion of children will inevitably fail, then we reduce the pressure to improve mainstream teaching - results from Solity using alternative methods of teaching early literacy in mainstream classrooms

suggest that much lower percentages can be achieved (Solity et al 1999). It is possible that mainstream teaching is capable of reducing the numbers of failing children significantly more. And screening and intervention for difficulties in phonological awareness (PA) at the Foundation Stage might begin to reduce the number of children who did not benefit sufficiently from "first wave" teaching.

A second assumption, that the majority of "failing" children will "catch up" as a result of additional wave 2 interventions, is the basis for one of the questions of the present research. Most teachers would expect that many children will actually need remedial small group intervention over a number of years using ALS in years 3 to 6 and FLS at Key Stage 3. Although there has been development in the effectiveness of these packages, there has been no serious re-evaluation of their overall cost benefit. It is not clear whether prescribed teaching packages can be as effective as individually administered programmes nor whether the intensity of such packages is sufficient to achieve "catch up" in sufficient numbers of cases.

The three waves model seems to assume that waves 2 and 3 are clearly distinguishable. Children in wave 2 are those who have "fallen behind", those at wave 3 are those with SEN. The former can benefit from small group interventions, the latter only from individually delivered help. Is this distinction made in practice? How are children assigned to different types of intervention at KS2?

The Three Waves model is a pragmatic approach intended to target additional help cost effectively, but it is important not to assume indefinitely that any of the waves represent the only and best solutions possible, or that they must inevitably strike the shore in the order the model prescribes.

2.4 Interventions to help children with literacy difficulties

2.4.1 The recent American Approach

The volume of research on reading is such that for some years now synthesis has become as important or even more important than original research. Perhaps the most important synthetic study in the early 90's was Marilyn Adams' Beginning to Read: Thinking and Learning about Print" (Adams 1990). 10 years later Adams was one of a panel of eminent researchers who contributed to "Preventing Reading Difficulties in Young Children" (Snow et al, eds, 1998) by the "Committee on the Prevention of Reading Difficulties in Young Children" of the US National Research Council (NRC). The NRC study led directly to a series of more detailed research syntheses funded by the US Congress which were published as the "National Reading Panel" (NRP) (National Reading Panel 2000) as an overall summary report and as reports from 5 linked subgroups, reporting on alphabetics (Phonemic Awareness and phonics), fluency, comprehension, teacher education and computers and reading.

Although these studies are not without controversy and problems of interpretation, they do provide a new level of coherence to research on

reading instruction. Their main focus is mainstream teaching, but there are some implications for literacy difficulties.

The United States' federal structure contrasts with the unified system of educational policy direction in England and Wales, where it has been possible to change the ways in which literacy is taught through the NLS, which has been almost mandatory in all English and Welsh primary schools. In the US research has been asked to show the way before states can be asked to consider changing the ways in which reading should be taught. In England and Wales researchers complain that policy constantly goes ahead of research, and research is therefore not able to play the part it might if we genuinely want "evidence based policy development". In the US, research on reading precedes and encourages reform of teaching methods, but cannot dictate change in the way that the government of England and Wales has done. As a result the emphasis on scientific rigour has perhaps been greater in the US (see Reid Lyon et al 2004).

2.4.2 Synthetic study of interventions for literacy difficulties in England & Wales: the aims and scope of Brooks' study

Brooks, Flanagan, Henkhuzens and Hutchison reviewed published interventions for children with literacy difficulties in England in 1998 (Brooks et al 1998), and Greg Brooks updated the review as "What Works for Children with Literacy Difficulties? The Effectiveness of Intervention Schemes" (Brooks 2002). Brooks deliberately chose to review only

intervention studies carried out in England & Wales, to avoid criticism that studies from other contexts might not be applicable in Britain. The main aim was to evaluate the effectiveness of intervention schemes using quantitative methods. This is the only modern study attempting to compare literacy interventions in England and Wales, and has been used almost directly to advise teachers about the most effective interventions at Wave 3. It is therefore important to consider whether its methodology and conclusions are adequate.

Brooks selected interventions from which effect size or ratio gain measures could be taken, and noted features of the study design which could contribute towards an analysis of the quality of research design. His main results are therefore a list of effect sizes and ratio gains for reading and spelling outcomes for the set of 25 studies.

A wide range of types of intervention are covered:

- training volunteers: Time for Reading, Better Reading Partnerships, Family Literacy and the Knowsley Reading Project
- paired work (either adult – child or child – child): Cued Spelling, Paired Reading and Paired Writing
- using ICT: AcceleRead AcceleWrite, Integrated Learning Systems, and RITA (Reader's Intelligent Teaching assistant, which is closely related to the Interactive Assessment and Teaching scheme
- comprehension training in small groups: Inference Training and Reciprocal Teaching

- individually delivered catch-up teaching: Catch-Up, Individual Spelling, Phonographix, Phonological Awareness Training, Reading Intervention (the Cumbria scheme also known as Sound Linkage), and Reading Recovery
- group delivered catch-up teaching: Interactive Assessment and Teaching, Multi-Sensory Teaching System for Reading, Somerset counselling with DISTAR, SPELLIT and THRASS

The size and number of studies of the schemes varies substantially, from small single studies, such as Reciprocal Teaching, to those studied several times using a large number of participants, such as Better Reading Partnerships.

2.4.3 Critique of Brooks' methods

Brooks draws attention to but does not question the current working assumptions about the Three Waves. He says that most of the interventions reviewed are intended for children with "moderate" literacy difficulties. From the primary research reports, participants were classified as either "having SEN", "low attainments" or "mixed ability". The majority appear to belong to the low attainers group, with about a quarter "mixed ability", and only two with SEN. It is difficult to tell even from the quite detailed reviews provided of each scheme how great were the literacy difficulties of the children upon whom the interventions were demonstrated. It is likely that most if not all of the children would be classified as wave 2, rather then wave 3. In only two cases were children

with SEN included (PAT and the Cumbria "Reading Intervention"), and in neither were effect sizes large enough to be worth testing for significance. Most importantly, the studies reported took place over more than 2 decades. The majority were done before the NLS began, and none involved children who had been involved in previous small group interventions. The three waves model did not therefore apply at the time of the studies reported, and in no cases were interventions a third rather than a second wave of intervention.

Brooks uses two outcome measures, effect size and ratio gain. Effect size is a measure of the difference between means and standard deviations of outcomes between experimental and control groups. Although these differences may be statistically significant, they may yet be so small as to have little effect. Effect sizes below 0.2 are weak. Those above 0.4 are moderate. Those above 1.0 are strong. Effect size is the most reputable measure and can be used to compare between studies testing different interventions. The US National Reading Panel (NRP) studies used effect size comparisons to make a clear and detailed analysis of PA interventions (Ehri et al 2001).

Ratio gain measures progress on a particular reading or spelling test over time. A ratio gain of 1.0 means, for example, that 10 months progress on a test has been achieved over 10 actual months, and 0.5 means 6 months progress has occurred over 12 months. There are two main problems with ratio gain: (1) a dramatic gain over a short time is indistinguishable from a

larger gain over a longer time, so for example 3 months gain in a one month intervention (ratio gain = 3.0) looks better than a 12 months gain in 12 months (ratio gain = 1.0). A further standardisation of ratio gain is needed, so that gains over the same interval are compared. (2) gains on one test might not be equivalent to gains on another, because reading and spelling tests are not straightforwardly comparable. Frequent revisions and comparisons between tests are needed. But restandardization of tests is the exception rather than the rule.

It is very surprising to note that in a very recent DfES summary of Brooks' study (DfES 2003) it is described as providing "an overview of the evidence on the impact of different types of Wave 3 intervention." It goes on to quote Brooks' conclusion that "ratio gains of at least 2....were achievable". This is seriously misleading as applied to children with identified SEN (wave 3). Brooks' review only considers data on children some of whom could perhaps be described as at wave 2, with none at wave 3.

2.4.4 Conclusions from Brooks about Intervention Schemes in England on which this study provides evidence:

1. **"The evidence on ordinary teaching therefore proves the need for early intervention schemes; in general, ordinary teaching does not enable children with literacy difficulties to catch up." (2.3, p 11)** Brooks draws this conclusion from the fact that most control groups were reported to be making normal progress, while most schemes

gave ratio gains of between 1.4 (Catch Up Project) and 8.3 (PhonoGraphix). In other words, Brooks is suggesting that children will actually benefit more from remedial intervention than from standard class teaching. This is a surprising conclusion from the evidence presented. The studies he reviews are very mixed and alternative intervention groups were available in only 10 of 27 schemes. Not all studies involved control (no intervention) groups, and those that did included some where normal progress was assumed rather than measured. It is therefore difficult to rule out Hawthorne effects, as NRP was able to do in the face of Troia's criticisms (Ehri et al 2001). It is also difficult to be sure that there was comparability between children selected for intervention and those in the control groups. The highest ratio gains were reported for relatively short duration studies (eg gain of 16.1 – 8.3 for Acceleread-Accelewrite over 4 weeks, 8.3 – 2.2 for Phonographics over 12 to 26 weeks, 5.0 –3.0 for Inference Training over 4 weeks), which raises doubt that gains over a standard time interval would be as impressive.

Brooks explicitly ruled out of this reanalysis schemes which sought improvements for all children by improving mainstream class teaching, such as Early Reading Research (Solity et al 1999). Brooks appears to be speaking here to government policy makers rather than researchers: he wants to support additional literacy interventions. The evidence from his analysis does not really justify any conclusion about

the relative effectiveness of mainstream teaching and additional interventions.

2. **"For greatest impact with children who struggle with spelling, highly structured schemes work best." (2.5, p12)** It is difficult to see what is meant by "highly structured" here. There does not seem to be a comparison between more and less structured schemes available, from the schemes included. There was a comparison between schemes which focused exclusively on spelling (Cued Spelling and Individual Spelling) and combined reading and spelling schemes, but neither ratio gain nor effect size data appear in Brooks' tables for Individual Spelling, so it is difficult to make it. This conclusion needs to be clarified.

3. **"Given the financial investment that all ICT approaches require, technology used to boost literacy attainment deserves to be targeted as precisely as possible" (2.8, p14)**
   Brooks bases this conclusion on only two very different schemes, the small scale AcceleRead AcceleWrite, and the much larger ILS. The conclusion he draws is not justified by the evidence. The NRP found few experimental studies of the use of computers in reading instruction, and was only able to conclude that the use of speech synthesis, word processing and hypertext all seemed "promising". It wanted to see more research generally and especially into the use of the internet and

speech recognition. It did not examine Brooks' hypothesis that the use if ICT needs to be very closely targeted.

4. **"Large-scale schemes, though expensive, can give good value for money." (2.9, p14)**

   In general, Brooks argues that some of the schemes he reviews (Family Literacy, Phonographix, Reading Intervention and Reading Recovery) were expensive but they also produced good results. In the longer-term, he points out that intervention to remediate reading difficulties may be less expensive for society than consequent social dysfunction and employment limitations. The crucial argument, however, is whether more expensive schemes produce gains which cheaper but ones do not. Thus, is Reading Recovery markedly better in the longer-term than Better Reading Partnerships or ALS? There is a lack of evidence of longer-term gains in Brooks' synthesis.

5. **"Therefore success with some children with the most severe problems is elusive, and this reinforces the need for skilled intensive, one-to-one interventions with these children. (2.11, p16)**

   Many researchers have argued (eg Torgeson 2001) that for "treatment resistors" (who are quintessentially at "wave 3") better training and more intensive interventions are likely to be required, hence greater expenditure. But there remains a case for continuing to use less individualised and more straightforward interventions as a first line of attack on such children's difficulties until we have reliable methods of

assessing which children have more severe difficulties, and therefore whether we can use more expensive interventions earlier.

5. **"Interventions longer than one term do not necessarily produce proportionately greater benefits."(2.12, p16)**This conclusion is based on the finding that there were more moderate to weak outcomes for longer term than for shorter-term interventions. The NRP study of PA training was the only subgroup to report on length of intervention, and they similarly found evidence that medium term duration might be optimal.

6. **"Good impact – sufficient to double the standard rate of progress – can therefore be achieved, and it is reasonable to expect it." (2.13, p17)** This is a very optimistic conclusion based on evidence that intervention schemes typically report ratio gains of 2.0 or more or effect sizes of 0.5 or more. The crucial question here is about which children were involved in the primary studies. Brooks' report says it was mostly "low attainers". This means they were almost certainly children in wave two. We have no grounds to expect children with more severe difficulties, those in wave three, to achieve ratio gains of 2.0 from Brooks' evidence.

2.4.5. What can be concluded about intervention schemes?

The greatest concern stems not from Brooks' own conclusions but from the use that has been made of his analysis by the DfES. His review does

not cover wave 3 interventions (except in the sense that they are mostly targeted at individuals, rather than groups), and yet it has been presented as a guide to choosing which wave 3 intervention to use. The DfES' rationale for selection of 15 out of 24 schemes is also puzzling: The DfES says in "Targeting Support" that the criterion was availability, but 2 of the omitted schemes are available through DfES Research Reports ("Individual Spelling" and "SPELLIT"), PAT has been readily available from Buckinghamshire LEA for several years, and Integrated Learning Systems (ILS) are very widely available (from most Educational supplies catalogues). Perhaps PAT and ILS should *not* be recommended on the current available evidence, but that is not what DfES says.

Questioning some of Brooks' conclusions is not intended to diminish the need for the sort of synthetic study he presents, nor to doubt the importance of the questions he attempts to answer. Particularly in the context of England and Wales, it continues to be vital to clarify how interventions relate to the National Literacy Strategy, or whatever alternative literacy curriculum is followed in the relevant mainstream classrooms, because limited time to carry out interventions is a crucial issue. New intervention materials, such as Additional Literacy Support, also need to be evaluated, alongside older materials.

The government of England & Wales has very recently published a new "Strategy for SEN" called "Removing Barriers to Achievement" which

comments particularly on how research on SEN and achievement might be improved:

> "3.6 There is a wealth of research evidence surrounding children with SEN and strategies for raising achievement. However, much of this research is short-term and small scale, giving rise to disputes about its validity, and mainly based on the learning of younger pupils. There is a lack of research evidence about the learning of pupils with SEN at secondary level. We are determined to learn from the available evidence, commissioning new research where there are gaps in our knowledge.
>
> 3.7 We have already begun to develop a framework of evidence-based strategies and effective teaching approaches for pupils with particular needs, across all age groups. The Universities of Cambridge and Manchester have recently completed a major scoping study to inform this. We will work collaboratively with the national strategies, building on the themes emerging from the scoping study, to further develop the knowledge base and capacity of schools to improve the quality of teaching and learning of children with SEN." (DfES 2004, p54)

These initiatives suggest that there will be an increasingly focused and coordinated national approach to research on teaching methods and outcomes for children across the range of SEN, including children with literacy difficulties. The "framework of evidence-based strategies and effective teaching approaches" sounds very valuable. It will be interesting to see whether it might include "standards" for evaluation studies which

would enable future evaluations following Brooks' lead to reach more reliable and precise conclusions than Brooks could be expected to have reached.

3. Methodology

3.1 Implications from the literature for research on literacy interventions:

In 2003, when this research began, most schools had put in place ways of working with parents on their children's literacy, and nearly all schools had incorporated the main school-based intervention, the National Literacy Strategy (NLS), into their teaching programmes, during the previous 4 years. English schools had used a wide variety of methods of helping children with literacy difficulties, often with strong local preferences for particular methods and materials. In 1999, two intervention packages intended to be used at Wave 2 were introduced, Early Literacy Support (ELS) and Additional Literacy Support (ALS), at Key Stage 1 and 2 respectively, which were intended not only to be consistent with the NLS but to be based on recent research about children's difficulties in learning to read and spell (DfES 1999).

In the previous chapter, it was argued that there is an implicit hierarchy of interventions, the scope of which vary from the national to the individual level. Although further change at the most general levels, National Curriculum and class teaching methods, should not be ruled out, it seemed unlikely in 2003 that further substantial change at these most general levels would be contemplated for a while.

It was therefore logical and apposite to consider the role of additional interventions within schools. It was not practical as far as this research was

concerned to study the effectiveness of interventions across a wide age range. The early Key Stage 2 was chosen because the relevant intervention, ALS, was the first to be produced, and there had been time for its use to become established in many schools. Although children in Years 1 and 2, for whom ELS was developed, are perhaps at a more critical stage of development, it seemed likely that more schools would have arranged additional interventions for the Y3-4 group, and most crucially, it was more likely that schools would be prepared to contribute data to a study in the early part of KS2 than during KS1 because of the pressure on schools to achieve good results in Year 2 SATs.

The design of the ALS materials was based on the main research findings of the previous 15 years (DfES 1999), for daily or near daily small group delivery, mainly by a Teaching Assistant with some input from the class teacher. The lessons were planned in detail, with little or no scope to vary their content or pace of delivery. There were guidelines, involving very short assessment tasks, to determine the most appropriate children to be taught using ALS. In general, ALS was expected to be most helpful for children who had fallen below expected levels on SATs but whose difficulties were likely to mild. It was not expected to be most appropriate for children with severe literacy difficulties (DfES 1999). In terms of the 3 waves model, ALS was clearly wave 2.

3.2 Research Aims

In the previous chapter, the main focus upon literacy intervention research in England and Wales has been the synthetic study of Brooks (Brooks 2002) which compared studies of 24 different interventions. It was suggested that Brooks' conclusions were not always as clear and well grounded as was claimed, and that to a considerable extent this was a result of a number of key weaknesses in the data itself, coming as it did from a wide variety of different sources. These included:

- Lack of clarity about the qualities of the participants (children) about whom the research was carried out (eg the main classification was SEN vs low attainments vs mixed ability)
- Lack of data about mainstream teaching experiences of the participants, while making strong claims about the role of intervention compared with class teaching;
- Lack of data about previous interventions with the participants, such that the subsequent interpretation of these studies as Wave 3 is misleading;
- Use of ratio measures of reading and spelling gains where the period over which gains were compared were very variable and some times quite short;
- Claims about value for money from large scale schemes without evidence on costs;

- Claims about reasonable expectations of ratio gains of 2.0 or higher from intervention studies, in the absence of clarity about the qualities of the participants – this is one of a number of conclusions which seem to go beyond the evidence actually available.

Although it had some weaknesses, Brooks' study was seen as an invaluable first step in the development of a framework for evaluative research into literacy interventions which was previewed in the government's strategy for SEN, "Removing Barriers to Achievement" (DfES 2004). The present study was intended to try to overcome at least some of the weaknesses identified from the analysis of Brooks' work, and more generally to explore the possibility that data routinely gathered by schools about the progress of their children might be supplemented without extreme effort from teachers and TAs to the point where the "everyday" interventions used in schools might be measured and analysed alongside data from interventions set up by researchers which are typically intended to break new ground, and often depend on the involvement of researchers from outside the school.

A number of the Brooks studies, most notably "Better Reading Partnerships", were "everyday" interventions about which data was collected in ways which enabled results to be published. But in general it has always been a striking and very surprising anomaly that Local Education Authorities and schools typically collected very detailed data on individual children in pupil files which they were unable to analyse because of both the complexity and

heterogeneity of the data itself. With the advent of modern computer databases, it would seem possible that this vast reservoir of information might be unlocked. The government and especially OFSTED are keen for schools to use data about pupil achievements more effectively in order to further raise achievement. Unfortunately, this has led to a serious increase in pressure for information upon schools, which since around 2003 the government has recognised and sought to limit (DfEE 2001, IRU 2005). New requests to schools to provide data about pupils has to be very carefully scrutinised by Local Authorities, with most requests turned down. A background aim of this research was, therefore, to test whether very modest requests for additional information about pupil qualities and how interventions were delivered might be easy enough to gather and productive enough when analysed to justify the Local Authority requesting other or even all its schools supplying data about pupils in similar ways.

The main aims of this study were therefore:

- To evaluate the progress in literacy of children identified as having difficulties and requiring additional help in Years 3 and 4 in mainstream schools;
- To compare the progress of children receiving additional help with those receiving only mainstream class teaching, and to try to evaluate the contributions of each to progress;
- To compare the progress made by children using the "Additional Literacy Support" (ALS) materials with those using other materials;

- To investigate whether differences in children's qualities interact with differences in teaching materials and methods.
- To investigate whether a number of factors, such as the ordinary teaching of literacy in the classroom to all the children, the amount of additional help, where help was delivered (in or out of class), the size of groups and the qualifications and experience of those who deliver additional help make a difference.
- To investigate whether particular materials or combinations of materials led to greater progress;
- To consider any implications for national educational policy for children with literacy difficulties;
- To consider whether limited additional data about pupils and intervention programmes could contribute to a rigorous and effective evaluation of "everyday" literacy interventions in schools;

Progress in literacy can be evaluated in a variety of ways. On the one hand, the reader's satisfaction, enjoyment of what is read and grasp of new meaning may mark vital aspects of the quality of learning experiences. On the other hand, the number of words read or spelled, or the number of questions successfully answered, may show the extent to which important skills have been learned. In this study, the amount of progress achieved by children was considered a better measure of the effectiveness of the teaching materials than the quality of their experiences. The main outcome variables were

therefore reading and spelling ages and National Curriculum levels. Some information was also collected about children's enjoyment of their learning in relation to particular materials.

The ALS materials were the first type of materials intended by the National Literacy Strategy implementation team to help children who had fallen behind at Key Stage 1. In the year in which they were introduced (1999), funding was provided to train LS staff to all schools in the use of ALS materials. Many schools trained staff and began to use the new materials. At the same time, many schools continued to use existing materials or other new materials intended to do the same job. Some schools used ALS exclusively or predominantly, others did not use the materials and some used a mixture.

The National Literacy Strategy (NLS) implementation team put ALS forward as the most effective way to help children who had fallen behind at Key Stage 1. The central hypothesis to be tested by this research was that ALS materials help children to make more progress in reading and writing than other types of help.

The research set out to evaluate actual practice, not the pure and original intentions of the NLS. By 2002, some schools had begun to modify their use of ALS so that it was not delivered as originally intended by the NLS. For example, it was no longer common for the class teacher to lead one of the

ALS teaching sessions. LS staff normally lead all sessions. It was assumed that changes made by schools would serve to make ALS more useful and helpful to children.

3.3 Research Design:

A quasi-experimental design was therefore adopted. The essential features of a quasi-experimental design are (a) the comparison of changes in the dependent variables as a result of different interventions upon two or more groups of participants, and (b) the inability of the researcher to assign participants randomly to intervention groups (Campbell & Stanley 1966). The main interventions were ALS, a range of other materials and methods for helping children with literacy difficulties, such as Phonological Approaches to Teaching (PAT) and Speaking Starspell, and a combination of ALS and other materials. The children were all those identified by their schools in Years 3 and 4 as requiring additional help, other than differentiation within their normal literacy lessons. Schools were asked to record the children's attainments in reading and spelling at the beginning and end of the year. It was not considered possible or desirable to ask teachers to assign children randomly to ALS, other interventions or a combination. Some schools had taken the decision that ALS was the most suitable intervention for all children, some that it was the most suitable intervention for some children only, and some that other interventions were always preferable. This research was

intended to look at the consequences of those decisions for children's reading and spelling progress.

An experimental design would have required children to be randomly assigned to either ALS, non-ALS or a combination of interventions. This would have meant a loss of control by teachers over the teaching methods used with their children. The teachers and non teaching assistants (TA's) would have required extensive pre-training to ensure fidelity to standardised teaching protocols for ALS and the other interventions. Monitoring and support throughout the year would also have been required, perhaps quite extensively, to ensure that methods continued to be delivered in accordance with protocols, in the face of doubts about their effectiveness or appropriateness to the students concerned. This would have created highly untypical contexts and instructor motivations. Paradoxically, in order to try to ensure sufficient fidelity to prescribed methods, high levels of experimenter monitoring and (perhaps) control would have threatened to reduce validity. By using an explicitly quasi-experimental design, it was possible to maximise teacher motivation and effectiveness, by allowing teachers to retain full control over the matching of teaching methods to children. It has been suggested that in any comparison between teaching methods, teaching should ideally be conducted by those who are in favour of the methods they are using (Troia 1999), perhaps because commitment is expected to be such an important component of motivation and thus effectiveness.

A true experimental design would also have been much more expensive than a quasi-experimental design, because extensive training and monitoring of existing school staff, or the introduction into schools of specially trained teachers or TA's would have been required. Such interventions would have generated a number of problems for both existing and experimental staff of settling in and familiarity with the children, and although longer interventions would be expected to be more reliable, they would have also been steadily more expensive. To obtain results from 8 schools and 16 classes, as this research did, would have required between £250,000 and £500,000 in salaries, because it would have required simultaneous intervention in each of the classrooms on a near daily basis. In fact, the use of a quasi-experimental design was achieved through teachers' interest and goodwill, with some small ex-gratia payments for the time required to complete the data collection forms.

Fidelity in the use of materials was not therefore a variable which it was considered practical to monitor extensively in this research. Teachers and TA's were asked to evaluate the materials they used through brief questionnaires and open-ended comments.

The independent variable which was considered most critical to the comparability of groups in this research was individual differences between

children. ALS was designed for those children who had fallen behind because of socio-cultural disadvantage. Other interventions than ALS were expected to be used with children with more severe SEN, some of which were used by the teachers and TA's in this research, with some overlap.

Unfortunately no clear and unambiguous method of measuring the extent of SEN is available. There is as yet no firm and reasonably short list of tests which can be administered to 7 to 8 years olds to ensure that individual differences are fully and reliably captured. Testing all children at the beginning of the school year would also be both intrusive and expensive, and would have contradicted the first aim (above) of asking for a minimum of extra data which schools could gather without substantial extra effort. Instead, the approach adopted was to ask teachers to rate children's individual difficulties on 14 X 3 point scales. It was then possible to investigate whether there were significant differences between those chosen for different interventions, and to then use statistical methods to compensate for and balance any differences found.

A quasi-experimental design was therefore considered to be not only the most cost-effective method of testing differences in effectiveness between ALS and other interventions for children with literacy difficulties, but also to have advantages in validity over a true experimental design. All data was collected by teachers and TA's, using forms prescribed by the researcher,

and teacher control over the methods of intervention was not affected. Children themselves would have almost no awareness that they were participating in a research project, hence they would have been unlikely to be affected by Hawthorne effects. Teachers themselves were aware of their participation, but this was confined to the period before the beginning of the year and at the end of the year when data was collected and sent to the researcher. During most of the experimental period, they were free to concentrate on their primary task, of attempting to help children progress in literacy.

3.4 Key Variables:

3.4.1 Reading and spelling levels.

Almost all primary schools are now routinely testing children at least once a year in reading and spelling. From a series of preliminary visits, it was clear that schools were used to collecting measurements of children's progress using tests which they had become familiar with, and it would therefore be an extra and potentially burdensome activity to request that they collected a further set of test data. The majority of schools in Northamptonshire use the original version of the Salford Sentence Reading test (Bookbinder 1976) to measure individual progress. This is an individually administered test in three equivalent forms. Children are asked to read a set of sentences from a card, until six errors are made. The child's score is the point at which the sixth error

is made and normative tables are available to translate this into a reading age.

Although a revised version of the test is available (Bookbinder et al 2000), most Northamptonshire schools continue to use the older version. Similarly, one school used only the Suffolk Reading test (Suffolk CC & Hagley 1987). Use of the new version would have prevented some year on year comparisons, and the difference between test administrations was more important than absolute levels of reading attainment. It was decided not to request the use of the new version of the Salford test. In two schools only Qualification & Curriculum Authority (QCA) Tests were used. The QCA tests report results in terms of raw score, a standard score and a National Curriculum (NC) level (from Working below Level 1, Level 1, 2c, 2b, 2a, 3c, 3b, 3a, to 4, 9 points in all). Unfortunately only NC levels were consistently available from the previous year. Schools were expected to make comparisons on the basis of NC levels, and closely monitor changes in levels year to year.

Some data was available from children who had been tested on both the Salford and QCA tests. It was therefore possible to investigate whether reliable comparisons were possible between these measures, and therefore whether it would be necessary to make separate comparisons of children

whose progress was measured using NC levels or whether the results could safely be combined into a single set of comparisons.

The measurement of spelling progress also involved some unexpected differences in practices between schools. Most schools used the Vernon test, which can be administered either as a group or individual test (Vernon 1977). One school used the Blackwell spelling test, which is associated with Blackwell materials for the teaching of spelling (Sadler & Page 1975). The test gives spelling age scores based on children's performance on a sample of words dictated individually, in the same way as the Vernon test. It was decided that spelling age scores could be used to make comparisons between children, even though the actual levels might not be straightforwardly comparable.

The main dependent variables, progress in reading and spelling, were thus measured in either the difference in months between scores at the beginning and end of the observation period, or in differences in NC levels using a 9 point scale (from below Level 1 to Level 4). The period of observation from the Salford, Suffolk, Vernon and Blackwell tests was not consistent. Some schools relied on test results from the previous July as their baseline measure, while others retested in September. Final measurements were sometimes made in June and sometimes July. It was therefore necessary to correct all age differences to provide a single measurement interval. One year

was chosen as the most straightforward standard time interval, and results from different time intervals were multiplied by 12 and divided by the actual interval between tests used to obtain a comparable measure. QCA tests are only used on an annual schedule and therefore there were no differences in time period for which adjustment might be needed.

3.4.2 Qualities of children:

Review of the literature on the relationship between pupil qualities and methods of teaching reading and spelling did not find reliable interactions. However, the NRP groups showed that phonics instruction was not equally effective for all groups of children:

Table 3.1 Effect sizes for phonics instruction for children in different categories

Category of learner	Effect size (d)
At risk kindergarteners	0.58 significant at $p<0.05$
At risk 1st graders	0.74 significant at $p<0.05$
"reading disabled" children	0.32 significant at $p<0.05$
Low achieving children 2nd to 6th grade	0.15 not significant

"Reading disabled" means children whose reading attainment is significantly below the levels expected from their cognitive abilities (from intelligence testing), while "low achieving" means children whose attainments are below

average but in line with expectations from cognitive testing. The NRP Phonic Instruction subgroup did not feel able to draw a clear conclusion from these results, but considered further research was needed (NRP 2000).

Brooks used mixed ability vs low attainment vs SEN, but had insufficient results to analyse the effectiveness of interventions by pupil types.

In spite of the lack of evidence for reliable pupil x method interactions, it remains difficult to be confident that pupil qualities might not make a difference. A previous unpublished study on literacy outcomes for statemented children did show significant effects of type of disability over the longer term (Bunn 2002). There is other evidence that particular disabilities do affect literacy outcomes (eg Fletcher-Campbell 2000). It would therefore be unsafe not to allow for the possibility that children's qualities might make a difference to the outcomes of interventions.

In this research, differences between official designations of the level of SEN (i.e. School Action, School Action Plus, and Statement) were considered to be too arbitrary to form a reliable basis for differences between children in terms of difficulty with literacy. Whether a child has or has not got a statement of special educational need can be affected by whether the school has been able to prioritise the child for assessment by their educational psychologist. Instead, a set of 14 descriptive scales was used, based on the work of

Skårbrevik in Norway (Skårbrevik 2001) and Foreman and colleagues (Foreman et al 2001) in Australia. Teachers were asked to rate children on 3 point scales on each of 14 dimensions, for example:

Cognitive level:

1. moderate learning difficulties, IQ 70 or below (about 2% of pupils)
2. below average general ability, IQ 71 to 85 (about 15% of all pupils)
3. average or above average, IQ 85 and above, (about 83% of all pupils).

Attention & Activity Level:

1. very short concentration and/or marked overactiviity, can't keep still (about 2%)
2. concentration short and often restless (about 15%)
3. concentration and activity not a problem most of the time.

The 14 X 3 structure and the level descriptions used were slightly modified from those used by Skårbrevik, and it would have been desirable to pilot these scales to check inter-observer reliability, but in practice there was insufficient time. Teachers did not comment on any difficulties in making these judgements. Levels of identification of difficulty between schools have subsequently been checked statistically (see below, Results).

The 14 dimensions on which children were rated were:

Cognitive level

Attention and activity level

Motivation to learn

Language delay or disorder

Additional language

Medical difficulties

Physical difficulties

Hearing impairment

Visual impairment

Social Communication difficulty

Attendance

Family

Emotional difficulty

Behavioural difficulty

There is evidence that difficulties in each of these areas can have significant effects upon progress in literacy. It seems possible that combinations of difficulties may also be important. An aim was therefore to look for interactions between children's qualities and methods of teaching.

One of the strongest associations between slow progress in literacy and children's difficulties is with "specific learning difficulties/dyslexia" (SpLD). This was not included in the list of children's difficulties because of the unreliability of teacher's judgements about which children have SpLD, because definitions of SpLD remain controversial, and because some definitions (such as the British Psychological Society definition now widely adopted in England & Wales, BPS 1999) rely heavily on lack of progress in

literacy despite appropriate teaching as a criterion for identification. This would introduce circularity into the use of SpLD as a type of difficulty. Although measures of the subtle language, visual, short-term memory, speed of processing or automatization factors that underlie current theories of SpLD (Fawcett 2004) are available or under development, they require expert individual administration, which was impractical in a research design of this type. The types of difficulties identified were relatively gross compared with the subtle difficulties identified by research into SpLD, but this was considered to be an advantage in terms of reliability and the subsequent interpretation of data.

3.4.3. Qualities of the intervention programmes:

There were a large number of interventions used by teachers and TAs in this study. Interventions can be grouped in a number of different ways. At the first level, and in accordance with the main hypothesis, there were 4 groups:

Additional Literacy Support (ALS)

Other interventions

ALS with other interventions

No additional help.

At a second level, there were groups of similar interventions, which are defined by the materials involved:

Comprehensive literacy programmes:

Beat Dyslexia

Hear it, Say it,

Read & Do,

Phonological Approach to Teaching (PAT)

Alpha to Omega,

Jumpstart,

Reading Recovery

Published reading schemes, which also involve activities designed to complement the reading scheme:

Fuzzbuzz,

Oxford Reading Tree

Wellington Square.

Spelling Programmes:

Spelling Made Easy,

Spelling for Literacy

Stile Early Spelling

Ashgate Dictation,

Blackwell Spelling Workshop.

Computer based interventions:

Oxford Reading Tree,

Speaking Starspell,

Other interventions designed by teachers which were not identifiable by the published materials they used:

Reading scheme books,

Individual speech and language programmes,

Learning high frequency words,

Other individual reading activities,

Other individual spelling activities

Effectiveness can therefore be investigated by named intervention, or by type of intervention:

Comprehensive literacy programme

Reading scheme based programme

Spelling programme,

Computer based programme,

Individually designed programme..

The numbers of children using some of these interventions was small, so not all comparisons were expected to be possible. Further information was obtained about some of these materials by examining the materials, by observing the materials in use with children, and by interviewing experienced teachers about their views on the strengths and weaknesses of the materials.

In addition to investigating the outcomes of these programmes in terms of progress in reading and spelling ages, children and teachers were also asked to rate the materials they used. Teachers and TAs were asked to rate the programmes on 5 point scales (1=very poor, 2=poor, 3=satisfactory, 4=good, 5=very good) in terms of ease of use, child interest and effectiveness. Children were asked to rate materials in terms of enjoyment ("Did you enjoy using them?") and effectiveness ("Did they help you?"), on a 3 point scale ("Yes, a lot", "A bit", "No", and "Don't know"). They were also asked to use the same 3 point ratings with "don't know" to answer the question: "How much progress did you feel you made this year?". They were also asked "Next year, how much extra help will you need?", with possible answers, "More than this year", "About the same as this year", "Less than this year", and "None at all". Ratings about effectiveness can be compared with outcomes to see whether judgements of adults and children and outcomes are related. Effectiveness can then be further investigated in terms of enjoyment, interest and ease of use.

3.4.4 Delivery variables

(1) Time and sessions: The amount of time during which additional literacy interventions are delivered is likely to be an important factor in whether the interventions are effective or not. The actual time each week was taken as the key measure, but how time was spread over the week (i.e. how many

sessions) was also expected to be important. ALS, for example, was designed to be used over 4 sessions of about 20 minutes each.

(2) Duration of intervention: Interventions were measured in terms of start and end points. Most continued throughout the year, but a small number lasted for about a term. In view of the controversial issue of the optimum length of interventions, it was considered to be important to measure this variable.

(3) Group size: Group size and especially the difference between 1:1 and small group teaching is important both because of the possible differences in effectiveness but also because of the costs. The majority of interventions other than ALS were delivered 1:1 in this study, but it would be desirable to compare the effectiveness of interventions across a small range of interventions by size of group as well as by type of intervention, to consider whether group size has any effect independently of the type of intervention.

(4) Location: The decision to withdraw children or provide additional help within their classroom can be controversial in English schools. If children remain within the classroom, they are more likely to be seen as being "included" in the Literacy Hour, and there is a considerable expectation on teachers to be "inclusive". However, concentration may be improved by withdrawal to somewhere other than the classroom, and children may prefer other children to be less aware of their different teaching programmes.

Although there is perhaps of continuum of distance between whole class teaching and 1:1 withdrawal, this study recorded only whether interventions were carried out within or outside the classroom.

(5) Staff status and experience: It was expected that differences between teachers and TAs would be important, and that the experience of both could also make a difference. It was possible that some types of intervention would be taught by staff with more or less experience, and that this might influence any findings about effectiveness. Staff status was recorded as either teacher or Learning Support Assistant (LSA), since this is the usual name for TAs in Northamptonshire. Years of working in schools was also recorded, and teachers and TAs were asked to record additional training in literacy. This was coded in terms of

1= No specific training in literacy interventions,

2= At least one course,

3= More than one course.

The main focus of this study was on the effectiveness of ALS and other literacy interventions. How interventions are delivered is likely to influence their effectiveness, and this study attempted to capture the most likely influences in intensity, grouping, location and staff qualities in broad terms, in order to consider the possibility that particular types of influence, such as staff status or experience, location or group size might affect outcomes to

significant extents. If particular factors appeared to have relatively strong influences, this might qualify conclusions about the interventions themselves, and might also suggest further more intensive investigations of these delivery features.

3.4.5 Class and school factors.

Studies of literacy interventions typically assume that how and what children are taught in their classrooms for the great majority of their school week is equivalent, and that any changes in literacy attainments result from the interventions used with the children. While it is very difficult to compare the quality of teaching in classrooms, and to evaluate the relative impacts of class teaching and additional help, it is not safe to assume that the influence of the class is equal. This study attempted to guard against this assumption by asking schools also to provide data on the outcomes for the classes from which children receiving additional help were drawn, and to provide an outline of the class composition and how literacy was taught. It was considered essential to make recording this information as simple as possible. A single page form was provided which asked for information about staff, class and teaching methods.

Differences in schools were also expected to have some influence on outcomes. Perhaps the most obvious measure of the effectiveness of schools' literacy teaching would be attainments in SATs at either the end of

Key Stage 1 or Key Stage 2. Headline figures (percentages achieving level 2 or 4 respectively) might not capture key qualities of the school's impact on children with literacy difficulties, and of course the attainments of children with difficulties also influence SATs results. And neither would directly measure the effectiveness of teaching in Years 3 and 4. Unfortunately, some schools in the study were Junior schools (Years 3 – 6) and some were Lowers (Nursery to Year 4). It was likely that there would serious doubts about the comparability of judgements on school effectiveness if data from KS1 and KS2 SATs were both used. So SATs data were not collected. This decision, it should also be acknowledged, was also a result of a wish not to risk withdrawals by some schools because of hostility to SATs data and published league tables.

It was considered that school size and socio-economic status were likely to have some influence on the overall impact of the school on literacy attainments. School size was taken from published figures of "Number on Roll". Socio-economic status was taken from the Northamptonshire "Social Advantage Factor" statistic. Northamptonshire collects data on schools in terms of the percentages of pupils claiming free-school meals, pupils with SEN and with additional language teaching needs. These figures are combined in a formula to produce a "Social Advantage Factor" (SAF). The formula calculates a "disadvantage factor" which can vary from 0 to 100, and this is subtracted from 100 to form the SAF.

3.5 The sample of schools

As the author of this study lives in Northamptonshire and works for the education service of Northamptonshire County Council, access to Northamptonshire schools was straightforward. Northamptonshire is a relatively large local authority, in which none of the towns or districts have become separate education authorities. The East Midlands regional economy and population have grown in recent decades, and the county contains a large town, Northampton (population just over 200,000), a number of middle-sized towns (Kettering, Corby, Wellingborough, Daventry) and a large number of smaller towns and villages. There is a mixture of industrial, agricultural, leisure, service and retail employment, with a relatively small number of private schools. Within the towns and in occasional rural communities there are areas of significant deprivation. Northamptonshire's overall educational achievements in terms of GCSE and SATs league tables are close to the English national average. At the time of this study, there was a mixture of school types: Northampton itself had Lower (5-9), Middle (9-13) and Upper (13-19) schools, while most of the rest of the county had primary (5-11) and secondary schools (11-19), with some areas retaining separate Infants (5-7) and Junior (7-11) schools. It was considered therefore that a sample of Northamptonshire schools would not be untypical of English school generally, lying as it does close to educational, economic and social averages.

SAF tables were used as a basis for selecting a sample of schools. It was expected that children with literacy difficulties would be likely to belong to schools with lower SAF ratings. The SAF table was divided into quartiles. The lowest two quartiles would provide a third of all schools each, with the upper quartiles together providing a third. A sample size of 30 schools was expected to provide around 300 pupils with literacy difficulties in Years 3 and 4. 10 schools were therefore required from the 3rd and 4th quartiles, and 5 from the 2nd and 1st. Schools were assigned numbers according to their placing on the SAF table and the required numbers were selected from each quartile.

The 30 schools selected were identified by area. For educational administrative purposes, Northamptonshire is divided into 4 areas, with the following spread:

Daventry &SW	Wellingborough & NE	Kettering/Corby	Northampton
9	10	5	6

Letters were sent to the schools explaining the project and asking them to join in. Two schools gave reasons why they should not be included (major changes in the schools or new staff in Years 3 and 4), and two further schools were selected as replacements within the same quartiles.

3.6 The data collection process.

The 30 schools were sent a letter and a booklet containing an explanation of the purpose of the project, a set of forms to be used for data collection, and explanations of how to use the forms. Schools were invited to contact the researcher if any aspects of the project were unclear, and all schools were contacted by phone to check if they were able to carry out the project. Some schools were a little reluctant and needed some persuasion to take part. Visits were made to about 7 schools. The schools visited were all positive about being involved and did not have any major concerns about their ability to collect the data. It was expected that most of the data collection would take place either at the end of the summer term 2003, or at the beginning of the Autumn term, to allow for schools which normally tested children once a year in September to complete that part of the outcome data in their normal way.

It is in retrospect very clear that it would have been advantageous to have made more contacts during the school year. In some schools, unknown to the researcher, staff changes had taken place and information about the project had been forgotten. Others were not routinely collecting information in a form that was relevant to the project. Visits could also have helped to ensure that data collection forms were fully understood by those who would have to complete them (rather than Headteachers), and would have provided opportunities to check fidelity of delivery of intervention programmes. When data was requested at the end of the summer term, there were an

unexpectedly high proportion of schools that were unable to provide any or who could only provide very limited data. Visits were made to a number of schools at this point to try to clarify what information might be obtainable, and some of these schools' data was usable. 11 schools submitted data, of which 9 were usable. By quartile, there were, with pupil numbers

1st quartile	3 schools	244
2nd quartile	2 schools	271
3rd quartile	1 school	257
4th quartile	3 schools	707.

The sample was therefore less well balanced than intended, but there was a spread across the range of SAF, and there was a greater representation of children from the lower half of the SAF range. Data on 126 children who received additional help with literacy was collected, and data from 256 children in 11 classes in 7 schools were available for whole class and school comparisons.

3.7 Ethical considerations:

The study was designed so that there would be no direct contact between the children involved and the researcher, other than through visits to schools to discuss the project with staff and observations in a small number of lessons where the researcher would be introduced only as a visitor to the school. Children would not work directly with the researcher. Data would be collected entirely by staff working normally with the children, and data collection would

occur at the end of the study period, so that staff would for most of the time be "getting on with their normal work" rather than taking part in a research project. Outcome data would only be that normally collected by the school.

Data was kept securely by the researcher, and was analysed on stand-alone computers at the researcher's office and home.

It was left to schools to inform parents of the school's involvement in the project if they wished. Data on outcomes would normally be available to parents through SEN review meetings. Ratings of children's qualities was considered to be a summary form of conversations that would normally be expected to have taken place already between parents and teachers. Ratings by children of the intervention programmes they had experienced were made with the staff who worked with them, and it was expected that staff would agree with the children whether a copy would be made of their views for the school's own records, and what use would be made of their views on the need for further help.

The data collected and the subsequent analysis was expected to be of interest and relevance to the teachers, TAs and schools concerned. The researcher undertook to send a summary of the results of the project to the schools concerned. The LEA would also be informed of the results.

Overall, the research was considered to present no significant ethical risks in terms of worrying or negative information about pupils to themselves or to their parents. Information to schools was expected to focus on the effectiveness of intervention programmes, and to have very little implications for staff competence or self-esteem. It was hoped that the focus on objective evaluation of teaching materials would contribute to an approach to teaching which emphasises the importance of empirical data as important in judging optimum teaching methods.

3.8 Specific questions.

Within the context of the overall aims of the study outlined in 3.2 above, the specific questions listed below act as an 'advance organiser' for the particular statistical analyses reported in Chapter 4 [with references to sections and the database used for analysis]:

- Did children (both those that did and did not receive extra help) catch up in reading and spelling? [4.2.1, Classes]
- Can the two types of measure (age-score and NC level scores) be used to predict each other reliably and so be used as a combined measure of progress? [4.2.2, Classes]
- Are gain scores using age-score and NC levels measuring different features of reading? [4.3, Classes]
- Were there differences in gains between children that did and did not receive extra help in reading and spelling? [4.4, Classes]

- To what extent were these differences a result of differences in their classes? [4.4, Classes]
- Did children who received extra help catch up equally in reading and spelling, and were gains as great for children with the least and greatest difficulties? [4.5, Additional Help]
- Were there differences in gains between ALS alone, ALS and other methods, or Other Methods alone? [4.6, Additional Help]
- Did class differences or types of help make more difference to gains? [4.6, Additional help]
- Was selection for different types of intervention based on severity of difficulty? [4.7, Additional Help]
- What were the differences between children in gender and pupil qualities? [4.8, Additional Help]
- How influential were gender and pupil qualities in explaining reading and spelling gains? [4.8, Additional Help]
- Were children with the greatest needs allocated equally or unequally to different types of help and to different classes? [4.8, Additional Help]
- Do broad school factors (school size, overall SES, SEN levels and class sizes) make a difference to reading and spelling gains? [4.9 Additional Help]
- Do schools or classes make more difference to reading and spelling gains? [4.9, Additional Help]

- Delivering additional help: does the amount of time make a difference? [4.10.1, Additional Help]
- Delivering additional help: is amount of help (in time given) related to the initial severity of difficulties? [4.10.1, Additional Help]
- Delivering additional help: do the number of programmes and the number of sessions make a difference to outcomes? [4.10.1, Additional Help]
- Does whether additional help is delivered in class or by withdrawal make a difference? [4.10.1, Additional Help]
- Delivering additional help: does group size make a difference? [4.10.1, Additional Help]
- Is ALS more efficient (in staff hours) in delivering reading and spelling gains than other types of help? [4.10.1, Additional Help]
- If other interventions beside ALS are compared by broad types, which are most effective for reading and spelling? [4.10.2, Additional Help]
- How did programmes compare in ease of use and interest? [4.10.3 and 4.10.4, Staff Ratings and Children's ratings databases]
- Delivering additional help: were there differences between Qualified Teachers and Teaching Assistants? [4.10.5, Additional Help]
- Delivering additional help: did years of experience and previous training make a difference? [4.10.5, Additional Help]

- Do differences between schools, between children between delivery methods or between types of intervention make greater differences to gains? [4.11, Additional Help]

Chapter 4: Results

4.0 Introduction

The results reported come almost entirely from data collected from the sample of schools using prescribed data sheets about individual children, children's rating of materials, staff details and ratings, and class records including printouts of attainments. Some observations and interviews were undertaken but not sufficiently systematically or rigorously to include reliably into these results.

Data was recorded onto 4 databases using the Statistics Package for the Social Sciences (SPSS) version 10.0. The 4 databases were:

- Individual child records of children who had all received extra help and for whom an individual record had been completed; this is called the "Additional Help" database, and contains information about the child's school, the child's individual strengths and difficulties (special needs), attainments in reading and spelling, what interventions were used with that child and the ways in which those interventions were delivered;
- Individual child records for all children in classes in which some children received individual help; this is called the "Classes" database; the class and school were identified but no other school details were recorded on this database (such as number on roll); no details of children's strengths and difficulties were recorded, nor any information about interventions or delivery details; details of reading and spelling attainments were recorded;

- Teacher and Teaching Assistant (TA) ratings of programmes;
- Children's ratings of programmes and of their own progress.

4.1 Data inclusion and exclusion policy

4.1.1 Additional Help database

As expected where data was collected from a number of sources, some data was incomplete. The key criterion for inclusion into the Additional Help database was that there should be sufficient information to specify what additional help had been given. This meant that one or more programmes (such as ALS) had been used, the size of the group and its frequency of use over a specified period of time should be known. Ideally there should also be data on reading and spelling attainments at the beginning and the end of the year, with dates of testing so that the interval between tests was known. Ideally also there should be reasonably complete information about the child's strengths and difficulties from the 14X3 child ratings.

15 cases, all those from one school, had no information at all about the additional help the children received. These cases could not be included. 5 other cases were not included into the Additional Help database, all because information about the amount of additional help they received was unclear. Two children had remained within the school, but had been withdrawn from ALS at some point. It was not possible to clarify when this change had taken place. Three children had moved away. Although one returned, it was not clear how long the absence had been. For the two

who did not return, there was also uncertainty about when this had occurred. For these three, there were only initial reading and spelling ages.

In 5 other cases there were either no end of year scores (three) or no beginning of year (two). However in these cases, there was satisfactory information about the delivery of help, and about pupil qualities. It was decided to include these cases. They would not be included by SPSS in any analysis which linked reading or spelling gains with qualities of additional help or pupil qualities, but they would be included in analysis of relations either within each of these areas or between them. In view of the smaller sample than intended, it was considered important not to exclude cases which could have some value.

Of the cases that were included (126), there was only National Curriculum level data on 8, and a further 16 had only reading age but not spelling age information. Of 126, 10 did not have complete records of strengths and difficulties, and so were not recorded in the summary variable "sensum" which gave the overall score from the pupil ratings. Similarly, there were 7 cases where information about the time required for all the programmes used was incomplete, so the variable "tot-time" could not be calculated. There were 5 cases where the summary variable for location of additional help "locsum" could not be calculated. Very few cases involved multiple omissions. If all those missing some key measures were omitted, 48 cases

would be lost, leaving a sample of only 78. It seemed better to exclude only the 20 cases where no data about additional help could be included. The database thus consisted of 54 girls and 72 boys, 81 in Year 3 and 45 in Year 4 (see further breakdown in Table 4.8.2 below).

4.1.2 The Classes database

The classes database was intended to provide a comparison between children who received extra help and those who did not from the same class. The classes database recorded child, school and class identification, reading and spelling results in either age-score terms or national curriculum levels, and what type of help the child received, ALS, ALS and other help, other help only or no extra help. Gender was not identified.

Mainly the information arrived in class lists or printouts of test results. All the schools from which individual children were included in the Additional Help database were also included in the Classes database, except one. In that case, of 48 children in Year 3 only for 19 was there sufficient National Curriculum level data to obtain beginning and end of year results. It seemed likely that using a sample of only about 40% of the year group could be subject to unknown selection and bias.

One other school said that the teacher who took that class had left and they were unable to trace the class records for the year. They were able to

send in records for children who received help, so those children were included in the Additional Help database only.

Only some class records could be used from one other school. Only the year 3 records could be used. The year 4 records did not give reading and spelling age scores at the beginning and end of the year, and the Qualification and Curriculum Authority (QCA) tests used were printed out in ways that made it impossible to say which column referred to which year.

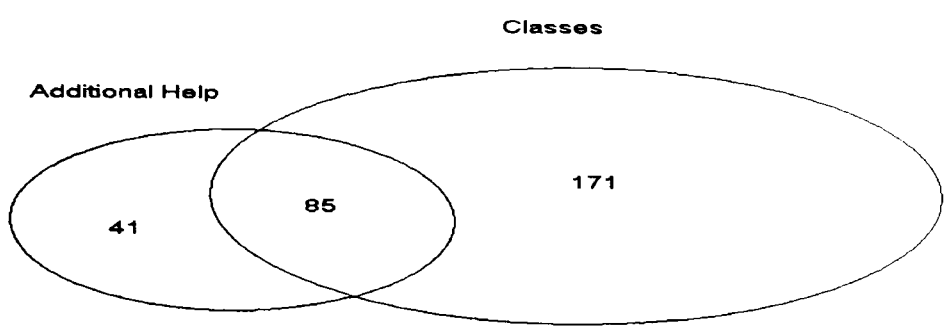
In all, of 256 records on the Classes database, 88 gave reading and spelling age scores (and of these 8 gave only reading scores), 78 gave National Curriculum levels but, with two exceptions, these were only for reading, and 81 gave both age score and NC level results, but only 43 gave both NC reading and spelling. 8 further records did not give two data points of either type and so were effectively excluded from any analysis of results.

4.1.3 Overlap between the Classes and Additional Help databases:

All the 85 children who received extra help in the Classes database were also included in the Additional Help database. The Classes database also included 171 children who did not receive additional help, 256 in all. Children were from 12 classes in 7 schools. A further 41 children included in the Additional help database could not be included in the Classes database, and neither could their classmates, because data about their

classes was incomplete or missing. These 41 children came from 7 classes within 3 schools. In the Additional Help database there were 126 children from 20 classes in 9 schools. Diagram 4.1.3.1 illustrates the overlap:

Diagram 4.1.3.1: Pupil Representation in Classes and Additional Help Databases



4.1.4 Staff ratings of programmes:

Staff ratings of programmes were relatively complete. 24 staff provided ratings, only 6 did not. However, most ratings were of only 3 or 4 programmes, and some were of only one (where only one was used, most often ALS). So the majority of programmes were rated by only one or two staff and only 6 were rated by 5 or more staff. This severely limited the meaningfulness of the ratings.

4.1.5 Children's ratings of the programmes:

Children's ratings of programmes were not intended to be obligatory, partly because of doubts about their likely validity, and partly to limit the overall demands on staff for records. The children's self-ratings were provided as a way of recognising their point of view if teachers and

teaching assistants considered that children's views should be taken into account. Only 61 of the 126 children who received extra help provided ratings of programmes and these often covered only some of the programmes they had worked on. 8 programmes were rated 5 times or more.

4.1.6 Questionnaires about class teaching:

Fewer than half the classes contributing to the Additional Help database provided returns about class teaching methods. The questionnaire was also much too brief to provide useful data about the quality of class teaching.

4.2 Attainments in reading and spelling ages and National Curriculum levels: can these two types of data be unified?

4.2.1 Measures of reading and spelling gain from age scores from the Classes database:

Gains in National Curriculum (NC) levels were based on tests done at the same time each year, but gains in reading and spelling age scores were based on different intervals. Within each school, intervals were generally the same. In order to correct for this variation, and enable comparisons to be made between schools, all reading and spelling age scores were converted into an annualised score by dividing by the interval in months between measurements and multiplying by 12. The first tables (Tables 4.2.1a and 4.2.1b) show the effects of annualization for reading and spelling.

Table 4.2.1a: Reading age scores at start and end, raw gain and annualised measures, in months, from the Classes database

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
RA at start	180	51	140	92.32	17.13
RA at end	176	52	140	105.25	18.04
Raw RA gain	173	-8	36	13.08	7.51
Interval between RA start and RA end	177	7	14	9.35	1.81
Ann. RA gain	173	-10.67	54.86	17.0326	9.8696
Valid N (listwise)	173				

Table 4.2.1b: Spelling age scores at start and end, raw gain and annualized measures, in months, from the Classes database

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SA at start	166	59	138	91.06	15.48
SA at end	165	71	156	105.94	20.28
Raw SA gain	163	-9	40	14.99	9.23
Interval between SA measures	166	8	9	8.70	.46
Ann. SA gain	163	-12.00	54.00	20.8221	12.9697
Valid N (listwise)	163				

Overall, reading age scores gained about 13 months, while spelling gained just under 15 months. Intervals between reading tests were between 7 and 14 months with the mean at just under 9 months, with spelling measured on a shorter and narrower interval of about 8½ months. The minimum and maximum columns show the extremes the data covered. Actual reading age fell by 8 months in one case, with the maximum gain of 40 months. Annualized gains were about 17 months for reading and nearly 21 months for spelling.

4.2.2 Comparing age score and NC data

At the outset, it was hoped that reading and spelling age scores could be correlated with National Curriculum based scores to enable the highest possible number of children's records to be available for comparisons between the main methods of help used. Some schools were able to provide data in terms of reading and spelling ages, some in terms of National Curriculum level data, and some in terms of both.

Correlations between initial and final reading age and the corresponding initial and final NC reading measures and between the corresponding spelling measures were high, as shown in Tables 4.1.2a and 4.1.2b, so it appeared that it might be possible to convert one measure into the other and so unify the data from the two types of measure:

Table 4.2.2a: Correlations between age-score and National Curriculum (NC) variables for reading

Correlations

		RA at start	NC-read-start	RA at end	NC-read-end
RA at start	Pearson Correlation	1.000	.783**	.911**	.785**
	Sig. (2-tailed)	.	.000	.000	.000
	N	180	87	173	93
NC-read-start	Pearson Correlation	.783**	1.000	.745**	.851**
	Sig. (2-tailed)	.000	.	.000	.000
	N	87	158	86	157
RA at end	Pearson Correlation	.911**	.745**	1.000	.767**
	Sig. (2-tailed)	.000	.000	.	.000
	N	173	86	176	93
NC-read-end	Pearson Correlation	.785**	.851**	.767**	1.000
	Sig. (2-tailed)	.000	.000	.000	.
	N	93	157	93	166

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.2.2b: Correlations between age-score and NC variables for spelling

Correlations

		SA at start	NC-spell-start	SA at end	NC-spell-end
SA at start	Pearson Correlation	1.000	.852**	.903**	.841**
	Sig. (2-tailed)	.	.000	.000	.000
	N	166	50	163	49
NC-spell-start	Pearson Correlation	.852**	1.000	.850**	.903**
	Sig. (2-tailed)	.000	.	.000	.000
	N	50	52	50	51
SA at end	Pearson Correlation	.903**	.850**	1.000	.868**
	Sig. (2-tailed)	.000	.000	.	.000
	N	163	50	165	49
NC-spell-end	Pearson Correlation	.841**	.903**	.868**	1.000
	Sig. (2-tailed)	.000	.000	.000	.
	N	49	51	49	51

**. Correlation is significant at the 0.01 level (2-tailed).

However, correlations between annualized reading and spelling gains and NC gains were very different (see Table 4.2.2c below). They show that, although there is a small significant correlation between spelling age gain and NC spelling gain, there is only a small relationship between the two reading gain measures, which is not significant. This strongly suggests either that reading and spelling age tests and NC tests are measuring rather different things, or that different interventions have different effects upon what are being measured (aspects of children's skills), or both. Whichever is the case, the two types of measure cannot validly be unified when trying to understand gain scores. Reading age and spelling gains and NC gains therefore need to be treated as 4 separate dependent variables. It was not therefore possible to unify the main outcomes as hoped.

Table 4.2.2c: Correlations between annualized reading age and spelling age gain scores and NC reading and spelling gain scores

Correlations

		Ann. RA gain	Ann. SA gain	NC-read-gain	NC-spell-gain
Ann. RA gain	Pearson Correlation	1.000	.198*	.155	-.167
	Sig. (2-tailed)	.	.012	.155	.250
	N	173	161	86	49
Ann. SA gain	Pearson Correlation	.198*	1.000	.252*	.321*
	Sig. (2-tailed)	.012	.	.022	.024
	N	161	163	83	49
NC-read-gain	Pearson Correlation	.155	.252*	1.000	.227
	Sig. (2-tailed)	.155	.022	.	.110
	N	86	83	157	51
NC-spell-gain	Pearson Correlation	-.167	.321*	.227	1.000
	Sig. (2-tailed)	.250	.024	.110	.
	N	49	49	51	51

*. Correlation is significant at the 0.05 level (2-tailed).

4.3 Differences between gain scores: are age-score and NC levels measuring different things?

The relationships between initial (red) and final (green) reading and spelling results from age-score and NC measures is shown in the following scatterplots, with regression lines with very similar slopes:

Chart 4.3.1a Reading ages and NC levels at the start and end from the Classes database (Y axis in months, X axis in NC points score 1-9):

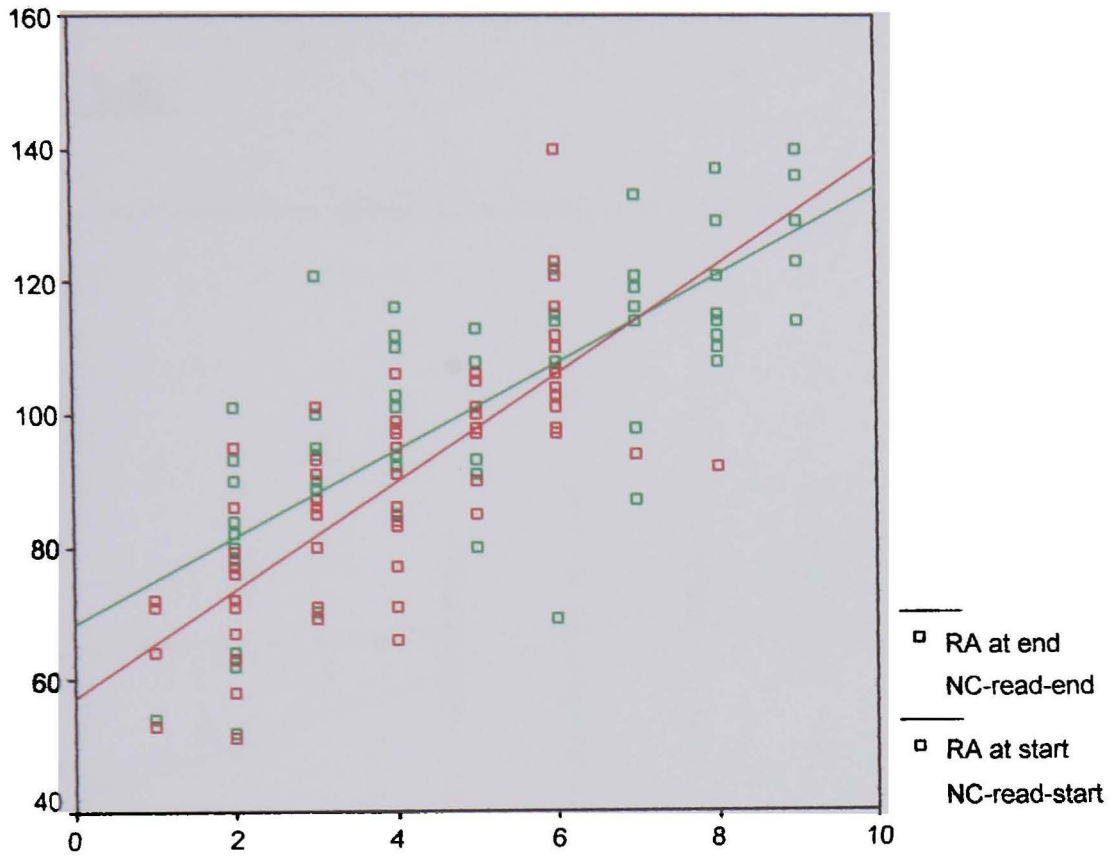
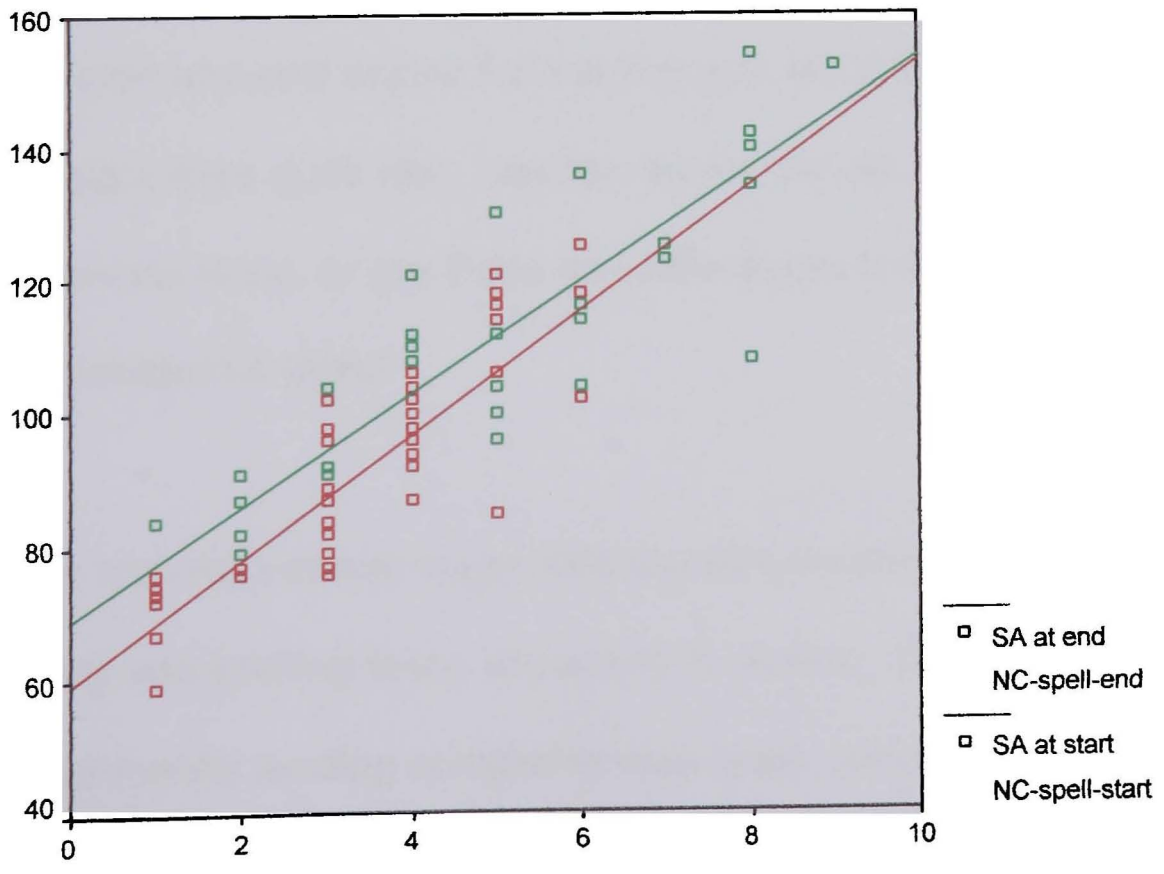
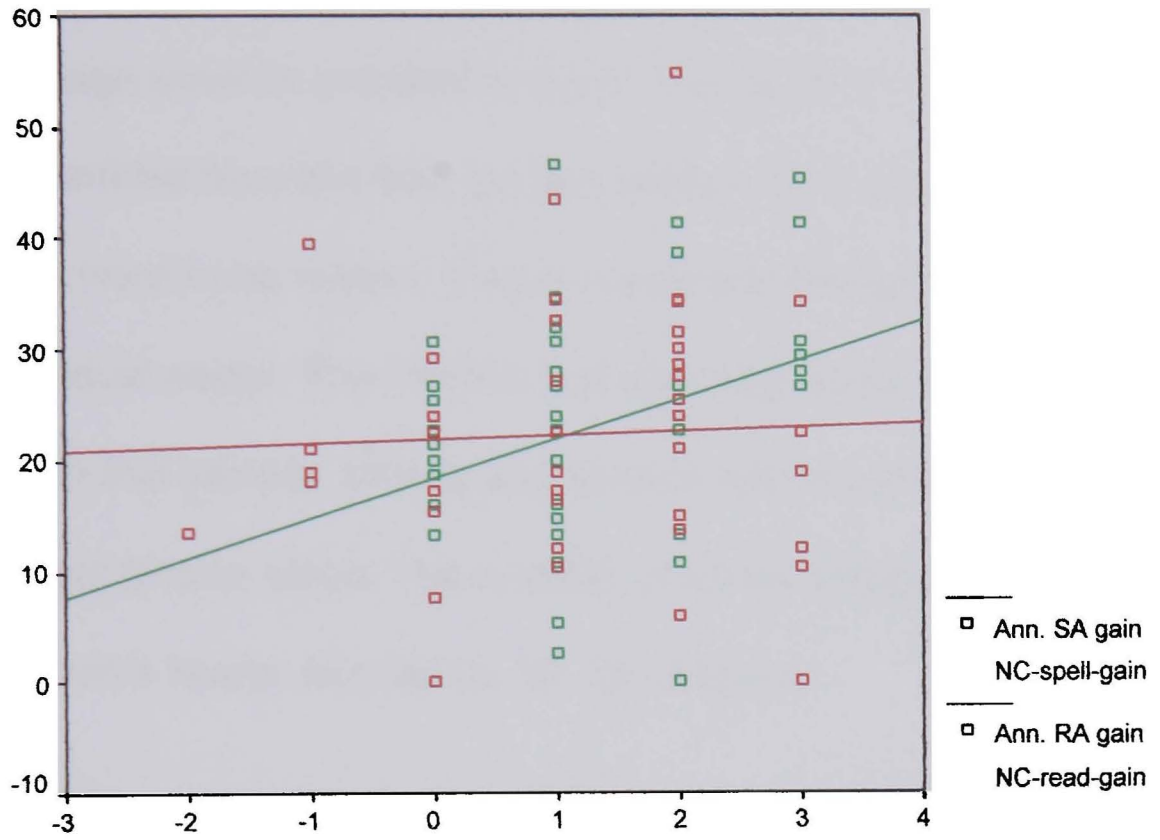


Chart 4.3.1b Spelling ages and NC levels at the start and end from the Classes database (Y axis in months, X axis in NC points score 1-9) :



However, the regression lines are very different when we look at gain scores:

Chart 4.3.1c:Reading and spelling annualised gains in terms of reading and spelling ages and NC levels (Y axis in months, X axis in NC points score 1-9):



The differences in the plots are very striking: gains are widely scattered, while start and end scores for reading and especially spelling approximate to straight lines quite well. Can this scatter be attributed to measurement differences alone, or are there real differences in the ways interventions affect children's skills?

There are undoubtedly major differences between NC and standardised reading and spelling tests, especially in reading. NC reading tests are predominantly reading comprehension tests, presented in a written format which children take in a class situation. The main standardised reading test used in this study, the Salford Sentence Reading test, is designed

only to test decoding of text, by reading aloud a series of sentences in a one to one situation with an adult. The differences between reading tests are therefore very marked.

Spelling tests are more alike. NC tests involve 20 words, with about half shown in pictures under which they write the word, and the other half in a passage which is dictated to them. The words to be written by the children are omitted from the text, but are spoken by the teacher, who pauses at each word to be written. Target words are omitted in the children's individual paper. The Vernon test was very much the most commonly used test in this sample. Words are dictated with a sentence context, but without picture clues. The number of words dictated varies a little with the children's levels, but can be 10, 20 or more.

In spite of these marked differences, the correlations between the two measures on initial and final scores were both quite high. Gain scores are differences between start and end scores, and some reduction in correlation would be expected in any case, because error variation is necessarily counted twice. As would be expected if measurement differences were the only source of variation, the correlations between spelling gain scores were higher than for reading, with spelling r=0.852 (initial) and r=0.868 (final) while reading r= 0.783 (initial) and r= 0.767 (final). But the reduction in correlation is much larger than expected if measurement was the variation's only source. The difference must lie

more in different responses to interventions; children's skills were changed following interventions.

4.4 Comparing outcomes between those who did and did not receive additional help (from the Classes database).

For each class, the average reading age at the start of the year was used to identify three groups:

- those below average for the class who received additional help,
- those below average for the class who did not, and
- those who were above average for the class (none of whom received additional help).

The results for these three groups were then pooled:

Table 4.4.1: mean gains in reading for children who did and did not receive additional help

	Below average with help	Below average, no help	Above average, no help
Ann. RA gain	16.38	17.92	17.40
N=	78	30	65
NC reading gain	1.00	1.63	1.37
N=	51	23	83

(RA gains are given in months, NC spelling gains are given in point scores where W=1, Level 1=2, 2c=3, 2b=4, through to Level 4=9)

T-tests (independent samples) were used to compare these means. The differences between reading gains did not reach significance. On NC reading, the group who received no additional help did better than the group who received help, but the difference was not quite significant.

When the average spelling age at the start of the year is used to identify three groups, a small number of children move from below to above average or from above to below average.

The results for the three groups for spelling were:

Table 4.4.2: gains in spelling for children who did and did not receive additional help

	Below average with help	Below average, no help	Above average, no help
Ann.SA gain:mean	15.53	22.93	26.06
s.d.	11.62	10.45	13.27
N =	72	27	63
NC spell gain:mean	0.71	0.83	1.60
s.d.	.73	.94	.91
N =	14	12	25

(SA gains are in months, NC gains in points as above Table 4.3.1)

Results were:

Table 4.4.3: Results from t-tests (independent samples) comparing 3 groups from the Classes database in spelling gains:

	Ann. SA gain	NC spell gain
Help vs below average	t= -2.896 df= 97 p= .005	t=-.365 df= 24 p= .719
Help vs above average	t= -4.912 df= 133 p< .001	t= -3.114 df= 37 p= .004
Below average vs above average	t= -1.087 df= 88 p= .280	t= -2.371 df= 35 p= .023

The group receiving help were significantly lower in spelling age gain than both the other groups, while in NC terms, both the below average groups were significantly lower than the above average group.

These results imply there were differences between reading and spelling. The differences between all three groups were not large in reading, so here the effects of extra help were not enough to bring the group who received help above the below average group who did not. And so the effects of class teaching appear to be satisfactory in enabling the below average children who did not receive extra help to make progress. Indeed in NC terms, this group made the greatest gains. In spelling, class teaching appears to have been more effective than extra help, because

both the groups who only received class teaching made greater gains. The difference was greatest in spelling age terms, and less in NC terms.

We would like to know how much of the progress of the children who received extra help was a result of this extra help and how much was a result of their ordinary class teaching. There are perhaps three main possibilities:

- progress was mainly a result of extra help
- progress was largely a result of ordinary class teaching
- or, progress was a result of an interaction between class teaching and extra help.

Some indications can be obtained from further analysis of results by classes.

12 classes were involved but from 4 only children who received help were scored in reading and spelling age terms. The following table is thus a little more limited in scope than it first appears. The numbers of cases meant comparisons using NC scores were not helpful.

Table 4.4.4: Reading and spelling age gains by classes and groups receiving help, below average who did not receive help, and above average who did not

	Reading age gain				Spelling age gain			
Class	Help	Below average	Above average	All reading	Help	Below average	Above average	All spelling
1	24.34	38.86	19.61	23.23	21.20	28.80	28.19	25.89
2	24.14	18.00	21.92	20.89	15.11	21.71	24.29	22.39
3	22.67	7.67	10.17	11.23	16.67	15.33	17.83	16.95
4	15.49	-	-	-	17.73	-	-	-
5	8.24	-	-	-	4.85	-	-	-
6	13.67	-	-	-	12.17	-	-	-
7	18.18	-	-	-	9.47	-	-	-
8	8.57	20.50	15.43	17.11	12.00	23.05	17.90	19.91
9	16.29	21.43	20.36	19.61	-	-	-	-
10	13.68	8.40	12.00	11.50	24.30	17.25	32.10	26.38
11	10.77	15.64	13.09	12.76	10.75	33.00	32.32	25.58
12	25.53	-	23.66	24.70	39.00	-	30.75	35.33

(All figures show annualised reading and spelling age gains in months)

Two-way analysis of variance was used to try to estimate the relative importance of class and extra help. Results for reading were:

Table 4.4.5: 2-way analysis of variance with annualised reading gain as dependent variable and class and whether or not help was given as "treatments":

Tests of Between-Subjects Effects

Dependent Variable: Ann. RA gain

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	5937.187	26	228.353	3.082	.000	.354
Intercept	35289.490	1	35289.490	476.305	.000	.765
CLASS	4552.945	11	413.904	5.586	.000	.296
HELPLEV	113.349	2	56.675	.765	.467	.010
CLASS * HELPLEV	1581.576	13	121.660	1.642	.080	.128
Error	10817.161	146	74.090			
Total	66943.020	173				
Corrected Total	16754.348	172				

a R Squared = .354 (Adjusted R Squared = .239)

The results show that, although the overall effect size is modest (adjusted R squared = 0.239), the effect of class differences is highly significant, whereas groups (with help, no help below average, & no help above average, whose variable name was "helplev") were far from significant. However, the interaction of class and group was close to being significant.

For spelling, the results were:

Table 4.4.6: 2-way analysis of variance with annualised spelling gain as dependent variable and class and whether or not help was given as "treatments":

Tests of Between-Subjects Effects

Dependent Variable: Ann. SA gain

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	11742.744	23	510.554	4.554	.000	.432
Intercept	48535.654	1	48535.654	432.969	.000	.758
CLASS	5488.761	10	548.876	4.896	.000	.262
SPELLEV	572.790	2	286.395	2.555	.081	.036
CLASS * SPELLEV	1955.025	11	177.730	1.585	.109	.112
Error	15469.751	138	112.100			
Total	97705.667	162				
Corrected Total	27212.495	161				

a R Squared = .432 (Adjusted R Squared = .337)

For spelling, the overall effect size is higher (adjusted R squared = 0.337) and, as with reading, the effect of classes is highly significant. The effects of the three groups (for spelling the variable was called "spellev") was closer to significance, but the interaction between class and group was a little smaller.

The results of these analyses suggest that differences between classes are greater than differences between those who received help and those below and above average who did not. But there is some interaction between class teaching and extra help. Of course, we do not have a comparison with classes where no children received extra help. Classes 8 (1 child received help) and 3 and 9 (2 in each) were closest to this condition, but they show different patterns – in class 3 the help group did much better than the next group up, while in classes 8 and 9 the extra help group did less well. In general, because differences between classes are greater than differences between those who received help and those who did not for both reading and spelling, it is likely that the quality of class teaching is more important for gains than the quality of extra help, but that there is some interaction between class teaching and extra help.

4.5 Do children who receive extra help catch up equally quickly?

We now consider only children who received additional help, and therefore turn to results from the smaller but more detailed "Additional Help" database. Annualized gains in reading and spelling were lower in the Additional Help database than in the Classes database:

Table 4.5.1 Reading and spelling age scores at start and end, raw gain and annualized scores in months, from the Additional Help database:

	Reading		Spelling	
	Mean	Std deviation	Mean	Std deviation
At start	82.32	12.33	79.52	7.06
N =	116		93	
At end	93.99	14.54	89.80	10.90
N =	115		94	
Raw gain	11.10	7.84	10.22	7.82
N =	113		92	
Interval	9.62	1.67	8.82	1.19
N =	113		92	
Annualized gain	14.12	10.09	14.49	11.41
N =	113		92	

Overall reading gains were about 11 months over an interval of around 9 to 12 months. Spelling gains were a little lower in this database at about 10 months over an interval averaging just under 9 months. Thus annualized gains for reading were just over 14 months and for spelling around 14½ months for children who received extra help.

Do children with greater or smaller difficulties in reading and spelling catch up equally quickly? We consider the proportion of children whose reading and spelling ages were above and below expected levels from the "Additional help" database:

Table 4.5.2 Percentages of children above and below actual ages at the start and end of the measurement period:

	Reading		Spelling	
	At start (%)	At end (%)	At start (%)	At end (%)
At or above age	21.6	23.9	3.2	15.2
Up to 6 mths below age	14.6	19.5	18.3	13.1
7-12 mths below age	19.0	16.8	24.7	18.4
13-18 mths below age	19.8	17.7	23.7	16.3
19-24 mths below age	6.9	6.2	12.9	16.3
Over 24 mths below age	18.1	15.9	17.2	20.7

There were increases in the proportion of children at or above their age level in reading (from 21.6% to 23.9%) and a bigger increase in spelling (from 3.2% to 15.2%). If we take the proportion of children no more than 6 months behind, there was a larger increase in reading, from 36.2% to 43.4%, while in spelling the change was from 21.5% to 28.3%. However, if we look at the proportions more than 12 months behind, there were 44.8% at the start in reading, and this had fallen to 39.8% at the end. In spelling, though, there was almost no change, from 53.8% to 53.3%. These figures show that there was greater "catch-up" in reading than in spelling.

4.6 Comparisons between types of help

The central research question posed at the end of Chapter 3 was whether there are differences in effectiveness between ALS, ALS with other interventions and other interventions alone. The table below shows that there were no clear differences between the 3 types of intervention.

Table 4.6.1: Reading and spelling gains by type of help from the Additional Help database:

	ALS	ALS + other	Other
Ann. Reading age gain	13.01	13.04	15.84
N =	31	38	44
Ann. Spelling age gain	14.31	13.89	14.91
N =	14	36	42

None of these comparisons reached statistical significance, using t-tests (independent samples). There were too few cases to test differences on NC measures on this database.

From this result, is it possible to generalise about the effectiveness of ALS? We need to look in more detail at the results class by class.

Table 4.6.2: Annualised reading and spelling gains from the Additional Help database by class and main types of help.

	Annualised reading gain in months					Annualised spelling gain in months				
Classes	ALS only	ALS & other	Other only	Total	Cases	ALS only	ALS & other	Other only	Total	Cases
1	8.67	15.67	17.33	15.49	13	12.00	17.78	17.71	17.21	11
3	13.07		11.00	12.15	9	14.67		6.00	10.81	9
3H	13.33			13.33	9	0.00			0.00	1
3P	28.50	25.13	21.48	24.34	10	21.00	24.37	25.67	24.21	10
3S	28.50		15.43	24.15	3	13.33		24.00	16.89	3
3W	9.5		7.43	8.46	6	15.44		9.14	12.29	6
4BL	7.25			7.25	8					
2		4.00	12.27	7.44	12		7.05	0.00	4.11	12
2W		14.57	18.00	16.29	2					
3/4C		15.00	8.40	13.68	5		26.25	16.50	24.30	5
3/4H			24.87	24.87	5			39.00	39.00	5
3/4I			8.57	8.57	1			10.80	10.80	1
4		18.22	18.13	18.18	11		12.67	4.67	9.47	10
4D		6.00	29.33	17.67	2		18.00	19.00	18.50	2
4H			11.25	11.25	6			10.75	10.75	6
3LH		6.00		6.00	3		5.33		5.33	3
4B		12.00		12.00	3		9.5		9.50	3
4K		22.67		22.67	2		16.67		16.67	2
4LH		5.00		5.00	3		7.67		7.67	3
Total	13.01	13.03	15.84	14.12	113	14.31	13.89	14.92	14.42	92

For reading, Table 4.6.2 shows that in the 5 classes using ALS alone where there are comparisons with other methods, it produces best results in 4/5 cases. In the 7 comparisons between ALS & others and Others alone, in 4/7 cases Other alone are better; in other words, the comparison is finely balanced. In spelling, of 5 comparisons, only 2/5 show ALS alone as better. ALS & others is better in 4/6 comparisons with Other only.

Two-way analyses of variance were used to obtain measures of the relative importance of class and type of help. The results for reading were:

Table 4.6.3: 2-way analysis of variance on annualised reading gain with class and type of help (ALS alone, ALS & other, and other alone) as "treatments".

Tests of Between-Subjects Effects

Dependent Variable: Annualised RA gain

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	4826.655	30	160.889	2.007	.007	.423
Intercept	14212.298	1	14212.298	177.334	.000	.684
HELPTYPE	102.225	2	51.113	.638	.531	.015
CLASSNAM	3641.165	18	202.287	2.524	.002	.357
HELPTYPE * CLASSNAM	757.827	10	75.783	.946	.497	.103
Error	6571.843	82	80.144			
Total	33932.212	113				
Corrected Total	11398.498	112				

a R Squared = .423 (Adjusted R Squared = .213)

Table 4.6.3 shows that the overall effect size for reading gain is modest, with adjusted R squared =0.213. Class differences were highly significant,

but type of help and the interaction between type of help and class were not.

Table 4.6.4: 2-way analysis of variance on annualised spelling gain with class and type of help (ALS alone, ALS & other, and other alone) as "treatments")

Tests of Between-Subjects Effects

Dependent Variable: Annualised SA gain

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	7775.300	27	287.974	4.526	.000	.656
Intercept	9067.435	1	9067.435	142.495	.000	.690
HELPTYPE	143.629	2	71.815	1.129	.330	.034
CLASSNAM	6851.684	16	428.230	6.730	.000	.627
HELPTYPE * CLASSNAM	486.837	9	54.093	.850	.574	.107
Error	4072.541	64	63.633			
Total	30983.978	92				
Corrected Total	11847.841	91				

a R Squared = .656 (Adjusted R Squared = .511)

For spelling, the overall effect size is moderate, with adjusted R squared =0.511. Class differences are again highly significant, while type of help and the interaction between type of help and classes is not.

These results suggest the tentative conclusion that the sample of schools was very important in affecting the overall outcomes on the effectiveness of methods. Class differences are more important than method differences or interactions between classes and methods. These effects are stronger for spelling than reading.

4.7 Selection for the different types on intervention

It is important to ask on what basis children were selected for different interventions. Policies differed somewhat across schools. Most but not all schools followed the NLS advice that ALS should be used with children who had fallen behind but who did not have significant learning difficulties. Results from the Additional Help database were:

Table 4.7.1: Differences in starting reading and spelling measures by type of help using the Additional Help database

		ALS	ALS + other	Other only
Initial reading age	mean	86.13	87.13	75.67
	s.d	9.598	9.085	13.502
	N =	31	39	46
Initial spelling age	mean	82.64	81.86	76.53
	s.d.	6.594	7.072	6.068
	N =	14	36	43
Initial NC reading	mean	2.50	-	1.00
	s.d.	1.225		.000
	N =	6		2
Initial NC spelling	mean	3.33	-	1.000
	s.d.	1.614		.000
	N =	12		2

Results are shown in the following table:

Table 4.7.2: Results of t-tests (independent samples) comparing initial reading and spelling levels between interventions:

	Initial RA	Initial SA	Initial NC read	Initial NS spell
ALS vs ALS+	t=-.446 df=68 p=.657	t=.357 df=48 p=.722	-	-
ALS vs Other	t= 3.721 df= 75 p<.001	t= 3.203 df= 55 p= .002	t= 3.000 df= 5 p= .030	t= 5.007 df= 11 p< .001
ALS+ vs Other	t= -4.645 df= 83 p<.001	t= -3.603 df= 77 p= .001	-	-

Initial reading and spelling measures were very consistent and showed no differences between the two ALS groups. They were both significantly higher than the group which received other types of help.

These results show that children were selected for other types of help than ALS because they were considered to be weaker in reading. ALS was used for children who were not starting from the lowest points.

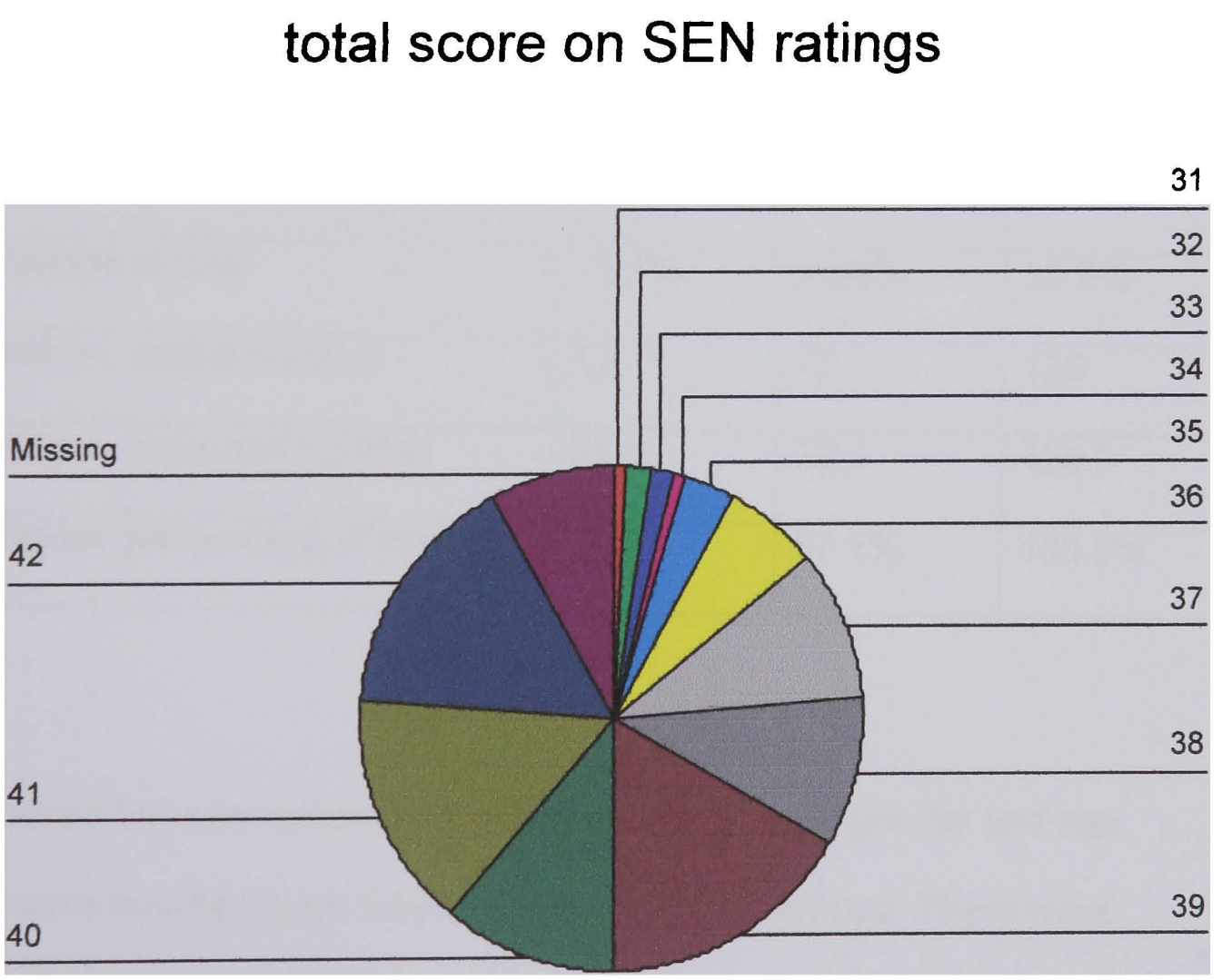
4.8 Individual child and gender differences: do these differences help to explain differences in outcomes?

The extent and type of SEN for each child on the Additional Help database was measured using a 14X3 rating scale. These ratings can be used to

explore general characteristics of SEN and literacy difficulties, and specific links with methods of teaching, starting reading and spelling levels, and gains.

The total possible rating on the 14 dimensions was 42, indicating no SEN. The ratings were made using higher numbers to mean less difficulty. This seemed a more "natural" assumption than the alternative, to make higher ratings indicators of greater difficulty. The pie chart shows how the overall amounts of SEN were distributed:

Chart 4.8.1 Pie Chart showing the distribution of SEN ratings from the Additional Help database



The median score on the SEN rating was 39. Nearly half the children scored above this level. In other words they had one moderate difficulty, two moderate difficulties or one severe difficulty. The pie chart illustrates that SEN as measured by this scale is not normally distributed.

As usually expected, there was a gender imbalance, across both years:

Table 4.8.2: Gender distribution on the Additional Help database:

	girls	boys	total
Year 3 actual number	35	46	81
Year 3 expected number	34.7	46.3	81.0
Percentage of total	27.8%	36.5%	64.3%
Year 4 actual number	19	26	45
Year 4 expected number	19.3	25.7	45.0
Percentage of total	15.1%	20.6%	35.7%
All children, actual number	54	72	126
All children, expected number	54.0	72.0	126.0
All children, percentage of total	42.9%	57.1%	100.0%

Differences between actual and expected levels when gender and age group were counted were very small for both year groups. There were nearly twice as many children in year 3 receiving extra help as in year 4. This does not of course show that help in year 3 is always effective in

reducing the numbers requiring help in year 4. Much more longitudinal data would be needed to establish whether this is typically the case.

The percentage of boys in the whole sample (57.1%) identified as needing extra help with reading was not as high as typically found, but expected numbers of boys in both year groups were very close to actual numbers, suggesting that in this sample at least the gender difference was stable over two year groups.

The distribution of SEN by type of need is best shown by the following table, which gives percentages of children identified at each of the three levels (no difficulty, some or moderate difficulty, severe difficulty affecting education):

Table 4.8.3: Percentages of SEN in sample at 3 levels by type of need:

Type of need	Not a problem	Some difficulty	More severe difficulty	Number
Cognitive level	57.1	37.3	1.6	121
Concentration	35.7	59.5	4.8	126
Motivation	44.4	47.6	5.6	123
Language delay/disorder	78.6	16.7	1.6	122
English as additional language	92.9	2.4	0	120
Medical condition	94.4	5.6	0	119
Movement, balance	96.8	3.2	0	122
Hearing impaired	94.4	4.0	0.8	125
Visually impaired	96.0	3.2	0.8	121
Autism	92.9	5.6	0	124
Attendance	89.7	7.9	0.8	124
Family loss, change	80.2	16.7	0.8	123
Emotional	74.6	23.8	0	124
Behavioural	69.0	29.4	1.6	126

Some of these qualities or factors are much more to do with the child's environment than with the child's own qualities, especially attendance, family change and a need for additional teaching in English. Some are not usually understood as special educational needs. They were included because they were possible "barriers to learning" which might explain differences in progress in reading and spelling.

These figures suggest that EAL, medical and physical difficulties, hearing and visual impairments and autism were not very common. Attendance affected learning in about 9% of the sample to some extent. The qualities that were considered by staff to be most commonly affecting education were concentration and motivational difficulties, overall cognitive level (where quite a small percentage were seen as having major difficulties) and then emotional and behavioural difficulties. Family loss or change affected education in about 17% of the sample.

Analysis of the number of statistically significant correlations between variables suggests these figures are very similar to the results of surveys of the extent of SEN (eg Skårbrevik 2001) in other populations, and the extent of interrelations between different types of SEN. They suggest that this sample was typical of the primary school population in terms of SEN.

If we consider correlations with reading and spelling measures, however, we find that SEN and gender have surprisingly little relationship with starting measures or outcomes:

Table 4.8.4: Statistically significant relationships between SEN variables and gender and starting and gain measures of reading & spelling

Kendall's tau b	RA start	RA gain (Ann)	SA start	SA gain (Ann)	NC read start	NC read gain	NC spell start	NC spell gain
gender	-.172 (p<.05)	.190 (p<.05)						
Cognitive	.233 (p<.01)		.235 (p<.01)					
Attention								
Motivation								
Language	.203 (p<.01)			.219 (p<.05)				
EAL								
Medical								
Movement								
Hearing								
Visual								
Autism								
Attendance	.177 (p<.05)			.214 (p<.05)				
Family								
Emotional								.531 (p<.05)
Behaviour								
Total SEN	.162 (p<.05)							

Girls had higher reading ages at the start of the year, but gains were associated with boys. None of the SEN variables had a significant relationship with reading gain on either age score or NC measures.

Spelling gain was associated with language delay or disorder and attendance, and emotional difficulties were associated with NC spelling gain (these associations were positive, in other words more SEN was linked with lower gains). The total SEN score was only associated with reading age at the start, but not with any of the gain measures.

However, there were differences between the extent of SEN between the three broad types of help (ALS, ALS + other, other help only). This can be illustrated using a scatterplot.

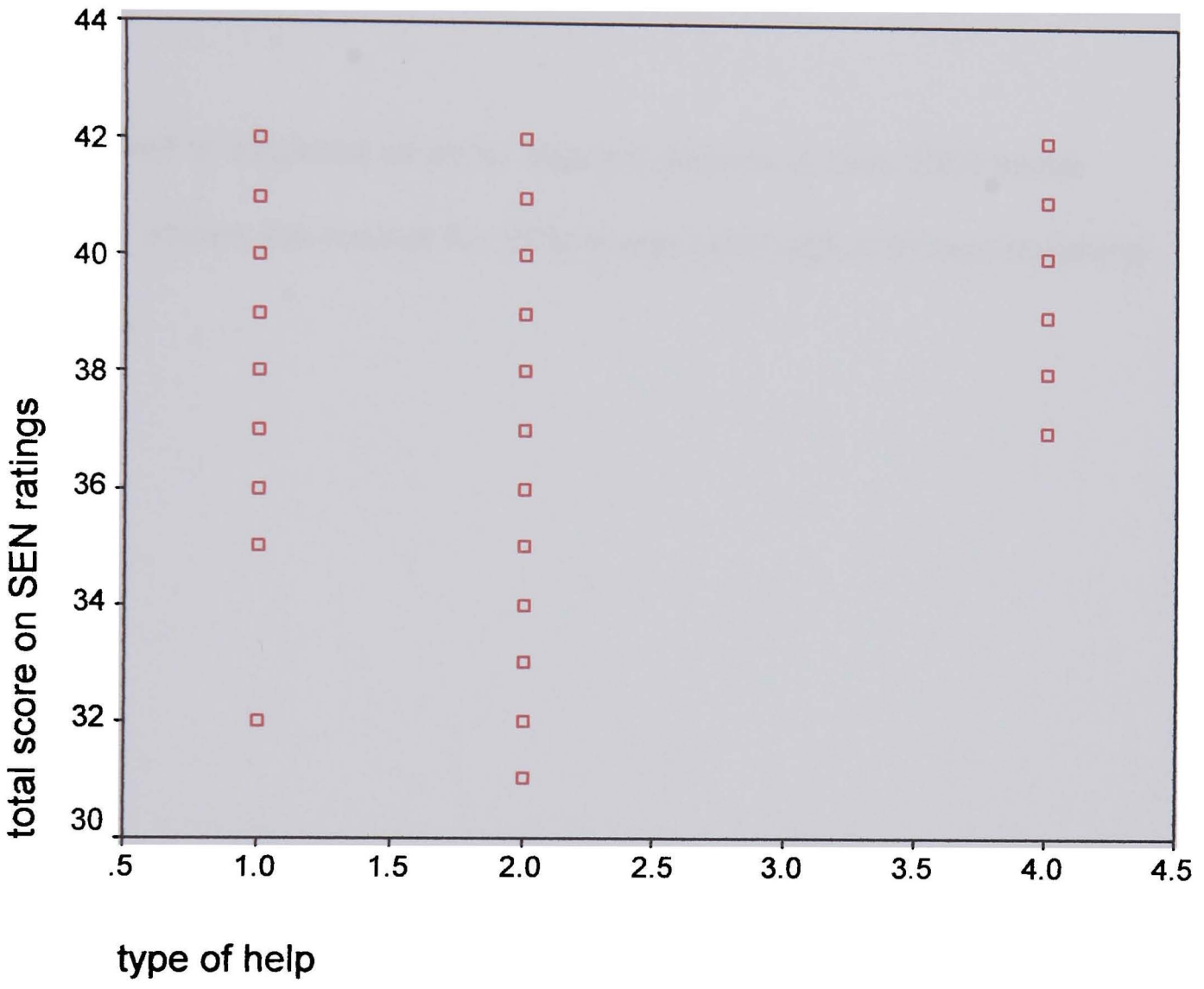
Chart 4.8.5: Scatterplot showing children at each SEN level by type of help:

On the x-axis,

1 = ALS

2 = other help

4 = ALS + other help



The plot shows that more children who received other help had higher levels of SEN (shown by lower scores on the y-axis). The ALS alone group included some children with more SEN. A Kruskal-Wallis test was used to compare the SEN levels for the three groups and gave a significant result, with χ^2=16.412 (2 df) ($p<.01$). There were relatively few students with total SEN ratings in the lowest range (31-34): 1 received ALS, 5 received other help and 0 received ALS + other.

This suggests that schools tended not to select the children with the greatest SEN for ALS. They tended to use a combination of individual programmes for the children with greatest SEN. This is consistent with the earlier conclusions that, while outcomes from the three broad types of help were not statistically different, selection of children for type of help by initial reading age was significant.

We also need to examine whether classes differed in their SEN levels. Table 4.8.7 shows the means for SEN levels (with higher scores meaning lower SEN):

Table 4.8.6: Distribution of SEN by classes:

Dependent Variable: total score on SEN ratings

class name	Mean	Std. Deviation	N
1	39.23	2.13	13
2	38.53	2.90	15
2W	39.60	3.05	5
3	37.44	2.74	9
3/4C	40.60	1.34	5
3/4H	38.60	2.70	5
3/4I	42.00	.	1
3H	37.67	1.58	9
3L	38.83	1.60	6
3LH	40.67	1.53	3
3P	40.50	1.43	10
3S	38.00	2.65	3
3W	40.33	1.21	6
4	39.00	2.65	3
4B	37.50	4.12	4
4BL	36.88	2.70	8
4D	42.00	.	1
4H	40.20	2.17	5
4K	37.00	.00	2
4LH	42.00	.00	3
Total	38.97	2.51	116

One-way analysis of variance with total score on SEN ratings as the dependent variable produced the following result:

Table 4.8.7: One-way analysis of variance with SEN ratings as dependent variable and classes as "treatment";

Tests of Between-Subjects Effects

Dependent Variable: total score on SEN ratings

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	207.051	19	10.897	2.020	.014	.286
Intercept	106537.400	1	106537.400	19749.279	.000	.995
CLASSNAM	207.051	19	10.897	2.020	.014	.286
Error	517.872	96	5.394			
Total	176927.000	116				
Corrected Total	724.922	115				

a R Squared = .286 (Adjusted R Squared = .144)

The effect size is small (adjusted R squared = 0.144). Nevertheless there was a significant effect of class upon SEN levels ($F_{(19,96)} = 2.020$, $p < 0.05$)). There were therefore small differences in the amounts of SEN in classes.

4.9: Do school factors help to explain differences in outcomes?

There was a wide range of numbers on roll (NOR), Social Advantage Factor (SAF) and SEN scores, as illustrated in Table 4.14:

Table 4.9.1: Numbers on roll, SAF and mean and median SEN scores for each school in the Additional Help database:

School	B	C	H	I	K	LH	O	S	W
NOR	277	86	177	103	86	185	253	257	60
SAF	47.78	99.58	82.10	99.68	98.08	98.47	89.77	96.29	99.87
Mean SEN	37.70	40.60	39.40	(42)*	(37)**	41.33	38.55	39.71	39.60
Median SEN	38	40	39.50	(42)*	(37)**	42	39.00	40.50	41

*one case only; ** two cases only.

It was possible to establish class sizes reliably for only some schools, but because of possible importance of class size as a variable which might affect outcomes, it was included for those the 80 children whose class size was known. For the database as a whole, the number on roll, SAF, total SEN rating scores and class sizes are given in table 4.9.2 below:

Table 4.9.2: Means, minima and maxima for number on roll, SAF and SEN scores

		Number on school roll	School's social advantage factor	size of child's class	total score on SEN ratings
N	Valid	126	126	80	116
	Missing	0	0	46	10
Mean		230.17	84.1441	26.46	38.97
Std. Deviation		59.548	17.87877	6.330	2.511
Range		217	52.09	18	11
Minimum		60	47.78	13	31
Maximum		277	99.87	31	42

It is important to ask how these school measures are related to reading and spelling gains. Table 4.9.3 shows Pearson r correlations:

Table 4.9.3: Correlations between Number on Roll, Class Size, Social Advantage Factor, SEN ratings and reading and spelling gains from the AH database:

Pearson r	Number on roll (NOR)	Class size	Social Advantage Factor (SAF)	SEN rating score (SEN)
NOR Pearson r	1.000	.973**	-.436**	-.207*
Sig(2-tailed)		.000	.000	.026
N	126	80	126	116
SAF Pearson r	-.436**	-.439**	1.000	.289**
Sig (2-tailed)	.000	.000		.002
N	126	80	126	116
SEN Pearson r	-.207*	-.165	.289**	1.000
Sig (2-tailed)	.026	.170	.002	
N	116	71	116	116
Ann. RA gain Pearson r	-.046	.004	.140	-.046
Sig (2-tailed)	.632	.974	.138	.647
N	113	74	113	103
Ann. SA gain Pearson r	-.239*	-.201	.065	.173
Sig (2-tailed)	.022	.098	.539	.118
N	92	69	92	83
NC read gain Pearson r	.504	Insufficient	-.504	.322
Sig (2-tailed)	.203	cases	.203	.437
N	8		8	8
NC spell gain Pearson r	-.027	Insufficient	.027	.252
Sig (2-tailed)	.928	cases	.928	.385
N	14		14	14

These results show that smaller schools are associated with smaller class sizes, higher SAF and lower SEN ratings. Of the 4 outcome variables, only spelling age gain is associated with school size (smaller schools have bigger gains) but not with class size. SAF and SEN ratings are not associated with outcomes.

Generally, it is schools which determine which methods of intervention to adopt. Where there are choices, class teachers may have some say in which children should receive what type of help. The larger schools with more children requiring help chose to use ALS. Only B used ALS exclusively, O and S using a combination of ALS and other methods.

Table 4.9.4: Types of intervention favoured by schools

School	B	C	H	I	K	LH	O	S	W
Classes	(3L), 3H 4BL	3/4C	4H, 3/4H	3/4I	4K	3LH, 4LH	1, 2, 3, 4	3S, 3P, 3W, 4D, 4B	2W
ALS	23						7	7	
Other only		1	11	1			24	10	3
ALS & other		4			2	6	17	8	2

We have already examined outcomes by classes (see tables 4.6.2 to 4.6.4) and found that classes differed significantly in both reading and spelling gains; the effect size for reading was small but for spelling it was moderate. Are school effects as strong, or do differences between classes continue to be more important than school effects?

Annualized reading and spelling gains by schools (from the Additional Help database) were:

Table 4.9.5 Annualized reading and spelling gains by school:

School	Annualized reading gain					Annualized spelling gain				
	Mean	Std Dev	Median	Min	Max	Mean	Std Dev	Median	Min	Max
B	10.47	9.19	8.00	1	36	-	-	-	-	-
C	13.68	3.94	15.6	8.4	18	24.30	8.04	22.5	16.5	36
H	17.44	8.47	19.64	6.6	30.5	22.59	15.80	19.5	1.5	45
K	22.67	3.77	22.67	20	25.3	16.67	4.71	16.67	13.3	20
LH	5.5	1.22	5.0	4	7	6.50	2.26	6.00	4	10
O	13.33	10.30	12.0	-11	41.3	10.25	9.63	9.33	-12	37.3
S	18.25	11.70	16.29	-3	43.5	18.00	9.75	16.5	0	41.3
W	16.29	2.42	16.29	14.6	18	-	-	-	-	-

It was not possible to perform a two-way analysis of variance between subjects on reading or spelling gains with both classes and schools as "treatments". However, we have already reported the results from one-way ANOVA. Results for schools and classes of one way ANOVA were:

Table 4.9.6: Results of one-way ANOVA on reading and spelling gains by classes and schools:

	Annualized reading gain				Annualized spelling gain			
	F value	Degrees of freedom	Significance level	Adjusted R squared	F value	Degrees of freedom	Significance level	Adjusted R squared
Schools	1.847	8,104	P=0.077	0.057	4.174	7,84	P=0.001	0.196
Classes	2.821	18,94	P=0.001	0.226	6.861	16,75	P=0.000	0.508

These results show significant effects of both schools and classes on spelling, and of classes on reading. Effect sizes for reading are small, but greater for spelling. Classes have stronger effects than schools. It is not possible to say definitely that school effects are simply an aggregate of class effects. But there is no evidence from these figures that schools have additional effects over and above those of the classes of which they are composed, other than through their size, which is linked to Social Advantage and SEN ratings, and affects spelling outcomes.

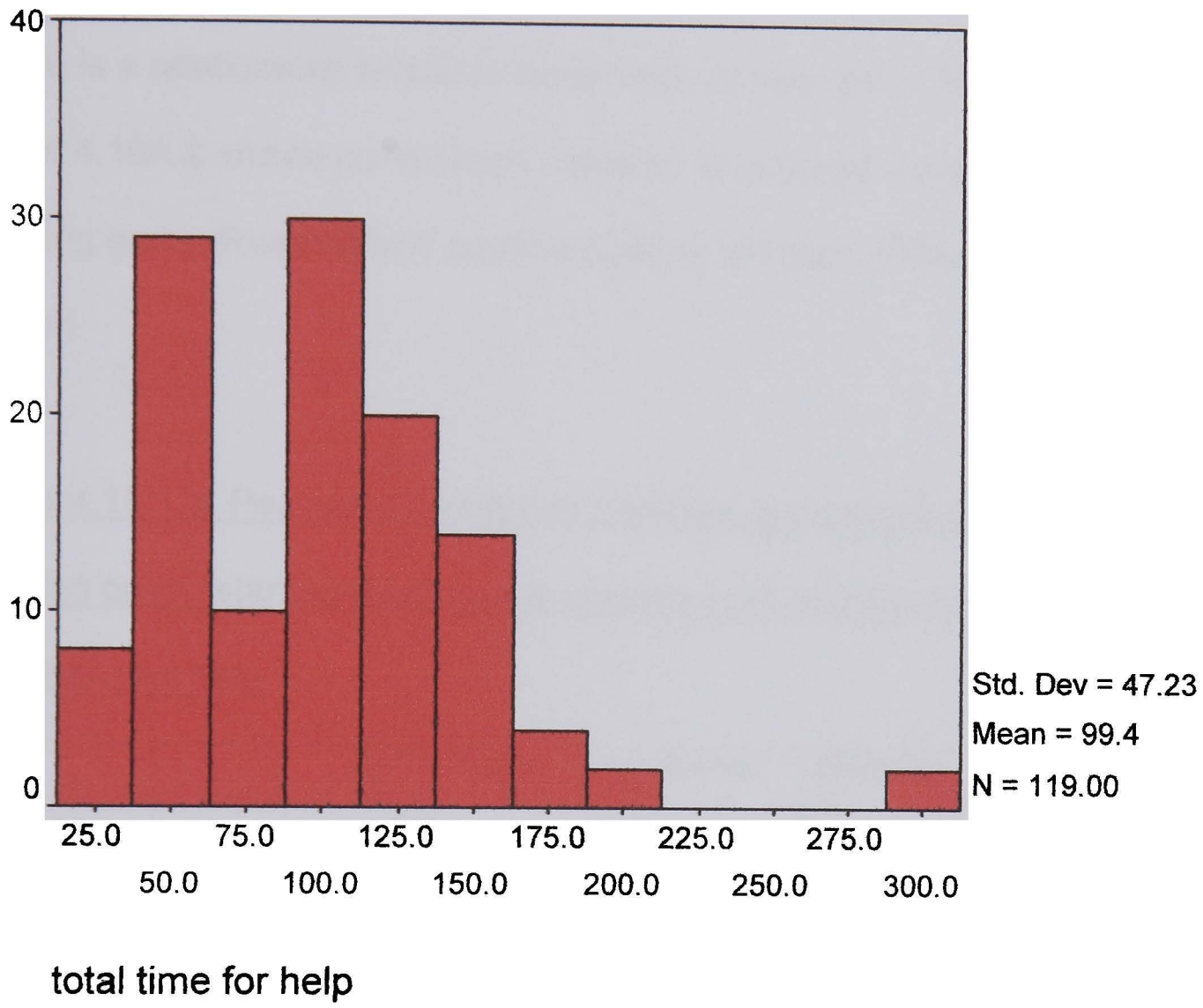
4.10 The delivery of additional help

Although outcomes have been analyzed in terms of the three broad types of help used, the data allow more detailed analysis of aspects of the delivery of help.

4.10.1: Summarizing the amount of help, the number of programmes and where help was delivered.

The amount of help was measured in minutes per week. The distribution is illustrated in the following graph:

Chart 4.10.1.1: The amount of help in minutes per week:



The total time in minutes per week of additional help can be compared between the three main types of help.

Table 4.10.1.2: Mean amounts of help by type of help

Type of help	ALS	ALS + other	Other only
Total minutes/week	82.84	122.83	95.00
N =	37	35	47

ALS alone took the least time, while ALS and other programmes took the most. The differences between ALS + other programmes and the two other conditions were significant using t-tests (independent samples) (ALS+other > ALS, t= -3.910, df= 70, p<.001, ALS + other > Other only, t= -2.558, df= 80, p= .012).

There is a relationship between the amount of help given and outcomes. Table 4.10A.2 shows correlations between annualised reading and spelling gains. Reading and spelling ages at the start of the year are also shown.

Table 4.10.1.3: Pearson Correlations between annualised reading and spelling gains, starting reading and spelling ages and the total time given for additional help

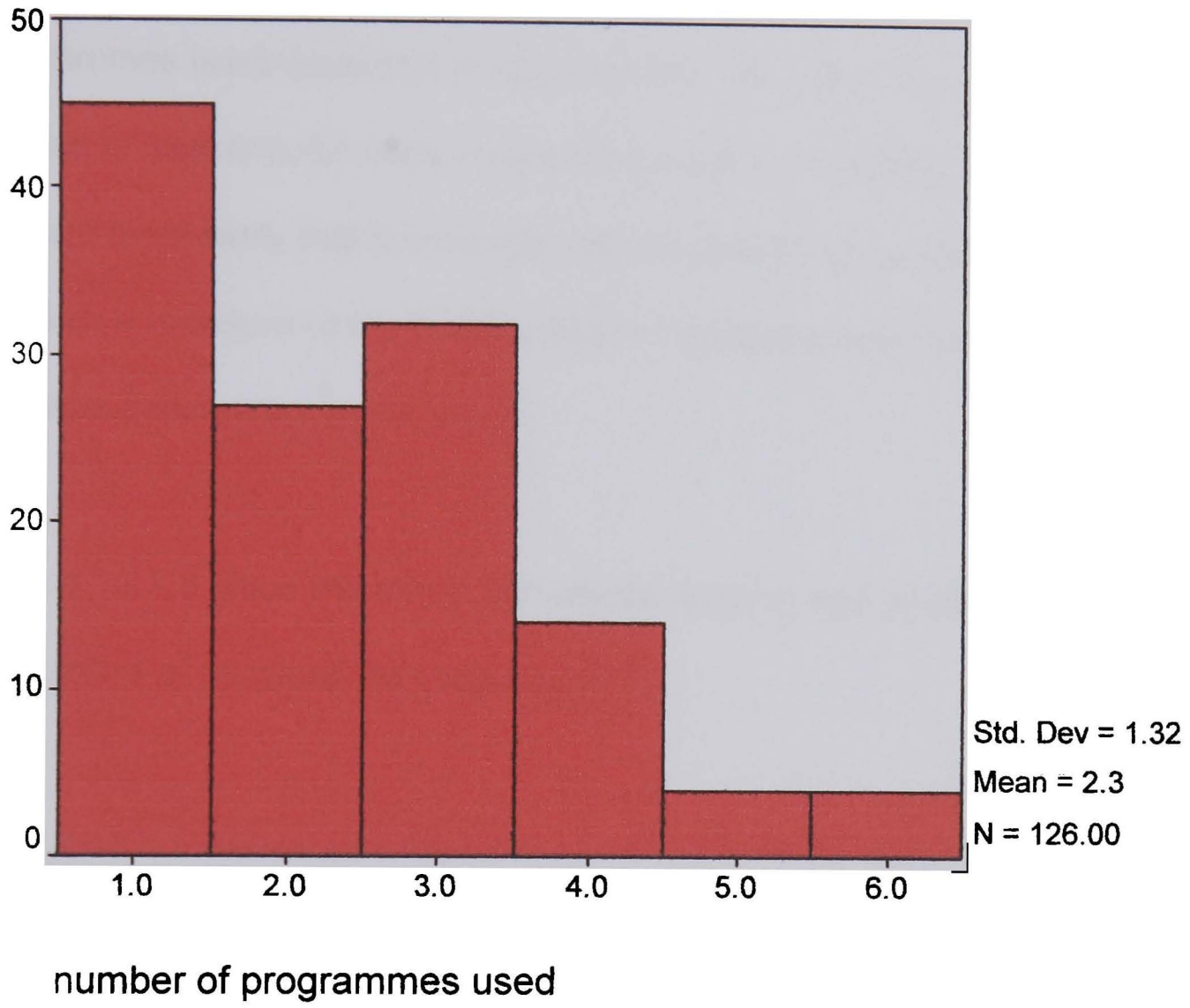
		Ann. Reading gain	Ann. Spelling gain	Starting RA	Starting SA
Total time for help	Pearson r=	0.201*	0.096	-.221*	-.143
	Significance	P=0.039	P=0.382	P=.021	P=0.382
	No. of cases	106	85	109	85

*significant at the p<.05 level

Reading gain was significantly related to the amount of help given, but spelling was not. But starting reading and spelling levels were also related to the amount of help given, but this time negatively. In other words, lower initial reading and spelling ages were associated with more time for help. This was only significant for reading.

The number of programmes used varied between 1 and 6. The distribution is shown in the following graph:

Chart 4.10.1.4: Number of programmes used:



The number of programmes used and the total time for additional help were closely related. The Pearson correlation between these two variable was high at $r=.585$ and was highly significant ($p<.01$). In other words, if more programmes are used, they take more time.

Most commonly, a single programme was used. Where more than one programme was used, the median number of programmes was 3. One way analysis of variance was used to look for differences in annualized

reading and spelling gains from the number of programmes used, but the results were far from significant.

Each programme was also recorded in terms of the number of sessions per week in which it was delivered. Where more than one programme was delivered, it seemed from the records very likely that a number of programmes were delivered in each session. So rather than treating the number of sessions for each programme each as separate sessions, it seemed most likely that the number of sessions for programme 1 would stand as a measure of the total number of sessions over which additional help was delivered over the week.

Table 4.10.1.5 gives the mean annualized reading and spelling gains by the number of sessions for programme 1:

Table 4.10.1.5: Mean annualized reading and spelling gains by the number of sessions over which programme 1 was delivered:

No. of sessions	1	2	3	4	5
Ann RA gain	14.63	13.90	13.57	9.53	23.16
No. of cases	26	14	26	30	15
Ann SA gain	15.61	12.90	12.81	9.56	21.58
No. of cases	25	14	25	12	4

One way analysis of variance comparing mean reading gains by the number of sessions over which programme 1 was delivered was significant ($F_{(4,106)}=5.219$, $p<.01$). Spelling gains were just above the $p<.05$

significance level: $F_{(4,85)}=2.322$, $p=.06$. These results suggest that the best results arose from daily sessions on programme 1. The gradual decline as the number of sessions increases from 1 to 4 , with a more marked falling off on 4 sessions per week, is puzzling, but cannot be readily explained from the data.

Each session was also recorded as located either in-class or withdrawal. Where more than one programme was used, there was a possibility of a child receiving help in both types of location. In order to look for effects of location, a summary variable was used. For each child, all programmes were delivered in class, or all were by withdrawal, or there was a mixture. Annualized reading and spelling gains were as follows:

Table 4.10.1.6: Mean annualized reading and spelling gains by the location of additional help

Location	All in-class	All withdrawal	Mixed
Annualized RA gain	12.81	13.45	17.32
N =	25	60	23
Annualized SA gain	14.67	12.54	17.98
N =	9	55	23

The differences between mixed locations and the other conditions just failed to reach significance in spelling, and failed more clearly in reading. There is a small but interesting difference suggesting that withdrawal is more relevant to reading and less to spelling. These results may suggest

that delivery needs to be suited to conditions and types of programmes – a one-size-fits-all may not be the most effective.

There were a range of group sizes used, from 1 to 9. The maximum number of programmes used with any child was 6. In order to examine whether there were effects of group size, the group size for the programme or programmes which took the largest amount of time for each child were calculated. For example, if a child followed 4 programmes over a total teaching time of 90 minutes, with programmes using groups of 1 for 20 minutes, groups of 2 for 20 minutes, and a group of 6 for 50 minutes, the dominant group size was considered to be 6. Only in 2 cases were there two equal group sizes, and these cases were omitted. In the great majority of cases, one size of group took the majority of teaching time. Groups of 1,2,3,5,6,7,and 9 emerged as "dominant" group sizes. The mean annualized reading and spelling gains were then calculated for these "dominant" group sizes. The number of programmes of a given size, the percentage of all programmes, and the mean annualized reading and spelling ages are given in table 4.10.1.7:

Table 4.10.1.7: Group size and mean reading and spelling gains (in months):

Size of group	Number of programmes	Percentage of all programmes	Mean Annualized RA gain	Mean Annualized SA gain
1	163	55.82	15.68	13.17
2	35	11.98	17.31	12.00
3	14	4.79	8.03	12.94
4	7	2.40		
5	9	3.08	11.93	13.31
6	54	18.99	14.37	14.12
7	3	1.03	1.67	-
8	0	0		
9	7	2.40	17.89	29.10
Total	292			

Differences between means did not reach statistical significance using one-way analysis of variance. Clearly the one case where a group of 9 was used achieved particular success in spelling gains, but there were too few cases for this difference to reach significance. These figures can therefore only be suggestive: groups of about 6 seem to achieve similar results to groups of 1 (that is, individual work) or 2 (pairs). But of course, types of intervention are not spread equally across all group sizes. ALS was nearly always delivered in groups of 6, while other interventions were most often delivered individually.

It is clear from this section that the total number of programmes used is closely related to the total time given, and time is related to outcomes. Although none of the 3 main types of help (ALS alone, ALS and other programmes, and other programmes alone) has emerged as

unambiguously superior in terms of annualized reading or spelling gains, it may be important to consider whether there are differences in efficiency between them. Measures of months of reading (and spelling) gain per hour of help given were calculated:

Months of reading gain per hour of help given=

(Annualized reading gain/total staff minutes)*60

There were large differences between the three main types of help, as table 4.10.1.8 shows:

Table 4.10.1.8: months of reading and spelling gain per staff hour of help by type of help:

	Reading gain in months per staff hour			Spelling gain in months per staff hour		
	Mean	Standard dev	Number	Mean	Standard dev	Number
ALS	53.32	44.48	31	37.48	19.14	14
Other	18.68	19.84	40	18.95	21.20	38
ALS & other	15.27	13.78	34	16.39	12.24	32
Total	27.80	32.49	105	21.06	19.22	84

One-way analysis of variance with reading gain per staff hour of help gave significant differences between types of help, with $F_{(2,102)}= 18.176$, $p<0.001$), as did spelling gain per staff hour, where $F_{(2,81)}= 7.218$, $p<0.01$. Effect sizes were modest, with adjusted R squared for reading = 0.248, and for spelling = 0.130. In terms of efficiency, ALS alone is superior to ALS combined with other types of help and to other types of help alone.

4.10.2: Interventions classified into 6 types

Interventions were classified into broad types (see section 3.4.3).

Table 4.10.2.1: The spread of interventions classified into 6 broad types:

Type of intervention	Number of programmes	Percentage of all programmes
ALS	77	26.37
Comprehensive Literacy programmes	38	13.01
Reading scheme based programmes	53	18.15
Spelling programmes	18	6.16
Computer based programmes	39	13.36
Individually designed programmes	67	22.95
	292	100.00

The figures show that, after ALS, the most common "extra" is an individually designed programme, including work on high-frequency words and handwriting. Additional work based on reading scheme books is the next most commonly chosen extra, with computer-based work and comprehensive literacy programmes the next most commonly chosen type of work. Programmes which focused entirely on spelling were relatively uncommon. Spelling comprises a part of most other programmes, so it seems possible that generally spelling is commonly not taught as a separate topic, but more often as a part of a more general literacy programme, whether computer based or linked to a reading scheme.

It is not possible to analyze the relationships between outcomes and this broad classification of programmes definitively, because the majority of children were taught using more than one programme, and it is not possible to isolate the contribution each programme made to the overall

outcome. However, means associated with each type of intervention can be given, and analysis of variance can be used to suggest possible differences in their effectiveness. Means for reading are given in table 4.10.2.2:

Table 4.10.2.2: Mean annualized reading gain by type of intervention for each of the 6 programmes in months:

	ALS	Comprehensive Literacy	Reading scheme	Spelling programme	Computer based	Individually designed
Programme 1	13.19	14.25	14.65	10.37	17.64	22.40
Number	*60*	*20*	*12*	*7*	*9*	*5*
Programme 2	14.45	13.21	14.04	17.41	8.76	16.85
Number	*4*	*13*	*28*	*9*	*6*	*6*
Programme 3	4.0	6.0	16.15	6.79	16.27	14.38
Number	*1*	*1*	*9*	*2*	*12*	*24*
Programme 4	7.0	24.57	8.4	0	11.28	14.21
Number	*2*	*3*	*1*	*0*	*5*	*12*
Programme 5	15.6	0	0	0	8.4	22.67
Number	*2*	*0*	*0*	*0*	*1*	*2*
Programme 6	8.4	0	22.67	0	0	15.60
Number	*1*	*0*	*2*	*0*	*0*	*1*
Mean	12.95	14.50	14.77	13.49	14.28	15.79

The means in this table were calculated as a simple average. This method perhaps adds extra variance to outcome figures where interventions are spread evenly across a larger number of programmes. So individually designed programmes appear most often as programme 3 but are spread

across all the programme slots. On the other hand, ALS is generally given as programme 1 when it is used, so the spread is probably quite low. In fact the overall mean for ALS alone was 13.01 months (table 4.6.1), so the "spread average" here is quite close (13.01-12.95=0.06). Individually designed programmes achieved the highest outcomes in reading, but none of the differences were significant using one-way analysis of variance.

Similarly, means for spelling outcomes are given in table 4.10.2.3 below:

Table 4.10.2.3: Mean annualized spelling gain by type of intervention for each of 6 programmes in months:

	ALS	Comprehensive Literacy	Reading Scheme	Spelling Programme	Computer based	Individually designed
Programme 1	13.52	18.21	7.90	18.40	9.40	23.80
Number	*41*	*20*	*12*	*7*	*7*	*5*
Programme 2	16.41	9.13	12.17	25.96	1.52	19.08
Number	*4*	*13*	*26*	*9*	*13*	*12*
Programme 3	5.0	18.00	15.65	10.40	10.92	13.63
Number	*1*	*1*	*9*	*2*	*11*	*24*
Programme 4	5.50	26.11	16.50	0	16.43	14.72
Number	*2*	*3*	*1*	*0*	*5*	*12*
Programme 5	32.25	0	0	0	16.50	16.67
Number	*2*	*0*	*0*	*0*	*1*	*2*
Programme 6	16.50	0	22.67	0	0	28.50
Number	*1*	*0*	*2*	*0*	*0*	*1*
Mean	14.06	15.65	12.04	21.29	8.26	16.31

The comparison between the mean for ALS here (14.06 months) and from table 4.6 (14.31months) is even closer. For annualized spelling gain, analysis of variance was significant for programmes 1 and 2 (programme 1: $F_{(5,86)}=2.603$, $p<0.05$; programme 2: $F_{(5,64)}=5.238$, $p<0.01$). Spelling programmes appeared to produce a much better outcome for spelling than any other type of intervention, with individually designed programmes (which included some work targeted at specifically at spelling) also a little better than other interventions.

4.10.3: Staff Ratings of Programmes

A different approach to programme analysis is to obtain ratings of them by their users. Only very limited numbers of schools returned programme ratings. Staff who used these programmes (nearly always TAs) rated them on 5 point scales in terms of ease of use, how interesting they were to the children, and effectiveness. Only six programmes were rated at least 5 times. These results are shown in Table 4.10.3.1:

Table 4.10.3.1: Ratings of 6 programmes by teaching assistants in terms of ease of use, interest and effectiveness, using a 5 point scale with 1=very poor and 5=very good

Programme	Number of ratings	Mean ease of use	Std dev ease of use	Mean child interest	Std dev child interest	Mean effectiveness	Std dev effectiveness
ALS	14	3.5	.85	3.9	.73	3.9	.77
Wellington Sq	5	4.4	.55	4.2	.45	4.4	.55
Oxford RT	11	4.6	.52	4.5	.69	4.4	.67
Fuzzbuzz	9	4.3	.71	4.0	1.12	3.7	1.32
PAT	13	4.4	.51	4.0	.58	4.0	.60
Starspell	7	4.4	.53	4.1	.90	4.0	.89

This very limited sample did not find ALS as easy to use as the other programmes, which were all rated at about the same level between good and very good. The only programme to stand out as better at holding children's interest was that based on "Oxford Reading Tree"; the others were all rated as good. Both Oxford Reading Tree and Wellington Square were seen as more effective than the other 4 programmes.

4.10.4: Children's ratings of Programmes

Ratings by children were also available, but were similarly used by only a very limited number of participants. Children used 3 point ratings, of whether they enjoyed the programme, and whether it helped. 8 programmes were rated more than 5 times. It was advised that these ratings should be done with their TA. The results are shown in Table 4.10.4.1:

Table 4.10.4.1: Ratings of 8 programmes by children in terms of enjoyment and how helpful they were, using a three point scale, where 1= No [it wasn't enjoyable/helpful] to 3= Yes [it was enjoyable/helpful]

	Number of ratings	Mean enjoyment	Std dev enjoyment	Mean helpfulness	Std dev helpfulness
ALS	32	2.7	.52	2.7	.71
Wellington Sq	7	2.7	.76	2.4	.79
Oxford RT	10	2.9	.32	-	-
Fuzzbuzz	6	3.0	.00	-	-
PAT	22	2.6	.50	2.4	1.10
Starspell	31	2.6	.77	2.4	.78
Look Cover Say Write Check	9	1.9	.78	2.0	1.00
Blackwell Spelling	9	2.1	.60	1.7	.87

The children's ratings of programme enjoyment were similar to those from their TAs, in that Oxford Reading Tree and another reading scheme based programme, Fuzzbuzz, were rated highest. The "Look Cover Say Write Check" procedure for learning spellings was rated lowest for enjoyment, with Blackwell Spelling, with which it is associated, not much above it. Both of these programmes were also rated lowest in terms of helpfulness by the children, who rated ALS in this respect the highest.

Two ratings of overall progress were also made by the children. These enable the relationship between ratings and actual progress to be explored. Children were asked to tick one of four boxes, in answer to the question

"How much progress do you feel you have made this year?"

□ A lot □ Some □ Not much □ Don't know

Mean actual reading gains (ie not annualized reading gains) and spelling gains were

Table 4.10.4.2: Mean actual reading and spelling gains in months for children responding to questions about how much progress they felt they had made:

	RA gain	N =	SA gain	N =
Don't know	8.28	7	10.71	7
Not much	14.00	4	8.00	4
Some	15.00	18	10.47	17
A lot	10.73	26	8.84	25

None of the differences reached statistical significance. Perhaps the only clear implication is that this kind of rating is not reliable as a way of measuring actual progress. Children who felt they had made the most progress had not actually done so.

The other self-rating was through the question,

"Next year, how much extra help will you need?"

More than this year □

About the same as this year □

Less than this year □

None at all □

This was intended as an indirect way of rating how much had been accomplished this year, as well as a useful discussion point between child and TA. However, this rating correlated only to a very limited extent with reading gain (r= -.208) and was not significant (p=.091), and less with spelling gain (r=-.134), which was substantially further from significance (p=.295). The two ratings ("How much progress..." and "How much extra help...") were not themselves closely related (r=-.104, p=.425).

In general, children's self-ratings of their own achievements in literacy and their views on the enjoyment and usefulness of the programmes they followed did not seem useful instruments in the form they were employed here. At best there were some similarities between their enjoyment ratings of programmes and the ratings of ease of use and child interest by TAs teaching the programmes.

4.10.5: Staffing factors in the delivery of additional help

Staff who gave additional help were asked to record their status. 94% of programmes were delivered by Teaching Assistants, and the remaining 6% were by qualified teachers.

As it happened, qualified teachers were involved in only programmes 1 to 4. Some indication of the contributions to children's reading and spelling of the status of staff is given by a comparison between means for these outcomes for each of programmes 1-4. Of course, the contributions of

other programmes have not been removed, so these results are no more than suggestive.

Table 4.10.5.1: Mean annualized reading and spelling gains associated with Teaching Assistant or Qualified Teacher delivery

	Number	Teaching Assistants		Number	Qualified Teachers	
Measures		Ann. RA gain	Ann. SA gain		Ann. RA gain	Ann. SA gain
Programme1	121	13.96	13.86	5	17.52	24.27
Programme2	69	14.40	13.19	8	14.77	24.26
Programme3	43	14.15	10.14	11	15.61	23.35
Programme4	18	13.41	14.59	4	16.64	23.63

Only two comparisons were statistically significant, using t-tests (independent samples), spelling gain for programme 1 (t= -2.017, df= 111, p= .047) and spelling gain for programme 3 (t= -3.548, df= 46, p=.001). In general, teacher delivered programmes were associated with slightly higher reading outcomes, though none were statistically significant. Teacher delivered programmes were associated with substantially higher spelling outcomes, and in 2 out of 4 cases these differences were significant.

Two other background factors were recorded: years of experience in schools and additional training in literacy. Training was divided into 3 categories: none, at least one course, and more than one course. For both variables, there was no straightforward way to aggregate experience and

training factors across all the programmes that children received. A child might receive programme 1 from a TA with no extra training and 10 years' experience, programme 2 from the same TA, and programme 3 from a teacher with substantial training and 15 years' experience. The contributions made by experience and training of different staff to the overall results in terms of reading and spelling gains are impossible to separate.

Teachers and TAs were generally experienced in the sample schools. Table 4.10.5.2 shows the distribution by programmes:

Table 4.10.5.2: Mean, median and modal years of experience in schools for teachers and TAs by each of 6 programmes they taught:

	Programme 1	Programme 2	Programme 3	Programme 4	Programme 5	Programme 6
Numbers of teachers and TAs	126	76	49	20	8	4
Mean years	10.97	12.67	12.98	10.60	9.00	4.00
Median Years	9	15	15	13	7	4
Modal years	9	16	16*	1	7	1*

*more than one mode available, the lowest figure is given.

The overall mean was calculated by averaging means across the 6 programmes. It was 11.59 years. The range was from 1 to 27 years. This suggests quite an experienced TA/teacher force in the sample schools.

The impact of years of experience can be analyzed indirectly. One-way analysis of variance was used to assess the effect of years of experience, and correlation was used to check the direction of this effect.

Table 4.10.5.3: Results of one-way analysis of variance on annualized reading and spelling gains compared across years of experience of teachers and TAs:

		Programme 1	Programme 2	Programme 3	Programme 4	Programme 5	Programme 6
Ann. Reading gain	F=	2.263*	1.452	0.873	1.299	10.558*	5.669
	D.of Freedom (between,within)	14 98	12 56	13 30	8 11	3 4	1 2
	Probability p=	0.010	0.171	0.588	0.335	0.023	0.140
Ann. Spelling gain	F=	2.450**	2.399*	1.399	1.982	4.515	0.722
	D.of Freedom (between,within)	14 77	12 53	13 29	8 11	3 4	1 2
	Probability p=	0.007	0.015	0.219	0.145	0.090	0.485

*significant at the p<.05 level, **significant at the p<.01 level

These results show that years of experience do seem to have an effect on reading in programmes 1 and 5, and spelling on programmes 1 and 2. However, the correlation figures show this is not in the expected directions, as table 4.10.5.4 shows:

Table 4.10.5.4: Pearson correlations between annualized reading and spelling gains and years of experience of teachers and TAs:

		Programme 1	Programme 2	Programme 3	Programme 4	Programme 5	Programme 6
Ann. Reading gain	R=	-.089	-.091	-.180	-.547*	-.401	-.860
	Number of cases	113	69	44	20	8	4
	Probability p=	0.351	0.459	0.242	0.013	0.324	0.140
Ann. Spelling gain	R=	-.225*	-.185	-.151	-.451	-.042	0.515
	Number of cases	92	66	43	20	8	4
	Probability p=	0.031	0.136	0.333	0.046	0.922	0.485

*significant at the p<.05 level.

These results unexpectedly show that for almost every programme, reading and spelling gains are negatively associated with years of experience. In other words more experience is less effective across the age range as a whole.

The same type of analysis was used to examine the effects of training on outcomes. Here the pattern is different.

Table 4.10.5.5: Results of one-way analysis of variance on annualized reading and spelling gains compared across three categories of training

		Programme 1	Programme 2	Programme 3	Programme 4	Programme 5	Programme 6
Ann. Reading gain	F=	2.746*	1.450	0.105	0.999	0.570	5.669
	D.of Freedom (between,within)	2 110	2 66	2 40	2 18	1 6	1 2
	Probability p=	0.069	0.242	0.901	0.388	0.479	0.140
Ann. Spelling gain	F=	4.936**	6.736	4.235*	0.337	0.593	0.722
	D.of Freedom (between,within)	2 89	2 63	2 39	2 18	1 6	1 2
	Probability p=	0.009	0.002	0.022	0.719	0.471	0.485

*significant at the p<.05 level, **significant at the p<.01 level.

Table 4.10.5.6 shows that in most cases the associations between outcomes and training are in the expected direction.

Table 4.10.5.6 Kendall's tau correlations between annualized reading and spelling gains and 3 categories of training

		Programme 1	Programme 2	Programme 3	Programme 4	Programme 5	Programme 6
Ann. Reading gain	T=	0.189*	0.179	0.128	0.093	0.248	0.816
	Number of cases	113	69	43	21	8	4
	Probability p=	0.010	0.060	0.292	0.602	0.453	0.121
Ann. Spelling gain	T=	0.239**	0.290**	0.353**	0.186	-.342	-.408
	Number of cases	92	66	42	21	8	4
	Probability p=	0.004	0.003	0.004	0.297	0.297	0.439

*significant at the p<.05 level, ** significant at the p<.01 level

A number of these correlations are highly significant, especially in the first three programmes, for spelling, and with the exception of programmes 5 and 6, they are all positive. So in most cases, additional training in literacy seems to make a difference, especially in its effects on spelling outcomes.

4.11: Explaining outcomes: do a combination of school, child and delivery factors explain outcomes?

One of the main aims of the study was to investigate whether background school, child and delivery factors made important differences to how much progress children made in reading and spelling. Although relationships among these variables were found, the general lack of relationship between them and the reading and spelling outcome measures was unexpected.

Linear regression was used to try to give some sense of the relative importance of these background factors, separately for both the annualized reading gain and spelling gain outcomes. For this analysis, the type of help given had to be reclassified into three "dummy variables". Type of help was classified as "ALS", "ALS and other" or "other alone"; it was reclassified into three either/or categorical variables, ALS alone versus all other types, ALS and other methods versus ALS alone or other methods but not ALS, and other methods versus ALS alone or ALS with other methods. The other variables entered were:

Social advantage factor

Number on roll

Total score on SEN ratings

Total time given for additional help

Staff status teaching programme 1, and

Reading (or spelling) age at the start of the year.

Linear regression does not allow nominal variables such as school or class to be used.

For reading gain, the adjusted R square was .060. This means that only 6% of total variance was explained by these variables.

In the table below, the Beta coefficients give an indication of the relative importance of the variables entered:

Table 4.11.1: Results of linear regression to predict annualized reading gain from school, child and delivery factors

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4.586	24.277		-.189	.851
	School's social advantage factor	5.384E-02	.083	.097	.652	.516
	Number on school roll	6.030E-03	.024	.036	.249	.804
	total score on SEN ratings	-5.71E-02	.473	-.014	-.121	.904
	Other vs ALS or mixed	-.831	3.360	-.041	-.247	.805
	Mixed vs ALS or other alone	4.455	3.273	.210	1.361	.177
	total time for help	7.623E-02	.027	.376	2.874	.005
	staff status teaching prog 1	-3.495	6.366	-.062	-.549	.584
	Reading age at start of year	4.756E-02	.104	.057	.458	.648

a. Dependent Variable: Annualised RA gain

The only variable which is significant (at the p<.01 level) is total time for help. Of the three "type of help dummy" variables, mixed (that is ALS and other help) emerges as the second largest ß factor, but is not statistically significant.

The results for spelling are similar, with an even smaller adjusted R square of 0.027. The table shows the ß coefficients:

Table 4.11.2: Results of linear regression to predict annualized spelling gain from school, child and delivery factors

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-86.375	53.090		-1.627	.108
	School's social advantage factor	9.300E-02	.192	.060	.485	.630
	Number on school roll	-1.18E-02	.032	-.057	-.364	.717
	total score on SEN ratings	1.133	.618	.241	1.833	.071
	ALS vs not ALS alone	4.069	4.448	.140	.915	.364
	Mixed vs ALS or other alone	6.440	3.553	.276	1.813	.074
	total time for help	3.396E-02	.032	.152	1.072	.288
	staff status teaching prog 1	10.586	8.001	.183	1.323	.190
	Spelling age at start	.229	.236	.131	.969	.336

a. Dependent Variable: Annualised SA gain

Here the order of factors is different, with total time for help relatively unimportant, while SEN ratings are more important. The ALS & other method again emerges as the stronger factor, and staff status is the third most important. SEN ratings were the closest to statistical significance, but none of the factors achieved the lower level of significance.

The results of these tests mean that we cannot make useful predictions about reading and spelling gains from the set of background and intervention variables we have examined.

The relative strengths of the variables help to summarize what previous analyses have already suggested. Reading and spelling results were different, and for reading intervention variables, especially total time and one of the "dummy" intervention variables were strongest. Perhaps surprisingly, initial reading level was quite weak, as was staff status (perhaps because there was so little teacher input). School background factors were also very weak. For spelling, the pattern was somewhat different, with one of the intervention variables still quite strong, but pupil qualities more important than in reading and total time less important. Initial spelling level was stronger than initial reading was in the other comparison. Staff status also becomes more important in spelling. Background school factors were again relatively unimportant.

For both reading and spelling, interventions were therefore relatively important (bearing in mind that the total variance explained is small). For reading, the amount of input was also quite important, but pupil qualities were not. For spelling the reverse was true. Initial reading level was less important for reading gain than initial spelling level was for spelling gain. Teachers made more of a difference in spelling.

Chapter 5: Discussion of results:

There are three main areas for discussion:

- A commentary on the results obtained and implications for our understanding of the effectiveness of interventions;
- In the light of the results, the justification of the research methods and design is examined, whether changes might be expected to answer questions which were not well answered here, and what are the implications for further research;
- Are there implications from this research for educational policy in England and Wales, and for the evaluation of interventions?

5.1 Commentary on the results:

5.1.1: Which is more important, class teaching or additional interventions?

The first conclusion drawn by Brooks was that "Ordinary teaching ('no intervention') does not enable children with literacy difficulties to catch up." (Brooks 2002, p17).

The present study specifically included results from all children in most of the participating classes. The interventions used in this study were more typically wave 2 than 3 (although see below for further discussion of the 3 waves model). The overall outcome results show that children receiving no additional help made better than expected gains (see tables 4.2.1a and b).

The group who were below average at the beginning of the year but who did not receive extra help made more progress than those who did receive

extra help. This is direct evidence and it contradicts Brooks' claim. The above average group (who received no extra help) also made more progress.

Further analysis of the relative importance of classes and interventions (tables 4.4.4 to 4.4.6) led to the conclusion that classes make considerably more difference than interventions. This is perhaps the most significant and surprising result of the research and needs to be considered carefully.

It was suggested in Chapter 2 that we should not too readily give up on the possibility that the quality of mainstream class teaching can be further improved (2.3.2). Unfortunately, apart from the work of Solity and his colleagues (Solity at el 1999, 2000) there has been very little research in England and Wales intended to investigate the scope for improvement in the effectiveness of mainstream teaching upon literacy. Solity's research did not include evidence of the effects of additional interventions with children (who were at Key Stage 1). Research on the effects of high quality teaching on literacy at Key Stage 2 (such as the major study by Wragg et al 1998) has not considered the effects of additional interventions. New research from Texas (Mathes et al 2005) appears to be the first to examine the interaction between classroom teaching and interventions. It shows that interventions can add significant value and enable most children identified as "at risk" to catch up in their first year of formal schooling, where class teaching is of uniformly high quality, and

interventions are delivered in withdrawal by motivated and specially trained teachers.

Most children spend the great majority of their time in their classrooms doing work organised by their class teacher. Only about 10% of the school week (in the present study) is spent working on additional interventions for literacy (from table 4.10A.1), and then only some help is delivered by withdrawal. A considerable amount of teaching which is not labelled as literacy but which does involve reading and writing actually takes place during the 90% of the week children spend in class. It is therefore plausible to suggest that there may be more variability and perhaps scope for improvement within class teaching than in additional interventions.

Two alternative explanations seem possible:

1) The quality of interventions is relatively uniform, but the quality of class teaching is more variable; children are likely to make some progress as a result of interventions but they only make good progress if they also receive good main class teaching;
2) The quality of interventions depends mostly on how well they are coordinated with class teaching; interventions make some difference but they make much more difference if what children learn through their additional interventions is effectively linked with what they learn in class.

These two alternative views could be described as 1) the class teacher only hypothesis, or 2) the coordination hypothesis.

Whenever the effects of class teaching and interventions were compared in this study, differences emerged between reading and spelling. The effects of class on reading and spelling gain were highly significant, but whether or not children received extra help did not have much effect on reading gain, but it did more clearly effect spelling gain (although not quite enough to reach statistical significance, see tables 4.4.4 and 4.4.5). Class teaching may, therefore be capable of making more of a difference to spelling than reading gain. Perhaps spelling is a more "public" activity, in which what you know can be seen (in what you write, and in the results of spelling tests) while reading is more a matter for individuals and small groups (children less often read aloud to each other). Spelling is perhaps more prominent in work which is not labelled as "literacy".

There were differences in the interaction terms in the two-way analyses of variance that show major class effects, but they were relatively small. These results suggest that the class teacher only hypothesis is more likely than the coordination hypothesis.

A great deal more evidence is needed to support and substantiate these conclusions. There are currently no very satisfactory ways to measure the quality of class teaching in literacy (pace Wragg et al 1998, see Ch 2), or the quality of additional interventions, other than through outcomes.

Relevant and important indications of the amount of coordination in content (words studied, for example) and method (reading strategies) would need to be identified, before the coordination hypothesis could be properly tested. Perhaps the extent and content of communication between class teacher and TAs would be a good starting point. On the other hand, the extent and importance of classroom differentiation (which was explicitly ruled out of consideration in this study) may be a major factor in how effective class teachers are in helping children across the range of literacy abilities.

5.1.2: Comparing different interventions, and different types of intervention:

There were no significant differences between ALS, ALS combined with other interventions, and other interventions alone. Other types of help were a little ahead, particularly in spelling.

Simple comparisons between classes (using the Additional Help database, see table 4.6.2) suggested that ALS alone or in combination with other materials might be best in the majority of cases for reading. For spelling, ALS seemed not to have any advantage; if anything, other methods alone or combined with ALS seemed more likely to be better.

However, comparisons between starting reading and spelling levels shows that children were selected for interventions according to their initial reading and spelling levels. Children with lower reading and spelling levels

(on both types of measure) were selected for "Other only" types of help, while children with better reading and spelling levels were selected for ALS or ALS and other help combined. Differences between ALS and ALS combined with other interventions were so small that there was no obvious prior attainment rationale for the selection.

Of course, this finding is not surprising. ALS was intended by NLS to be used with children who had fallen behind but who did not have major literacy difficulties. It seems that most of the schools in this sample were following the original NLS plan, although some were supplementing ALS for some children.

In terms of reading and spelling gains, then, ALS alone or in combination with other methods is most effective in only some class situations (perhaps the majority, but we cannot be sure from this data). Its effectiveness is also qualified by the policy of selecting children who have the least learning difficulties for it. However, in value for money terms, ALS is most commonly delivered in groups of 5 or 6, while most other interventions are delivered 1:1 or 1:2. When outcomes were reanalysed in terms of gain per man hour of help given (table 4.10A.6) a striking advantage emerged for ALS alone, for both reading and spelling (with a substantially greater effect size for reading). ALS was almost entirely delivered by TAs, with very few schools using the original model of 1 session per week delivered by the class teacher. TAs are paid at between half and a third the rate of qualified

teachers. The cost-benefit of ALS is therefore particularly strong, even if it is not clearly more effective in reading and spelling gains alone.

Faced with an unexpected variety of programmes, with children receiving combinations of up to 6 programmes, it was impossible to reliably separate which components of these combinations were most effective. The actual classification into 6 types was itself crude, and needs further consideration (see below). However, because the results offer comparisons across types of programmes which are usually compared even more indirectly, they are worth considering. In table 4.10B.2, there are no clear advantages in reading gain for any type of programmes. The two highest averages are, surprisingly, for individually designed programmes and reading scheme based programmes. By contrast with Brooks' findings, computer based interventions do not score especially well. However, the picture for spelling (from table 4.10B.3) was different: surprisingly, computer based schemes were rather unsuccessful, while specific spelling programmes were successful. This finding tends to support Brooks' conclusion (which I argued is not well related to his evidence) that "highly structured schemes work best" for spelling.

5.1.3: Why did differences in children's qualities have very little bearing on outcomes?

The consistent finding that class differences are more important than differences between methods of intervention might have been a result of

differences between children. In other words, results from classes may vary because they are composed of children with different qualities.

The qualities of children were examined using gender and ratings of the children's special educational needs (SEN). Examination of the distribution of SEN variables (table 4.8.3) suggested that the amount and type of SEN in the sample was as expected from other surveys and typical of direct experience. But reading and spelling outcomes from initial, final or gain scores had very little relationship with SEN scores, with some relatively minor exceptions (spelling age gains were associated with language delay or disorder and poor attendance, table 4.8.4).

The absence of relationships between SEN qualities and outcomes meant there was no point in comparing outcomes between the three main types of intervention and SEN. However, there were differences between the groups selected for the three types of intervention by SEN. As Chart 4.8.5 shows, the small number of children with most severe SEN were placed mainly in the Other help group. This difference is probably related to the differences in selection between the three types of intervention. Those with more severe SEN tend to receive more individualised help. There were also small but significant differences between classes in the amount of SEN they contained. Ratings of SEN were made by teachers and TAs who worked with the children, so it is possible that these differences are a result of differences in the way ratings were made, rather than in the children themselves. Independent observations of at least part of the

sample would have been needed to rule out such differences. However, it would be surprising if there were not some differences between classes in the extent of SEN. Experience in schools suggests that variation in the extent of SEN is normal.

The absence of effect of SEN ratings upon gain outcomes is consistent with the apparent absence of interactions between methods and differences between children (Adams 1990, Stanovich et al 1997, Vellutino et al 1997), but Fletcher et al 1997 and Snowling and Griffiths 2004 have continued to raise doubts about this conclusion.

Can this finding be taken as support for the typical finding that there is no clear evidence for interactions between differences between children and methods of intervention?

There seem to be three possibilities:

- Official SEN classifications are after all more relevant than teacher ratings, and if they had been used differences between interventions would have emerged more strongly;
- The SEN ratings used were not sufficiently comprehensive, and some (missing) qualities might have explained more of the differences between interventions;
- SEN qualities, however described, are not of much importance in understanding the effects of interventions on children with literacy difficulties;

The first explanation seems implausible, because of the substantial variations between official SEN ratings between areas in England and Wales, and because of the chance factors that seem to influence whether a child is classified as School Action, School Action Plus or Statement, but it cannot be ruled out by this study. It would be valuable in any case to compare official SEN levels with the 14X3 ratings used here, as a means of examining the relationships between SEN measures and teacher ratings.

It may be that more severe difficulties have longer lasting effects, and therefore that a study such as this with a one year scope covering children across a broad but not complete spectrum of severity of difficulty was unlikely to pick up much effect of difficulty upon literacy gain. No children in special schools or units were included. It remains unclear to what extent generalisations about literacy difficulties for the majority apply to the quite small numbers of children with exceptionally severe difficulties. But the focus of this study is upon the effectiveness of interventions for the broad majority. The result that children's qualities do not seem to have much effect upon the outcomes of interventions may remain true for all but the most extreme difficulties. Analysis of outcomes by broad type of help did show that a small number of cases involving more severe difficulties were significantly different (section 4.8).

The second possibility is that some qualities of children which are important were not rated. Muter and Snowling (Muter & Snowling 1998) have identified key factors at age 9 which are strongly associated with reading achievement : phonemic awareness (but not rhyme awareness), non-word repetition which Muter and Snowling argue is closely related to speech rate, rather than verbal short-term memory, and grammatical awareness. Some of these difficulties are associated with qualities that were rated here. Attention and concentration would be expected to be associated with short-term memory. But it was not related to gain outcomes. It is possible that specific measures would be needed to tap key qualities of children at 7-9, such as those measured in the Phonological Abilities Test (Muter,Hulme & Snowling 1997).

The strongest reason for not including direct measures of short-term memory, phonological difficulties, and grammatical awareness was that they could not be rated reliably by teachers, without performing additional tests. Tests would have been expensive to conduct. Unless every child in all the classes involved (those with literacy difficulties and those without) were tested, the possibility of a Hawthorne effect would have been introduced. However, it is important to try to establish whether children with literacy difficulties respond differentially to different types of intervention, and so cost effective and minimally obtrusive measures of these missing ratings are needed. Research currently suggests that the most likely finding would be that severity of phonological difficulty would be the strongest indicator of poor responsiveness to interventions (Stanovich

et al 1997, Velluntino et al 1997). However, the very recent evidence from the Texas study of interventions and class teaching (Mathes et al 2005) suggests that there may be no significant interactions between methods of intervention and children's qualities. Comparison of the Proactive programmes, which was a scripted behaviourally based programme using a "bottom-up" methodology, and Responsive, an unscripted meta-cognitively based programme using a more "top-down" approach, did not find that the expected differential benefits for children with greater phonological or vocabulary and comprehension difficulties.

These American results were obtained in first grade students. We cannot rule out the possibility that there is an interaction between interventions and children's qualities in English Year 3 and 4. So we cannot validly conclude that children's qualities are unimportant.

5.1.4: How important is the school?

This study did not set out to examine in detail whether schools as units are able to significantly influence the effectiveness of interventions for children with literacy difficulties. Experience in education suggests that there are differences between schools and that some schools are better able to help children make progress than others. It may be that there are some exceptional schools which are very effective in this aspect of their work. Most schools as such probably do not have a strong effect upon children with literacy difficulties; rather, class teachers are what make the difference.

As expected, small schools in this study were associated with smaller class sizes, higher social advantage and lower SEN ratings. But the only outcome measure significantly associated with school size was spelling. Class size did not appear to have any effect independent of school size. This result is not a strong confirmation of the findings of Blatchford et al (2003), but it is consistent with their finding that once children pass Reception and Year 1, class size effects on literacy outcomes cannot be detected in English classrooms.

By comparing effect sizes in separate analyses of variance, classes appear to have greater influence than schools upon outcomes in reading and spelling. Multi-level modelling would be invaluable, across a larger sample of schools, in clarifying the extent of school effects on literacy interventions.

School processes which might mediate greater effectiveness include the role and time available to the school's Senco, assessment procedures and data collection and analysis, the levelling of reading books and other materials, working with parents, the effectiveness of class teaching on literacy, communication between class teachers and TAs about children and programme content, choice of intervention programmes, how much time is allocated to interventions, and aspects of school ethos which might affect children's motivation to overcome difficulties and improve their skills.

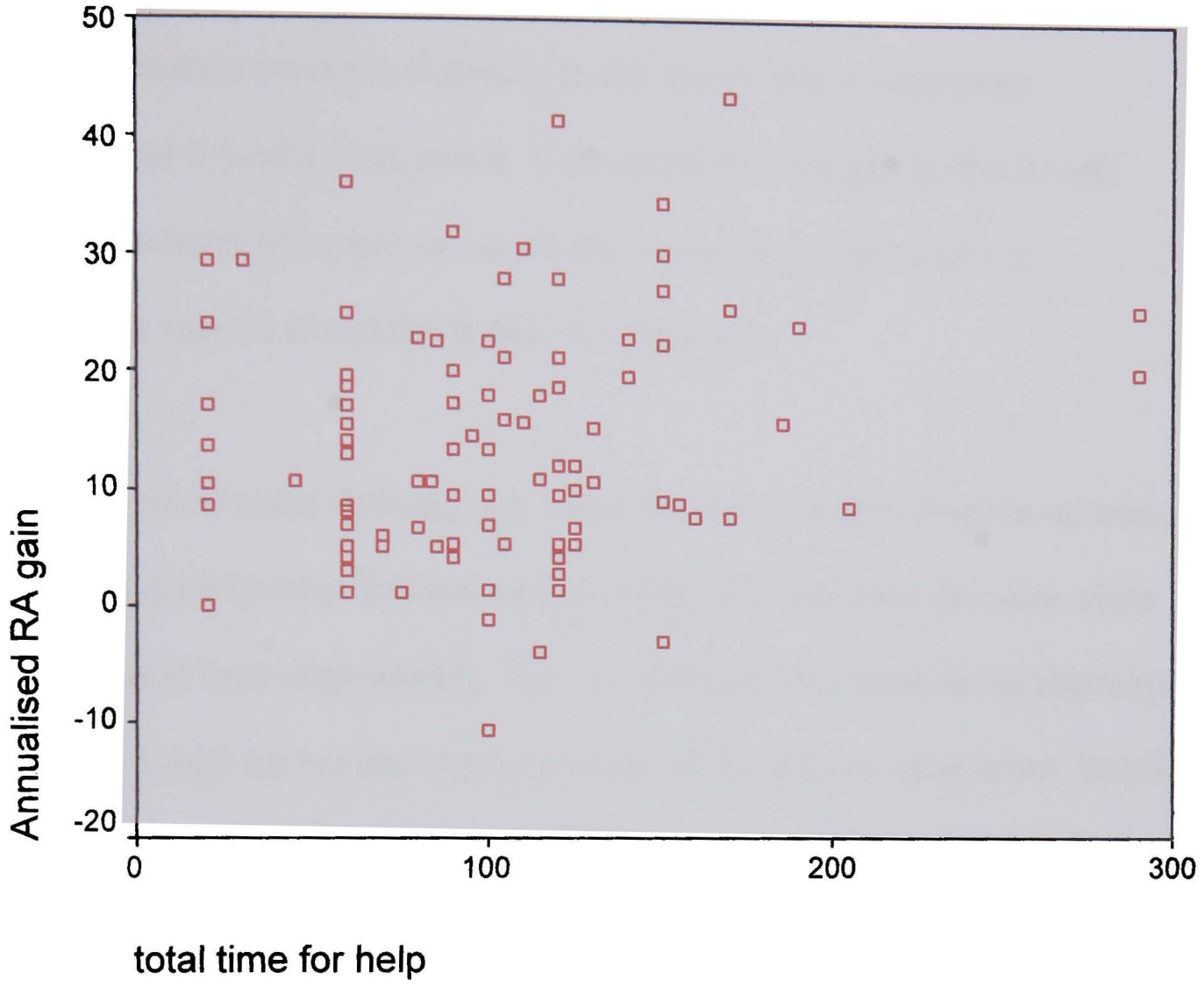
5.1.5: How important are summary quantitative measures of additional help? Do location, number of programmes and time per week matter?

If additional help is to be given to children, it is not only important to identify the most effective programmes, but also to identify the optimum amount of help, the best location (whether within the classroom or in a separate room) and the most effective frequency for delivery, daily, 3 or 4 times a week, or weekly.

This study found a range of help from 25 minutes per week to 200 (with some exceptional cases of 300 minutes). There was a small but significant relationship between amount of help and reading gain but not spelling gain. The difference between reading and spelling in this respect is similar to other aspects of the difference – spelling seems to be more to do with class teaching, while reading depends more on additional help.

Would even more help have led to even more progress? A scatterplot of reading gain against amount of time suggests not:

Chart 5.1.5.1: Annualized reading gain against total time for additional help in minutes per week:



The relationship between more time and greater progress is not strongly evident from Chart 5.1.5.1, which implies that the amount of time given is not necessarily the most important factor. This is consistent with the finding from other studies, notably the US National Reading Panel (NRP 2000) and Brooks (Brooks 2002), that longer interventions do not lead to greater gains.

Different aspects of delivery interact – the number of sessions, how they are spread, how many programmes are used and how and what combinations of programmes are deployed. So it is extremely difficult to

identify any one factor separately. By making some assumptions about how multiple sessions were delivered, it appeared that there were differences between the different patterns (in table 4.10.1.5), with a quite striking difference between 4 sessions per week (which was least effective) and 5 (which was most). Commonsense suggests this is odd. Surely 4 sessions per week is nearly the same as 5. Because it is surprising, it seems to call for further investigation.

Withdrawal vs in-class delivery has been for some time in English schools a source of controversy. Delivery within class is in the view of many more "inclusive" and less stigmatising. Yet, for children who have some learning difficulties, it may be harder to concentrate within a busy classroom. Small differences between reading and spelling are again suggestive of the relatively greater importance for reading of extra help by withdrawal and of class teaching for spelling; but the clearest result showed that a combination of withdrawal and in-class help was probably the most effective. Here the figures do seem to support commonsense.

Research has paid considerable attention to group size. The complexity of interactions between aspects of delivery make this a factor which is difficult to isolate. But this study's results suggest that spelling outcomes are less affected by group size (apart from one very successful group of 9 children), probably because extra help is less important and class teaching more, while in reading 1:1, 1:2 and 1:6 are about equally successful, with the group of 9 again an intriguing leader.

These aspects of delivery have some cautious implications for teachers. The amount of time given makes a little difference to reading, the number of programmes does not, the spread of sessions over the week favours either daily or weekly, and group size favours 1:1 or 1:6 (or 1:9), but these relationships are not very clear. But when we look at efficiency, in terms of the amount of progress achieved in proportion to the amount of staff time given, at last we have a clear difference: ALS alone is three times as efficient as the other main types of help for reading, and twice as efficient for spelling. If ALS alone had been the least effective method of help, then this would have provided us with a major dilemma. But, depending on class differences, it is sometimes the most effective. So its efficiency gives good reason to prioritise efforts to improve the effectiveness of comprehensive, highly scripted small group programmes further and try to work out why they are not always as effective as they can be.

5.1.6: Does it matter that most additional help is delivered by Teaching Assistants?

These results suggest that it does matter. Additional programmes in spelling delivered by teachers were markedly more effective, while those for reading achieved better outcomes but the differences did not reach statistical significance.

The great majority of staff involved in additional help were TAs not teachers. Our finding raises questions about the policy of using TAs to

deliver additional help. It may be more cost effective (because of the large differential between teacher and TA salaries) but in some circumstances it may not lead to the greatest gains. Small groups led by teachers with additional training may be the most effective way of delivering some kinds of help.

Two other important qualities of staff also mattered. Training did make a difference to outcomes, more so in spelling than reading. And years of experience also mattered, but not in the expected way: the greatest experience was associated with poorer outcomes. The difference was not great but suggests that perhaps some element of youth or enthusiasm is a little more important than age alone.

5.1.7: Why were the variables used in this study relatively powerless to explain reading and spelling gains?

The conclusion that class teaching was probably more important as a source of progress in reading and even more so in spelling than additional help was unexpected. If it had been anticipated, it might have been possible to obtain more information about qualities of teachers and of the mainstream teaching programmes they used. This might have helped to explain why the variation between classes was so strong. Instead, the focus was on broad aspects of children, schools, the delivery of interventions and staff qualities.

Data about interventions was collected on the Additional Help database. The findings therefore relate only to children who were receiving recognised additional help. Extra support in class through additional TA support and through differentiated programmes of work for below average children were not recorded, and might have played a significant role.

The findings for children receiving extra help did nevertheless contain some important surprises, especially that the SEN ratings used seem to have little correlation, either as a summary score (crudely, the "amount" of SEN a child has) or as a series of measures of severity of particular types of difficulty, with outcomes. Three possible explanations were suggested, and all probably merit further consideration, but the most likely seems to be that measures of qualities of children that research has identified as particularly important to literacy difficulties were not used.

Differences between schools were measured in very broad ways, using numbers on roll, class size and Social Advantage Factor. It is a possibility that schools might have effects on the outcomes of interventions over and above the effects of class teachers: what kind of additional help and how much is given depend more on the school than on the individual teacher. But for the small sample available in this study such effects did not seem as strong as class effects. Perhaps a greater school effect might occur in some schools, but this might itself be an uncommon phenomenon.

Some qualities of intervention programmes were related to outcomes. Total amount of help was significantly related to reading gain, but not spelling. The pattern of sessions over a week seemed surprisingly to favour 1 or 5 sessions, with 4 particularly low. For reading withdrawal was more effective than work in class, but for spelling the reverse was true; for both reading and spelling a combination of withdrawal and in class work was more successful. And group sizes had some relationships to outcomes, with groups of 9,1,2,and 6 appearing best.

Staffing also mattered: although there were very few teachers involved in delivering help, where they did better outcomes were seen. Training was also important, and perhaps surprisingly, years of experience were not necessarily related to better outcomes.

The final set of results using linear regression found that, of the variables entered, for both reading and spelling, interventions were the most important (as measured by their beta scores). However, it has to be remembered that the effect sizes were both below 0.1, which is small, and that the effects of classes could not be entered into this type of analysis. Multi-level modelling would have been needed to establish the relative importance of school, class, group and individual child factors.

The effectiveness of interventions in relation to background child, class and school factors is not well understood. It may be that differences between interventions, especially at the middle primary stage, depend to

only a very limited extent on correctly identifying differences between children's qualities, or on differences in the qualities of schools in which they are educated, or even in differences in the classes in which they are taught most of the day. Instead, interventions for their success may depend mainly on their own design qualities, their content and structure, pace, interest and ease of delivery, and hence to issues in a more general theory of instruction (for example, Englemann & Carnine 1981).

I have suggested that the effects of class teaching and additional interventions may either add together, or interact. The results of this study suggest that they are more likely to add, because the evidence from two-way analysis of variance pointed to quite small interaction terms. If the effects of interventions are relatively constant, then variations in the value added by class teaching may explain the main results of this study.

5.2: Methodological Issues:

From a methodological point of view, there is an inevitable tension between naturalistic "quasi-experimental" and "experimental" methods. This is not only a matter of differences in allocation of participants to groups (random or otherwise), but also of costs and ownership. In this study costs had to be minimal, and schools could be asked but not required to participate. This meant that additional data collected had to be as simple and easy to collect as possible. We must ask whether the limits of such an approach can be stretched to overcome the shortcomings in this study, or whether other approaches are needed. We consider

- The appropriateness of annualization of data;
- Problems of harmonization of outcome measures;
- Sample size;
- Assessing pupil qualities;
- Assessing school and class teaching qualities;
- Delivery factors, and drawing a line between differentiation and additional interventions;
- Assessing the qualities on interventions themselves;

5.2.1 Rates of progress and the appropriateness of annualization

It was inevitable in a naturalistic study such as this that there would be some variation in the intervals over which gains were measured. Interval standard deviations were relatively small, but the maximum and minimum figures show that there were cases in the reading measures where intervals of more than 12 months were used.

It is likely that the resulting annualised gains exaggerate the actual gains made, because children often seem to regress after the summer holiday, and most intervals were from September to June or July. For the children whose reading interval was more than 12 months, annualization actually reduces gains for these children. But it may have further reduced their gain, since the summer holiday relapse is already present in their data. Fortunately there were only two cases based on intervals of more than 12 months, both at 14 months.

It was necessary to annualise reading and spelling ages in order to make fair comparisons between children measured over slightly different intervals. The likelihood of summer holiday regression suggests it is most realistic to quote overall gains in terms of actual gains. But actual gains were above expected levels for both children who received help and those who did not.

5.2.2: Can reading and spelling outcome measures be harmonized? What are the most appropriate measures of reading and spelling progress?

Until very recently and during 2002-03 when this data was collected, National Curriculum measures were based on materials newly developed every year. From 2005, Key Stage 1 assessment is based on teacher assessment, with tests or tasks used to support these judgements. New or past year's materials can be used. Optional assessments for years 3 and 4 use prescribed tests which are now intended to be reusable. The contrast between standardised tests which could be used repeatedly with the same or different children, and NC tests which were different every year, is now reducing.

The problem with standardised tests is that their norms become out of date, especially if national standards change (as they did in the first two years following the introduction of the National Literacy Strategy). Age norms therefore become unreliable, although differences between one test administration and the next should remain valid. In other words, gain measures should not be affected.

In this study, overall gains for all children on standardised tests were (from the Classes database) at above expected levels: average reading gain was 13 months in 9 months, and average spelling gain was 15 months in just under 9 months. Comparisons between groups remain valid because any errors should be randomly distributed. But there is no reason to suppose that the children in the sample were achieving so much above national averages. It is more likely that the test items and norms have become outdated.

The problem with NC tests was that they changed every year, and there was substantial suspicion that the difficulty levels of the tests were capable of adjustment to show progress. NC tests are now moving towards stability over time, at least from Year 2 to 4. Ideally, reporting primarily in NC level terms needs to be replaced by reporting in standardised score terms. It should be only a technical matter to convert standardised scores into NC levels. Indeed, if the figures were treated in this way, normal error measurement statistics could be used to demonstrate how reliable NC levelling is.

On the other hand, individual or group standardised tests now need to be restandardised with high and predictable associations with NC tests. They continue to be useful, because they are quicker to use and the variety of test formats and purposes mean they can be used diagnostically. Ideally a study such as this would continue to use both group and individual tests of

reading accuracy and comprehension, and would be able to compare differences in outcomes in different types of test reliably and easily.

5.2.3: Was the sample large enough?

Tables comparing outcomes by classes (tables 4.4.3, 4.5.2 and 4.6.2) show that even with 19 classes participating (in the Additional Help database) and 12 (in the Classes database), the number of comparisons between methods by classes is quite small. The sample was drawn from a single large Local Education Authority, in which it is likely there has been some narrowing of the range of experience of interventions, though shared training and contact with neighbouring schools.

A larger sample (of 30 schools, perhaps including up to 60 classes) was originally intended, but a high number of schools dropped out or delivered unusable data. Except at the extremes, gross school differences (Social Advantage and size) were relatively unimportant. This study has suggested that differences in the quality of class teaching may have been more important than any other factor in influencing the effectiveness of annual gain scores.

A study involving a number of schools serving the lower to middle SES range, in which intervention strategies varied from reliance on additional small group interventions to mixed and mainly individual interventions, would probably not need to involve more than 18 schools (3 levels of SES X 3 intervention strategies). Class teaching qualities could be best

controlled by extending the study to two years, so that data on the outcomes from most teachers could be averaged, and most children could be followed in two different classes, with some receiving help in one year and not the next, and some receiving help over two years.

5.2.4: Assessing pupil qualities:

Ratings of pupil qualities were designed to be as simple and reliable as possible. Brief guidance notes were provided, and each scale asked for judgements on which of only 3 levels to place a child. Completion rates suggested that there was very little difficulty in reaching judgements. However, the rating scales were piloted only in a single study (Bunn 2002), and no data on inter-observer reliability were available. It was therefore not possible to exclude systematic variation between raters or schools, and the comparability of levels of SEN between schools was unproven.

The rating system was derived from a large scale study of SEN categories and systems in Norway (Skårbrevik 2001). The version used here was trialled in a small scale study examining outcomes of children with statements of SEN in a single area (Bunn 2002). The range of SEN and of outcomes was much greater than in the present study. It may therefore not have had the sensitivity to deal with differences in a study mainly of children requiring wave 2 interventions in mainstream schools over a single year.

As has already been discussed, the most likely omission was any ratings of phonological, rapid naming and verbal short-term memory differences. These qualities were not included because it seemed unlikely that ratings about them would have been reliable without the use of tests. Perhaps the most likely test to use in England would be the Phonological Abilities Test (Muter, Hulme and Snowling, 1997).

A first step would be to examine how accurate teacher and TA judgements of children's phonological abilities are in comparison with test results, and then whether some further observations short of actual testing could be used to improve accuracy. An intermediate possibility would be to examine whether some form of group of individual computer based assessment could be used instead of a more costly individually administered test.

Striking the right balance of costs and benefits between structured observations (teacher assessment) and tests is a very familiar dilemma in English education. The optimum in this case is probably to make tests and more detailed classroom observations available for some judgements. It is likely that the development of a reliable system for rating pupils' qualities would have other benefits, especially in assessing longer term cost-benefits of the whole range of SEN arrangements and intervention strategies, and more broadly as part of the "Common Assessmnent Framework" initiative. Substantial investment in further development of a system for reliably rating children's qualities would therefore be worth considering seriously.

5.2.5: Assessing school and teacher qualities

School qualities were assessed only through very broad, background measures, such as Social Advantage Factor and number on roll. The results have suggested that class teaching, and therefore perhaps whole school factors, may be more important than anticipated. Better measures of teacher and school qualities may be needed.

There has been widespread support for the use of "value-added" ratings of schools' performance since before school performance league tables were introduced into England & Wales. Unfortunately, the system introduced by the government of England & Wales in January 2006, "contextual value-added" (CVA) is controversial (Times Education Supplement 12th May 2006), because similar systems, Fischer Family Trust and Yellis, can reach very different conclusions about the same schools. At present, it may be necessary to use at least 2 of these measures.

The current school inspection arrangements expect schools to complete a "self-assessment" process (OFSTED 2006). At the primary stage this does not contain specific references to literacy development, except in terms of SATs results and overall attainments targets. However, it may be possible to offer schools a "literacy self-assessment", which would cover both main class teaching and additional interventions, and other crucial factors, such as parental involvement, use of assessment data, size and use of school

and class libraries, training and support for literacy teaching. Schools might consider that it would be in their interest to have completed the "literacy self-assessment" as a way of demonstrating a willingness to evaluate and if necessary develop the quality of what the school offers to "learners".

If the assessment of schools is a sensitive issue, the assessment of teacher quality is even more so, and data is very hard to collect. Typically studies which involve the quality of mainstream literacy teaching focus only on schools or teachers expected to be of high quality (eg Wragg et al 1998, Mathes et al 2005). It is possible that some form of moderated self-rating of effectiveness in literacy teaching might be acceptable, but anxiety about access to such information would be high. Results in the form of reading and spelling gains over more than one year, compared with national and school averages are likely to be the most easily available data. It might be possible to add some contextual data at the pupil level from questionnaires for parents, such as the PIRLS study used (Twist et al 2003), and some data about choice of interventions, time allocation to interventions, communication between teachers, TAs and the school SEN Coordinator by questionnaires to teachers. SEN data was available in this study, and with some further development (see above, 5.2.4) could also contribute substantially to class-level value-added. Open-ended questions about methods of class teaching were used in this study, but were not completed consistently enough to be useful. Structured interviews may be a more reliable method for gathering this type of data.

5.2.6: Delivering additional interventions: how should the boundaries between additional interventions and classroom differentiation be drawn?

This study showed that a number of aspects of the delivery of additional interventions were important. Combinations of programmes presented serious problems in analysing the effects. It would be desirable both to seek user judgements about the relative importance of programmes from a multi-programme mix as a way of ordering the input of additional programme data more rationally, and to rotate programme order systematically to clarify the extent to which order makes a difference to overall gain results. This means that the results from tables 4.10.2.2 and 4.10.2.3 would be resorted systematically to check what the effects of placing, say computer based programmes in programmes 1, 2, 3, 4 or 5 were.

This study explicitly ruled out classroom differentiation as an "additional intervention". Additional interventions were defined as those involving additional staff with nominated pupils over a significant duration where an Individual Education Plan had been written. They did not therefore include support for learning in classrooms or arrangements organised by the class teacher involving variations in materials or levels of difficulty. Differentiation is notoriously difficult to define reliably (Weston 1992)) and is probably better seen as part of class teaching than additional interventions. Grouping, support strategies and differentiation are probably best assessed as part of a structured interview with class teachers and

TAs. In other words, they are part of the "classroom processes" which mediate learning (as understood in Blatchford 2003), rather than distinct inputs or outputs. Although in practice the borderline between additional interventions and classroom support may not always be clearly drawn, it is probably as significant as the borderline between learning at home and in school, and should therefore be maintained.

5.2.7 How should we analyse intervention programmes?

It would also be valuable to be able to understand more about the features of programmes. There is some evidence that, given certain essential ingredients, the actual mix within intervention programmes is not as important as expected. Both Torgeson (Torgeson 2001) and Mathes et al (2005) found in comparisons between well-constructed interventions with substantially different proportions of direct instruction on phonics and supervised practice in reading continuous text that there were no major differences in effectiveness between the interventions. Mathes et al (2005) suggest (p179) the list of key ingredients (at first grade) should be: explicit instruction and practice in phonemic awareness, decoding, fluent word recognition and text processing, spelling, and comprehension strategies applied to connected text. They argued that "if schools are allowed to choose from among the most effective choices an approach that best aligns to personal philosophy and theory, then there is likely to be less resistance, higher quality implementation, as well as sustainability over the longer term."

The Brooks meta-analytic study (Brooks 2002) was not able to go beyond the same kind of broad features we have considered. Only the NRP studies in the United States have attempted to go further. In the study on Phonemic Awareness (PA) training, differences between training in one, two, three or more skills was analysed. In the Phonics Instruction study, differences between analytic and synthetic phonics were examined (albeit without a conclusive result).

This study attempted 2 further approaches to programmes evaluation: a breakdown of programmes into 6 broad types (table 4.10.2.1), and user ratings by staff and children. The children's rating were not found to be very helpful in this study. This does mean that they do not have uses, particularly at the school level. And it may be salutary to take this kind of information into account, especially because the children rated unfashionable reading scheme programmes as best in ease of use. The staff ratings were valuable partly by confirming some of the children's judgements about ease of use and interest, and partly by suggesting why some programmes may not have been as effective as hoped. For example, the low "ease of use" rating for ALS may suggest ways it could be improved. Attempts to improve the effectiveness of programmes probably need systematic staff ratings. Ideally ways of rating programmes need to be developed which are as widely applicable as possible.

The breakdown into 6 types of programmes remains broad. Many different programmes fall under each heading, especially "comprehensive literacy"

programmes and "individually designed" programmes. The results of the further analysis were at best suggestive, mainly because each was one among several interventions used with each child. Better evaluation of programmes requires more rigorous feature analysis, for example synthetic vs analytic phonics, scripted vs adaptable delivery, percentage of time on supported text reading.

5.2.8: What would an experimental design have achieved that a quasi-experimental design could not?

It was argued in 3.3 that a true experimental design for this research would have required significant funding (to pay for interventions) and would still have raised significant ethical and methodological questions. Analysis of the shortcomings of the research and how it could have been improved does not suggest that the choice of a quasi-experimental design was mistaken. Indeed, if a larger sample size and data on the schools' and especially class teachers' effectiveness had been obtained, the costs of an experimental approach would have been higher. Random assignment of children to interventions would perhaps have produced greater differences between outcomes from interventions, but questions would have remained about the appropriateness of some interventions to some children (ALS was explicitly designed for children at wave 2 without significant SEN). In other words, questions of validity would have reduced the utility of the results, even if they were sharper. The gains from a true experimental design would have been far outweighed by the costs.

5.2.9. Studying literacy interventions in context: two directions for future research

Some additional types of data could have been collected, and some data could have been collected in different and more effective ways than was actually done. Putting these additions and corrections together, there seem to be at least two possible directions for development. The first involves extending and improving the present study, and the second conducting a different type of study to address some of the questions to which no answers are currently available.

The study set out to compare the effectiveness of ALS and other interventions for children with literacy difficulties. The results from this study suggest that in some (perhaps the majority) of class contexts ALS is able to produce the greatest gains in reading, and that in value-for-money terms, ALS is clearly ahead of the alternatives with which it was compared. A further study of the same type would need to seek a more stable result: is it possible to be more definite about the effectiveness of ALS and other interventions?

This study also reached an unexpected conclusion about the relative importance of class teaching and individual interventions, but the data available was not sufficient in quantity or quality to give firm answers about whether the quality of class teaching is more important than additional interventions. At best a tentative theory has been suggested, that class

teaching and additional interventions both contribute separately to literacy gain. This theory needs to be further tested.

The first step would be to present the results obtained in this research to the participating schools. This would involve discussion with key staff about how some of the weaknesses identified could be overcome, and about what further or different data might be collected.

A number of technical improvements have been suggested. The main conclusions from the discussion in 5.2.1- 5.2.7 are:

- Individual and class reading tests need to be restandardised in such a way that reading and spelling attainment scores from group and individual tests are fully comparable;
- Schools should be asked to carry out beginning and end of year testing within common time bands to improve the comparability of gain scores;
- Scores on all literacy achievement tests would be best presented in standard score form for this type of research;
- The reliability of the system for rating pupil qualities needs to be checked and it needs to be extended to include at least some ratings of phonological processing abilities;
- A measure of the overall effectiveness of schools would need to be included, probably using a value-added measure;

- It may be possible to develop a "literacy self-evaluation" which schools could use as part of the overall self-evaluation processes they are obliged to carry out by OFSTED;
- It may also be possible to include some teacher self-evaluation as part of the school self-evaluation, although this data may not be accessible to research;
- A framework for describing and evaluating the key features of literacy interventions needs to be further developed, as a basis for comparison between programmes and combinations of programmes.

Perhaps the most important practical difference which would improve a study of this type would be to collect data over two years from the same classes and children. This would have considerable advantages: there would be more data about classes, children and teachers; average performances over two years would be available from children and teachers, and in some cases (where children remain with the same classmates) from classes. Data about classes in Year 2 would also need to be collected so that data about children in Year 3 in the second year of the study was available; and data about Year 5, so that children in Year 4 in the first year of the study were also followed for one year. But individual ratings and intervention data would not be needed for those children in Years 2 and 5.

If a slightly narrower range of schools were selected, excluding schools with single form entry (which would generally also exclude very small

classes) more comparisons within the same school would also be possible, and thus there would be opportunities, using multi-level modelling techniques, to compare class and school effects. Some range of SES would continue to be important. The overall sample size would probably need to be increased only to about 18 schools, which should ensure that enough comparisons between methods of intervention were available.

Another direction for research is, however, to examine more closely the processes involved. How do class teaching and additional interventions affect children's progress in literacy? Is it correct that there is little interaction between their effects? On the face of it, it appears more likely that there would be some important interactions between class teaching and additional interventions: if a TA is able to remind children in small group work of something explained in a whole class lesson, some strengthening of associations would be expected. What features of class teaching have most effect on progress for children with literacy difficulties? Can additional interventions complement class teaching, under some conditions (perhaps only if significant time is allocated to discussion between class teacher and TA)? These types of question can only be answered by a very different type of study, which examines the detail of day by day and even minute by minute interactions. This would involve a major qualitative element, perhaps in similar ways to the recent study by Peter Blatchford and his team reported in "The Class Size Debate".

The Blatchford study suggests itself as a model in a number of ways: children were followed over three years in different schools and LEAs. Attainment data were collected using a combination of group tests of reading and maths and specially adapted protocol for recoding raw scores on SATs tests on a continuous scale. Background data such as gender, free school meals status, additional language and SEN status were collected for each child. Quantitative data were also collected on class grouping. Complementary data was also collected from questionnaires and semi-structured interviews, and case studies of smaller sub-sample of the classes using whole class and individual child observations using experienced teachers as observers. The study deliberately combined quantitative and qualitative data in order to try to answer its key question, "does class size matter?" but it also tried to explain what processes within classes mediate differences in class size in their effects on attainments. It seems likely that a similar combination of methods would be more likely to answer questions about the effects of class teaching and additional interventions upon attainments in literacy.

5.3.1 Implications of this study for national education policy;

The review in Chapter 2 of national policy development in England and Wales on literacy difficulties suggested that successive governments adopted a rational policy of dealing with the most far-reaching options first (the National Curriculum), and in successive stages have narrowed the focus of change from class teaching through to individual interventions at wave 3. It was suggested that there may now be some danger of loss of

focus, because further improvements on a number of levels remain possible.

This study was intended to evaluate one aspect of the spectrum of national policy initiatives, the use of a small group intervention largely delivered by TAs for children at wave 2. The results show that ALS may not be significantly more effective than alternative interventions in terms of gains in children's' attainments, but it is significantly more cost effective. It is not significantly less effective in most classes. While superior effectiveness in gains and costs would have been desirable, the findings of this study show that this type of intervention delivered in small groups by TAs is worthwhile. Further development is therefore both justified and desirable. It was not considered very easy to use by TAs in this study. This study was not intended to examine which units or activities were most effective, but further analysis would be valuable as a basis for development.

The unexpected finding that the quality of class teaching may have had more impact on gains than additional interventions reinforces the questions raised in Chapter 2 about the importance of class teaching. In fact, recent guidance from the government of England & Wales indicates that class teachers are now much less restricted in the methods and approaches they are expected to follow in delivering literacy teaching in primary schools. This means that it should be possible to develop through research in collaboration with practising teachers models of class teaching

which are more effective than those available through the National Literacy Strategy.

The possibility that the development of class teaching might be more effective does not mean that the development of additional interventions is less important. Ideally a competition between developments in class teaching and the development of more effective interventions may be the most healthy way forward.

The analysis in Chapter 2 of the 3 waves strategy suggested that it was based upon 3 key assumptions, all of which this study has some bearing on. The first was that some children will not succeed through mainstream teaching alone. The new Texas study of the combined effectiveness of class teaching and additional interventions (Mathes et al 2005) found that enhanced classroom instruction (without additional interventions) was able to reduce "failure" at the end of Grade 1 (defined as a result below the 30th centile on the Woodcock-Johnson III Basic Reading test) by extrapolation to only 3% of the total population. With additional interventions a further 2% were helped to achieve above the critical level. The Texas study took place in successful schools, and its authors are clear that it will be important to see how its findings generalize to other contexts. 3% may represent an optimum for Grade 1. However, the important point for national policy here is that research and practice should seek to reduce "failure" through class teaching to a similar optimal low level. It is likely that as much can be achieved for most children through good class teaching

as through additional interventions, and that their effects supplement each other. It is therefore important to continue to research more effective ways of teaching children in main class lessons, and to disseminate through training the results of the research.

The second assumption was that children will "catch up" as a result of the use of small group teaching interventions such as ALS. In the Texas study most children who received additional interventions and enhanced class teaching did catch up, as the figures above show, but not quite all. The results of this study were much less positive. Class size figures were difficult to establish clearly and completely. The best estimates available from this data suggest that about 31.5% of children in 11 classes for which class size was available received additional help. Extrapolating from this figure, 14.1% of all children in the population had a reading age more than 12 months below expected level, and 16.9% in spelling at the start of 2002-03, while 12.5% and 16.8% had reading and spelling ages more than 12 months below, about 9 months later. These figures suggest that children are catching up in reading (but less so in spelling), but 12.5% more than 12 months below expectation in reading is not at all an encouraging figure.

It seems very unlikely from the current study and from the national SATs results that the need for additional interventions will be reduced as a result of wave 2 interventions to the point that only children with more severe difficulties will need them. A gradual reduction in numbers significantly

below average is probably the best that can be expected. This conclusion leads to further questions about the trigger for wave 3 intervention.

However, this study also raises questions about whether there is a clear distinction to be made between wave 2 and 3 interventions, except perhaps between small group and 1:1 delivery. Most schools appeared to mix methods. Children with lower attainments received either mainly 1:1 or a combination of 1:1 and small group work. Perhaps what is most questionable is the assumption in the 3 waves model that children must graduate from class teaching to small group work, and from small groups to 1:1 only when they have not progressed sufficiently on the previous wave. Effective screening tools at the Foundation Stage to enable children who are at significant risk need to be developed and used, followed by early preventative intervention. Torgeson (Torgeson 2001) and others have shown how much more likely this strategy is to achieve lasting "catch-up". More readily available screening at later stages is also vital, as we have suggested above (section 5.2). Children who appear likely to have more serious difficulties need to be identified and the most appropriate mix of 1:1 and individual help provided.

A practical aim of this study was to explore whether it would be possible for Local Authorities to assist schools within their areas by collecting the types of information used in this study, analysing it along similar lines to this study and feeding back to schools about which interventions work best. In considering whether this research could be modified and rerun to

give clearer results, I suggested a number of technical developments are necessary, improving the rating scales and especially including measures of phonological abilities, using better tests reported in standard scores, and more thorough and explicit feature analysis of interventions. The same conclusion also applies to local or even national collection of data on interventions. It should be a major aim of local and national education services to ensure that interventions are used because of their proven effectiveness. There is perhaps some room for debate about what we mean by "proven" (see Reid-Lyon 2004 and Cunningham 2001 for differing views). But there can be no doubt that reputable evidence of effectiveness is essential if we wish to see progress in the development of literacy interventions. And there is good reason, from this research, to suggest that evidence can be provided by schools doing their everyday jobs. They need to be asked to provide information about the key elements of what they are doing, and this information needs to be analysed and fed back. Hopefully this research has shown only a little further development is needed to turn intervention research into everyday good practice.

Chapter 6: Conclusions

6.1 What sort of conclusions are possible?

This study can be seen as an attempt to bring together work from a number of areas, research on reading difficulties, research on interventions and policy initiatives on reading difficulties on several levels, and evaluation leading to feedback to teachers and Local Education Authority advisors about what works best, into a research methodology which has not been used as a whole in these areas before. Along a spectrum of research types from exploratory to definitive, it is much closer to the exploratory end, even though it has been framed in "hard-science" quantitative terms. The results reported in chapter 4 and discussed in chapter 5 therefore lead to conclusions that are not surprisingly tentative and very far from the definitive that a scientific approach normally expects.

At the end of chapter 5, a replication of this study using a very similar methodology was suggested, alongside a complementary study, using a different methodology, which would look for greater clarity about the processes involved in deciding how to deal with and actually working with children with literacy difficulties within schools. As well as conclusions, therefore, there is some consideration of the kinds of data, and hence the kinds of studies, that would be needed to further test and solidify the results of this study.

Conclusions are presented in relation to each of the main aims from chapter 3.

6.2.Evaluation of progress in literacy of children identified as having difficulties

The 126 children in this study who received extra help made progress in reading and spelling at slightly better than expected rates overall. But those with least difficulties caught up most and those with the greatest difficulties (ie those at least 12 months below expectation) reduced only a little in reading and not at all in spelling.

There are a number of important issues about the measurement of reading and spelling attainments which are relevant to future studies. It would perhaps not be too strong to say that without significant improvements in the instruments available, it may not be worth trying to replicate this study. After almost 10 years' of Standard Assessment Tasks, their reliability as measures of progress over time remains uncertain. The most commonly used reading and spelling tests in this study have not been restandardised since the NLS began. A replication would be significantly more effective if these two types of measure could be compared directly as a result of properly conducted comparative studies which provided standard scores from both types of test.

6.3 Comparing progress of those receiving additional help with below average children not receiving additional help:

Results from 256 children showed that children who were below average in reading or spelling attainments at the beginning of the year and who

received no additional help made slightly more progress in reading than those who did receive extra help although the difference was not statistically significant. Differences in spelling in age score terms between these two groups were significant, but they were not significant in National Curriculum terms. These findings provide direct evidence against Brooks' claim that ordinary teaching does not enable children with literacy difficulties to catch up (Brooks 2002).

However, the children who did receive extra help were very likely also the children with greatest difficulties (in terms of their attainments). In the middle of the range of literacy difficulties and at the most severe end, those with greater difficulties make slower progress, even when extra help is given. Differences between progress in reading and spelling were large, suggesting that additional help has more impact on reading than spelling, and that spelling progress is more a result of what happens in class than is reading progress.

Replication of these results would also benefit very significantly from better reading and spelling measures. But here especially a larger sample running over two years would provide greater certainty, by providing direct comparisons between below average children receiving extra help in one year and not receiving help in another year, and simply by providing more cases.

A two year study would also provide some ability to assess the effectiveness of class teaching by comparing overall outcomes. It would be valuable to try to obtain more direct measures of the effectiveness of literacy teaching but this would take a replication into much more complex territory.

6.4 Comparing progress using ALS with other materials:

In a three-way comparison between ALS alone, ALS and other materials, and other materials alone, there were no statistically significant differences between the three types of intervention. When differences between gains are compared between classes, it seems likely that only in some class situations is ALS alone superior to other forms of help. Putting this finding another way, in general ALS with and without other interventions is about equally as effective as other interventions alone in terms of gains in reading and spelling.

When we consider the effectiveness of interventions in gains per hour of additional staff help, ALS is strikingly more efficient than other forms of intervention, about three times more so for reading and twice as much for spelling. This is because ALS is always delivered in small groups, usually of 5 or 6, and most other interventions are delivered in smaller groups or 1:1.

ALS was the first of the wave 2 interventions to be produced, and TAs and teachers commented that it is relatively "unfriendly" to use. This study

therefore strongly supports the case for further research and development effort to be invested in improving the ease of use and the general effectiveness of these materials, because current materials are cost-effective.

6.5 Do differences in children's qualities interact with differences in methods and materials?

The study found only very small associations between some qualities: poor spelling age gains were correlated with language delay or disorder, and with poor attendance. There were small but perhaps important differences in the ways that children with the greatest SEN were allocated to types of intervention (they were mainly given other types of help than ALS). But the main finding was of a general lack of relationship between children's qualities and outcomes.

This result was not expected but is broadly consistent with most of the research literature (Foorman et al 1997, Mathes et al 2005).

This study did not include any measure of children's phonological abilities, and any replication of the study would require a satisfactory measure of them. The research literature suggests that a measure of phonological abilities would not show up interactions between methods and children's qualities (in other words some methods work better with some children), except that greater phonological difficulty would be linked with greater difficulty in making progress (Vellutino et al 1997).

A priority for further development work is therefore to try and create ways of rating children in terms of phonological abilities which do not require extensive and technically sophisticated testing.

6.6 Do class teaching and the ways in which additional interventions are delivered make a difference?

Analysis of the relative importance of class teaching and additional interventions in this study shows that classes make more difference than interventions. The literature comparing methods of class teaching, especially methods other than the (until very recently prescribed) "literacy hour" and other ways of teaching is very sparse. Only Solity and colleagues (Solity et al 1999,2000) have compared outcomes from a standard literacy hour with their own ERR (Early Reading Research), and only at Key Stage 1. Very recently, a new Texas study (Mathes et al 2005) has compared the effects of effective (but not fully described) mainstream teaching and additional interventions.

The findings of this study suggest variability in class teaching may contribute more to differences in gains than do additional interventions. Class teaching seemed to have more effect on spelling than reading, and (as we have already said) reading gain seemed to be more influenced by additional interventions.

Two alternative explanations are put forward: the effectiveness of additional interventions depend on how well they are coordinated with class teaching, and that class teaching acts to multiply learning from additional interventions; or, class teaching and additional interventions have largely separate effects on learning, but class teaching is both potentially more variable and more powerful. The results of this study favour the second hypothesis.

These tentative conclusions imply a need for substantially more research comparing methods of teaching literacy in Year 3 and 4 classrooms. The ability of good class teaching to reduce the need for additional interventions by quite impressive amounts at least in the first year of schooling seems clear (Solity et al 1999, Mathes et al 2005). Whether this is or could also be the case in Years 3 and 4 remains to be seen. It would be important in such research to continue to be able to compare pupil qualities, and to have some direct measures of teacher qualities. A replication which ran over two years would provide significant data on teacher effectiveness. Studies of classroom and school processes might also contribute significantly to answering questions about communication and interaction between staff about pupil progress and especially about how much difference it makes.

Some of the possible influences of the school formed part of this study. Small schools are associated with smaller class sizes, higher Social Advantage Factor and lower SEN ratings. But only school size was

significantly related to outcomes, and then only spelling age gain. Larger schools tended to use ALS more than small schools.

A replication of the study would focus particularly on the effectiveness of class teaching, but the sample of schools would also be important. A focus on larger schools while retaining a range of Social Economic Status levels would be desirable. School processes, such as the role of the Senco, communication between TAs and teachers about pupils, the choice of intervention programmes, time for interventions and the overall school ethos would probably be important.

Aspects of the delivery of additional help were found to be important. More time for extra help was associated with more progress in reading but not in spelling. This may be because class teaching has more influence on spelling. But the data did not suggest that the range of help given in this study (from 25 minutes to 200 per week) was itself limiting. It cannot be concluded from these results that more intensive help (as Torgeson has argued, see Torgeson 2000) would be more effective. However, nearly all help delivered in this study was by TAs. The results of this study did support the possibility that more teaching help would be more effective: the number of comparisons was small, but additional help delivered by teachers was more effective than by TAs.

The results on numbers of sessions are puzzling, suggesting as they do that 1 or 5 sessions a week are best, but 4 are worst. Groups of 1,2 or 6

were found to be about equally successful, but of course children were selected for 1:1 in the light of the severity of their literacy difficulties and to some extent their other special needs.

Training of TAs did make a difference but best outcomes were not associated with greatest experience. Most TAs had been working in schools for a considerable time. Those with less than 10 years experience were more effective than those with more than 10 years. Replication of this result would have practical implications.

Results did not strongly support either delivery by withdrawal or in class. However, reading seemed to benefit more from being taught by withdrawal. There was a clear advantage for a combination of withdrawal and in class work, over either alternative exclusively.

Replication would involve collecting further data on delivery. Ideally more additional help would be delivered by teachers. A direct test of Torgeson's intensity hypothesis in an English context would be very desirable. This would perhaps mean a funded project within a larger study.

6.7 Did particular materials or combinations of materials lead to greater progress?

Comparisons between additional interventions call for a feature based analysis of interventions (for example, whether teaching involves synthetic or analytic phonics), so points of similarity and difference can be more

readily compared. In the absence of a systematic schema, this study relied on very crude categorizations of interventions, whose reliability and validity was untested.

Using these categories, ALS was compared to other comprehensive literacy programmes, reading scheme based interventions, specific spelling programmes, computer based programmes, and individually designed programmes. Significant differences were not found for reading outcomes (though individually designed programmes had the best results), but there were for spelling (where individually designed programmes were also effective, but specific spelling programmes produced good outcomes).

6.8 The implications of this research for national policy on literacy difficulties.

This study shows that ALS, an intervention designed for delivery by TAs in small groups on a near daily basis with children at Key Stage 2, is justified in cost-effectiveness terms, and may be the most effective intervention in terms of reading gains in many classrooms. There is therefore a strong case for upgrading it to improve its ease of use.

This study also suggests that class teaching may be more important than additional interventions in its influence on reading and spelling gains. Given the priority rightly attached to improving class teaching at the beginning of the National Literacy Strategy, and the present relaxation of

prescription on teachers about how they should now teach literacy, it would now be particularly relevant for new or improved methods of class teaching of literacy to be developed through research.

It has also become possible to ask how class teaching and additional interventions combine. Research both on classroom processes and on outcomes of combinations of methods is needed. A race between those who advocate priority for class teaching and those who say additional interventions are more effective, to see which approach can improve catch-up rates most, is a potentially imaginative and exciting possibility.

The 3 waves strategy, with its emphasis on priority for good class teaching followed by cost-effective interventions for children who have fallen behind largely because of social disadvantage, has merit. It has given us the NLS itself and the ALS, ELS FLS interventions. However, there is now a danger that it will hinder rather than foster further progress. The government wave 3 initiative muddled wave 2 and 3 interventions; Brooks' results were used inappropriately to justify certain interventions. There has been no comparable effort by the government to intervene very early with children who may have phonological difficulties, perhaps because this type of intervention does not fit conveniently into the 3 waves model. The results from this study suggest that catch-up at waves 2/3 is not sufficient to reduce the numbers needing more intensive interventions to affordably small numbers. In schools a neat distinction between waves 2 and 3 does not seem to be made. In general, there is a danger that too many levels of

intervention are operating, and results from any one will be difficult to identify. The government should recognise that there may be a gap between rhetoric and reality, commission research to explore what might improve matters and redraft policy to make it genuinely helpful and realistic.

6.9 Using easily available data to evaluate "everyday" literacy interventions

This study has shown that there need not be a huge demand for extra work by schools to create databases of sufficient power to obtain useful answers to questions about additional interventions. However, there are some further technical developments needed before this kind of research can become a matter of routine. Better measures of reading and spelling are needed, slightly more extensive methods of rating pupil qualities, and much better schemas for comparing interventions.

This study has also shown, hopefully, that a naturalistic "quasi-experimental" methodology can make a significant contribution, and that there is some advantage in comparing multiple methods of intervention within the same study. While there may still be a very long way to go before we have the answers the introductory quotation from Foorman et al sought, this study has at least begun to seek out one path that might lead, by successive accumulations of further data and results, to the kinds of answers we need if all children are to experience success in mastering the essential foundations of reading and writing.

References:

Adams, M.,J., (1990) Beginning to Read: Thinking and Learning about Print, Cambridge (Mass):MIT Press

Aldrich-Blake, P., Bunn, T., Dunbar, R., Headley, M., (1971) "Observations on baboons, *Papio Anubis,* in an arid region of Ethiopia", Folia Primatologica 15, p 1-35.

Allegria, J., (2004) "Deafness and Reading" in Nunes, T., & Bryant, P., (eds) Handbook of Children's Literacy, p 459-492, Dordrecht, Netherlands: Kluwer Academic Publishers

Beard, R., (1998) National Literacy Strategy: Review of Research and other Related Evidence. London:DfEE

Bishop, D.,V.,M. (2001) "Genetic Influences on Language Impairment and Literacy problems in Children: Same or Different?" J. of Child Psychology and Psychiatry, 42, 2, p 189-198

Blachman, B., A. (1997) "Early Intervention and Phonological Awareness: A Cautionary Tale", in Blachman, B.A., (ed) Foundations of Reading Acquisition and Dyslexia", Mahwah N.J & London: LEA

Blatchford, P. (2003) The Class Size Debate: Is Small Better? Maidenhead: Open University Press

Bookbinder, G.,E. (1976) Salford Sentence Reading Test, London: Hodder & Stoughton.

Bookbinder, G.,E., Vincent, D. & Crumpler, M. (2000) Salford Sentence Reading Test (Revised), London: Hodder & Stoughton.

Bradley, L., & Bryant, P., (1983) "Categorizing Sounds and learning to read: A causal connection," Nature, 30, p 419-421.

B.P.S.(1999) Dyslexia, Literacy and Psychological Assessment, Report of the Working Party of the Division of Educational and Child Psychology, Leicester: British Psychological Society,

Brooks, G., (1998) "What works for slow readers?" Support for Learning, 14, 1 p 27-31

Brooks, G., Flanagan, N., Henkhuzens, Z., & Hutchinson, D., (1998) What works for Children with Literacy Difficulties? The Effectiveness of Early Intervention Schemes. Slough: National Foundation for Educational Research

Brooks, G., (2002) What Works for Children with literacy Difficulties? The Effectiveness of Intervention Schemes. (Research Report 380) London: DfES

Bryant, P., & Nunes, T. (2004) "Morphology and Spelling" in Nunes, T., & Bryant, P., (eds) Handbook of Children's Literacy, p 319-360, Dordrecht, Netherlands: Kluwer Academic Publishers

Bryant, P., Nunes, T. & Bindman, M. (2000) "The relations between children's linguistic awareness and spelling: the case of the apostrophe". Reading and Writing, 12, p 253-276

Bunn, T. (1987a) "Sociological Perspectives on Special Education – Part 1", Educational Psychology in Practice, 2, 4, p 5-9.

Bunn, T. (1987b) "Sociological Perspectives on Special Education – Part 2", Educational Psychology in Practice, 3, 1, p3-10

Bunn, T., (2002) Rates of Progress in Literacy and Numeracy in Children with Special Educational Needs in the UK, unpublished essay,

Bus, A.,G., van Ijzendoorn, M.,H. & Pellegrini, A.,D. (1995) "Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy", Review of Educational Research, 65, 1, p1-21.

Campbell D.,T. & Stanley, J.,C. (1966) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally

Clay, M. (1985) The Early Detection of Reading Difficulties (3rd edition)". Portsmouth N.H.:Heinemann

Conrad, R. (1977) "The Reading Ability of deaf school-leavers", Br. J. of Educational Psychology, 47, p 138-148.

Conrad, R. (1979) The Deaf School Child, London: Harper Row.

Cunningham, J.,W. (2001) "The National Reading Panel Report", Reading Research Quarterly, 36, 3, p326-335.

Dearing, R., (1994) The National Curriculum and its Assessment: Final Report, London: School Curriculum and Assessment Authority.

DfE (1995) English in the National Curriculum, London: HMSO

DfES (1999) Additional Literacy Support (ALS), DfES0337-2004CD, Nottingham: DfES Publications Centre.

DfEE (2001) Code of Practice on Local Education Authority – School Relations Annexes 1-5, DfEE 0036/2001, Nottingham: DfES Publications Centre

DfES (2001) Special Educational Needs Code of Practice. London: Department for Education & Skills.

DfES (2003) Targeting support: choosing and implementing interventions for children with significant literacy difficulties. Management Guidance. London: Department for Education and Skills

DfES (2004) "Removing Barriers to Achievement", Nottingham: DfES Publications Centre

Di Francesca, S. (1972) "Academic Achievement test results of a national testing program for hearing impaired students", Office of Demographic Studies, Spring 1971, Series D, Number 9, Washington DC: Gallaudet College.

Dockerell, J.,E., & Lindsay, G., (2004) "Specific Speech and Language Difficulties and Literacy", in Nunes, T., & Bryant, P., (eds) Handbook of Children's Literacy, p 319-360, Dordrecht, Netherlands: Kluwer Academic Publishers

Ehri, L.,C., Nunes, S.,R., Willows, D.,M., Schuster, B.,V., Yaghoub-Zadeh, Z., Shanahan, T., (2001) "Phonemic awareness Instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis", Reading Research Quarterly, 36, 3, p 250-287

Elbro, C., Borstrom, I. & Petersen, D.,K. (1998) "Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items", Reading Research Quarterly, 33, 1, p 36-60

Elbro, C., & Scarborough, H.,S. (2004) "Early Intervention" in Nunes, T., & Bryant, P., (eds) Handbook of Children's Literacy, p 319-360, Dordrecht, Netherlands: Kluwer Academic Publishers

Engelman, S. & Carnine, D. (1981) Theory of Instruction: principles and practice, New York: Irvington

Fawcett, A.,J. (ed) (2001) Dyslexia Theory and Good Practice, London: Whurr

Finn, J.,D. & Achilles, C.,M. (1999) "Tennessee's Class Size Study: findings, implications, misconceptions, Educational Evaluation and Policy Analysis, 21, 2, p 97-109.

Fletcher, J., M., Shaywitz, S., E., Shankweiler, D., P., Katz, L., Liberman, I.,Y., Stuebing, K.,K., Francis, D., J., Fowler, A.,E., & Shaywitz, B.,A., (1994) "Cognitive Profiles of Reading Disability: Comparisons of Discrepancy and Low Achievement Definitions", Journal of Educational Psychology, 86, 1, p 6-23

Fletcher, J.,M., Morris, R., Reid Lyon, G., Stuebing, K.,K., Shaywitz, S.,E., Shanweiler, D.,P., Katz, L. & Shaywitz, B.,A. (1997) "Subtypes of Developmental Dyslexia: Differences in Phonological and Orthographic Coding" p115-142 in Blachman, B.A., (ed) Foundations of Reading Acquisition and Dyslexia", Mahwah N.J & London:LEA

Fletcher-Campbell, F., (2000) Literacy and Special Educational Needs: A Review of the Literature, London:DfEE

Foreman, P., Bourke, S., Mishra, G. & Frost, R. (2001) "Assessing the Support Needs of Children with a Disability in Regular Classes", Int. J. of Disability, Development and Education, 48, 3, p 239-252.

Foorman, B.R., Francis, D.J., Shaywitz, S.E., Shaywitz, B.A., & Fletcher, J.M. (1997) "The case for Early Reading Intervention" in Blachman, B.A., (ed) Foundations of Reading Acquisition and Dyslexia", Mahwah N.J & London:LEA

Frederickson, N., Reason, R and Frith, U., (1997) The Phonological Abilities Battery (PhAB), Windsor: NFER-Nelson

Grigorenko, E., L. (2001) "Developmental Dyslexia: An Update on Genes, Brains, and Environments", J. of Child Psychology and Psychiatry, 42, 1, p 91-125.

Hatcher, P.,J., Hulme, C. & Ellis, A.W., (1994) "Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis." Child Development 65, p 41-57

Hulme, C., & Snowling, M., J. (eds) (1997) Dyslexia: Biology, Cognition and Intervention, London: Whurr

IRU (2005) Reducing Bureaucracy in Schools: Gatekeeping, Impact assessments and data collection, 1594-2005DOC-EN, Nottingham: DfES Publications Centre.

Iverson, S. & Tunmer, W.,E. (1993) "Phonological processing skills and the Reading Recovery Programme", J. of Educational Psychology, 85, 1, p 112-126.

Kemp, N. & Bryant, P. (2003) "Do bees buzz? Rule-based and frequency based knowledge in learning to spell the plural-s". Child Development, 74, p63-74

Landerl, K. & Wimmer, H. (2000) "Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children." Applied Psycholinguistics, 21, p 243-262

Mathes, P.,G., Denton, C.,A., Fletcher, J.,M., Anthony, J.,L., Francis., D.,J. & Schatschneider, C. (2005) "The effects of theoretically different instruction and student characterisics on the skills of struggling readers", Reading Research Quarterly, 40, 2, p 148-182.

Miles, T.,R. (1983) Dyslexia: the pattern of difficulties. London: Granada

Muter, V. & Snowling, M., (1998) "Concurrent and longitudinal predictors of reading: The role of metalinguistic and short-term memory skills", Reading Research Quarterly, 33, 3, p 320-337

Muter, V., Snowling, M. & Hulme, C. (1997) The Phonological Abilities Test, London: The Psychological Corporation

National Reading Panel (2000) Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Report of the subgroups. Washington D.C.:National Institute of Child Health and Human Development.

OFSTED (2002) The National Literacy Strategy: the first four years 1998-2002, London: Office for Standards in Education

OFSTED (2003) Excellence in cities and education Action Zones: management and impact, (HMI 1399) London: Office for Standards in Education

OFSTED (2006) Self-Evaluation Form, www.ofsted.gov.uk/publications/

OISE (2003) Watching & Learning 3: Final report of the External Evaluation of England's National Literacy and Numeracy Strategies, London: DfES

Reid, G. & Fawcett, A.J. (eds) (2004) Dyslexia in Context, London: Whurr

Reid Lyon, G., Shaywitz, S.E., Chhabra, V., & Sweet, R. (2004) "Evidence-based reading policy in the US", in Reid, G. & Fawcett, A.J. (eds) Dyslexia in Context, London: Whurr

Reynolds, D. (1997) "School Effectiveness", National Children's Bureau Highlight, 157. London: National Children's Bureau

Sainsbury, M. (1998) Evaluation of the National Literacy Project, Sudbury: DfEE

Sandler, B. & Page, E.,G. (1975) Blackwell Spelling Workshop, Oxford: Blackwell

Scarborough, H.,S., (1998) "Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors" in Shapiro, B.,K., Accardo, P.,J., Capute, A.,J., (eds) Specific Reading Disability: A view of the Spectrum, Timonium, MD: York Press

Scarborough, H.,S. & Dobrich, W. (1994) "On the efficacy of reading to pre-schoolers", Developmental Review, 14, p 245-302.

Shanahan, T. & Barr, R. (1995) "Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners", Reading Research Quarterly, 30, 4, p 958-996.

Skårbrevik, K.,J. (2001) Disabilities Among Norwegian Students Found Eligible for Special Education", Int. J. of Disability, Development and Education, 48, 3, p 253-268.

Slavin, R.,E., Madden, N.,A., Karweit, N.,L., Dolan, L., Wasik, B.,A., (1992) Success For All: A Relentless Approach to Prevention and Early Intervention in Elementary Schools, Arlington, VA: Educational Research Service.

Snow, C.,E., Burns., & Griffin, P., (1998) Preventing Reading Difficulties in Young Children, Washington D.C.: National Academy Press.

Snowling, M., J., & Nation, K.,A., (1997) "Language, Phonology and Learning to Read" in Snowling, M., and Hulme, C., (eds) Dyslexia: Biology, Cognition and Intervention", London: Whurr

Snowling, M., J. & Griffiths, Y.,M. (2004) "Individual Differences in Dyslexia" in Nunes, T., & Bryant, P., (eds) Handbook of Children's Literacy, p 383-402, Dordrecht, Netherlands: Kluwer Academic Publishers.

Solity, J., Deavers, R., Kerfoot, S., Crane, G., & Cannon, K., (1999) "Raising Literacy Attainments in the early Years: the impact of instructional psychology", Educational Psychology, 19, 4, p 373-397

Solity, J., Deavers, R., Kerfoot, S., Crane, G. & Cannon, K. (2000) "The Early Reading Research: The impact of instructional psychology", Educational Psychology in Practice, 16, 2, p 109-129.

Stanovich, K.,E. (1986) "'Matthew effects' in reading: some consequences of individual differences in acquisition of literacy", Reading Research Quarterly, 4, p 360-407

Stanovich, K., E. (1991) "Discepancy definitions of reading disability: Has intelligence led us astray?" Reading Research Quarterly, 26, p 7-29.

Stanovich, K.E., (2000) Progress in understanding reading: scientific foundations and new frontiers, London: The Guildford Press

Stanovich, K.,E., & Siegel, L.,S., (1994) "Phenotypical performance profile of children with reading disabilities: a regression-based test of the phonological-core-variable-difference model", Journal of Educational Psychology, 86, 1, p 24-53

Stanovich, K.,E., Seigel, L.,S., Gottardo, A., Chiappe, P. & Sidhu, R (1997) "Subtypes of Developmental Dyslexia: Differences in Phonological and Orthographic Coding", in Blachman, B.A., (ed) Foundations of Reading Acquisition and Dyslexia", Mahwah N.J & London:LEA

Suffolk County Council & Hagley, F.,J. (1987) Suffolk Reading Scale, Windsor: NFER Nelson.

Torgesen, J.,K., (2001) "The theory and practice of intervention: comparing outcomes from prevention and remediation studies", in Fawcett, A., J., (ed) Dyslexia: Theory and Good Practice, London: Whurr

Troia, G., (1999) "Phonological awareness intervention research: A critical review of the experimental methodology", Reading Research Quarterly, 34, 1, p 28-52

Tunmer, W.,E. (1989) "The role of language related factors in reading disability" in Shankweiler, D. & Liberman, I.,L. (eds) Phonology and Reading Disability: Solving the reading puzzle, International Academy for Research in Learning Disabilities Monograph, (6). Ann Arbour, MI: University of Michigan Press.

Twist, L., Sainsbury, M., Woodthorpe, A., & Whetton, C., (2003) Progress in International Reading Literacy Study: National Report for England, Slough (Bucks): National Foundation for Educational Research

Velluntino, F.,R., Scanlon, D.,M. & Sipay, E., R. (1997) "Towards Distinguishing Between Cognitive and Experiential Deficits as Primary Sources of Difficulty in Learning to Read: The Importance of Early Intervention in Diagnosing Specific Reading Disability" in Blachman, B.A., (ed) Foundations of Reading Acquisition and Dyslexia", Mahwah N.J & London: LEA

Vernon, P.,E. (1977) Graded Word Spelling Test, London: Hodder & Stoughton.

Wearmouth, J. (2002) "Dilemmas in the choice of responses to students' difficulties in literacy development", in Wearmouth, J., Soler, J. & Reid, G. (2002) (eds) Addressing Difficulties in Literacy Development, London: RoutledgeFalmer

Weston, P. (1992) "A Decade of Differentiation", Br. J. of Special Education, 19, 1, p 6-9

White, K.,R. (1982) "The relation between socio-economic status and academic achievement", Psychological Bulletin, 91, p 461-481.

Wolf, M. & O'Brien, B. (2001) "On issues of time, fluency and intervention" in Fawcett, A., (2001) Dyslexia Theory and Good Practice, London: Whurr

Wragg, E.,C., Wragg, C.,M., Haynes, G.,S.,& Chamberlin, R.,P., (1998) Improving Literacy in the Primary School, London: Routledge