THE UNIVERSITY OF HULL

Development and Evaluation of Computer-based Techniques for Assessing Children in Educational Settings

Being a Thesis submitted for the Degree of Doctor of Philosophy

in the University of Hull

By

Joanna Kathryn Horne (BSc)

January 2002

Contents

List of tables

List of figures

Abstract

This thesis reports and discusses an integrated programme of research on computerised assessment in education, focussing on two themes. The aim of the first study was to develop and evaluate a computerised baseline assessment system for four to five year olds (CoPS Baseline). The aim of the second study was to develop and evaluate a computerised dyslexia screening system for the secondary school age group (LASS Secondary).

CoPS Baseline was shown to be a reliable and valid assessment of pupils' skills in literacy, mathematics, communication and personal and social development on entry to school at age four or five. It was also found to be predictive of children's later reading, spelling, writing and mathematics ability up to three years after the initial testing.

LASS Secondary was shown to be a reliable and valid assessment of students' reading, spelling, reasoning, auditory memory, visual memory, phonological processing and phonic skills from the ages of 11 to 15. It was also seen to be a good indicator of dyslexia, with significant differences between the scores of dyslexic students and non-SEN students on the sentence reading, spelling, auditory memory, non-word reading and syllable segmentation tests.

CoPS Baseline and LASS Secondary were also found to be more objective than conventional assessment administered by a person, time-saving in their test administration and scoring, and more enjoyable and motivating for children, particularly children who have specific difficulties.

Computer-based techniques have been shown to be beneficial in the assessment of children in educational settings. However, further research is proposed in the areas of:

gender and ethnic differences in computerised versus conventional assessment; the addition of reading comprehension, verbal intelligence, mathematics and motor skills tests to the LASS Secondary system; follow-up tests of students assessed on LASS Secondary to provide information about teaching outcomes; and the development of tests suitable for use with deaf / hearing-impaired individuals in order to assess literacy skills and identify dyslexia.

Acknowledgements

I should like to thank my supervisor, Dr Chris Singleton, for all his support and advice over the last six years. I would also like to thank the other members of the dyslexia research team, particularly Kevin Thomas, Rik Leedale, Brenda Smith, Jared Holwell and Jan Trotter. Rik Leedale carried out the programming of CoPS Baseline and LASS Secondary. Jared Holwell carried out the graphic design and artwork involved in CoPS Baseline and LASS Secondary. CoPS Baseline is copyright © Lucid Research Limited, 1998. LASS Secondary is copyright © Lucid Creative Limited, 1999. The test design and creation of items for CoPS Baseline and LASS Secondary was carried out by the author in collaboration with Dr Chris Singleton and Kevin Thomas. All the data collection was carried out directly, or supervised, by the author. All statistical analysis was carried out by the author. Thanks to the students who helped in some of the data collection, including Magda Szpala, Daniel Horne, Kerry Park and Jess Woodcock. Special thanks to my parents, brothers and friends for all their support. Finally, I would like to thank the teachers and pupils at the 69 schools that took part in the studies, without whose help this research could not have taken place.

1 Introduction

"If computers are in future going to interpret (as well as deliver) tests, what need have we of highly qualified (and expensive) professionals who hitherto have been the 'experts' on psychological and educational assessment?" (Singleton, 1997a, page 274).

The use of educational assessment in the United Kingdom has increased greatly since the 1944 Education Act, and more so with successive Acts, despite some criticism of assessment (Brady, 1997). Some group differences have been reported in educational assessment with respect to gender (Tizard et al, 1988; Fergusson and Horwood, 1997; Strand, 1999), ethnicity (Lindsay, 1988; Plewis, 1991; Brady, 1997; Strand, 1999) and language (Strand, 1997; Strand, 1999; Lindsay and Desforges, 1999).

Computerised assessment is becoming more common and has been seen to have a number of advantages over conventional tests (Hunt and Pellegrino, 1985; French, 1986; Greewood and Rieth, 1994; Woodward and Rieth, 1997). These advantages include that computerised tests are less time-consuming, more objective, enable adaptive testing, can record additional data and, in some circumstances, can be more enjoyable.

Baseline assessment may be defined as the assessment of children's developmental attainments when they first enter school. The National Framework for Baseline Assessment (SCAA, 1997b) requires that children's language, literacy, mathematics and personal and social development be assessed at the start of compulsory education and this became statute in the 1997 Education Act. Since September 1998 all maintained primary schools in England and Wales have been required to assess children on entry using a baseline assessment scheme accredited by the Qualifications and Curriculum

Authority. This statutory requirement presents challenges to teachers, schools and Local Education Authorities, and to designers of assessment instruments[1].

Developmental dyslexia is a syndrome that can affect written language, number skills, memory, phonological skills, organisational skills and motor function. There are a number of different methods of assessing dyslexia including the traditional ability versus attainment discrepancy model and the more recent phonological deficit hypothesis. In recent years there has been increasing concern about using valid and reliable methods for identifying children with dyslexia in school, so that their educational needs can be addressed under the requirements of the Code of Practice on the Identification and Assessment of Special Educational Needs (DfE, 1994; DfES, 2001). Since the Code places responsibility for this on teachers and schools in the first instance, there is therefore a need for effective screening systems suitable for use by teachers rather than just psychologists.

1.1 Aims and rationale of the study

If assessment is seen to be a necessary part of the education process then it is important that any assessment is acceptable to pupils, is fair to all groups of the population, is standardised nationally across all the relevant age groups and has satisfactory evidence of reliability and validity. Consistent with the SEN Code of Practice, there must be assessments available for teachers to use, rather than just educational psychologists.

---

[1] Arrangements for baseline assessment are set to change from 2003 with the introduction of a single national assessment scheme at the end of the Foundation Stage.

This thesis reports and discusses an integrated programme of research on computerised assessment in education, focussing on two themes. The aim of the first (Study 1) was to develop and evaluate a computerised baseline assessment system for four to five year olds. The aim of the second (Study 2) was to develop and evaluate a computerised dyslexia screening system for the secondary school age group. These systems should be shown to be appropriate for the relevant age groups and standardised nationally. They must also have evidence of significant reliability and concurrent validity. The baseline program would be expected to be predictive of later abilities in literacy and mathematics. The secondary program must be capable of distinguishing between dyslexic and non-dyslexic students. It is important that there are not a high number of false negatives (students not identified as having difficulties when they are actually dyslexic) or false positives (students identified as having dyslexia when they do not). Neither assessment system should show any evidence of bias in gender, ethnicity or language. By selecting two completely different computerised assessment systems, one for use at the very beginning of schooling, and the other for use in the later stages of schooling, it was aimed to derive broad and general conclusions about the validity and functionality of computerised assessment in educational settings.

1.2 Scope

There are six main areas of investigation in this research:

- To develop a standardised computerised baseline assessment system and provide evidence of its reliability, concurrent validity and predictive validity

- To develop a standardised secondary school dyslexia screening system and provide evidence of its reliability, concurrent validity and effective identification of dyslexia
- To examine whether computerised tests are more objective than conventional tests, particularly with respect to bias in gender, ethnicity and language
- To examine whether computerised assessments are less labour-intensive and less time-consuming than conventional educational assessment
- To examine pupils' enjoyment of computerised tests compared to conventional assessment
- To consider more widely the role of computerised assessment in education and the challenges raised for future research.

1.3 Structure of the thesis

The thesis is divided into four parts and contains 13 chapters in total. **Part 1** gives the introduction and research background to the field and includes three chapters. *Chapter 1* introduces the research area, giving information about the aims and rationale and the scope of the projects. *Chapter 2* gives an overview of the definition, purposes and history of educational assessment, the role of teachers and psychologists and the challenges facing this field. *Chapter 3* reviews the difficulties with conventional educational assessment, the use and advantages of computerised assessment in education and the challenges of producing computerised educational assessments.

Part 2 covers Study 1 and includes work on the development and evaluation of a computerised baseline assessment system and consists of four chapters. *Chapter 4* reviews the definition, purpose, educational history and statutory framework of baseline assessment as well as the difficulties with it. *Chapter 5* provides the methodology used in the baseline assessment research, including the test development and follow-up projects. *Chapter 6* contains the results from all of the baseline assessment studies. *Chapter 7* discusses these results with reference to test development, longitudinal studies, group differences, technical issues and feedback from users.

Part 3 covers Study 2 and includes work on the development and evaluation of a secondary assessment system and consists of four chapters. *Chapter 8* reviews the definition, purpose and alternative theories of dyslexia assessment as well as the challenges in this area. *Chapter 9* provides the methodology used in the secondary assessment studies, including the test development, standardisation and reliability, validity and dyslexia projects. *Chapter 10* contains the results from all of the secondary assessment studies. *Chapter 11* discusses these results with reference to test development, group differences, technical issues, identification of students with dyslexia and feedback from users.

Part 4 contains the general discussion and conclusions to be drawn from the research and includes two chapters. *Chapter 12* discusses the issues of gender bias in educational tests, how computers address the challenges of educational assessment, the advantages and disadvantages of computerised assessment and the changes facing baseline assessment and dyslexia screening. *Chapter 13* contains a summary of the thesis, the overall conclusions and proposals for future research.

2 Educational assessment – issues and challenges

2.1 Aims of the chapter

This chapter aims to:

- provide a definition of educational assessment
- state the purposes of educational assessment
- provide an overview of the history of educational assessment
- provide an overview of the role of teachers in educational assessment
- provide an overview of the role of psychologists in educational assessment
- provide evidence of bias in educational assessment
- identify the challenges facing educational assessment

2.2 Definitions and purposes of educational assessment

Dyer (1972) states that:

> The primary task of those of us who are persuaded of the ancient truth that good testing really is essential to good teaching is to get others to see it in the same light, for the impact of testing on education is a function of the perceptions of the consumers of tests. If tests are perceived as peripheral to the education process, their impact will be peripheral. If they are perceived as devices for putting constraints on the curriculum, they will indeed put constraints on the

> curriculum. If they are perceived as instruments for sorting children into ironbound categories, they will be used for that purpose. But if they are perceived as supplying basic data needed for helping children learn to cope more and more effectively with the world into which they are growing, then they will be used to provide that kind of help (page 326).

According to Kibby (1995), educational diagnosis comprises assessment, evaluation and decision making. He states that diagnosis is a process of acquiring information about an individual's educational performance, strategies, skills and instructional needs through careful observations in order to improve instruction. Kibby distinguishes between assessment and evaluation, suggesting that assessment involves making an estimation or measurement, whilst evaluation involves judging the quality of a measure or comparing it to some criterion.

Harding and Beech (1991) maintain that there are three types of assessment commonly used in primary schools. The first is informal evaluation, which is made by teachers based on their experience. These evaluations are beneficial as they can pinpoint individual needs, abilities and deficits which can be further investigated. The second is norm-referenced assessment, in which an individual's score is compared with a distribution of scores from a sample of the population. The third is criterion-referenced assessment, which involves identifying the position of individual pupils in the learning process. Results are given in terms of whether or not individuals can meet the criterion. Criterion-referenced assessments inherently relate to norm-referenced processes, as the criteria must be compared to the expected level for a child at a particular age. Ingenkamp (1977) argues that education has been dominated by norm-referenced assessment rather than criterion-referenced assessment for three main reasons: (a) it is easier to define the level of a behaviour by comparison with that of other pupils, than to make a precise analysis and description of the objectives; (b) comparative testing has been used in the past for selecting pupils with high ability; (c) there are various statistical problems with criterion-referenced testing.

Harding and Beech (1991) suggest that the purposes of assessment fall into two categories: those in the interests of society or a wider group, and those in the interests of the child. There are a number of purposes of assessment that are external to the school. LEAs often need information on individual children in order to organise transfer from one school to another. Assessment aids the identification of children who require special needs teaching. LEAs require information on numbers of children with special educational needs, to aid in the allocation of resources. LEAs may request schools to carry out screening procedures to identify children who may be 'at risk' educationally or who may be in need of special education. Finally, schools are increasingly required to give information on standards in order to make them more accountable to parents and the local community. There are also a number of purposes of assessment that are internal to the school. Brady (1997) advocates that one reason why so much money and effort is devoted to assessment is that as education has become more expensive, it is sensible to ensure that money is being spent reasonably. Hence, assessments aid in the organisation of the school and help to monitor standards. Assessments also allow teachers to identify the attainments of children and plan their teaching more effectively. Furthermore, Ingenkamp (1977) suggests that assessments provide information on the effectiveness of specific teaching practices. Assessment for the primary purpose of benefitting the pupil includes diagnostic assessment, enabling a potential, and criterion-referenced assessment to plan an educational program to suit individual children. Ingenkamp (1977) and Brady (1997) argue that assessment provides students with information about their progress that aids further development and increases motivation.

Lindquist and Hieronymus (1972) point out that educational assessment is only useful if it enhances teaching, improves supervisory practices, and leads to more effective educational guidance of pupils. Ingenkamp (1977) adds that examinations aid in the selection of students with higher ability for further studies or specific professions. However, Dyer (1972) maintains that the use of tests as predictors is not enough. He suggests that tests give us knowledge of the student's learning style so that we can help him or her overturn the predictions.

Woodings (1997) asserts that, although the National Curriculum has undergone a number of changes since it began in 1988, assessment still follows the principles proposed by the Task Group on Assessment and Testing (Black, 1987). Assessment was viewed as having diagnostic, formative, summative and evaluative purposes. Diagnostic assessment allows learning difficulties to be scrutinised and classified so that appropriate remedial help and guidance can be provided. Formative assessment involves recognising the positive achievements of pupils so that the appropriate next steps can be planned. Summative assessment is the recording of the overall achievement of pupils in a systematic way. Finally, evaluative assessment allows aspects of the work of a school, LEA or other educational service to be assessed and reported on. Kelly (1992) views formative assessment as having the positive function of taking the pupil's education forward, whilst diagnostic assessment has the negative purpose of identifying learning difficulties. According to the Task Group on Assessment and Testing, the diagnostic and formative functions were to be managed by teachers and related to the teaching process. These would provide the teacher with information concerning what the child has learnt and help in planning the curriculum. The summative process involves national tests (Standard Assessment Tasks), which are set to test the National Curriculum at the end of each key stage. TGAT proposed that the scores produced by these tests could provide information on the performance of pupils, teachers, schools and the educational system as a whole. This would show any increase or decrease in educational standards as well as assist parents in choosing a school for their child. It has been argued that, due to the conflict between diagnostic / formative and summative / evaluative assessments there is a prominence on 'teaching to the test' and a disregard of formative assessment (Troman, 1989; Gipps, 1990). However, Black (1995) states that "good formative assessment can be a powerful tool for raising standards of learning, but that it is in general badly underdeveloped in schools. There is therefore a tremendous opportunity for improvement" (page 7).

2.3 History of educational assessment

The modern emphasis on examinations in schools has its origins in the middle of the nineteenth century (Brady, 1997). In 1850 the College of Preceptors set external examinations for grammar and private schools and the Science and Art Department introduced examinations in 1854. This national assessment provided a common target for educators. During the nineteenth century the government began to support schools financially, which resulted in central control of the curriculum. The Revised Code of 1862 gave the government further control by increasing the accountability of schools.

Aaron and Joshi (1992) argue that reading assessment before 1900 was primarily qualitative, usually involving listening to pupils reading. From 1910 standardised tests were developed to assess reading comprehension and spelling, allowing pupils' reading to be assessed with a degree of objectivity which had not previously existed. Furthermore, these tests were designed to aid the development of teaching practices and improve reading.

In the UK, it was Burt who initially promoted the use of intelligence tests (within the London County Council) to assist in allocating children to different forms of education (Lindsay, 1991). The 11-plus test later made use of intelligence tests as well as assessment of attainment. The 11-plus was instituted in the 1940s to assess each child's aptitude and ability at age 11 in order to determine the nature of his or her secondary education (MacKenzie, 1989). Kelly (1992) argues that when assessment has been used as a basis for making predictions, for making far-reaching decisions about educational provision, such as the use of the 11-plus for allocating pupils to different forms of secondary education, these decisions have been inaccurate. Lindsay (1991) reports that local authorities and schools have used intelligence tests to group children within schools and identify those who need special provision. According to Ingenkamp (1977), the three purposes for which intelligence tests have most often been used in education are selection, identifying learning difficulties and guidance in individual learning.

The Binet Intelligence Scales, published in French in1908 and 1911, were revised by Terman and published in English in 1916 as the Stanford-Binet Intelligence Scale. The Otis Group Intelligence Scale (1918) was the first group test of mental ability designed to be used by schools and educational institutions. According to Stake and Hastings (1972), the Stanford Achievement Battery (1923) was developed at a time when the advantages of standardised testing had just become apparent. They suggest that, typically, achievement test norm groups comprised those students most available for testing, and usually the composition of the reference group was not specified to test users. However, Stake and Hastings maintain that the Stanford Achievement Tests were based on a thorough item selection and developed according to psychometric principles and educational practices. They provided norms that allowed comparisons across subjects as well as with a nationally representative sample of students.

In 1944, in England and Wales, LEAs were made responsible for establishing which children in their authority required special educational provision and specified procedures for identifying and placing such pupils (1944 Education Act; see also Eaton, 1991). The Education Act defined several categories of children who required significantly different forms of education to the majority of children. Every school-aged child was required to attend a school "suitable to his age, ability and aptitude and to any special educational needs he may have" (Section 36). Woods (1973) argues that the 1944 Education Act put too much stress on doctors' opinions and not enough on assessment by educational psychologists and teachers, who are better qualified to recommend that a child have special education and the form that this should take.

The 1970s saw the development of alternative ability tests. Whilst intelligence tests provide general assessments of cognitive ability, new tests were developed to assess more specific abilities. During this time, there was a great deal of interest in specific language, perceptual, motor and perceptuo-motor abilities. Research and experience had shown that many children failed at basic educational tasks such as reading, despite good levels of general intelligence as measured by IQ tests. During the 1970s there was also an increase in the use of detailed individual assessments, enabled by an increase in the numbers of specially trained staff, particularly educational psychologists (Lindsay, 1991). The development of the Stanford-Binet Intelligence Scale, Wechsler Intelligence Scale for Children and the British Ability Scales illustrate the heightening importance of assessing individual children's specific cognitive abilities using standardised assessments. The Stanford-Binet Intelligence Scale was first published in 1916 (Terman, 1916) and was revised in 1937 (Terman and Merrill, 1937), 1960 (Terman and

Merrill, 1960) and 1972 (Terman and Merrill, 1972). The Wechsler Intelligence Scale for Children (WISC) was first published in 1949 (Wechsler, 1949) and has since been updated twice (Wechsler, 1974; Wechsler, 1991). The WISC assesses the general intellectual ability of children aged six to 16 years, 11 months. The British Ability Scales (BAS) were first published in 1979 (Elliott, 1979) to measure a range of cognitive abilities in two and a half to 17 and a half year olds. The second edition of the scales was published in 1996 (Elliott, 1996). However, Lindsay (1991) suggested that, during the late 1980s, the prominence of intelligence testing decreased. Gipps et al (1983) reported on the introduction of LEA testing programmes in the 1970s, which appeared to be a response to calls for greater accountability. Murphy (1991) commented that, during the 1970s and 1980s, although there was a rise in the amount of assessment taking place in primary schools, there was a lot of variation in the assessment procedures followed by individual schools.

Explicit procedures for the identification of, and provision for, children with special educational needs, in England and Wales, was established by the 1981 Education Act, which was a response to the Warnock report of 1978 (Harding and Beech, 1991; Eaton, 1991). Assessment was further extended by the 1988 Education Reform Act with the introduction of the National Curriculum. The Task Group on Assessment and Testing (DES, 1988) proposed the use of SATs (Standardised Assessment Tasks) which were designed to reflect the range of attainment targets within the National Curriculum. The Education Reform Act (1988) states that 'the curriculum ... shall ... specify the arrangements for assessing pupils at or near the end of each key stage, for the purpose of ascertaining what they have achieved in relation to the attainment targets for that key stage' (Section 2, 2).

The 1993 Education Act required the Secretary of State to issue the 'Code of Practice on the Identification and Assessment of Special Educational Needs' (DfE, 1994). The Code of Practice places on schools a legal duty to identify and address all special educational needs as early as possible in the child's schooling. This principle is maintained in the revised Code of Practice (DfES, 2001).

2.4 The role of teachers and psychologists in educational assessment

Burt (1921) states that "there is no standard of comparison that can surpass or supersede the considered estimate of an observant teacher, working daily with the individual children over a period of several months or years" (page 211). However, findings of research studies have not always borne out Burt's assertion. In a study by Ingenkamp (1972), marks given by teachers were compared with the results of standardised tests. Significant differences between classes were found. For example, pupils graded II in arithmetic in one class had an average test score of 53.5, but in another class the average was only 30.9. He concludes that marks given by teachers in different classes could not be compared in any way. Ingenkamp suggests that a class-based marking system is to be expected when teachers have no regional standards to guide them. In addition, Fuchs and Fuchs (1984) studied the accuracy of special education teachers' judgements of student performance based on unsystematic observation. They found that, despite confidence in their own judgements, special education teachers and teaching trainees tended to be inaccurate and to overrate their students' performance. Teaching experience did not affect the accuracy of the assessments, although it made teachers more confident about their assessments. Furthermore, Saint-Laurent et al (1996) examined the use of teacher ratings and curriculum-based achievement tests in the identification of students with educational difficulties. They found that, although achievement scores and teacher ratings were significantly correlated, only half of the students with the lowest achievement test results were identified by their teachers as having difficulties.

Pole (1998) argues that the enhanced role of testing in education, due to the 1988 Education Reform Act, has meant a different role for teachers in assessment procedures. The Task Group on Assessment and Testing (DES, 1988) established that pupils would be assessed by Standard Assessment Tasks as well as by teacher assessments based on ongoing classroom observations. According to Pole (1998), the greater emphasis on the paper-and-pencil tests reduced the role of the teacher to that of an administrator and indicated that teacher judgements were not trusted.

Not all types of assessment are available to teachers. Harding and Beech (1991) report that tests which require familiarity with norming procedures and psychometric principles are generally only available to psychologists, whereas, tests which do not require such knowledge (for example, many reading tests and criterion-referenced assessments) are available to teachers. Some research has found that teachers believe that achievement tests are of little use (Salmon-Cox, 1980) and so teachers use a range of non-standardised assessments in their day to day teaching (Yeh, 1978; Salmon-Cox, 1980). Cameron (1991) argues that teachers administer curriculum-based assessment for two purposes. First, to identify the pupil's current knowledge and, secondly, to plan an individual programme for each pupil. However, Butterfield (1995) contends that, although greater involvement of teachers in the assessment of students enhances the professional role of teachers, it also greatly increases their workloads.

Burt was Britain's first educational psychologist with the London County Council in 1913. He maintained that the priority of educational psychologists is "to ascertain educationally sub-normal pupils using psychological tests and other scientific procedures" (Burt, 1957, page 31).

In a survey carried out by Topping (1978), it was found that most children seen by educational psychologists were assessed outside the classroom and half of the teachers did not know what the educational psychologist did with the child.

Dessent (1992) argues that, since the 1981 Education Act, the role of educational psychologists has changed from defining what is a significant special educational need to defining which individuals deserve additional resources. The 1981 Education Act requires that a statement of a child's special educational needs be based on educational, psychological and medical advice (Norwich, 1988). There is no guidance on how psychological advice is distinct from medical or educational advice. Norwich (1988) reports that the 1981 Education Act gave recognition to the work of educational psychologists not given to other professional psychologists and increased the number of positions for educational psychologists in LEAs.

According to Booth and Statham (1982), the traditional role of educational psychologists has been as assessors of special educational provision. Bryans (1992) argues that this meant "that psychologists appeared at times to be friendly police guarding the special school gates and escorting pupils away from the mainstream school into special educational provision" (page 144).

Psychologists have had more extensive higher education than teachers and may be seen as more expert about pupil's difficulties. Norwich (1988) argues that this view may be attributable to the assessment methods that psychologists use being seen as scientific, because they are standardised. Teachers are often seen as being able to assess pupils' attainment in literacy and numeracy, whilst psychologists are able to assess pupils' learning potential. Traditionally, education placement decisions have been relatively long-term. However, since the 1970s, individuals have not been seen as having unalterable learning characteristics (Clarke and Clarke, 1976) and there has been a move towards curriculum-based measurement by teachers. Norwich (1988) suggests that the traditional distinction between educational assessment (past achievements) and psychological assessment (future potential) is no longer clear. However, it is still important that a pupil's assessment by a teacher is compared to that of an outside educational professional, such as an educational psychologist.

Huebner and Cummings (1986) investigated the importance of psychological test scores in psychologists' decisions concerning student placement in special education. Test scores were found to differentiate between students who were placed in special education and students who were not, especially the IQ-achievement discrepancy. Huebner and Cummings concluded that educational psychologists can establish eligibility of individual students very accurately.

Bryans (1992) suggests that, to ensure professional integrity, educational psychologists should: participate in LEA policy making on special needs issues; ensure they are not coerced into discriminatory practices; ensure parental co-operation at all stages of informal and formal assessment; and, give unbiased evidence of assessments. Elliot (1996) argues that the knowledge and skills of school psychologists are needed to reform educational assessment. He concludes that school psychologists should be encouraged to assume leadership in the assessment of all pupils.

2.5 Biases in educational assessment

Blatchford and Cline (1992) suggest that the assessment process should operate without bias with respect to gender, social class, ethnicity, language use and religion. However, many researchers have demonstrated the existence of bias in educational assessment.

Flaugher (1978) describes two meanings of the term test bias: 'test bias as mean differences' and 'test bias as content'. As far as 'test bias as mean difference' is concerned, he argues that:

> knowing what we do about the relative status, socio-economic and otherwise, of ethnic minorities...it would be surprising if most kinds of tests didn't show mean differences in favour of the majority group...many critics of testing merge the concept of equality of *opportunity*, which is certainly a legitimate goal to be sought, with the concept of equality of *result*; but it is only results that the tests in fact measure (page 673).

Flaugher suggests that 'test bias as content' exists if a test is prejudiced against a subgroup of a population because of the content of the questions. It has been argued that tests should enable individuals to show the knowledge and skills they possess regardless of its nature. It could, therefore, be said that a fair test is one that only asks questions that the test taker can answer. According to Beck and St George (1983), there have been two approaches for removing test content bias. First, the detailed evaluation of test content, and, second, item analysis for different sub-groups of the population. Two main aspects of potential testing bias have been researched: gender bias and cultural (ethnic) bias.

Tizard et al (1988) report that girls in Britain out-perform boys in reading and writing by the age of seven but there is only a small gender difference in mathematics.

According to Fergusson and Horwood (1997), the traditional educational disadvantage of girls has been replaced by a male disadvantage. They studied a group of children from school entry to age 18 and found that males achieved less well than females. These differences were not due to differences in intelligence since boys and girls had similar IQ scores. Fergusson and Horwood suggest that the gender differences were explained by boys being more disruptive and inattentive in class, which impeded their learning. Wright and Payne (1979) carried out an evaluation of a psychological service and found that the overall gender difference in referrals was just over two boys to each girl. This ratio was larger for children of primary school age, but at secondary school age the rates of referrals were nearly equal for boys and girls. In addition, Male (1994) states that in one LEA, two-thirds of the children receiving a statement of special educational needs under the 1981 Act were boys. Furthermore, Vardill (1996) reports that more boys than girls are identified as having special educational needs and, in one LEA, the referral ratio of boys to girls was 1.86:1. He suggests that one of the reasons for this is that boys experiencing difficulties are more likely to be a problem to the teacher because of associated behavioural difficulties.

Many researchers have argued that most assessment is culture-biased and discriminates against certain ethnic groups (Brady, 1997). Anastasi (1972) reports that, when psychologists began developing cross-cultural tests, they hoped it would be possible in theory to measure intellectual potential independently from cultural experience. An individual's behaviour was thought to be overlaid with a cultural veneer, whose penetration was the objective of culture-free testing. However, Anastasi (1972) argues that it is not productive to attempt to develop tests that are 'culture-free' (free from cultural influences) and, instead, there should be efforts to develop tests that are 'culture-fair' (common to different cultures). She suggests that culture-fair tests include those that do not use language (or use spoken language but no reading) and do not include culture-specific objects. Nevertheless, Anastasi maintains that no test can be universally fair as any test favours individuals from the culture in which it was developed. She concludes that the purpose of tests is to identify what individuals are capable of and that concealing cultural disadvantage by using a culture-fair test is not a solution. Joyce (1988) agrees that all educational assessments are culture bound and suggests that this is due to the content, standardisation, population and situations in which they are administered and the fact that norm-referenced tests are biased in favour of the majority group. He argues that knowledge of the cultural background of the child should be taken into account before assessing the pupil's needs and that criterion-referenced tests should be used which are specific to the child's own progress so children are not being compared to other pupils. Other researchers agree that some cultural bias is inevitable in the content of any assessment tools (Wood, 1991; Lam, 1993). However, Cline and Shamsi (2000) argue that it is possible to remove the worst sources of bias and produce fairer assessments to all children. Plewis (1991) reported educational differences between ethnic groups, with African Caribbean pupils, especially boys, performing less well than white pupils in language and mathematics. In addition, Bryans (1992) purports that the majority of educational psychologists are apprehensive about the use of standardised tests with certain groups because they tend to show that pupils from certain ethnic groups have low ability. Frederickson (1992) reports that in norm-referenced assessment, a significant proportion of the test items may be outside the cultural experience, customs or values of particular groups of children. Furthermore, Cummins (1984) claims that most psychologists regard IQ tests as preserving the 'anglocentric educational status quo'. Helms (1995) agrees that norm-referenced tests disadvantage children from ethnic minority communities. Nevertheless, Bryans (1992) reports that most educational psychologists still use standardised assessments with these groups of pupils.

Plewis (1997) suggests that the reasons for gender and ethnic differences in educational attainment are not well understood. It has been argued that teachers have lower expectations for boys and pupils from ethnic minority groups. He suggests, however, that to conclude that teachers are biased against these groups it is necessary to show that these low expectations are inaccurate and that teachers behave differently towards these pupils. Plewis did not find teachers' assessments to be less accurate for certain groups, although he did find teachers' expectations to be biased against boys. A number of researchers suggest that expectations could be affected by the way boys behave (Tizard et al, 1988; Bennett et al, 1993; Delap, 1995). Plewis also found that for spelling, expectations for black pupils are too low, whereas for Pakistani pupils they are too high and for mathematics, expectations for Pakistani pupils were too low.

2.6 Challenges of educational assessment

According to Ebel (1972), the increase in the use of educational tests has been accompanied by an increase in criticism of the practice. Tests vary in quality, with some being particularly poor. In 1933, Ruch contended that out of approximately one thousand different educational and mental tests available at that time, evidence of the validity, reliability and standardisation procedures was available in less than ten percent. Today, although more educational test manuals provide technical information, the evidence is not always sufficient for teachers or psychologists to make an informed decision concerning the accuracy of a test or its relevance to a specific group.

Researchers have argued that educational testing may be socially detrimental for a number of reasons (Ebel, 1972). First, it labels a child, which may damage his or her self-esteem and decrease motivation. Brady (1997) suggested that there is a need to balance the necessity of assessment against the risk of it producing undesirable competition at an early age. He acknowledges the well-documented effects of 'failing' the 11-plus. However, Ebel (1972) comments that tests should not be evaluated in terms of how accurately they predict later achievement, but in terms of how much they increase achievement by motivating and directing the efforts of students and teachers. A second criticism is that assessment encourages development of a single ability, and reduces the diversity of talent within society. On this matter, Ebel (1972) replies that the difficulty of advancing the development of various abilities is mostly due to the demand that all students study the same courses for most of their education. A third argument is that assessors assume control of the educational curriculum. Brady (1997), for example, suggested that, although assessment is intended to support the curriculum, there is a risk that it may come to dominate the curriculum because what is assessed is taken as an indication of what is important. He argues that the curriculum may then change because there is a tendency to assess those subjects that are easy to assess. By contrast, Ebel (1972) argues that tests generally lag rather than lead curricular change. Finally, critics of educational assessment suggest that testing imposes courses of actions on individuals. However, Ebel (1972) responds that tests should be used to provide a reasonable basis of choice and not to impose decisions.

Ebel (1972) concludes that:

> if the use of educational testing were abandoned, the distinctions between competence and incompetence would become more difficult to discern... the encouragement and reward of individual efforts to learn would be made more

difficult... decisions on important issues of curriculum and method would be made less on the basis of solid evidence and more on the basis of prejudice or caprice (page 13).

2.7 Summary and conclusions

Educational assessment involves making an estimation or measurement of a child's ability in a particular area. The three main types of assessment used in schools are informal evaluations, norm-referenced tests and criterion-referenced tests. The main purposes of educational assessment are:

- to identify children with special educational needs
- to aid in the planning of individual and class teaching
- to increase motivation
- to provide information on the effectiveness of certain teaching practices
- to increase the accountability of schools
- to guide LEAs in the allocation of resources
- to monitor educational standards nationally.

Formal educational assessment began in the middle of the nineteenth century with the introduction of academic examinations. With an increase in government control of schools came an increase in the assessment of reading and spelling standards. Research into intelligence testing also allowed general ability tests to be carried out in educational settings. Educational assessment has been legislated for, most importantly within the 1944 Education Act, the 1981 Education Act, the 1988 Education Reform Act, the 1993 Education Act, the Code of Practice (1994), the 1996 Education Act and the Code of Practice (2001).

The individuals primarily responsible for assessing children are teachers and educational psychologists. Teachers are in a position to observe children on a daily basis but are generally limited in the types of tests that they can use. Teacher evaluations are not always reliable and accurate tests are necessary. However, such instruments are often only available to psychologists. Educational psychologists are able to assess children, referred to them by teachers, with a greater level of accuracy using the appropriate assessment tools.

The two most commonly reported differences in educational assessment concern poorer performance of boys and ethnic minority groups. These differences may be due to actual differences between groups, bias on the part of the assessor or bias within the actual test, due to norming procedures or content. There has also been seen to be a bias in the referral of children to educational psychologists, with more boys than girls being referred.

There are a number of challenges currently facing educational assessment. Tests may not show sufficient evidence of validity and reliability. Researchers and teachers are also wary of the social and educational effects of labelling children, reducing the diversity of abilities within society, producing an assessment-led curriculum and imposing decisions on individuals.

Although there are a number of difficulties associated with educational assessment, it is only through assessment that a child's difficulties can be properly identified and dealt with. Labels for different special educational needs have been unpopular in the past but there are signs of a change in opinion amongst educationalists. Although all SEN students are individuals, there are broad categories that are useful in teaching and SEN labels are often necessary to ensure that the student receives the right support in learning. However, there is still a need for differentiation of teaching and learning activities within a single category, particularly dyslexia. A bigger problem in educational assessment is mislabelling children. There is a need for tests to be produced that can be administered by teachers rather than psychologists, that do not have a problem of bias on the part of the assessor and that are well standardised and have good evidence of validity and reliability. Most importantly, teachers and pupils should benefit from the results of any tests. As stated by Brady (1997) "it is easier to generate figures than it is to make clear what they signify or how they should be used" (page 9).

3 The potential of computerised educational assessment

Surtton (1992) argues that:

> Assessment is a human process, conducted by and with human beings, and subject inevitably to the frailties of human judgement. However crisp and objective we might try to make it, and however neatly quantifiable may be our 'results', assessment is closer to an art than a science (page 2).

3.1 Aims of the chapter

This chapter aims to:

- provide a definition of computerised educational assessment
- provide an overview of the difficulties with conventional educational assessment methods
- identify the advantages of computerised educational assessment
- provide an overview of the use of adaptive testing methods
- provide an overview of studies using computerised educational assessment
- identify the challenges facing computerised educational assessment

3.2 Definition and growth of computerised educational assessment

According to the British Psychological Society (1999) "computer-based assessment may be considered to include any psychological assessment that involves the use of digital technology to collect, process and report the results of that assessment" (page 1). Computer-based assessment encompasses four areas: assessment generation; assessment delivery; assessment scoring and interpretation; and, storage, retrieval and transmission.

Hunt and Pellegrino (1985) write that, traditionally, psychological tests are either individually or group administered using a paper-and-pencil format. They suggest that psychologists prefer individually administered tests as "they are seen as more accurate or more complete evaluations of mental ability" (page 208). Hunt and Pellegrino give three reasons for this: the examiner can produce a variety of auditory, visual and tactile stimuli; can determine what questions are most likely to be informative about a particular examinee; and, can note any aspects of the testing (e.g. illness) that might affect the test score. Group-administered tests have none of these advantages but are much cheaper. According to Hunt and Pellegrino, computer-based test administration falls between the individual and the group testing situations. Almost any question that can be presented in paper-and-pencil format can be presented using a computer display.

During the 1970s and 1980s, advancements in technology (e.g. networked computer suites and portable personal computers) allowed for the computerised administration, scoring and interpretation of psychological tests (British Psychological Society, 1999). However, Woodward and Rieth (1997) report that studies in special education technology in the 1970s and early 1980s were primarily concerned with the potential of computer-assisted instruction (CAI). In the mid 1990s, computer-based assessments were commonly used in the occupational sector and, to a lesser extent, in further and higher education (Bartram, 1994). Computerised assessment has remained scarce in primary and secondary education, although there have been some developments in this area (NCET, 1994 and 1995; Singleton and Thomas, 1994; Singleton, 1996; Singleton, Horne and Thomas, 1999). McCullough and Wenck (1984) report that the earliest widespread applications of computer-based assessment in education were for scoring student test forms. Even ten years later, the greatest use of computer-based assessment in schools and colleges in the UK was to record or report students' results rather than for generation or delivery of assessment (NCET, 1994). During the 1990s in the USA, computer-based assessments were being used, in the field of special education, for recording and analysing classroom observation data, diagnosis of eligibility for special education services and ongoing curriculum-based assessment (Greenwood and Rieth, 1994; Woodward and Rieth, 1997).

3.3 Advantages of computerised educational assessment

There are a number of advantages of computerised assessment commonly reported in the literature.

One of the main advantages of computerised assessments is that they are very labour-saving and allow tests to be administered by less highly trained personnel (Greenwood and Rieth, 1994; Woodward and Rieth, 1997; British Psychological Society, 1999; Singleton, Horne and Thomas, 1999). As early as the 1970s, Elwood (1972) reports that the initial cost of purchasing a computer system is quickly compensated by the saving in skilled person-hours. Computerised assessments are often administered quicker than conventional tests and as scoring is automatic, the results are more accurate and can be immediately available (French, 1986; Singleton, Horne and Thomas, 1999).

For tests consisting of many items, this can lead to considerable savings in time (French, 1986). As well as being time consuming, the scoring of psychological tests may also involve complex calculations (e.g. comparisons of scores on different tests or different scores from the same test) and may require substantial expertise in the theory supporting the test (British Psychological Society, 1999).

According to Feldman and Sullivan (1971), test performance may be influenced by the examiner's personality. Furthermore, Paitich (1973) argues that human behaviour is difficult to observe reliably, difficult to evaluate and difficult to predict. Administration of tests by computer is much more standardised and controlled than conventional test administration (French, 1986; British Psychological Society, 1999). As computerised test items are usually presented one at a time, the possibility of the testee getting out of synchronisation with the test booklet and the answer sheet is eliminated (Byers, 1981).

Another principal advantage of computerised assessment is the possibility of adaptive testing (Hunt and Pellegrino, 1985; French, 1986; British Psychological Society, 1999; Singleton, Horne and Thomas, 1999). Adaptive testing involves tailoring a test to the individual test taker. It is relatively easy to program rules to make items given contingent upon the response to past items (Hunt and Pellegrino, 1985). Adaptive testing is a more efficient use of the testee's time than is evaluation using a fixed test. The issue of adaptive testing will be returned to in the next section.

As well as the usual data collected by traditional methods, the computer can also record data on the number of items attempted and response times (Stout, 1981). These data may have a number of uses (Hunt and Pellegrino, 1985; French, 1986; British Psychological Society, 1999). They may indicate testee fatigue or that the testee has not understood the instructions at a particular stage of the test. The pattern of responses (for example, rapid responses or the same response option for all items) may indicate that the testee is not taking the test seriously. In these cases, the computer may suggest a rest period or call for staff assistance. It is also possible that hesitancy on particular items in a personality test may be as revealing as the actual response finally given (French, 1986).

Wade and Moore (1993) found that some primary and secondary students with special educational needs recognised that conventional tests were useful for feedback and motivation but half of the sample disliked these tests and found them boring. These negative views were intensified by the physical and emotional effects during the testing session. However, a number of studies have indicated that children and adults (particularly if they feel they might perform badly) often prefer computer-based assessment to traditional assessment by a human assessor (Skinner and Allen, 1983; Moore, Summer and Bloor, 1984; French, 1986; Singleton, Horne and Vincent, 1995; Singleton, 1994 and 1997a).

Furthermore, it has been reported that testers prefer computerised assessment. In education, curriculum-based measurement computer programs administer tests, analyse results, depict progress graphically and advise teachers on how to adapt their teaching to meet individual needs (Fuchs, Fuchs and Hamlett, 1993). Using a curriculum-based measurement approach, direct and frequent measurements of student performance are made within the regular context of the school curriculum (Druckman, 1997). According to Fuchs and Deno (1991), curriculum-based measurement features the assessment of proficiency on global outcomes toward which the entire curriculum is directed and the reliance on a standardised methodology that produces critical indicators of performance. Computer-based versions of curriculum-based measurement were developed because teachers saw frequent, systematic measurement as being too time-consuming (Wesson et al, 1986). Early computer versions of curriculum-based measurement stored, graphed and analysed data that had been entered by teachers after paper tests had been administered and scored. Despite the fact that using such a system gave little or no saving in time, teachers generally preferred the computer version (Fuchs et al, 1987).

Wilson (1987) reports that there is a major difficulty in assessing severely physically disabled people. The more limited the ability to communicate because of the extent of physical impairment, the more difficult it is to assess. Wilson argues that this can lead to either underestimation of the person's mental abilities or non-recognition of a deficit in psychological function. However, the wide range of computer input and output devices currently available (for example, touch sensitive screens, speech recognition, speech synthesis, mouse, joystick, tracker ball, switches and graphics tablet) allows the testing of even severely handicapped individuals (Wilson, Thompson and Wylie, 1982; French, 1986; Wilson, 1987). Bennett (1999) suggests that computer-based testing could improve higher education admissions assessment for examinees with disabilities by increasing comparability. Computer-based tests can utilise the power of the technology to the full so that test items can include speech, sound and animation (Singleton, Horne and Thomas, 1999). This can be of great assistance in the assessment of individuals with visual and auditory impairments (French, 1986). Furthermore, the use of multimedia tools makes assessment more relevant to the current school generation as they are more similar to tasks encountered in an academic setting (Bennett et al, 1999).

Computers can generate detailed interpretative reports for tests and give recommendations (French, 1986; Woodward and Rieth, 1997). These reports show equal or superior validity and reliability when compared with traditional assessment (Space, 1981). Watkins and Kush (1988) found that similar instructional interventions were generated by computer and paper-based versions of a capitalisation test, but students preferred the computerised test. Furthermore, students completed the computerised version of the test more quickly and answers were automatically scored by the computer, reducing scoring time and eliminating the possibility of scoring errors. Watkins and Kush conclude that computer-assisted testing has potential for the assessment of educational skills and the design of instructional programs. Computer-based test interpretation software packages have been available for some time and are widely used to generate immediate reports or oral feedback (Bartram, 1994; British Psychological Society, 1999). In recent years, they have begun to incorporate sophisticated test score analysis and provide reports that appear to have been written by a human. Bartram (1994) suggests that by ensuring uniformity and consistency in how evidence is weighed, computer-based test interpretation should lead to an increase in the reliability of test interpretations.

Another advantage of computerised assessment is that whilst psychological tests have conventionally been human-produced, today they can be generated by item engines (British Psychological Society, 1999).

According to the British Psychological Society (1999), computerised test administration enables distance assessment, allowing greater access to psychological assessment. Furthermore, group administration can be carried out through local area networks on one site or, via the internet, on a wider basis (Bartram, 1994).

French (1986) argues that whilst it is difficult to update the content and norms of traditional tests, such revisions are easy with computerised tests. Norms can even be generated for specific institutions.

Computers allow for a more efficient method of storage, retrieval and data management than paper-based records (Paitich, 1973; British Psychological Society, 1999). This is cheaper and more flexible and enables results to be quickly and easily transmitted to clients at other locations.

Finally, Bartram (1994) argues that if tests are only published in electronic form, with controlled access, producers are able to control the use of their test materials and control the use of personal test data.

3.4 Adaptive testing

In traditional tests, all items are presented in the same order, even though many will not be relevant for specific individuals (French, 1986). In a test containing items of increasing difficulty, the testee will often start with items that he or she finds very easy (which may lead to boredom) and end up struggling with items that he or she finds difficult (which may lead to frustration). The term 'adaptive testing' refers to "any technique that modifies the nature of the test in response to the performance of the test taker" (Singleton, Horne and Thomas, 1999, page 69). It is generally taken to mean tests which use Item Response Theory (Lord, 1980; Stocking, 1987; Reckase, 1989; Wainer et al, 1990). Adaptive testing uses the flexibility of the computer to select from a large item pool and present those items that are most appropriate for a particular individual, based on initial item responses and the individual's ability (Vale, 1981; Johnson and Johnson, 1981). The testee's likely responses to items that are not actually presented can often be inferred from the responses obtained (Lord, 1980; French, 1986). As the computer can score performance at the same time as item presentation, it can effectively tailor the test to the capabilities of each individual. With conventional tests, much of the time the individual's abilities are not being assessed with great precision. In an adaptive test, individuals are moved quickly to the test items that will most efficiently discriminate his or her capabilities, making assessment shorter, more reliable, more efficient and often more acceptable to the person being tested (Singleton, 1997a). In an adaptive test, the programme usually starts by administering a number of items to estimate a person's ability level. Further items, varying in difficulty, are then selected for administration based on the responses to previous items, until some discontinuation rule is reached. Some individually administered conventional tests have adaptive features, such as discontinuation rules and starting points which vary according to age and/or performance on initial items. However, such features can be administratively complicated (Singleton, Thomas and Horne, 2000).

Adaptive testing often leads to a reduction in administration time (French, 1986; Singleton, Horne and Thomas, 1999). Bloxom (1989) argues that adaptive tests are more reliable and more efficient than conventional tests and adaptive tests of 10 to 15 items can be as precise as conventional tests with twice the number of items. Olsen (1990) compared paper-based and computer-administered educational tests with computerised adaptive tests. The computer-based non-adaptive version took 50% to 75% of the time taken to administer the conventional version, while the testing time for the adaptive version was only 25% of the time taken with the paper-based version. Research has also shown that adaptive testing is more motivating for the testee as items that are too easy or too difficult are generally avoided (British Psychological Society, 1999).

Day (1999) found the use of the Computer Adaptive Placement Assessment and Support System (COMPASS) to be a viable method of assessment for recognising and recommending students within higher education for a remedial / developmental programme.

Bartram and Bayliss (1984) suggest that in future, a rule-based approach will be applied to item construction, such as the generation of items for spatial ability tests and verbal reasoning tests. Once research has determined the parameters (for example, difficulty) of items generated by a set of rules, the rules can be embodied in an adaptive system. This would have enormous potential as it would effectively address a very large pool of potential items without the need for those items to be pre-specified or stored. French (1986) suggests that objective mathematical measures of item complexity can be devised and used to generate new items. Some computerised tests already use this procedure, for example, the Perceptual Maze Test (Elithorn, Mornington and Stavrou, 1982).

3.5 Studies of computerised educational assessment

Bartram (1994) has observed that computers have allowed us to move from paper-based tests, which are fixed in their item content, item order and duration, to the use of computerised adaptive tests. Most computerised assessment has involved the computerisation of existing printed tests rather than the development of new types of test. This has mostly involved attainment, ability and aptitude tests, where changes in response procedure or item format are likely to impact on scores. Self-report measures have been regarded as being less likely to be affected by changes in presentation medium. There has also been some interest in the use of computers to generate interpretative reports, especially of personality inventories. Bartram (1994) argues that "the development of computer-based interpretation has led us to question the need for highly qualified and expensive 'expert' test interpreters" (page 32). Federico (1991) suggests that a number of features are likely to affect equivalence. First, items cannot usually be omitted in computerised tests. Second, computerised tests generally do not allow back-tracking. Third, the computer screen has a limited capacity in relation to the printed page. Fourth, the quality of computer graphics relative to printed material. Finally, the nature of the response mode (e.g. a key press or mouse click, as opposed to ticking a box) is particularly important where speed of response is critical to the test. Furthermore, Dillon (1992) reports that reading from a computer screen is 20-30% slower than reading from paper. Bartram (1994) argues that different issues arise with new forms of computer-based tests which cannot be administered by people and which have no parallel with non-computerised forms.

Elwood (1970) reports that test-retest reliabilities of the automated WAIS are extremely high, that there is a significant correlation between the automated WAIS and the conventional WAIS, that subjects accept the automated procedure and that the automated system can accurately and reliably record several dimensions of responses, that are difficult to record in face-to-face testing. According to Gilberstadt, Lushene and Buegel (1976) the automation of intelligence testing is less extensive than the automation of personality tests, which are more easily automated. However, they argue that automated assessment of deficit functioning is feasible.

A number of researchers have looked at the equivalence of computer-based and traditional versions of the digit span task. Beaumont (1985) found that subjects performed better on the standard digit span test than on computer presentation. He showed that higher scores on the standard version were due to the use of computer-based response devices, rather than vocal responding, and the visual presentation of digits. Similar findings have also been reported by French and Beaumont (1992) who conclude that the use of computerised auditory presentation of digits, along with speech recognition technology for responding, should be encouraged to produce a computerised digit span task which is psychometrically parallel to the original WAIS version. Penney and Blackwood (1989) suggest that allocation of mental resources to the novel keyboard response takes resources away from keeping the digits in memory. French and Beaumont (1992) report that subjects receiving the standard version first performed significantly better on the computerised version than subjects receiving the computerised version first. They suggest that subjects do not find the cognitive load of computer-based response systems as disruptive when they are familiar with the digit span test. Wilson (1987) ascertained that an automated digit span test significantly correlated with the WAIS Digit Span and also had a higher retest reliability. This is possibly because a computer gives a more standardised presentation than a human.

A number of researchers have found that computerised and paper-and-pencil versions of Raven's Standard Progressive Matrices test are essentially equivalent (Calvert and Waterfall, 1982; Watts, Baddeley and Williams, 1982). However, Beaumont and French (1984) found that performance on a computerised version of Raven's Matrices was inferior to that on the pencil-and-paper version. They argue that this is probably a result of poor graphics resolution on certain items. Furthermore, French and Beaumont (1990) report that scores on a computerised version of Raven's were significantly lower than on the standard version. Subjects who received the standard version first scored significantly higher than those who received the computerised version first. The computerised version took significantly less time to complete than the standard version. Reliability between these forms of the Standard Progressive Matrices test was 0.84. French and Beaumont suggest that since this project, computers with much higher resolution graphics systems have become widely available and better results could now be obtained.

Beaumont (1981) and Wilson, Thompson and Wylie (1982) found significant correlations between computerised and standard versions of the Mill Hill Vocabulary test. Wilson, Thompson and Wylie reported that means for the standard and computerised forms were very similar, but that subjects who received the standard version first showed a strong practice effect. These subjects attempted more items, increasing the probability of getting higher scores. French and Beaumont (1990) reported that scores on the Mill Hill Vocabulary test did not differ significantly between the standard and computerised tests and reliability between these forms was 0.90. Subjects reported that the computerised version presented items more clearly than the standard version and those subjects who were administered the computer version first were more willing to take another similar test.

Evans, Tannehill and Martin (1995) obtained strong significant correlations between traditional and computerised versions of the Woodcock-Johnson Psycho-Educational Battery Revised Tests of Achievement, indicating that computer-administered reading tests can assess the same domains as traditional measures.

Some researchers have looked at the potential of computerised assessments that have no paper-based equivalent. Singleton, Horne and Vincent (1995) found significant correlations between a pilot version of a computerised reading comprehension test and scores on the Edinburgh Reading Test (Godfrey Thompson Unit, 1993) for children in Keystage 2.

Cisero et al (1997) developed CAAS (Computer-based Academic Assessment System), which assesses reading skills using computer-presented tasks that measure speed and accuracy of performance. Comparisons are made between scores on word and non-word reading and phonological processing (which are generally poor for students with a reading disability) and scores on category matching and semantic knowledge (which are generally poor for students with a learning disability). The CAAS system distinguished students with learning disabilities from those who did not have a learning disability, as well as differentiating among students with different types of disabilities (specific reading disability or general learning disability). Students' performance on CAAS was consistent with history information and standardised test performance. However, CAAS assessments provide more detailed information regarding the nature of the student's reading difficulty than standardised test scores. The information provided by the CAAS system appeared to be useful in designing effective intervention strategies.

3.6 Challenges facing computerised educational assessment

A number of researchers have put forward the view that psychological testing should be carried out by professionals, not computers (Moses, 1969; Groth-Marnet and Schumaker, 1989). Matarazzo (1983) states that "there is a danger that wholesale use of automated tests by people without any knowledge of their limitations will be a disservice to the public. Compounding this danger, the tests have a spurious appearance of objectivity and infallibility as a halo effect from the computer, and their ease of use may cause them to be more widely employed than are current tests" (page 323). Furthermore, Basden (1984) argues that expert systems should not be used by novices because 'in any specialist field not only are there phrases and jargon words of specialised meaning, but apparently ordinary words might have special meanings of which the novice might be dangerously unaware' (page 64). French and Hemmings (1984) report that a computerised version of the Cattell IQ test, available to the general public, was not favoured by psychologists. According to Eyde and Kowal (1985), computer-based test interpretations, programmed according to fixed rules, are not selfcorrecting and do not show the flexibility and intuition found in experts. Eyde (1987) suggests that as computer-based test interpretations are not tailored to deal with each individual's circumstances, it is important that they are not made available to untrained people. French (1986) replies to these arguments, suggesting that the fact that computerised tests can be administered by non-professionals is an advantage, in that routine aspects of test administration are taken away from the psychologist. However, many psychologists maintain that testing requires a high level of skill if it is to be properly applied. French argues that the areas of test administration and test interpretation can be separated. He suggests that the degree to which professionals need to be involved in the testing procedure should be resolved by consultation between the users and the system designers.

A number of psychologists are reluctant to use computerised assessment methods. However, French (1986) believes that computerisation of traditional tests with which professionals are familiar may lead to greater acceptance. Furthermore, the norms derived for standard versions of tests are often applied to the computerised version. However, French argues that it cannot be assumed that simply because the standard and computerised versions of a test bear a surface resemblance to each other, they will be psychometrically parallel. The modes of response input and item presentation differ between the two forms. Nevertheless, Bartram (1994) concludes that when care is taken to simulate essential features of the ergonomics of the paper-and-pencil test in the computer version and where the test is not highly speeded, computerisation is not likely to cause a problem.

French (1986) claims that, although there have been suggestions that the extensive data-storage capabilities of computers magnify the potential for unauthorised access to test results, data are at least as secure when stored on electromagnetic media as when stored by traditional methods. The Data Protection Act (1984) applies to any individual or organisation that holds personal data on a computer. The Act gives individuals access to personal data and regulates the holding and use of such data. It sets out eight principals of data protection, which require data to be: obtained fairly and lawfully; held only for one or more lawful purposes specified in the data user's register entry; used or disclosed only in accordance with the data user's register entry; adequate, relevant and not excessive for the purposes; accurate and where necessary up to date; not kept longer than necessary for the specified purposes; made available to data subjects on request; and properly protected against loss or disclosure.

Schoenfeldt (1989) stated that "the development of computer-based tests and test interpretations has become a cottage industry in the worst sense of the word. For those with even minimal skills, software development is an easy entry business that offers the chance for a large profit on a small investment" (page 20). It has been suggested that computer-based assessment gives an unwarranted impression of objectivity and accuracy, which arises from people's perceptions of computers as objective and accurate (Groth-Marnet and Schumaker, 1989). Farrell (1989) conducted a survey of psychological test software vendors and psychologists. He found that only three out of nine vendors of test software and only one out of 15 vendors of computer based test interpretation systems were able to give information concerning the standards that their programs met. Only one third of psychologists using computer-based assessment had information to determine whether it was accurate or not. However, Bartram (1994) argues that this does not mean that the products are unsatisfactory, but that without such information it is not possible to evaluate them. He suggests that it is more worrying that, even in the 1990s, so many psychologists were using tests without questioning their accuracy or validity.

Hunt and Pellegrino (1985) suggest that certain groups are selectively favoured or disfavoured by computerised testing. Computer-based tests can be presented in a game-like format and therefore be self-motivating. However, this could be specific to particular groups of people. Some individuals are motivated by computer game formats, whilst others find computers intimidating, boring or both. Computer games are usually associated with younger age groups (Loftus and Loftus, 1983). Furthermore, there is evidence that older people dislike situations in which rapid responding is required (Hunt and Hertzog, 1981). On the other hand, some studies have found that computerised tests are more acceptable to elderly and handicapped people than paper-based tests (Carr et al, 1982; Watts, Baddeley and Williams, 1982). In addition, different people have different experiences with computers, which is likely to affect performance on a computerised test. However, Taylor et al (1999) found no relationship between computer familiarity and level of performance on a computerised test of English as a foreign language, after controlling for English language ability.

Davis and Swezey (1983) consider that human factors issues are important in the development of computerised assessments. French (1986) argues that this is of particular importance with regard to clarity of instructions, wording of error messages and screen layout, in order to ensure that the system is user-friendly.

French (1986) reports that the development of computerised tests is likely to be more expensive than that of conventional tests, given that group testing for standardisation purposes would be difficult if not impossible. However, with the availability now of computer networks, computerised group testing is a possibility. There is also a saving in the time spent scoring tests.

Another difficulty with computerised tests is the possibility of copying the software. However, French (1986) suggests that the use of discs which auto-erase after a specified number of test administrations could be used.

The British Psychological Society Guidelines for the development and use of computer-based assessments (BPS, 1999) contain four main principles. First, users should be aware of what constitutes best practice in computer-based assessment so that they can make informed choices between the available computer-based tests. Second, that computer-based assessments should be supported by clear documentation of the rationale behind the assessment and the chosen mode of delivery, appropriateness and exclusions for use, and research evidence supporting validity and fairness. Third, requirements for administration of the assessment should be documented and should include the knowledge, understanding and skills required for competent administration. Finally, the knowledge, understanding and skills required for interpretation of computer-based assessment information and for the provision of such information to a third party should be stated.

Bartram (1994) argues that many of the above problems are eradicated if the use of test materials is restricted to those trained to use them. This competence ensures that results are not over-interpreted, tests with no evidence of validity are not used, materials are kept up to date and unqualified people are protected.

3.7 Summary and conclusions

Computerised assessment includes any use of computer technology in test administration, test scoring and recording and reporting of test results. The use of computerised assessment has grown since the 1970s and its use has spread from the area of occupational psychology to adult education and, more recently, to school-age education.

The main advantages of computerised assessment are:

- it is labour-saving
- it is quicker
- it provides immediate feedback
- it is objective and accurate
- it makes adaptive testing possible
- it can record additional data
- it is preferred by testees and testers
- it enables individuals with physical, visual and auditory impairments to be assessed
- it can generate interpretative reports
- it allows greater access to assessment through the internet
- it is easily revised and standardised
- it efficiently stores and manages test data
- it allows more control over the use of test materials.

Adaptive testing allows the test administration to be modified to suit each individual, avoiding the testee wasting time and effort attempting items that are far too easy or far too difficult for them. This makes the test time considerably shorter, and the whole process more acceptable to the testee.

A number of studies have looked at the computerisation of traditional paper-based tests, including the WAIS digit-span, Raven's Matrices and the Mill Hill Vocabulary test. These have generally found the different forms of test to be essentially equivalent.

The main criticisms of computerised assessment are that: they may be used by untrained people who will be unable to adequately interpret the results; they may not be psychometrically parallel to traditional versions of the test; they may allow unauthorised access to test results; they may be produced without research into their validity and reliability; some groups are likely to be more advantaged than others by the use of computerised tests; and, they may be more expensive and can be copied illegally.

However, many of these criticisms apply equally, if not more so, to paper-based and other conventional tests. All tests, whether paper-based or computer-based, should be used under the guidance of a trained professional and should show evidence of reliability and validity. Computers allow test results to be stored more securely and (with appropriate security mechanisms) the copying of software is more difficult than the photocopying of paper-based test materials. In conclusion, the advantages of computerised assessment far outweigh its potential disadvantages.

4 Baseline assessment

4.1 Aims of the chapter

This chapter aims to:

- provide a definition of baseline assessment
- state the purposes of baseline assessment from the perspective of the school and the LEA
- provide an overview of the educational basis of baseline assessment from the early 1970s onwards
- provide an overview of the statutory framework for baseline assessment
- provide evidence of differences on baseline assessment between different gender, ethnic, language and socio-economic groups
- provide evidence of the reliability and validity of baseline assessment schemes
- identify the challenges facing baseline assessment.

4.2 Definitions and purposes of baseline assessment

Various terms have been used alongside 'baseline assessment', including 'on-entry testing', 'initial assessment', 'entry profile', 'starting points', 'early screening', 'baseline profile', 'reception profile' and 'point of entry baseline'. 'On-entry' refers to the start of statutory schooling at the age of four or five.

Lindsay (1993) and Wolfendale (1993a) define baseline assessment as any evaluation of children's abilities and developmental level when they first enter school. It is stated in the 1997 Education Act (Part 4, 1:15) that "Baseline Assessment Scheme' means a scheme designed to enable pupils in a maintained primary school to be assessed for the purpose of assisting the future planning of their education and the measurement of their future educational achievements'.

According to Wolfendale and Bryans (1979), the original purpose of baseline assessment was to identify children with educational and developmental difficulties, to provide appropriate education and meet special educational needs. Lindsay (1984) argues that during the early 1980s the purpose of baseline assessment widened to include monitoring of all children.

Blatchford and Cline (1992) propose four main functions of baseline assessment. First, on-entry testing can be used as a basis for measuring future progress (this is commonly termed 'value added'). Second, getting a picture of the new intake to help decision-making about general teaching approaches. Third, obtaining a profile of each entrant in order to plan the curriculum to suit the needs of the individual pupil. Finally, identifying children who may have difficulties at school. Wolfendale (1993b) adds that baseline assessment is also an attempt to measure and estimate pre-school experience.

It is argued by Fisher (1995) that local education authorities (LEAs) and schools view the purpose of baseline assessment quite differently. He suggests that LEAs see baseline assessment as providing a basis for planning teaching, maintaining a partnership with parents, structuring teacher's assessments, making for consistency of practice in the LEA, helping to identify special needs, helping to determine distribution of resources and monitoring school effectiveness. Schools, however, see baseline assessment as enabling informal monitoring of children's progress, shaping provision for individual children, screening for language provision and special needs and providing material for discussion of a child's progress with parents.

Lindsay (1998) suggests that there are eight purposes of baseline assessment that can be put into two categories. Within the 'child focus' category are early identification of individual children with special educational needs, early identification of the nature of pupils' special educational needs, monitoring the progress of all pupils and identification of learning objectives and strategies for individual pupils. The 'school focus' category includes resource planning within the school, accountability (or value added), budget determination within the LEA and school improvement. According to Wolfendale and Lindsay (1999), a ninth purpose was allegedly added by the then Minister of Education, Charles Clarke, at the launch of the national scheme for England in September 1998. This ninth purpose concerns the setting of pupils into ability groups, which would belong to the 'child focus' category.

According to Lindsay (1993), the publication of the National League table of schools, based on SAT results, has caused concern over the effect of social background. He states that 'there is very strong support throughout the educational world for developing a system which assesses not the absolute levels of attainment of the children, but the progress they have made while under the influence of the school. This has come to be known as 'value added' assessment' (pages 62-63). Blatchford and Cline (1992) argue that the result obtained from a test at one point in time is educationally not as useful as comparison of results made at two different points in time. A fairer system would measure each child's educational progress during their time at a school. Desforges and Lindsay (1995a) purport that baseline assessment makes it possible to calculate the educational value added during a child's time at school. Originally, end of Key Stage 1 results were intended to be used as a baseline for measuring the value added of junior schools. This would mean that infant education would not have the accountability that is required of the subsequent stages of the education system.

Wolfendale and Lindsay (1999) argue that knowing the purpose of baseline assessment is important for two main reasons. First, it is necessary to be clear about the purpose if satisfactory procedures are to be developed. Second, the purpose of baseline assessment will determine the process. For example, if the aim is to identify children with special educational needs then the process must show greatest sensitivity at the levels reached by the lower 20% of children.

4.3 The educational basis of baseline assessment

According to Wolfendale (1993b), although the term 'baseline assessment' is a modern expression that is prevalent in the education field, the practice of screening children early in their school life is not new.

Early screening in the 1970s was concerned mostly with the identification of special educational needs. Wolfendale and Bryans (1979) argue that "early failure, if not recognised, inevitably leads to frustration and avoidance" (page 2) and that 'prevention at the earliest stages of schooling is always preferable to remediation at later stages' (page 2). Screening had traditionally involved the administration of reading and non-verbal intelligence tests to junior school children. Three problems with this approach are put forward by Wolfendale and Bryans. First, remedial education was shown to have limited effectiveness and they propose that early identification should therefore be taking place within the infant school rather than the junior school. Second, there was a move away from norm-referenced tests, which were standardised on indigenous English children and were not suitable for the multi-cultural composition of schools in the 1970s. Criterion-referenced tests, they suggest, are more appropriate as they do not compare the child's performance with that of other children. Finally, there was evidence of the detrimental effect of early deprivation on later educational attainment.

During the 1970s many LEAs initiated early screening procedures when children entered school. Most of these schemes involved the use of checklists or observation schedules and many concentrated on identifying skills associated with reading (Wolfendale and Bryans, 1972, 1979). Clay (1975) argued that baseline assessments at that time could not assist teachers in deciding what to teach next and would not, in themselves, raise standards. Examples of early systems of early identification in the UK include the Croydon Checklist (Wolfendale and Bryans, 1979), the Bury Infant Check (Pearson and Quinn, 1986), and the Infant Rating Scale (Lindsay, 1981).

Throughout the 1980s screening became more widespread in schools and by the mid 1980s over 70% of LEAs had some form of early screening in place (Lindsay, 1988).

However, there was also increasing awareness of the inaccuracies of checklists. According to Potton (1983) screening, in it's original meaning, was an acceptable but unrefined approach to classifying, sorting or identifying. He surveyed 48 teachers using the Croydon Checklist (Wolfendale and Bryans, 1979). Only a third of the teachers felt that the checklist was useful in assessing children, less than 30% said that it pointed to any specific problems and less than 10% said that it told them anything new about the children.

During the early 1990s there was a decline in the use of screening tests in schools and LEAs in the UK because the schemes in use did not show satisfactory levels of accuracy or give teachers useful information. Checklists are generally considered to be inadequate for the purposes of identifying children with special educational needs or for measuring value added by a school. Norm-based assessments would be necessary for this information to be yielded. Consequently, the use of baseline assessments using specific criteria on a numerical scale has increased during the 1990s (Desforges and Lindsay, 1995a).

Wolfendale (1993b) carried out two surveys involving 100 primary schools in two LEAs, asking if schools used 'baseline' or 'on-entry' assessment. She found that most schools had developed their own checklists of skill areas for teachers to complete and some schools used observation as their main method of getting to know children, with only a few schools making use of published assessment material. The majority of teachers said that they did not know a great deal about baseline assessment but the majority believed that it was relevant to their school. Participants of the study concluded that they needed to re-appraise their current practice. Desforges and Lindsay (1995a) argue that although many LEAs have developed programmes for early identification of educational difficulties over the last 20 years, there has been little research into the reliability and validity of these schemes.

Figures reported by the School Curriculum and Assessment Authority (SCAA) in their *Draft Proposals* (1996a) show that baseline assessment was already being used, in some form, by approximately half the LEAs in England and Wales at this time.

SCAA developed three baseline assessment schemes for consideration that involved either comparing the child with performance descriptions or using a checklist to assess children's abilities in reading, writing and mathematics. NFER were commissioned by SCAA (1996c) to independently trial these three schemes in more than 300 schools. Teachers expressed reservations about all three schemes as they found the performance descriptors to be too complex and difficult to apply consistently and the checklists too narrow and misleading because of the 'yes-no' format. Checklists were found to be more satisfactory when they were accompanied by a descriptive record. However this was very time-consuming and therefore not manageable for teachers. According to Sainsbury (1997) none of the three suggested schemes were entirely suitable. Sainsbury et al (1999) report that the SCAA research (1996c) showed children's attainments at baseline to be very varied, reflecting the variety of pre-school experiences and range of ages at which children started school. They argue that it is important that all children should be able to show some positive attainment and suggest that four-point scales are the most appropriate method for this purpose.

The Baseline Assessment Scales published by SCAA (1997a) now appear in a number of accredited baseline assessment schemes. They cover reading for meaning and enjoyment, letter knowledge, phonological awareness, writing, speaking and listening, number, using mathematical language, and, personal and social development. Each scale has four items, mostly arranged in order of difficulty. Children should be assessed on the basis of their typical performance in the classroom.

Ninety-one schemes were initially accredited by QCA (1999a). This was reduced to 90 in September 2000.

4.4 The statutory framework for baseline assessment

The starting point in the political history of baseline assessment lies with the Bullock Report (1975) which recommends that "LEAs and schools should introduce early screening procedures to prevent cumulative language and reading failure and to guarantee individual diagnosis and treatment" (page 514). However, there was little further legal input concerning baseline assessment until the late 1980s.

The 1988 Education Act aimed to increase the accountability of schools through publishing assessment results, public examination results and non-attendance rates (Desforges and Lindsay, 1995a; Lindsay and Desforges, 1998). All 7, 11 and 14 year olds in England and Wales were to be administered Standard Assessment Tasks in English, mathematics and science from 1991 onwards. The purpose of this Act was to raise educational standards and provide better parental choice. Many teachers were concerned about the publication of league tables ranking schools on their performance, especially those inner-city schools that were at a disadvantage. Many supported the use of baseline assessments to measure children's abilities on entry to school in order to ascertain the value added by the school (Lindsay, 1995). Until 1988, LEAs had mainly used eligibility for free school meals as the criteria for allocating special educational needs budgets to schools as this has been shown to be highly correlated with special educational needs (Lindsay, 1993). However, as a result of the 1988 Education Act, baseline assessment began to become more widely used for this purpose.

The Code of Practice for the Identification and Assessment of Special Educational Needs (DfE, 1994) further advocates early screening for special needs. It states that "the earlier action is taken, the more responsive the child is likely to be, and the more readily can intervention be made without undue disruption to the organisation of the school including the delivery of the curriculum for that particular child" (Section 2:16). Moreover, it is the responsibility of schools to identify children with special educational needs as early as possible using suitable screening or assessment tools (Section 2:17). The code also states that lack of proficiency in English should not be equated with learning difficulties, and demands that tests should be 'culturally neutral' (Section

2:18). The revised Code of Practice (DfES, 2001) upholds these requirements in Sections 5:11, 5:13 and 5:16.

SCAA (1996c) conducted a national survey on schools' practice and teachers' views on baseline assessment. Government proposals for a National Framework for Baseline Assessment were then announced in September 1996 (SCAA, 1996a). These proposals state that the National Framework would require accredited baseline assessment schemes to:

- ensure entitlement for all children to be assessed, including those children for whom English is an additional language
- be sufficiently detailed to identify individual children's learning needs, including special educational needs
- enable children's later progress to be monitored effectively
- involve parents or carers in partnership with the school
- take place in the first half-term of the child's entry to the reception class (or year one if the child enters school at that point)
- include assessment of early literacy and numeracy
- be unobtrusive for children (in that they form part of everyday classroom activities);
- be manageable for teachers
- and, provide outcomes which will contribute to value-added measurement.

An important property of these proposals was that baseline assessment should "be sufficiently detailed to identify individual children's learning needs, including special educational needs" (page 12).

These proposals were later revised and the National Framework for Baseline Assessment (SCAA, 1997b) asserted that "Most baseline assessment schemes...will not provide sufficiently detailed assessments on their own to lead schools to place a child on the school's register of children with special educational needs" (page 10). The key reason for this modification was the specification that baseline assessment should be 'manageable' for teachers. Manageability was defined principally in terms of time for administration and minimal disruption to normal classroom activities. For assessments to be manageable it is difficult to make them detailed enough to identify special educational needs. Nevertheless, baseline assessment schemes are expected to give teachers guidance on using the results to plan appropriately for children identified as having special educational needs.

The National Framework for Baseline Assessment requires that baseline assessment schemes must:

- cover aspects of language and literacy, mathematics, and personal and social development as specified in the Desirable outcomes for Children's Learning on Entering Compulsory Education (SCAA, 1996b)

- include clear guidance to teachers on how the outcomes of the assessment can be used to inform the planning both for a class and for the individual children

- provide one or more numerical outcomes capable of being used for later value-added analysis

- specify the period after a child has started primary school within which an assessment should be completed. This period should, in normal circumstances, be no longer than seven weeks after a child had started primary school.

In addition, the National Framework stipulates that if baseline tests are conducted in a way that prevents the teacher from teaching the rest of the class, then the maximum testing time should not normally exceed 20 minutes per child.

SCAA's Baseline Assessment Scales were designed for use within baseline assessment schemes, in accordance with the Desirable Outcomes for Children's Learning on Entering Compulsory Education (SCAA 1997a). The scales use a checklist approach and cover reading, writing, maths, personal and social development and speaking and listening. However, the technical quality of these scales had not been evaluated. This matter will be discussed in more detail later in this chapter.

The recommendations of the National Framework for Baseline Assessment became statute in the 1997 Education Act. From September 1998 all maintained primary schools in England and Wales were required to assess all children on entry using a baseline assessment scheme that conforms to the National Framework. Baseline assessment schemes are accredited and monitored by the Qualifications and Curriculum Authority (QCA; formerly SCAA).

A Department for Education and Employment Circular in June 1998 states that, from the start of the Autumn term 1998, teachers must assess all new 4-5 year old pupils within 7 weeks of starting primary school. The assessment should cover as a minimum the basic skills of speaking and listening, reading, writing, mathematics and personal and social development. Headteachers must give parents the opportunity to discuss how their children performed in the assessments. Schools must also pass on the assessment results to the LEA, which in turn will pass the information on to the QCA. When a pupil changes school, his or her baseline assessment results must be sent to the new school.

Arrangements for baseline assessment are set to change from 2003 with a single national assessment scheme at the end of the Foundation Stage.

4.5 Group differences in baseline assessment

Strand (1999) maintains that girls are typically found to score higher than boys in reading, writing and mathematics throughout Key Stage 1. Indeed, a number of studies using baseline assessment schemes have shown girls to achieve higher scores than boys (Lindsay, 1980; Strand, 1997; Lindsay and Desforges, 1999). In addition, the Annual Report on sample data from the QCA Baseline Assessment Scales for the Autumn 1998 and Spring and Summer 1999 periods (1999b) also shows that, from a nationally representative sample of 6,953 children, girls score higher than boys on all items. It has also been reported that girls make more progress than boys in reading and writing during Key Stage 1 and so increase the gender gap (Strand, 1997, 1999). Strand (1999) reports that girls make less progress in mathematics but still score slightly higher than boys at age seven. However, Tizard et al (1988) indicate that there are no significant differences between boys and girls at the end of their nursery education (average age 4 years 7 months) but that girls are significantly ahead of boys in reading by age seven.

Much research has also shown a diversity on baseline assessment scores between various ethnic groups. A number of studies report that white children score significantly higher than African, Caribbean, Indian, Pakistani and Bangladeshi children

(Everett, Farnsworth and Mitchell, 1997; Strand, 1999; Lindsay and Desforges, 1999) with Pakistani, Indian and Bangladeshi groups having the lowest scores (Strand, 1995; Lindsay, 1988). Strand (1999) also reports that Caribbean children make less progress and Chinese children make more progress during Key Stage 1 than white children. However, Lindsay (1998) found no significant differences between baseline assessment scores of white, Caribbean and Chinese children. Furthermore, a report by Birmingham LEA (1996) asserts that Caribbean children scored above the average for white pupils, and Lindsay and Desforges (1999) argue that Caribbean children have high literacy scores. Strand (1999) demonstrates that where English is the main home language, Indian, Pakistani and Chinese pupils achieve as well as white children. The Annual Report on the QCA Baseline Assessment Scales (1999b) shows that children of Chinese heritage achieved a higher score than other ethnic groups in all scales except phonological awareness, speaking and listening, and personal and social development. Children of white UK heritage scored higher on phonological awareness and speaking and listening than children of any other ethnic group. Nevertheless, Strand (1999) contends that when all pupil background factors are considered simultaneously there are interactions between ethnic group and other pupil background factors and therefore no simple conclusion can be drawn about differences across ethnic groups. This conviction was also advocated by an OFSTED (Office for Standards in Education) review carried out by Gillborn and Gipps (1996).

Children who have English as an additional language are reported to score significantly lower than children for whom English is the main home language (Strand, 1997; Strand, 1999; Lindsay and Desforges, 1999). Lindsay and Desforges demonstrate that scores for English speaking children are significantly higher than those for children from homes where the main languages are Punjabi, Urdu, Bengali and Somali. They also found that children for whom English is the main language perform better than children whose first language is Arabic on all measures except writing. Additionally, Strand (1997, 1999) suggests that children for whom English is an additional language make better progress through Key Stage 1 and so close the gap between them and their peers. He argues that low baseline assessment scores for this group merely indicate unfamiliarity with the English language when first entering school.

Entitlement to free school meals has customarily been taken to be an indicator of social class and it is widely reported that children entitled to free school meals score lower than other children on baseline assessment (Strand, 1997, 1999; QCA 1999b). Strand (1997, 1999) adds that progress in reading, writing and mathematics throughout Key Stage 1 is also slower for children entitled to free school meals, thus widening the gap between them and their peers. However, he also maintains that there is a highly significant interaction between ethnic group and entitlement to free school meals (Strand, 1999).

Many studies report that pupils with experience of early education have higher attainment on baseline assessment than their peers (Strand, 1997, 1999; Lindsay and Desforges, 1999). Furthermore, Strand (1999) claims that the benefits of nursery education are still apparent at age seven. However, Lindsay and Desforges (1999) contest that experience of pre-school education may improve children's abilities at school entry but this advantage may not be maintained to the end of Key Stage 1.

Strand (1997) argues that children in schools with a high percentage of children entitled to free school meals, a low percentage of girls, a high percentage of children for whom English is an additional language and a high average baseline score make significantly less progress than children in other schools. He concludes that schools obtaining high raw scores on Key Stage 1 tests are not necessarily the schools in which children make the most progress during infant education.

Lindsay and Desforges (1999) argue that the issues of group diversity mentioned here have important implications for baseline assessment. First, it is important not to assume that low scores necessarily indicate special educational needs as ethnicity and home language must also be considered. However, they claim that baseline assessment is still useful for determining each child's current level and need. Lindsay and Desforges also argue that as gender, ethnicity and home language differences are found on many baseline assessment scales, this should be used to provide evidence for the validity of the instrument. Wolfendale and Lindsay (1999) also believe that evidence on factors including gender, ethnicity, EAL and social disadvantage should become part of the accreditation process for baseline assessment schemes.

4.6 Technical issues in baseline assessment

Educational screening tests are increasingly being expected to have high levels of accuracy and reliability. Certainly, inaccurate results could have very serious consequences. However, the National Framework for Baseline Assessment does not require that tests are standardised or that there is any evidence of reliability and validity of the schemes. Indeed, none of the three suggested schemes in the Draft Proposals (SCAA, 1996a) had any standardised norms. Desforges and Lindsay (1995a) suggest that this lack of research into reliability, validity and fitness for the purpose makes it seem as if the professional views of those involved in the development of baseline assessments are sufficient to construct an appropriate checklist or screening device.

They conclude that poor quality instruments could lead to significant errors and harmful outcomes for children.

Many LEAs have developed programmes of early identification to be used within the Authority with little or no research into technical quality (Lindsay, 1995). According to the document 'Evaluation of the First Year of Statutory Baseline Assessment 1998/9' (QCA, 1999c), fewer than 50% of LEAs that are scheme providers stated how the reliability of the data on their baseline assessment scheme is ensured. One third of LEAs stated that the analyses of their recommended schemes included the use of standardised scores. However, fewer than 25% of the LEAs explained this in terms of, for example, an age-standardised score, whilst the others explained it in terms of the internal structure of their own schemes, and the fact that they included the QCA Baseline Assessment Scales. Just under 40% of LEAs stated that their recommended schemes included predictive validity in respect of end of KS1 assessments.

Reliability is the extent to which a test yields approximately the same results when used repeatedly under similar conditions. Three important measures of reliability are internal consistency, inter-rater reliability and test-retest reliability. Internal consistency is the degree to which test items measure the same variables. Strand (1999) reports that the baseline assessment scheme used in Wandsworth, shows an internal consistency of 0.90 for teacher-completed checklists and 0.85 for the LARR Test of Emergent Literacy (NFER-Nelson, 1993). Lindsay and Desforges (1999) report that the Infant Index (Desforges and Lindsay, 1995b) has very high internal consistency (0.92).

Inter-rater reliability is the degree to which a test administered by different testers produces approximately the same result. Lindsay and Desforges (1999) report that the Infant Index (Desforges and Lindsay, 1995b) shows a satisfactory inter-rater reliability (0.86).

Test-retest reliability is the degree to which a test administered on more than one occasion, with a reasonably long period of time in between, by the same tester produces approximately the same result. Lindsay and Desforges (1999) report that the Infant Index (Desforges and Lindsay, 1995b) shows satisfactory test-retest reliability (0.89)

Validity is the extent to which a test measures the variable that it is intended to measure. There are various different measures of validity. Construct validity is the degree to which test items assess the hypothetical construct that it was designed to measure. This is usually only used during the initial stages of test construction. According to Lindsay and Desforges (1999) the Infant Index (Desforges and Lindsay, 1995b) and Baseline-PLUS (Desforges and Lindsay, 1998) are shown to have satisfactory construct validity.

Predictive validity is the extent to which test scores predict actual performance. Strand (1997, 1999) reports that the baseline assessment scheme used in Wandsworth, consisting of a teacher-completed checklist and LARR Test of Emergent Literacy (NFER-Nelson, 1993), produces a multiple correlation of 0.60 with attainment at the end of Key Stage 1. The combination of these two baseline assessment instruments was shown to be a better indicator of future attainment than either assessment alone. Lindsay (1998) purports that the most favourable baseline assessments do show satisfactory levels of reliability and construct validity, but predictive validity is more problematic. For example, the Infant Rating Scale (Lindsay, 1981) shows highly significant test-retest reliability (0.96) and sound construct validity (Lindsay, 1980) but more moderate predictive validity (0.54).

Tymms' (1999) findings suggest that there is a limit to the predictability of reading and mathematics amounting to a correlation of about 0.7. According to Desforges and Lindsay (1995a) there is differential predictability of children at the extremes as opposed to the centre of the continuum. Children performing very well or very poorly at age five tend to have high and significant correlations with criterion measures at age seven, whereas those scoring in the borderline 'at risk' region show low, non-significant correlations. Correlation statistics on the whole sample mask this effect.

Wolfendale and Lindsay (1999) assert that the technical quality of schemes should be taken into consideration during the accreditation process.

4.7 The challenges facing baseline assessment

One of the major problems within the area of baseline assessment concerns the lack of evidence of technical quality of the schemes. Fletcher and Satz (1984) compared test-based and teacher-based predications of academic achievement in America. Both show good predictive validity with the overall hit rate (percentage correctly identified as 'at risk' or 'not at risk') for teachers being 74% and for the screening tests 77%. However, teachers made more false negative errors (children identified as 'not at risk' but are later found to have special educational needs) whereas tests made more false positive errors (children identified as 'at risk' but are later found not to have special educational needs). In fact, teachers made so many false negatives that they would have missed 87% of children who would have had severe difficulties in Grade 2, compared to 34% missed by tests. The solution to these difficulties would be the use of test-based rather than teacher-based assessments of children's abilities that are standardised using a national sample and can be shown to be valid and reliable.

A second challenge is that the National Framework demands that baseline assessment time should not exceed 20 minutes per child. Lindsay (1995) agrees that if screening tests are to be used, they should be simple, quick and easy to perform and interpret.

Singleton (1997b) refers to the 'practicality-accuracy dilemma'. In order to be useful, tests have to be fairly accurate which would mean that they would have to be quite complex. Complex assessments are not generally practical as they are not simple to deliver. A solution to the problem of assessing children within the 20 minute time limit whilst remaining fairly accurate would be the use of a computer-based assessment which can be both complex and practical. This could incorporate adaptive testing in order to test children as effectively as possible.

A third difficulty within baseline assessment is that schemes are required to fit a number of different purposes, including the identification of special educational needs, planning teaching and value added assessment. Nutbrown (1999) suggests that, for teaching purposes, details of achievement by each child are more useful than a simple score of 1 or 0. For example, knowing whether a child can write letters or how many letters a child can write is arguably less useful than knowing which letters he or she can write and where they might need help. She argues that baseline assessment as it stands will not improve standards, especially in the area of writing. An answer to this problem may again lie in the use of a computer-based assessment, which can automatically produce a standardised score for value added purposes and also give detailed test information to teachers, including items passed and items failed for the purposes of planning future teaching.

Finally, there are a number of difficulties in the actual measurement of value added by a school. Individual differences exist in early development and it is often the case that children found to be doing well at the age of five may not develop at same pace during Key Stage 1 as other children performing less well at the age of five. Furthermore, Lindsay and Desforges (1998) point out that teachers could mark children down at the initial baseline assessment and then mark them up at the end of Key Stage 1 assessment to increase the value added by the school. In addition, Lindsay (1998) claims that not only do baseline assessment measures not need to meet technical standards but also the end of Key Stage 1 assessments are themselves not of known and acceptable quality. In order to solve these problems the technical quality of Key Stage 1 assessments needs to be investigated. Additionally, the use of a computer-based baseline assessment would prevent teachers from marking children down at this stage.

4.8 Summary and conclusions

Baseline assessment is an evaluation of a child's abilities at the point when they start school. There are four main purposes of baseline assessment, as far as schools are concerned. First, to identify children with educational difficulties so that their special needs can be met. Second, to assess children's abilities at the start of schooling so that their future progress can be measured (value-added). Third, to assess the whole intake in order to plan teaching. Fourth, to help plan the curriculum to suit the needs of individual children. LEAs, however, may see baseline assessment as having other purposes, including helping to determine distribution of resources and monitoring school effectiveness.

Early screening has been practised in schools since the 1970s, with the original focus being on the identification of special educational needs. Early systems concentrated on reading ability. Throughout the 1980s screening became more widespread. However, in the 1990s there was a decline in the use of screening tests as they were not thought to be accurate or provide useful information.

The Bullock Report (1975) advocated early screening procedures and this had an impact on school practice. The 1988 Education Act made baseline assessment more widely used for the purpose of allocating special educational needs budgets to schools and also increased the accountability of schools. The Code of Practice (1994) further advocated early screening for special needs. The National Framework for Baseline Assessment (1997) set in place guidelines for compulsory baseline assessment within seven weeks of entry into primary schools, covering language and literacy, mathematics and personal and social development. These recommendations became statute in the 1997 Education Act and, from September 1998, all maintained primary schools in England and Wales were required to use a baseline assessment scheme accredited by QCA.

A number of researchers have observed significant differences between groups of children on baseline assessment on the basis of gender, ethnic group, home language, social class and experience of early education. However, these differences are not observed in all studies of baseline assessment.

Baseline assessment schemes are not required to provide evidence of reliability and validity. Indeed, many LEAs have developed baseline assessment schemes, for use within the authority, with little or no research into its technical quality. Researchers who do provide evidence of the technical quality of their schemes give correlation coefficients ranging from 0.85 to 0.96 for reliability and 0.54 to 0.60 for validity.

The main challenges facing baseline assessment are the lack of evidence of technical quality of the instruments, the QCA requirement that baseline assessment time should not exceed 20 minutes per child, the requirement for baseline assessment schemes to fit a number of different purposes, and the difficulties in the measurement of value added by a school.

In conclusion, it can be argued that the use of a computer-based system could be a solution to the difficulties facing baseline assessment.

5 Methodology for the baseline assessment studies

5.1 Overview

The baseline assessment research encompasses a number of phases. First, the test development phase, during which test items were developed and tested locally and the results analysed. The children tested during the test development stage were then followed up and re-tested on their literacy skills (after 12 months and again after 32 months) and on their mathematics skills (after 20 months and again after 32 months). Reliability and validity tests were conducted using a different sample of children on a national basis.

5.2 Test development - Summer 1997

5.2.1 Participants

The participants in this study attended seven different schools in the Hull area. Four institutions were primary schools (one was located in the inner city area and three in the outer suburbs) and three were nursery schools (one inner city state nursery, one suburban state nursery and one suburban private nursery). The inner city institution had a very high proportion of children of families of manual or unskilled workers and the unemployed. The suburban schools had a more even mix of children, including a fair proportion of those from white collar and professional families.

153 children (79 boys and 74 girls) were involved in the literacy assessment. These children's ages ranged from 48 months to 66 months, with a mean of 57.9 months (4 years 9 months) and a standard deviation of 4.89 months. 140 children (73 boys and 67

girls) were involved in the mathematics assessment. Their ages ranged from 48 months to 66 months, with a mean of 58.3 months (4 years 10 months) and a standard deviation of 4.98 months. 138 of the children were involved in both the literacy and mathematics assessments, 15 were involved in only the literacy assessment and 2 were involved in only the mathematics assessment.

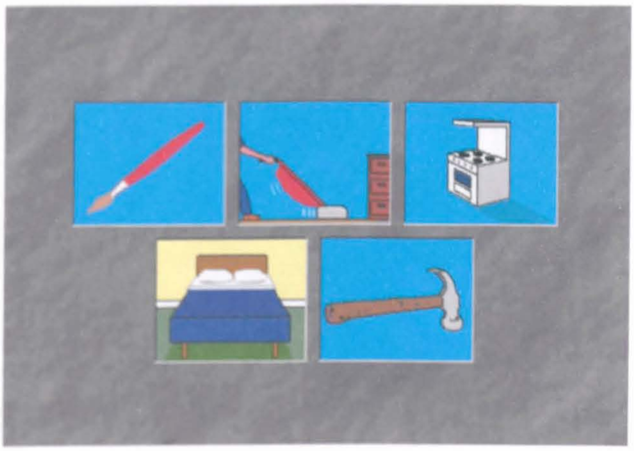
The children were selected at random from the reception classes of the primary schools and from those children in the nursery schools who had passed their fourth birthday. All children were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

5.2.2 Procedure

CoPS Baseline comprises four modules: literacy; mathematics; communication; and, personal and social development. However, this study only includes the literacy and mathematics modules.

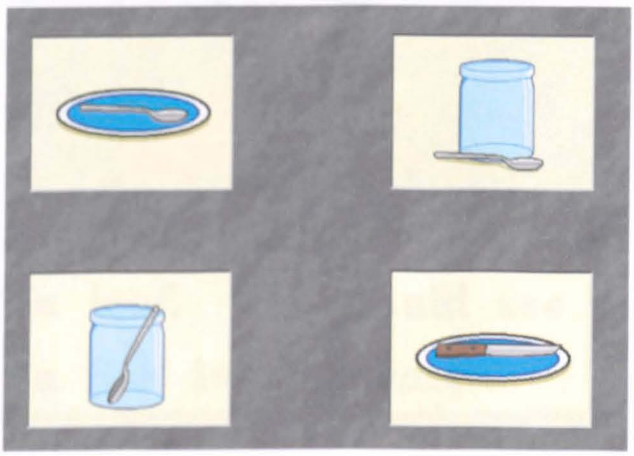
For these two modules, several skill or concept areas were derived from the Desirable Outcomes for Children's Learning on Entering Compulsory Education (SCAA, 1996b). For each of those areas, items were devised which were administered to the sample. The literacy module comprised 1 practice item and 61 test items within the areas of: verbal concepts (10 items); aural comprehension (7 items); rhymes (7 items); alliteration (8 items); knowledge about print (8 items); letter recognition (7 items); simple reading (7 items); and, simple spelling (7 items). Examples of each type of literacy item are given on the following pages (See Appendix 1 for a list of all items).

Figure 1: CoPS Baseline - Literacy - Verbal concepts



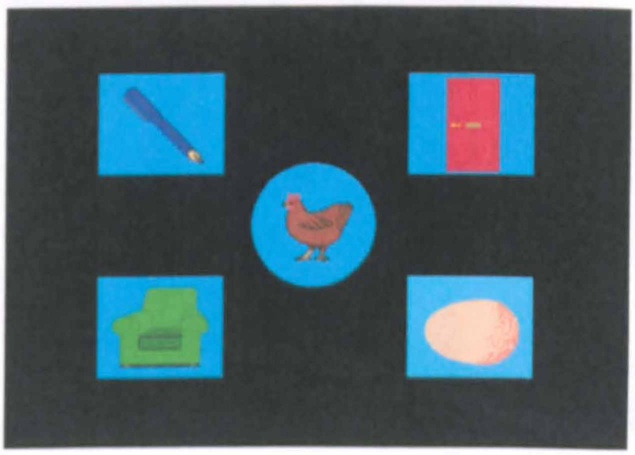
Show me the picture that goes best with the word cooking.

Figure 2: CoPS Baseline - Literacy - Aural comprehension



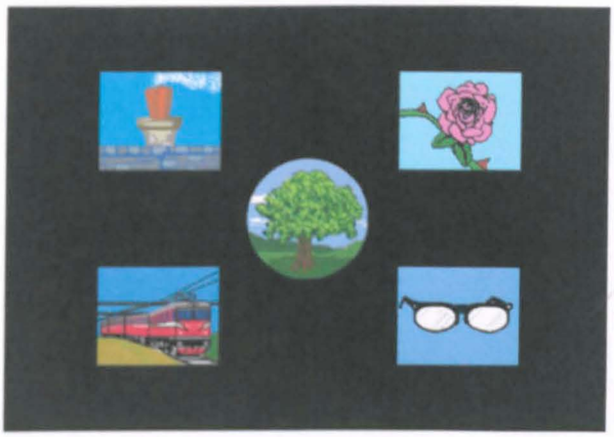
Show me the correct picture for 'The spoon is in the jar'.

Figure 3: CoPS Baseline - Literacy - Rhymes



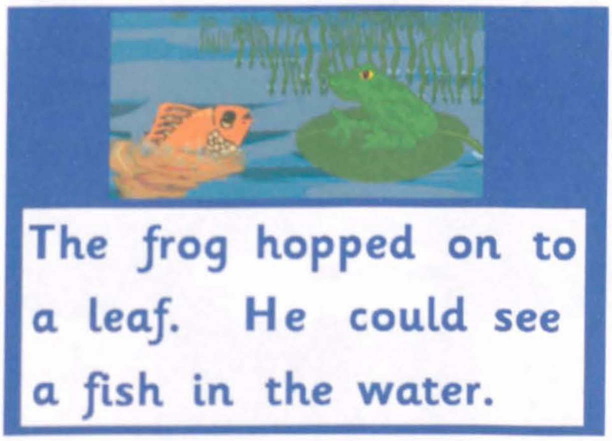
Listen carefully. Pen...door...chair...egg...hen. Which one sounds like hen?

Figure 4: CoPS Baseline - Literacy - Alliteration



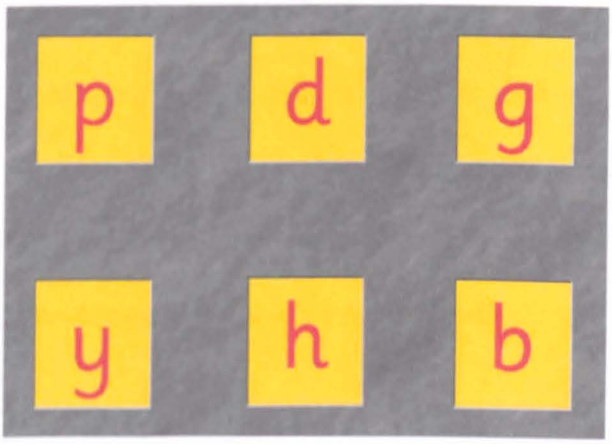
Listen carefully. Chimney...flower...train...glasses...tree. Which one begins with the same sound as tree?

Figure 5: CoPS Baseline - Literacy - Knowledge about print



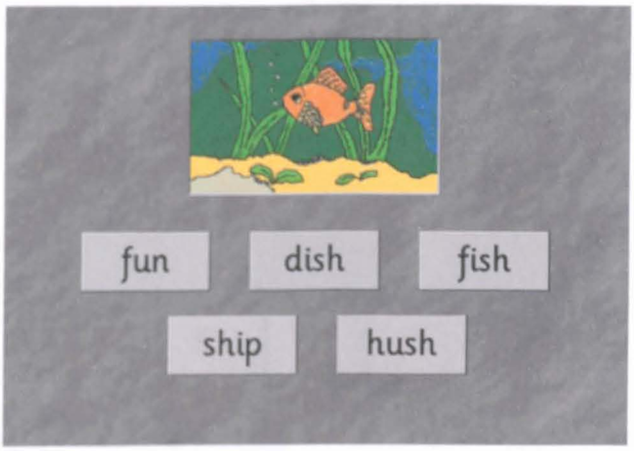
The frog hopped on to a leaf. He could see a fish in the water. Show me a capital letter.

Figure 6: CoPS Baseline - Literacy - Letter recognition



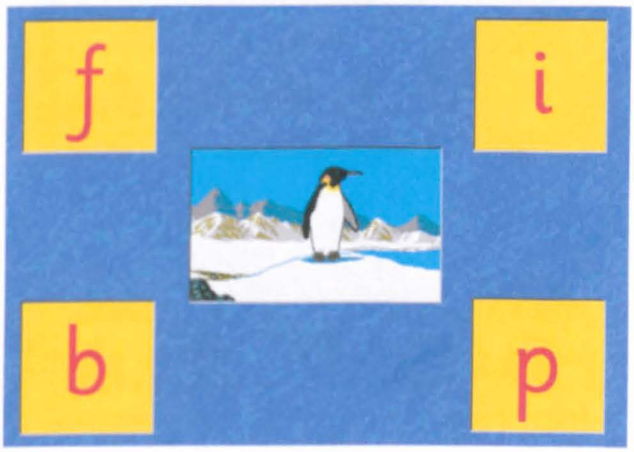
Show me the letter g.

Figure 7: CoPS Baseline - Literacy - Simple reading



This is a fish. Show me the word that says fish.

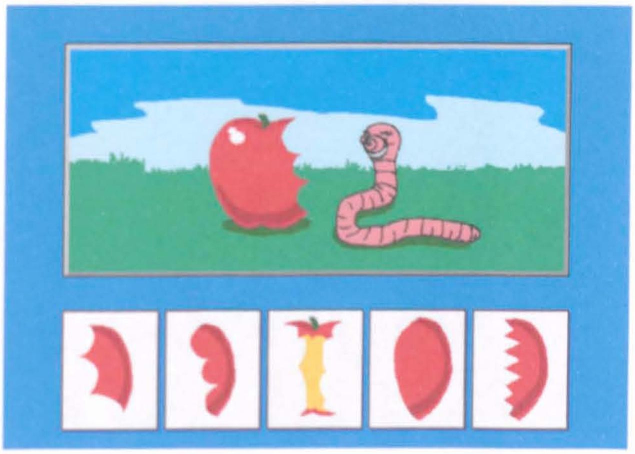
Figure 8: CoPS Baseline - Literacy - Simple spelling



This is a penguin. Which letter does the word penguin begin with?

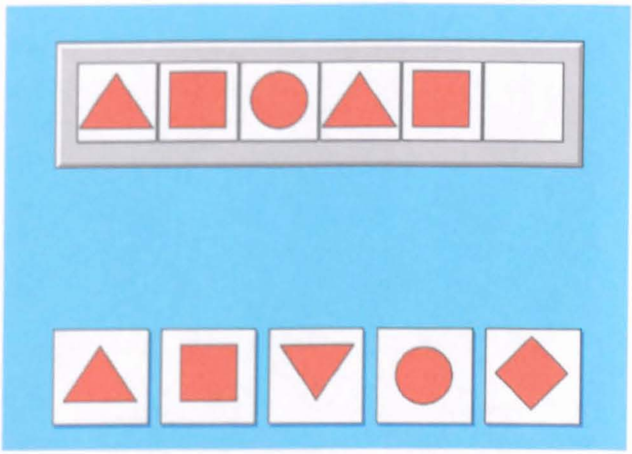
The mathematics module consisted of 1 practice item and 59 test items within the areas of: shape (6 items); pattern (7 items); relative size and quantity (7 items); sets (5 items); seriation (7 items); numbers (5 items); counting (6 items); addition (8 items); and, subtraction (8 items). Examples of mathematics items are given on the following pages (See Appendix 2 for a list of all items).

Figure 9: CoPS Baseline - Mathematics - Shape



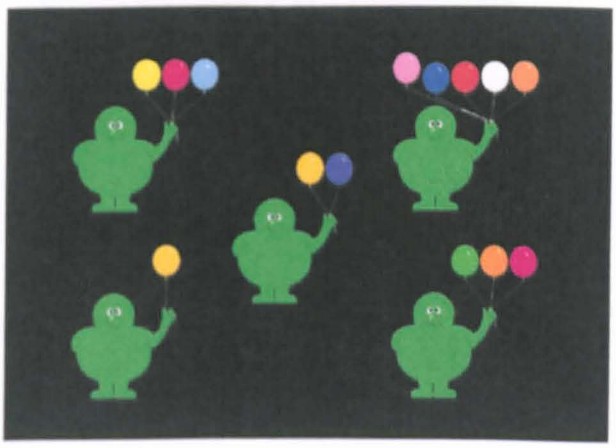
Wendy the worm has taken a big bite out of the apple. Show me which piece she had.

Figure 10: CoPS Baseline - Mathematics - Pattern



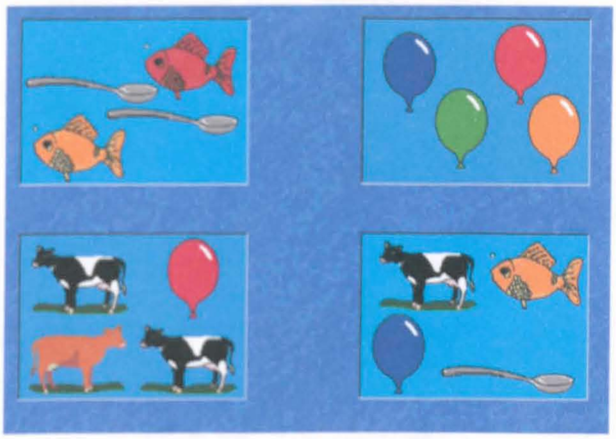
Here are some shapes in a row. Show me the shape which comes next.

Figure 11: CoPS Baseline - Mathematics - Relative size and quantity



Here are some of Zoid's friends. Two of them have the same number of balloons. Show me which two friends have the same number of balloons.

Figure 12: CoPS Baseline - Mathematics - Sets



Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.

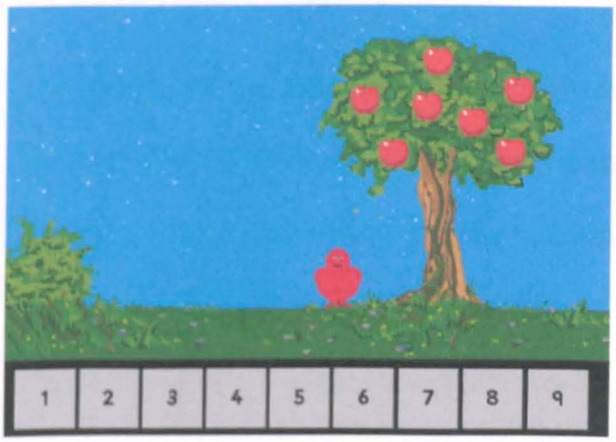
Figure 13: CoPS Baseline - Mathematics - Seriation

Zoid's friends are at the seaside. They are going to buy ice-cream. Can you see them waiting in a line? Which friend is last in the line?

Figure 14: CoPS Baseline - Mathematics - Number recognition

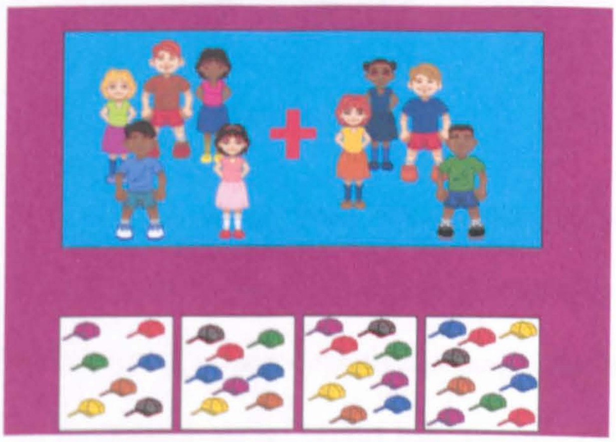
Here are some coins. Show me the 10p.

Figure 15: CoPS Baseline - Mathematics - Counting



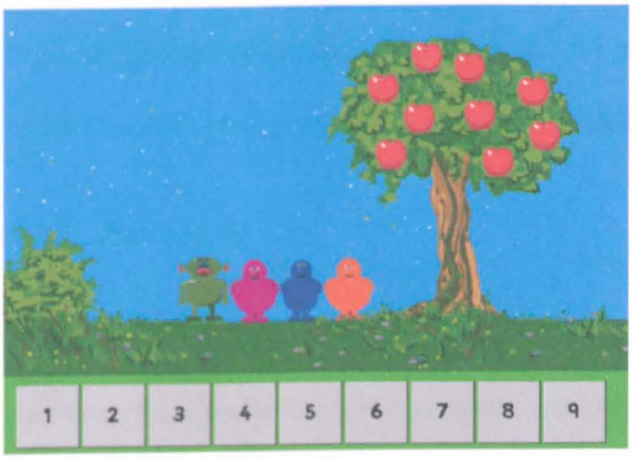
This is Zoid's apple tree. How many apples can you see? Show me the number on the number line.

Figure 16: CoPS Baseline - Mathematics - Addition



Here are some friends. There are five friends in one group and four friends in another group. They all want to wear hats. How many hats do they need altogether? Show me the picture with the right number of hats for all of them.

Figure 17: CoPS Baseline - Mathematics - Subtraction



There are nine apples on this tree. Zoid's friends are going to pick one apple each. How many apples will be left on the tree. Show me the answer on the number line.

The assessment was carried out individually in a quiet area of the school, using paper-based versions of the two modules. Paper-based versions of test items are commonly used in the early stages of development of computerised tests. This allows for large-scale data collection to enable reliable item analysis without the cost of unnecessary programming at this stage. The pupils' teachers were not informed of the pupils' results.

5.3 Literacy follow-up part 1 - Summer 1998

5.3.1 Participants

For practical reasons, data collection for this study was confined to the four primary schools that had participated in the test development study. Where the children in the three original nursery schools had progressed to one of these primary schools they were included in the study. If they had progressed to other primary schools, had moved out of the area or were absent from school on the days in which the literacy follow-up assessment was conducted, they were dropped from the study. A total of 92 of the original 153 children were assessed in the literacy follow-up (44 boys and 48 girls). The children in this group ranged in age from 60 months to 76 months, with a mean age of 69.3 months (5 years 9 months) and a standard deviation of 3.81 months.

5.3.2 Instruments

The British Ability Scales (Second Edition) Word Reading Test (Elliott, 1996) is a conventional, standardised, single-word, out-of-context oral reading test that comprises 60 items in increasing order of difficulty. The age range of the test is 5 years 0 months to 17 years 11 months.

The British Ability Scales (Second Edition) Spelling Test (Elliott, 1996) is a conventional, standardised, contextual spelling test that comprises 75 items in increasing order of difficulty. The age range of the test is 5 years 0 months to 17 years 11 months.

5.3.3 Procedure

Assessment using the BAS-II Word Reading Test and the BAS-II Spelling Test was carried out individually, in a quiet area of the school, by a psychologist who had no prior knowledge of the children's baseline assessment results. The teachers were not informed of the pupils' results.

5.4 Mathematics follow-up part 1 - Spring 1999

5.4.1 Participants

Again, data collection for this study was confined to the four primary schools that had participated in the test development study. Where the children in the three original nursery schools had progressed to one of these primary schools they were included in the study. If they had progressed to other primary schools, had moved out of the area or were absent from school on the days in which the mathematics follow-up assessment was conducted, they were dropped from the study. A total of 83 of the original 140 children were assessed in the mathematics follow-up (40 boys and 43 girls). The children in this group ranged in age from 70 months to 87 months, with a mean age of 80.9 months (6 years 8 months) and a standard deviation of 4.62 months.

5.4.2 Instruments

The Wechsler Objective Numerical Dimensions (WOND) Numerical Operations Test (Wechsler, 1996b) assesses the ability to write dictated numerals and solve calculation problems involving all basic operations (addition, subtraction, multiplication and division). It comprises 40 items in increasing order of difficulty. The age range of the test is 6 years 0 months to 16 years 11 months. The test is a group test and takes approximately 15 minutes to complete.

5.4.3 Procedure

The WOND Numerical Operations Test was administered to groups of children by a psychologist who had no prior knowledge of the children's baseline assessment results. The teachers were not informed of the pupils' results.

5.5 Vocabulary follow-up - Spring 1999

5.5.1 Participants

Once more, data collection for this study was confined to the four primary schools that had participated in the test development study. Where the children in the three original nursery schools had progressed to one of these primary schools they were included in the study. If they had progressed to other primary schools, had moved out of the area or were absent from school on the days in which the mathematics follow-up assessment was conducted, they were dropped from the study. A total of 82 of the original children were assessed in the vocabulary follow-up (40 boys and 42 girls). The children in this group ranged in age from 70 months to 87 months, with a mean age of 80.9 months (6 years 8 months) and a standard deviation of 4.64 months.

5.5.2 Instruments

The British Picture Vocabulary Scale - Second Edition (Dunn and Dunn, 1997) is a test of hearing or receptive vocabulary for standard English that comprises 168 items in increasing order of difficulty. The age range of the test is 3 years 0 months to 15 years 8 months. Test administration is on an individual basis. Participants are presented with four simple black and white illustrations on a page and asked to select the picture considered to illustrate best the meaning of the target word spoken by the tester. After the child has completed some practice items, his or her starting point and basal level are established. Testing carries on until the ceiling level is reached. The average administration time per child is ten to fifteen minutes.

5.5.3 Procedure

Assessment using the BPVS was carried out individually, in a quiet area of the school, by a psychologist who had no prior knowledge of the children's baseline assessment results. The teachers were not informed of the pupils' results.

5.6 Literacy follow-up part 2 - Spring 2001

5.6.1 Participants

The participants in this study included all those from the test development study who could be traced to their current school. If the children could not be traced, had moved out of the area or were absent from school on the days in which the literacy follow-up assessment was conducted, they were dropped from the study. A total of 70 of the original 153 children were assessed in the second literacy follow-up (31 boys and 39 girls). The children in this group ranged in age from 89 months to 110 months, with a mean age of 102.9 months (8 years 7 months) and a standard deviation of 4.87 months.

5.6.2 Instruments

The SPAR Spelling and Reading Tests - Second Edition (Young, 1987) were used to assess the participants reading and spelling abilities. The SPAR Reading Test consists of 15 single word reading items and 30 sentence completion items and takes about 15 minutes to complete. The SPAR Spelling Test is a conventional, standardised, contextual spelling test that comprises 40 items in increasing order of difficulty. The age range of the tests is 7 years 0 months to 12 years 11 months.

The Wechsler Objective Language Dimensions (WOLD) Written Expression Test (Wechsler, 1996a) assesses various writing skills: ideas and development; organisation, unity and coherence; vocabulary; sentence structure and variety; grammar and usage; and, capitalisation and punctuation. Participants are prompted to write a letter to someone who is designing a place for them to live, describing their ideal home. They are given 15 minutes to complete their writing. The age range of the test is 8 years 0 months to 16 years 11 months.

5.6.3 Procedure

The SPAR Reading and Spelling Tests and the WOLD Written Expression Test were administered to a group of children by a psychologist who had no prior knowledge of the children's baseline assessment results.

5.7 Mathematics follow-up part 2 - Spring 2001

5.7.1 Participants

Again, the participants in this study included all those from the test development study who could be traced to their current school. If the children could not be traced, had moved out of the area or were absent from school on the days in which the mathematics follow-up assessment was conducted, they were dropped from the study. A total of 84 of the original 140 children were assessed in the second mathematics follow-up (40 boys and 44 girls). The children in this group ranged in age from 92 months to 111 months, with a mean age of 103.57 months (8 years 7 months) and a standard deviation of 4.75 months.

5.7.2 Instruments

Mathematics 8 (Patilla, 1994) assesses children's mathematical skills, including: understanding number; non-numerical processes; computation and knowledge; mathematical interpretation; and, mathematical application. It consists of 35 items in increasing order of difficulty. The age range of the test is 7 years 0 months to 9 years 2 months. The test is a group test and takes about 40 to 60 minutes to complete.

5.7.3 Procedure

The Mathematics 8 Test was administered to a group of children by a psychologist who had no prior knowledge of the children's baseline assessment results.

5.8 Reliability and validity - Spring 2001

5.8.1 Participants

The participants in this study attended eight different primary schools in England and Wales. Two of the schools were located in an inner city area, five were located in a suburban area and one school was in a rural area. The inner city schools had a high proportion of children of families of manual workers and the unemployed. Three of the suburban schools had mostly children of manual and skilled workers, whilst two suburban schools had a more even mix of children, including a fair proportion of those from white collar and professional families. The rural school had a high proportion of children from skilled and professional families.

The inner city schools had an average of 306 pupils in the whole school, 40 per year group and 84 on the SEN register. The suburban schools had an average of 370 pupils in the whole school, 62 per year group and 53 on the SEN register. The rural school had 300 in the whole school, 56 per year group and 54 on the SEN register.

60 children (26 boys and 34 girls) participated in the study. Their ages ranged from 48 months to 67 months, with a mean of 57.5 months (4 years 9 months) and a standard deviation of 5.39 months.

15 children from each school were selected randomly from the reception class registers. All children were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

5.8.2 Instruments

CoPS Baseline (Singleton, Thomas and Horne, 1998) is a computer-based assessment instrument comprising four modules: literacy; mathematics; communication; and, personal and social development. Each module takes approximately five minutes to complete. Norms are built into the system, which generates numerical results as well as written reports for teachers and parents. The system was accredited by the Qualifications and Curriculum Authority in 1998, for use as an on-entry assessment system in schools in England and Wales. The package includes a CD-Rom containing the assessment program, a teacher's manual and a video programme to be used in training.

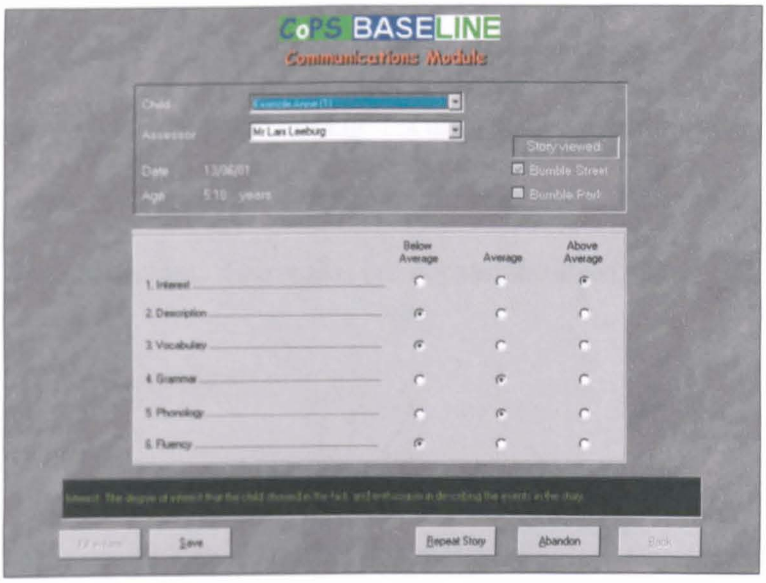
The Literacy module assesses the receptive language abilities which form the basis of effective literacy development, including verbal comprehension, awareness of print, letter recognition, phonological skills, basic reading and basic spelling. It is an adaptive computerised test of 16 items taken from an array of 56 items. See section 5.2.2 for examples of items and Appendix 1 for a full list of items.

The Mathematics module assesses the concepts and skills which young children need for mathematics learning, including grasp of fundamental mathematical language, recognition of shape and pattern, understanding of classification, seriation and ordinal position, basic number recognition, simple addition and simple subtraction. It is an adaptive computerised test of 16 items taken from an array of 56 items. See section 5.2.2 for examples of items and Appendix 2 for a full list of items.

The Communication module assesses the fundamental expressive language abilities required for good communication and learning, including vocabulary knowledge, maturity of grammar and phonology, fluency of expression and accuracy of description. The child views an animated story on the computer (from a choice of two stories) and then describes what he or she has seen and what happened in the story. The tester then answers six questions about the child's description, entering their answers into the computer. Screenshots of the animated story and the assessment form are shown below (See Appendix 3 for a full list of items).

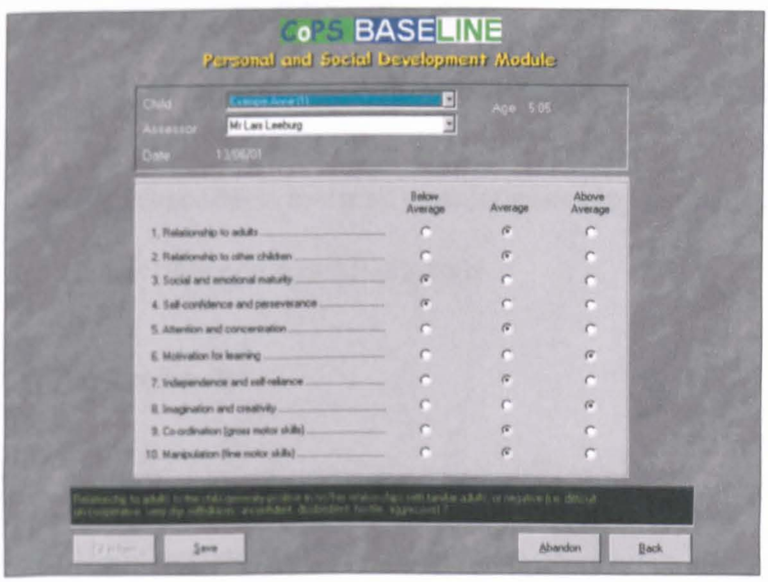
Figure 18: CoPS Baseline - Communication - Bumble Street

Figure 19: CoPS Baseline - Communication form



The Personal and Social Development module assesses aspects of the child's personal and social development that are particularly important for early learning and functioning in a school environment. These include maturity of social and emotional behaviour, the child's relationship with peers and adults, concentration and attention, motivation of learning, motor skills and co-ordination. It comprises an inventory of 10 items, the tester entering responses into the computer. A screenshot of the assessment form is shown below (See Appendix 4 for a full list of items).

Figure 20: CoPS Baseline - Personal and Social Development form



The QCA Baseline Assessment Scales (Qualifications and Curriculum Authority, 1997) cover reading, writing, speaking and listening, mathematics, and, personal and social development. There are eight scales, each containing four items increasing in difficulty. These are paper-based tests designed for use with children in their first seven weeks in a reception class. The scales are not standardised, providing only raw scores.

5.8.3 Procedure

The four CoPS Baseline modules were administered in a specified order, with each school completing the tests in a different order. Pupils were re-tested on the CoPS Baseline modules, in the same order, four weeks later. The QCA Scales were administered by the pupils' teachers, either before or after the first CoPS Baseline assessment.

5.9 Questionnaire study - Autumn 1997 - Spring 2001

5.9.1 Participants

29 teachers using CoPS Baseline, in England, Wales and Ireland, were involved in a questionnaire study between October 1997 and March 2001.

10 of the respondents returned questionnaires in October 1997, 13 in December 1998, 2 in October 1999 and 4 in March 2001.

5.9.2 Instruments

The CoPS Baseline Questionnaire (October 1997) consisted of 61 questions covering installation, registration, the communication module, the literacy module, the mathematics module, the personal and social module, the assessment in general, the reports and the system overall.

The questionnaire was revised in November / December 1998 to include a section on training and becoming familiar with the system. It now consisted of 72 questions. The questionnaire remained the same for the October 1999 administration.

In March 2001, the questionnaire was reduced to 50 items, although it still covered the same areas covered in the previous versions. The questionnaire is shown in full in Appendix 5.

5.9.3 Procedure

Questionnaires were sent to four random samples of teachers, using CoPS Baseline Assessment, on different occasions (October 1997, November / December 1998, October 1999 and March 2001).

6 Results for the baseline assessment studies

6.1 Overview

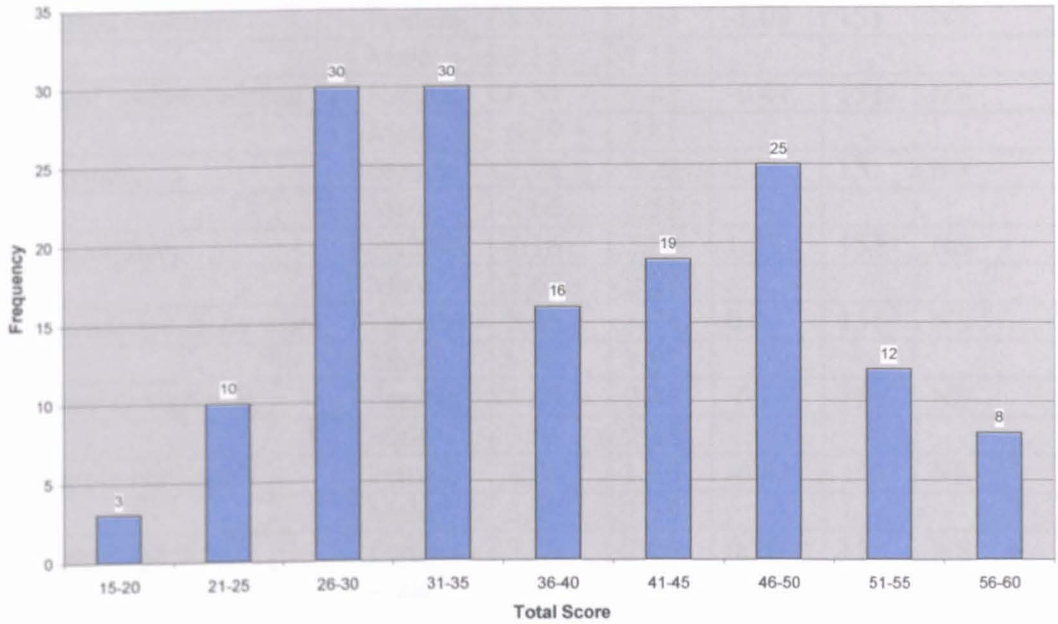
The baseline assessment research encompasses a number of phases. First, the test development phase, during which test items were developed, tested locally and the results analysed. The children tested during the test development stage were then followed up and re-tested on their literacy skills (after 12 months and again after 32 months) and on their mathematics skills (after 20 months and again after 32 months). Reliability and validity tests were conducted using a different sample of children on a national basis.

6.2 Test development - Summer 1997

6.2.1 Literacy

The baseline literacy data (see Figure 21) were distributed satisfactorily (mean 38.10; s.d. 10.02; range 19 to 59), without ceiling or floor effects, and no child obtained an overall score near or below chance level (15).

Figure 21: Distribution of total scores from the baseline literacy assessment



The means, standard deviations and ranges for each of the eight literacy skill / concept areas are given in Table 1.

Table 1: Descriptive statistics for the literacy skill / concept areas (n = 153)

	Minimum	Maximum	Mean	Standard Deviation
A) Verbal concepts	5	10	9.06	1.09
B) Aural comprehension	3	7	6.57	0.81
C) Rhymes	0	7	2.71	2.26
D) Alliteration	0	8	3.92	2.43
E) Knowledge about print	1	8	5.86	1.62
F) Letter recognition	0	8	3.59	2.42
G) Simple reading	0	7	3.32	1.60
H) Simple spelling	0	7	3.15	2.18

Analysis by gender revealed no significant differences in total score (boys: mean 37.89, s.d. 9.85, n= 79; girls: mean 38.26, s.d. 10.27, n=74; t = 0.23, not significant). Nor did the results of any of the separate skill / concept areas show any significant gender differences (see Table 2).

Table 2: Gender differences on the eight baseline literacy skill / concept areas (n=153)

	Gender	Mean	SD	t	df	Significance
A) Verbal concepts	Female	8.96	1.04	-1.09	151	NS
	Male	9.15	1.13			
B) Aural comprehension	Female	6.54	0.81	-0.41	151	NS
	Male	6.59	0.81			
C) Rhymes	Female	2.76	2.22	0.27	151	NS
	Male	2.66	2.31			
D) Alliteration	Female	4.16	2.58	1.22	151	NS
	Male	3.68	2.27			
E) Knowledge about print	Female	5.86	1.74	0.06	151	NS
	Male	5.85	1.52			
F) Letter recognition	Female	3.58	2.48	-0.07	151	NS
	Male	3.61	2.38			
G) Simple reading	Female	3.26	1.65	-0.43	151	NS
	Male	3.37	1.55			
H) Simple spelling	Female	3.20	2.05	0.29	151	NS
	Male	3.10	2.31			

Intercorrelations between the eight skill / concept areas were calculated using Pearson Product Moment Correlation Coefficient. The results (Table 3) indicate that all eight areas had significant correlations with the total score ($p<0.001$ in all cases) showing a high degree of cohesiveness of the components of the test. The two skill / concept areas producing the lowest correlations with total score were verbal concepts ($r = 0.31$) and aural comprehension ($r = 0.32$). Those producing the highest correlations were simple spelling ($r = 0.85$), alliteration ($r = 0.81$), knowledge about print ($r = 0.78$), simple reading ($r = 0.78$) and letter recognition ($r = 0.75$). Amongst the intercorrelations of the skill / concept areas, the only non-significant correlations were between detection of rhyme and aural comprehension ($r = 0.03$), between detection of rhyme and verbal concepts ($r = -0.01$), between verbal concepts and alliteration ($r = 0.14$) and between verbal concepts and simple reading ($r = 0.14$).

Table 3: Intercorrelations between the eight skill / concept areas of the literacy assessment (n = 153)

	A	B	C	D	E	F	G	H
B	0.37***							
C	-0.01	0.03						
D	0.14	0.16*	0.42***					
E	0.27**	0.26**	0.32***	0.53***				
F	0.16*	0.20*	0.22**	0.47***	0.64***			
G	0.14	0.16	0.38***	0.66***	0.53***	0.48***		
H	0.17*	0.22**	0.33***	0.68***	0.63***	0.61***	0.72***	
Total	0.31***	0.32***	0.57***	0.81***	0.78***	0.75***	0.78***	0.85***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

Internal consistency was calculated for each of the eight skill / concept areas using the Kuder-Richardson formula. The coefficients range from 0.42 to 0.82 for the individual skill / concept areas with an overall internal consistency coefficient of 0.91 (see Table 4).

Table 4: Kuder-Richardson coefficients for the eight baseline literacy skill / concept areas

Skill area	Kuder-Richardson
A	0.42
B	0.46
C	0.79
D	0.75
E	0.62
F	0.82
G	0.48
H	0.81
Total	0.91

A principal components analysis was run to ascertain the number of factors that could be separated out. For the literacy module, 19 factors had eigen values above 1. The first of these accounted for approximately 9% of the total variance (see Table 5).

Table 5: CoPS Baseline literacy module PCA - total variance explained

Component	% of Variance	Cumulative %
1	9.39	9.39
2	6.05	15.44
3	4.99	20.43
4	4.49	24.92
5	4.30	29.21
6	4.01	33.23
7	3.85	37.08
8	3.72	40.80
9	3.27	44.07
10	3.15	47.21
11	2.81	50.02
12	2.76	52.78
13	2.65	55.43
14	2.64	58.07
15	2.57	60.63
16	2.46	63.09
17	2.38	65.47
18	2.32	67.80
19	2.07	69.86

A rotated factor analysis was used to calculate the factor loadings. The first factor is represented largely by the items from the letter recognition skill / concept area but also from the knowledge about print and simple spelling areas (see Appendix 6). The second factor corresponded to the rhyme items, while the remaining 17 factors had contributions from items in various skill / concept areas.

Pass rates for each item were calculated and those items with appropriate difficulty for the age ranges were included within the final suite. Five items were removed because they were either too difficult or too easy (verbal concepts, 3 items; knowledge about print, 1 item; and, alliteration, 1 item) leaving a total of 56 items (7 in each of the 8 skill / concept areas). The easiest item in each skill / concept area was passed by approximately 75% of the children whilst the hardest item within each skill / concept area was passed by approximately 40% of the children.

The mean time taken to complete the items in the literacy module was 19.66 minutes (s.d. 3.93). The time available for assessment in each module (5 minutes) precluded conventional assessment using all these items. Therefore, experimentation with various adaptive models was carried out, checking the output of each model against the real data that had been obtained from the children. A fairly uncomplicated algorithm was most suitable. Children begin the test at a difficulty level appropriate to their age and then progress through each skill / concept area, being given harder items when they pass and easier items when they fail. For each skill / concept area the child attempts two items, making a total of 16 items for the adaptive form of the test (as opposed to 56 in the full form). The difficulty of the first item, within each skill / concept area, is based on performance on previous items (or upon age if the first item in the test). The difficulty of the second item is increased when the first item was passed or decreased when the first item was failed. In addition, the teacher is given the opportunity to modify the start point in the test if it is believed that the child is 'above average' or 'below average'.

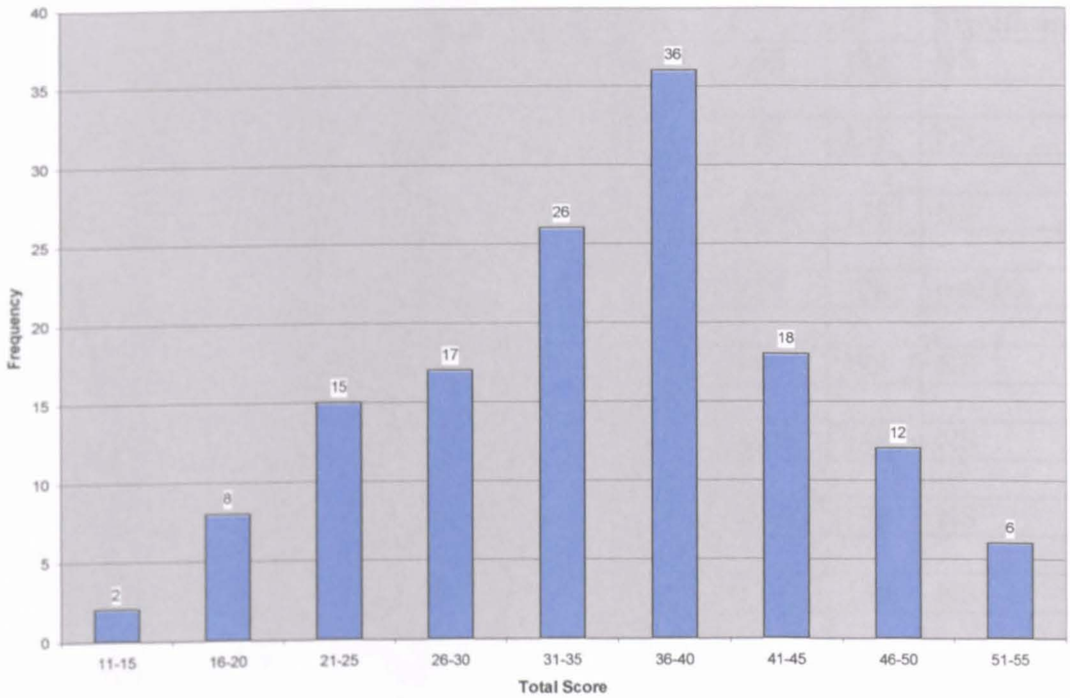
The validity of the adaptive algorithm was investigated by simulating the adaptive form of the CoPS Baseline literacy test with data that had been collected from the children in this study using the full form of the test. The two measures (original raw scores and adaptive form score) were correlated using the Pearson Product Moment Correlation, with a positive correlation coefficient of 0.81 being obtained ($n = 153$, $p<0.001$, 1-tailed).

The equivalency of the computerised literacy items were checked by retesting pupils, from two of the primary schools, using a computerised full form of the test. A significant Pearson Product Moment Correlation of 0.94 was obtained ($n = 19$, $p<0.001$). Item order (based on pass rates) remained the same as the pen and paper version.

6.2.2 Mathematics

The baseline mathematics data (see Figure 22) were distributed satisfactorily (mean 34.76; s.d. 9.07; range 12 to 54), without ceiling or floor effects. However, two children obtained an overall score below chance level (14).

Figure 22: Distribution of total scores from the baseline mathematics assessment



The means, standard deviations and ranges for each of the nine mathematics skill / concept areas are given in Table 6.

Table 6: Descriptive statistics for the mathematics skill / concept areas (n = 140)

	Minimum	Maximum	Mean	Standard Deviation
A) Shape	1	6	4.23	0.98
B) Pattern	0	7	3.30	1.64
C) Relative size and quantity	1	7	5.53	1.39
D1) Sets	0	5	2.54	1.57
D2) Seriation	1	7	3.83	1.22
E) Numbers	0	5	3.96	1.35
F) Counting	1	6	4.64	1.55
G) Addition	0	8	3.59	2.04
H) Subtraction	0	8	3.04	1.62

Analysis by gender revealed no significant differences in total score (boys: mean 34.52, s.d. 9.03, n=73; girls: mean 35.03, s.d. 9.18, n=67; t = 0.33, not significant). Eight of the nine separate skill / concept areas did not show any significant gender differences (see Table 7). However, girls did score significantly higher than boys on the component assessing knowledge of sets (t = 2.59, p<0.05).

Table 7: Gender differences on the nine baseline mathematics skill / concept areas

	Gender	Mean	SD	t	df	Significance
A) Shape	Female	4.37	0.97	1.68	138	NS
	Male	4.10	0.99			
B) Pattern	Female	3.32	1.61	0.20	138	NS
	Male	3.27	1.68			
C) Relative size / quantity	Female	5.42	1.59	-0.90	138	NS
	Male	5.63	1.18			
D1) Sets	Female	2.90	1.46	2.59	138	p<0.05
	Male	2.22	1.62			
D2) Seriation	Female	3.73	1.16	-0.90	138	NS
	Male	3.92	1.28			
E) Numbers	Female	3.91	1.32	-0.45	138	NS
	Male	4.01	1.38			
F) Counting	Female	4.61	1.59	-0.17	138	NS
	Male	4.66	1.52			
G) Addition	Female	3.61	2.07	0.11	138	NS
	Male	3.58	2.02			
H) Subtraction	Female	3.07	1.67	0.22	138	NS
	Male	3.01	1.58			

Intercorrelations between the nine skill / concept areas were calculated using Pearson Product Moment Correlation Coefficient. The results (Table 8) indicate that all nine areas had significant correlations with the total score (p<0.001 in all cases) showing a high degree of cohesiveness of the components of the test. The two skill / concept areas producing the lowest correlations with total score were shape (r = 0.45) and seriation (r = 0.58). Those producing the highest correlations were addition (r = 0.80), counting (r = 0.76), relative size and quantity (r = 0.72) and numbers (r = 0.70). Amongst the intercorrelations of the skill / concept areas, the only non-significant correlation was between shape and relative size and quantity (r = 0.16).

Table 8: Intercorrelations between the nine skill / concept areas of the mathematics assessment (n = 140)

	A	B	C	D1	D2	E	F	G	H
B	0.26 **								
C	0.16	0.47 ***							
D1	0.26 **	0.37 ***	0.42 ***						
D2	0.30 ***	0.42 ***	0.31 ***	0.29 **					
E	0.26 **	0.42 ***	0.38 ***	0.25 **	0.35 ***				
F	0.31 ***	0.41 ***	0.50 ***	0.35 ***	0.33 ***	0.66 ***			
G	0.31 ***	0.41 ***	0.52 ***	0.36 ***	0.36 ***	0.55 ***	0.59 ***		
H	0.18 *	0.33 ***	0.48 ***	0.34 ***	0.32 ***	0.36 ***	0.45 ***	0.53 ***	
Total	0.45 ***	0.69 ***	0.72 ***	0.61 ***	0.58 ***	0.70 ***	0.76 ***	0.80 ***	0.68 ***

* = p<0.05; ** = p<0.01; *** = p<0.001

Internal consistency was calculated for each of the nine skill / concept areas using the Kuder-Richardson formula. The coefficients range from 0.24 to 0.72 for the individual skill concept areas with an overall internal consistency coefficient of 0.88 (see Table 9).

Table 9: Kuder-Richardson coefficients for the nine baseline mathematics skill / concept areas

Skill area	Kuder-Richardson
A	0.24
B	0.51
C	0.55
D1	0.63
D2	0.41
E	0.72
F	0.66
G	0.63
H	0.52
Total	0.88

A principal components analysis was run to ascertain the number of factors that could be separated out. For the mathematics module, 21 factors had eigen values above 1. The first of these accounted for approximately 10% of the total variance (see Table 10).

Table 10: CoPS Baseline mathematics module PCA - total variance explained

Component	% of Variance	Cumulative %
1	9.63	9.63
2	5.04	14.66
3	4.00	18.66
4	3.55	22.21
5	3.34	25.55
6	3.33	28.87
7	3.26	32.13
8	3.17	35.30
9	3.14	38.44
10	3.07	41.51
11	3.04	44.55
12	2.92	47.47
13	2.79	50.26
14	2.68	52.94
15	2.65	55.59
16	2.58	58.17
17	2.55	60.71
18	2.54	63.25
19	2.50	65.75
20	2.50	68.25
21	2.38	70.62

A rotated factor analysis was used to calculate the factor loadings. The first factor is represented largely by the items from the numbers skill / concept area but also from the counting and addition areas (see Appendix 7). The second factor corresponded to the sets items, while the third factor had contributions from some items within the seriation skill / concept area. The remaining factors had contributions from items in various skill / concept areas.

Pass rates for each item were calculated and those items with appropriate difficulty for the age ranges were included within the final suite. Two of the skill / concept areas were combined (shape and pattern). Nine items were removed because they were either too difficult or too easy (shape, 2 items; pattern, 4 items; relative size and quantity, 1 item; addition, 1 item; and, subtraction, 1 item). Six new items with higher difficulty levels were trialled and added (relative size and quantity, 1 item; sets, 2 items; numbers, 2 items; counting, 1 item) giving a total of 56 items (7 in each of the 8 remaining skill / concept areas). The easiest item in each skill / concept area was passed by approximately 90% of the children whilst the hardest item within each skill / concept area was passed by approximately 30% of the children.

The mean time taken to complete the items in the mathematics module was 19.56 minutes (s.d. 3.40) and so an adaptive algorithm, using the same formula as the literacy module, was used. Again, the validity of the adaptive algorithm was investigated by simulating the adaptive form of the CoPS Baseline mathematics test with data that had been collected from the children in this study using the full form of the test. The two measures (original raw scores and adaptive form score) were correlated using the Pearson Product Moment Correlation, with a positive correlation coefficient of 0.80 being obtained ($n = 140$, $p<0.001$, 1-tailed).

The equivalency of the computerised mathematics items were checked by retesting pupils, from two of the primary schools, using a computerised full form of the test. A significant Pearson Product Moment Correlation of 0.93 was obtained ($n = 18$, $p<0.001$). Item order (based on pass rates) remained the same as the pen and paper version.

6.3 Literacy follow-up, part 1 - Summer 1998

6.3.1 Reading ability

The mean BAS-II word reading ability score obtained in the follow-up was 55.35 (s.d. 31.35, n = 92). Girls scored slightly higher than boys but not significantly so (boys: mean 53.86, s.d. 31.57, n = 44; girls: mean 56.71, s.d. 31.43, n = 48; t = 0.43, not significant).

The BAS-II word reading ability score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 12 months earlier. The result indicated a significant Pearson correlation coefficient of 0.79 ($p<0.001$), which became 0.73 ($p<0.001$) when age was statistically controlled for. The highest correlations with word reading 12 months later were produced by the baseline skill / concept areas assessing simple spelling (r = 0.76), letter recognition (r = 0.66), simple reading (r = 0.66), knowledge about print (r = 0.62) and alliteration (r = 0.61). The lowest correlations with word reading were produced by components assessing aural comprehension (r = 0.18) and verbal concepts (r = 0.21). These results are shown in Table 11.

Table 11: Correlation between BAS-II reading score and the skill / concept areas of the CoPS Baseline literacy module (n = 92)

A	B	C	D	E	F	G	H
0.21*	0.18	0.36***	0.61***	0.62***	0.66***	0.66***	0.76***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

A Stepwise Multiple Regression shows that simple spelling is the best predictor of BAS-II reading score (accounting for 57% of the variance), with letter recognition being the next best predictor (accounting for another 6% of the variance) and simple reading accounting for a further 2% of the variance (total $R^2 = 0.65$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in BAS-II reading scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in reading 12 months later. For the purposes of this analysis, 'poor reading' was defined as having a reading score more than 1 standard deviation below the mean. The results are shown in Table 12. Despite a fairly good prediction of the non-target group (i.e. 'not poor'), there was a relatively poor prediction of the target group of poor readers, with 8% false positives and 7% false negatives.

Table 12: Discriminant function analyses of the prediction of 'poor' readers from baseline assessment

Baseline Predictions	Tested reading ability (BAS-II)		
	Poor	Not poor	TOTAL
Poor	12	7	19
Not poor	6	67	73
TOTAL	18	74	92

Prediction of a target group of 'good readers' (i.e. those having a reading score more than 1 standard deviation above the mean) was no better, with 10% false positives and 10% false negatives. These results are shown in Table 13.

Table 13: Discriminant function analyses of the prediction of 'good' readers from baseline assessment

Baseline Predictions	Tested reading ability (BAS-II)		
	Good	Not good	TOTAL
Good	11	9	20
Not good	9	63	72
TOTAL	20	72	92

A one-way ANOVA was carried out on the data to see if there was a difference, on the BAS-II reading scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 14. A highly significant difference was found between the three groups on the BAS-II Reading test (F = 37.04, df = 2, p<0.001).

Table 14: BAS Reading mean scores for children classified by baseline testing as 'poor', 'average' or 'good' readers

	N	Mean	Standard deviation
Poor	19	20.42	15.05
Average	53	56.75	26.33
Good	20	84.80	21.39
Total	92	55.35	31.35

6.3.2 Spelling ability

The mean BAS-II spelling ability score obtained in the follow-up was 42.52 (s.d. 24.08, n = 92). There was very little difference between the mean scores for girls and boys (boys: mean 42.41, s.d. 24.59, n = 44; girls: mean 42.63, s.d. 23.87, n = 48; t = 0.04, not significant).

The BAS-II spelling ability score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 12 months earlier. The result indicated a significant Pearson correlation coefficient of 0.75 (p<0.001), which became 0.69 (p<0.001) when age was statistically controlled for. The highest correlations with spelling 12 months later were produced by the baseline skill / concept areas assessing simple spelling (r = 0.73), knowledge about print (r = 0.64), letter recognition (r = 0.62), simple reading (r = 0.61) and alliteration (r = 0.56). The lowest correlations with spelling were produced by components assessing aural comprehension (r = 0.18) and verbal concepts (r = 0.29). These results are shown in Table 15.

Table 15: Correlation between BAS-II spelling score and the skill / concept areas of the CoPS Baseline literacy module (n = 92)

A	B	C	D	E	F	G	H
0.29**	0.18	0.30**	0.56***	0.64***	0.62***	0.61***	0.73***

* = p<0.05; ** = p<0.01; *** = p<0.001

A Stepwise Multiple Regression shows that simple spelling is the best predictor of BAS-II spelling score (accounting for 54% of the variance), with knowledge about print being the next best predictor (accounting for another 6% of the variance) and letter recognition accounting for a further 2% of the variance (total $R^2 = 0.62$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in BAS-II spelling scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in spelling 12 months later. The results are shown in Table 16. Again, the prediction of the non-target group (i.e. 'not poor') was fairly good, but prediction of the target group of poor spellers showed 3% false positives and 11% false negatives.

Table 16: Discriminant function analyses of the prediction of 'poor' spellers from baseline assessment

Baseline Predictions	Tested spelling ability (BAS-II)		
	Poor	Not poor	TOTAL
Poor	16	3	19
Not poor	10	63	73
TOTAL	26	66	92

Prediction of a target group of 'good spellers' was poorer, with 11% false positives and 11% false negatives. These results are shown in Table 17.

Table 17: Discriminant function analyses of the prediction of 'good' spellers from baseline assessment

Baseline Predictions	Tested spelling ability (BAS-II)		
	Good	Not good	TOTAL
Good	10	10	20
Not good	10	62	72
TOTAL	20	72	92

A one-way ANOVA was carried out on the data to see if there was a difference, on the BAS-II spelling scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 18. A highly significant difference was found between the three groups on the BAS-II Spelling test ($F = 31.66$, $df = 2$, $p<0.001$).

Table 18: BAS-II Spelling mean scores for children classified by baseline testing as 'poor', 'average' or 'good' spellers

	N	Mean	Standard deviation
Poor	19	15.58	12.97
Average	53	44.68	21.99
Good	20	62.40	11.83
Total	92	42.52	24.08

6.4 Mathematics follow-up, part 1 - Spring 1999

The WOND Numerical Operations standard score obtained in the follow-up was 105.71 (s.d. 15.24, n = 83). There was very little difference between the mean scores for girls and boys (boys: mean 105.18, s.d. 15.27, n = 40; girls: mean 106.21, s.d. 15.38, n = 43; t = 0.31, not significant).

The WOND mathematics score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the mathematics module administered in the baseline assessment 20 months earlier. The result indicated a significant Pearson correlation coefficient of 0.68 ($p<0.001$), which became 0.66

($p<0.001$) when age was statistically controlled for. When vocabulary ability was controlled for (measured by the BPVS), the correlation between baseline mathematics and WOND score was 0.53. When both age and vocabulary were controlled for, the correlation became 0.49. The highest correlations with mathematics 20 months later were produced by the baseline skill / concept areas assessing addition ($r = 0.57$), subtraction ($r = 0.56$), numbers ($r = 0.54$) and counting ($r = 0.52$). The lowest correlation with mathematics 20 months later was produced by the component assessing shape ($r = 0.29$). These results are shown in Table 19.

Table 19: Correlation between WOND mathematics score and the skill / concept areas of the CoPS Baseline mathematics module (n = 83)

A	B	C	D1	D2	E	F	G	H
0.29**	0.44***	0.48***	0.39***	0.38***	0.54***	0.52***	0.57***	0.56***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

A Stepwise Multiple Regression shows that addition is the best predictor of WOND mathematics score (accounting for 32% of the variance), with subtraction being the next best predictor (accounting for another 10% of the variance) and numbers accounting for a further 6% of the variance (total $R^2 = 0.48$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in WOND mathematics scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in mathematics 20 months later. The results are shown in Table 20. There was a good prediction of the non-target group (i.e. 'not poor'), but a very poor prediction of the target group of children who are poor at mathematics, with 11% false positives and 10% false negatives.

Table 20: Discriminant function analyses of the prediction of children 'poor' at mathematics from baseline assessment

Baseline Predictions	Tested mathematics ability (WOND)		
	Poor	Not poor	TOTAL
Poor	3	9	12
Not poor	8	63	71
TOTAL	11	72	83

Prediction of a target group of children who are 'good' at mathematics was also poor, with 19% false positives and 6% false negatives. These results are shown in Table 21.

Table 21: Discriminant function analyses of the prediction of children who are 'good' at mathematics from baseline assessment

Baseline Predictions	Tested mathematics ability (WOND)		
	Good	Not good	TOTAL
Good	5	16	21
Not good	5	57	62
TOTAL	10	73	83

A one-way ANOVA was carried out on the data to see if there was a difference, on the WOND mathematics scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at mathematics. Means and standard deviations for the three groups are given in Table 22. A highly significant difference was found between the three groups on the WOND Numerical Operations test ($F = 20.81$, $df = 2$, $p<0.001$).

Table 22: WOND mean scores for children classified by baseline testing as 'poor', 'average' or 'good' at mathematics

	N	Mean	Standard deviation
Poor	12	6.75	2.83
Average	50	11.04	3.14
Good	21	13.91	3.03
Total	83	11.14	3.74

6.5 Vocabulary follow-up - Spring 1999

The BPVS standard score obtained in the follow-up was 105.56 (s.d. 9.95, n = 82). There was a significant difference between the mean scores for girls and boys (boys: mean 108.75, s.d. 9.72, n = 40; girls: mean 102.52, s.d. 9.30, n = 42; t = -2.97, p<0.01).

The vocabulary score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 20 months earlier. The result indicated a significant Pearson correlation coefficient of 0.67 (p<0.001), which became 0.59 (p<0.001) when age was statistically controlled for. The highest correlations with vocabulary 20 months later were produced by the baseline skill / concept areas assessing knowledge about print (r = 0.60), simple spelling (r = 0.58) and simple reading (r = 0.51). The lowest correlation with vocabulary 20 months later were produced by the components assessing aural comprehension (r = 0.21) and verbal concepts (r = 0.25). These results are shown in Table 23.

Table 23: Correlations between BPVS vocabulary score and the skill / concept areas of the CoPS Baseline literacy module (n = 82)

A	B	C	D	E	F	G	H
0.25*	0.21	0.46***	0.45***	0.60***	0.45***	0.51***	0.58***

* = p<0.05; ** = p<0.01; *** = p<0.001

The vocabulary score obtained on follow-up was also correlated with the overall score obtained by the same children on the full version of the mathematics module administered in the baseline assessment 20 months earlier. The result indicated a significant Pearson correlation coefficient of 0.65 (p<0.001), which became 0.56 (p<0.001) when age was statistically controlled for. The highest correlations with vocabulary 20 months later were produced by the baseline skill / concept areas assessing relative size and quantity (r = 0.57), numbers (r = 0.57), pattern (r = 0.55) and addition (r = 0.51). The lowest correlation with vocabulary 20 months later was produced by the component assessing shape (r = 0.14). These results are shown in Table 24.

Table 24: Correlations between BPVS vocabulary score and the skill / concept areas of the CoPS Baseline mathematics module (n = 82)

A	B	C	D1	D2	E	F	G	H
0.14	0.55***	0.57***	0.37**	0.47***	0.57***	0.42***	0.51***	0.36**

* = p<0.05; ** = p<0.01; *** = p<0.001

A Stepwise Multiple Regression shows that knowledge about print is the best predictor of BPVS Vocabulary score (accounting for 36% of the variance), with pattern being the next best predictor (accounting for another 10% of the variance) and rhymes accounting for a further 6% of the variance (total $R^2 = 0.51$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in BPVS Vocabulary scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline literacy module could predict children who would have a 'poor' vocabulary 20 months later. The results are shown in Table 25. There was a fairly good prediction of the non-target group (i.e. 'not poor'), but a very poor prediction of the target group of children who have a poor vocabulary, with 10% false positives and 14% false negatives.

Table 25: Discriminant function analyses of the prediction of children with 'poor' vocabulary from baseline literacy assessment

Baseline Predictions	Tested vocabulary ability (BPVS)		
	Poor	Not poor	TOTAL
Poor	3	7	10
Not poor	10	50	60
TOTAL	13	57	70

Prediction of a target group of children with 'good' vocabulary (i.e. those having a vocabulary score more than 1 standard deviation above the mean) was slightly better, with 9% false positives and 6% false negatives. These results are shown in Table 26.

Table 26: Discriminant function analyses of the prediction of children with 'good' vocabulary from baseline literacy assessment

Baseline	Tested vocabulary ability (BPVS)		
Predictions	Good	Not good	TOTAL
Good	8	6	14
Not good	4	52	56
TOTAL	12	58	70

A one-way ANOVA was carried out on the data to see if there was a difference, on the BPVS scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 27. A significant difference was found between the three groups on the BPVS test ($F = 5.00$, $df = 2$, $p<0.01$).

Table 27: BPVS mean scores for children classified by baseline testing as 'poor', 'average' or 'good' at literacy

	N	Mean	Standard deviation
Poor	10	100.50	8.14
Average	46	104.80	9.68
Good	14	112.07	8.97
Total	70	105.64	9.88

6.6 Literacy follow-up, part 2 - Spring 2001

6.6.1 Reading ability

The mean SPAR reading quotient obtained in the follow-up was 103.09 (s.d. 13.40, n = 70). Boys scored slightly higher than girls but not significantly so (boys: mean 105.19, s.d. 14.23, n = 31; girls: mean 101.41, s.d. 12.64, n = 39; t = 1.18, not significant).

The SPAR reading score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 32 months earlier. The result indicated a significant Pearson correlation coefficient of 0.66 ($p<0.001$), which became 0.65 ($p<0.001$) when age was statistically controlled for. The highest correlations with reading 32 months later were produced by the baseline skill / concept areas assessing simple spelling ($r = 0.60$), knowledge about print ($r = 0.54$), alliteration ($r = 0.52$), letter recognition ($r = 0.51$), simple reading ($r = 0.45$) and aural comprehension ($r = 0.42$). The lowest correlations with reading were produced by components assessing verbal concepts ($r = 0.29$) and rhymes ($r = 0.39$). These results are shown in Table 28.

Table 28: Correlations between SPAR reading score and the skill / concept areas of the CoPS Baseline literacy module (n = 70)

A	B	C	D	E	F	G	H
0.29*	0.42***	0.39**	0.52***	0.54***	0.51***	0.45***	0.60***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

A Stepwise Multiple Regression shows that simple spelling is the best predictor of SPAR reading score (accounting for 36% of the variance), with aural comprehension being the next best predictor (accounting for another 7% of the variance) and rhymes accounting for a further 4% of the variance (total $R^2 = 0.48$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in SPAR reading scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in reading 32 months later. The results are shown in Table 29. Despite a fairly good prediction of the non-target group (i.e. 'not poor'), there was a poor prediction of the target group of poor readers, with 10% false positives and 13% false negatives.

Table 29: Discriminant function analyses of the prediction of 'poor' readers from baseline assessment

Baseline Predictions	Tested reading ability (SPAR)		
	Poor	Not poor	TOTAL
Poor	8	7	15
Not poor	9	46	55
TOTAL	17	53	70

Prediction of a target group of 'good readers' was similar, with 11% false positives and 10% false negatives. These results are shown in Table 30.

Table 30: Discriminant function analyses of the prediction of 'good' readers from baseline assessment

Baseline Predictions	Tested reading ability (SPAR)		
	Good	Not good	TOTAL
Good	6	8	14
Not good	7	49	56
TOTAL	13	57	70

A one-way ANOVA was carried out on the data to see if there was a difference, on the SPAR reading scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 31. A highly significant difference was found between the three groups on the SPAR Reading test ($F = 11.00$, $df = 2$, $p<0.001$).

Table 31: SPAR reading mean quotients for children classified by baseline testing as 'poor', 'average' or 'good' readers

	N	Mean	Standard deviation
Poor	15	91.87	12.56
Average	41	104.12	12.16
Good	14	112.07	9.61
Total	70	103.09	13.40

6.6.2 Spelling ability

The mean SPAR spelling quotient obtained in the follow-up was 104.16 (s.d. 13.39, n = 70). There was very little difference between the mean scores for girls and boys (boys:

mean 103.77, s.d. 14.50, n = 31; girls: mean 104.46, s.d. 12.63, n = 39; t = 0.21, not significant).

The SPAR spelling score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 32 months earlier. The result indicated a significant Pearson correlation coefficient of 0.59 ($p<0.001$), which remained 0.59 ($p<0.001$) when age was statistically controlled for. The highest correlations with spelling 32 months later were produced by the baseline skill / concept areas assessing knowledge about print (r = 0.50), alliteration (r = 0.50), letter recognition (r = 0.46), simple spelling (r = 0.45) and simple reading (r = 0.43). The lowest correlations with spelling were produced by components assessing verbal concepts (r = 0.27), rhymes (r = 0.35) and aural comprehension (r = 0.37). These results are shown in Table 32.

Table 32: Correlations between SPAR spelling score and the skill / concept areas of the CoPS Baseline literacy module (n = 70)

A	B	C	D	E	F	G	H
0.27*	0.37**	0.35**	0.50***	0.50***	0.46***	0.43***	0.45***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

A Stepwise Multiple Regression shows that knowledge about print is the best predictor of SPAR spelling score (accounting for 25% of the variance), with alliteration being the next best predictor (accounting for another 6% of the variance) and aural comprehension accounting for a further 5% of the variance (total $R^2 = 0.36$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in SPAR spelling scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in spelling 32 months later. The results are shown in Table 33. There was a fairly good prediction of the non-target group (i.e. 'not poor'), but a poorer prediction of the target group of poor spellers, with 13% false positives and 14% false negatives.

Table 33: Discriminant function analyses of the prediction of 'poor' spellers from baseline assessment

Baseline Predictions	Tested spelling ability (SPAR)		
	Poor	Not poor	TOTAL
Poor	6	9	15
Not poor	10	45	55
TOTAL	16	54	70

Prediction of a target group of 'good spellers' was much poorer, with 13% false positives and 16% false negatives. These results are shown in Table 34.

Table 34: Discriminant function analyses of the prediction of 'good' spellers from baseline assessment

Baseline Predictions	Tested spelling ability (SPAR)		
	Good	Not good	TOTAL
Good	5	9	14
Not good	11	45	56
TOTAL	16	54	70

A one-way ANOVA was carried out on the data to see if there was a difference, on the SPAR spelling scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 35. A significant difference was found between the three groups on the SPAR Spelling test ($F = 6.96$, $df = 2$, $p<0.01$).

Table 35: SPAR spelling mean quotients for children classified by baseline testing as 'poor', 'average' or 'good' spellers

	N	Mean	Standard deviation
Poor	15	93.87	11.28
Average	41	106.15	12.93
Good	14	109.36	11.71
Total	70	104.16	13.39

6.6.3 *Writing ability*

The mean WOLD Written Expression standard score obtained in the follow-up was 95.84 (s.d. 13.22, n = 63). There was very little difference between the mean scores for girls and boys (boys: mean 94.18, s.d. 12.86, n = 28; girls: mean 97.17, s.d. 13.54, n = 35; t = 0.89, not significant).

The writing ability score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the literacy module administered in the baseline assessment 32 months earlier. The result indicated a significant Pearson correlation coefficient of 0.48 (p<0.001), which became 0.56 (p<0.001) when age was statistically controlled for. The highest correlations with writing 32 months later were produced by the baseline skill / concept areas assessing letter recognition (r = 0.41), rhymes (r = 0.40), knowledge about print (r = 0.40), simple spelling (r = 0.35) and alliteration (r = 0.34). The lowest correlations with spelling were produced by components assessing verbal concepts (r = 0.13) and aural comprehension (r = 0.18). These results are shown in Table 36.

Table 36: Correlations between WOLD writing score and the skill / concept areas of the CoPS Baseline literacy module (n = 63)

A	B	C	D	E	F	G	H
0.13	0.18	0.40**	0.34**	0.40**	0.41**	0.30*	0.35**

* = p<0.05; ** = p<0.01; *** = p<0.001

A Stepwise Multiple Regression shows that letter recognition is the best predictor of WOLD writing score (accounting for 16% of the variance), with rhymes accounting for a further 10% of the variance (total $R^2 = 0.26$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in WOLD writing scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in writing 32 months later. The results are shown in Table 37. There was a good prediction of the non-target group (i.e. 'not poor'), but a poor prediction of the target group of poor writers, with 10% false positives and 10% false negatives.

Table 37: Discriminant function analyses of the prediction of 'poor' writers from baseline assessment

Baseline Predictions	Tested writing ability (WOLD)		
	Poor	Not poor	TOTAL
Poor	6	6	12
Not poor	6	45	51
TOTAL	12	51	63

Prediction of a target group of 'good writers' was no better, with 10% false positives and 11% false negatives. These results are shown in Table 38.

Table 38: Discriminant function analyses of the prediction of 'good' writers from baseline assessment

Baseline Predictions	Tested spelling ability (WOLD)		
	Good	Not good	TOTAL
Good	6	6	12
Not good	7	44	51
TOTAL	13	50	63

A one-way ANOVA was carried out on the data to see if there was a difference, on the WOLD writing scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at literacy. Means and standard deviations for the three groups are given in Table 39. A significant difference was found between the three groups on the WOLD Written Expression test ($F = 7.03$, $df = 2$, $p<0.01$).

Table 39: WOLD writing mean scores for children classified by baseline testing as 'poor', 'average' or 'good' at literacy

	N	Mean	Standard deviation
Poor	12	86.25	11.01
Average	39	96.05	12.03
Good	12	104.75	13.29
Total	63	95.84	13.22

6.7 Mathematics follow-up part 2 - Spring 2001

The mathematics standard score obtained in the follow-up was 104.98 (s.d. 12.29, n = 84). Boys scored slightly higher than girls, but not significantly so (boys: mean 106.05, s.d. 14.52, n = 40; girls: mean 104.00, s.d. 9.91, n = 44; t = -0.76, not significant).

The mathematics score obtained on follow-up was correlated with the overall score obtained by the same children on the full version of the mathematics module administered in the baseline assessment 32 months earlier. The result indicated a significant Pearson correlation coefficient of 0.60 ($p<0.001$) when age was statistically controlled for. The highest correlations with mathematics 32 months later were produced by the baseline skill / concept areas assessing subtraction (r = 0.47), numbers (r = 0.43), addition (r = 0.40) and relative size and quantity (r = 0.39). The lowest correlations with mathematics 32 months later were produced by the components assessing shape (r = 0.08), seriation (r = 0.15) and sets (r = 0.18). These results are shown in Table 40.

Table 40: Correlation between Mathematics 8 score and the skill / concept areas of the CoPS Baseline mathematics module (n = 84)

A	B	C	D1	D2	E	F	G	H
0.08	0.31**	0.39***	0.18	0.15	0.43***	0.35**	0.40***	0.47***

* = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$

A Stepwise Multiple Regression shows that subtraction is the best predictor of Mathematics 8 score (accounting for 22% of the variance), with numbers accounting for a further 7% of the variance (total $R^2 = 0.29$). All other baseline skill / concept areas were removed from the analysis as they did not significantly add to the variance in Mathematics 8 scores, due to covariance with the other variables in the equation. See Table 45 for a summary of the regression analyses.

Discriminant function analysis was carried out to see if the baseline assessment measures could predict children who would be performing 'poorly' in mathematics 32 months later. The results are shown in Table 41. There was a good prediction of the non-target group (i.e. 'not poor'), but a poor prediction of the target group of children who are poor at mathematics, with 11% false positives and 8% false negatives.

Table 41: Discriminant function analyses of the prediction of children 'poor' at mathematics from baseline assessment

Baseline Predictions	Tested mathematics ability (Maths 8)		
	Poor	Not poor	TOTAL
Poor	6	9	15
Not poor	7	62	69
TOTAL	13	71	84

Prediction of a target group of children who are 'good' at mathematics was poorer, with 14% false positives and 8% false negatives. These results are shown in Table 42.

Table 42: Discriminant function analyses of the prediction of children who are 'good' at mathematics from baseline assessment

Baseline Predictions	Tested mathematics ability (Maths 8)		
	Good	Not good	TOTAL
Good	6	12	18
Not good	7	59	66
TOTAL	13	71	84

A one-way ANOVA was carried out on the data to see if there was a difference, on the Maths 8 scores, between children classified on the baseline assessment as 'poor', 'average' or 'good' at mathematics. Means and standard deviations for the three groups are given in Table 43. A highly significant difference was found between the three groups on the Maths 8 test ($F = 11.04$, $df = 2$, $p<0.001$).

Table 43: Maths 8 mean standard scores for children classified by baseline testing as 'poor', 'average' or 'good' at mathematics

	N	Mean	Standard deviation
Poor	15	95.33	10.77
Average	51	104.82	11.14
Good	18	113.44	10.90
Total	84	104.98	12.29

6.8 Follow-up studies - summary statistics

6.8.1 Intercorrelations between all tests

Intercorrelations between the CoPS Baseline literacy and mathematics modules and all of the tests administered in the follow-up studies are given in Table 44.

Table 44: Intercorrelations between CoPS Baseline and follow-up tests

	1	2	3	4	5	6	7	8	9	10
1) CB Literacy (4:9)		0.76 ***	0.79 ***	0.75 ***	0.61 ***	0.67 ***	0.66 ***	0.59 ***	0.48 ***	0.52 ***
2) CB Maths (4:9)	0.76 ***		0.58 ***	0.43 ***	0.68 ***	0.65 ***	0.61 ***	0.52 ***	0.46 ***	0.48 ***
3) BAS Reading (5:9)	0.79 ***	0.58 ***		0.83 ***	0.57 ***	0.38 **	0.78 ***	0.74 ***	0.53 ***	0.55 ***
4) BAS Spelling (5:9)	0.75 ***	0.43 ***	0.83 ***		0.53 ***	0.34 **	0.78 ***	0.70 ***	0.53 ***	0.56 ***
5) WOND Maths (6:8)	0.61 ***	0.68 ***	0.57 ***	0.53 ***		0.40 ***	0.72 ***	0.65 ***	0.63 ***	0.65 ***
6) BPVS (6:8)	0.67 ***	0.65 ***	0.38 **	0.34 **	0.40 ***		0.49 ***	0.35 *	0.35 *	0.41 **
7) SPAR Reading (8:7)	0.66 ***	0.61 ***	0.78 ***	0.78 ***	0.72 ***	0.49 ***		0.86 ***	0.65 ***	0.61 ***
8) SPAR Spelling (8:7)	0.59 ***	0.52 ***	0.74 ***	0.70 ***	0.65 ***	0.35 *	0.86 ***		0.66 ***	0.54 ***
9) WOLD Writing (8:7)	0.48 ***	0.46 ***	0.53 ***	0.53 ***	0.63 ***	0.35 *	0.65 ***	0.66 ***		0.44 **
10) Maths 8 (8:7)	0.52 ***	0.48 ***	0.55 ***	0.56 ***	0.65 ***	0.41 **	0.61 ***	0.54 ***	0.44 **	

6.8.2 Regression analyses for follow-up studies

A summary of the regression analyses for each of the follow-up tests is shown in Table 45. For further explanation of these results see sections 6.3, 6.4, 6.5, 6.6 and 6.7.

Table 45: Summary table of regression analyses

Age	5:9	6:8	8:7
Reading	$R^2 = 0.65$ [SS = 0.57] [LR = 0.06] [SR = 0.02]	N/A	$R^2 = 0.47$ [SS = 0.36] [AC = 0.07] [R = 0.04]
Spelling	$R^2 = 0.62$ [SS = 0.54] [KAP = 0.06] [LR = 0.02]	N/A	$R^2 = 0.36$ [KAP = 0.25] [Al = 0.06] [AC = 0.05]
Mathematics	N/A	$R^2 = 0.48$ [Ad = 0.32] [S = 0.10] [N = 0.06]	$R^2 = 0.29$ [S = 0.22] [N = 0.07]
Vocabulary	N/A	$R^2 = 0.51$ [KAP = 0.36] [P = 0.10] [R = 0.06]	N/A
Writing	N/A	N/A	$R^2 = 0.26$ [LR = 0.16] [R = 0.10]

AC = Aural Comprehension
Ad = Addition
Al = Alliteration
KAP = Knowledge About Print
LR = Letter Recognition
N = Numbers
P = Pattern
R = Rhymes
S = Subtraction
SR = Simple Reading
SS = Simple Spelling

6.9 Reliability and validity - Spring 2001

6.9.1 CoPS Baseline Scores

The median centile scores and ranges for each of the four CoPS Baseline modules and the overall CoPS Baseline centile score, for both testing sessions, are given in Table 46.

Table 46: Descriptive statistics (centile scores) for the CoPS Baseline modules

	N	Minimum	Maximum	Median
Literacy 1	60	2	99	66
Mathematics 1	60	2	97	54
Communication 1	60	2	98	45
Personal and Social Development 1	60	9	93	57
Overall (mean) Score 1	60	6	92	52
Literacy 2	45	10	99	66
Mathematics 2	45	1	97	65
Communication 2	45	2	98	55
Personal and Social Development 2	45	13	93	57
Overall (mean) Score 2	45	22	95	58

Analysis by gender revealed a significant difference on the first session overall z-score with girls outperforming boys (boys: mean -0.17, s.d. 0.60, n= 26; girls: mean 0.24, s.d. 0.57, n=34; t = 2.70, p<0.01). However, only two of the four modules (communication and personal and social development) showed significant gender differences (favouring girls) on the first testing session (see Table 47).

Table 47: Gender differences on the four CoPS Baseline modules (first testing session z-scores)

	Gender	N	Mean	SD	t	df	Sig.
Literacy	Female	34	0.59	0.95	1.50	58	NS
	Male	26	0.20	1.03			
Mathematics	Female	34	0.09	0.84	0.85	58	NS
	Male	26	-0.09	0.92			
Communication	Female	34	0.02	1.06	2.32	58	p<0.05
	Male	26	-0.56	0.80			
Personal and Social Development	Female	34	0.28	0.61	3.38	58	p<0.01
	Male	26	-0.28	0.65			
Overall (mean) Score	Female	34	0.24	0.57	2.70	58	p<0.01
	Male	26	-0.17	0.60			

Analysis by gender also revealed a significant difference on the second session overall z-score (boys: mean 0.05, s.d. 0.50, n= 17; girls: mean 0.38, s.d. 0.56, n=28; t = 2.62, p<0.05). Two of the four modules (communication and personal and social development) showed a significant gender difference (favouring girls) on the second testing session (see Table 48).

Table 48: Gender differences on the four CoPS Baseline modules (second testing session z-scores)

	Gender	N	Mean	SD	t	df	Sig.
Literacy	Female	28	0.78	1.02	1.82	43	NS
	Male	17	0.26	0.78			
Mathematics	Female	28	0.38	0.83	0.81	43	NS
	Male	17	0.16	0.92			
Communication	Female	28	0.28	1.15	2.21	43	p<0.05
	Male	17	-0.43	0.86			
Personal and Social Development	Female	28	0.31	0.62	2.87	43	p<0.01
	Male	17	-0.25	0.68			
Overall (mean) Score	Female	28	0.38	0.56	2.62	43	p<0.05
	Male	17	0.05	0.50			

The 60 children in the sample came from six ethnic backgrounds (see Table 49). However, as the majority of children were categorised as 'white', analysis by ethnic background was made difficult.

Table 49: Ethnic background of children

Ethnic background	N	Percent
White	55	91.7
Black African	1	1.7
Black other	1	1.7
Indian	1	1.7
Chinese	1	1.7
Other	1	1.7

Analysis comparing 'white' children to all other children revealed no significant differences on the first session overall z-score (white: mean 0.08, s.d. 0.60, n= 55; others: mean -0.17, s.d. 0.71, n=5; t = 0.89, not significant) or the second session overall z-score (white: mean 0.23, s.d. 0.57, n= 41; others: mean 0.15, s.d. 0.68, n=4; t = 0.24, not significant). None of the four modules showed significant ethnic differences (see Table 50) on either of the testing sessions.

Table 50: Ethnic differences on the four CoPS Baseline modules (z-scores)

	Ethnicity	N	Mean	SD	t	df	Sig.
Literacy 1	White	55	0.44	1.01	0.65	58	NS
	Other	5	0.14	0.81			
Mathematics 1	White	55	0.03	0.85	0.80	58	NS
	Other	5	-0.29	1.19			
Communication 1	White	55	-0.20	1.01	0.77	58	NS
	Other	5	-0.56	0.87			
Personal and Social Development 1	White	55	0.05	0.67	0.81	58	NS
	Other	5	-0.20	0.88			
Overall (mean) Score 1	White	55	0.08	0.60	0.89	58	NS
	Other	5	-0.17	0.71			
Literacy 2	White	41	0.57	0.96	0.21	43	NS
	Other	4	0.68	1.11			
Mathematics 2	White	41	0.30	0.88	0.00	43	NS
	Other	4	0.30	0.71			
Communication 2	White	41	0.03	1.07	0.32	43	NS
	Other	4	-0.16	1.52			
Personal and Social Development 2	White	41	0.10	0.70	0.05	43	NS
	Other	4	0.08	0.70			
Overall (mean) Score 2	White	41	0.23	0.57	0.24	43	NS
	Other	4	0.15	0.68			

6.9.2 QCA Scales Scores

The means, standard deviations and ranges for each of the four QCA Scales are given in Table 51.

Table 51: Descriptive statistics for the QCA Scales (n = 60)

	Minimum	Maximum	Mean	SD
Reading	1	12	6.37	2.89
Writing	0	4	2.77	1.23
Mathematics	0	8	4.32	2.27
Speaking and Listening	0	4	2.00	1.06
Personal and Social Development	0	4	2.43	1.21
Total Score	3	32	17.82	7.21

Analysis by gender revealed no significant differences on the total scores. However, two of the five QCA scales (reading and writing) showed significant gender differences, favouring girls (see Table 52).

Table 52: Gender differences on the five QCA Scales

	Gender	N	Mean	SD	t	df	Sig.
Reading	Female	34	7.03	2.69	2.09	58	p<0.05
	Male	26	5.50	2.97			
Writing	Female	34	3.06	0.98	2.18	58	p<0.05
	Male	26	2.38	1.42			
Mathematics	Female	34	4.65	2.16	1.30	58	NS
	Male	26	3.88	2.39			
Speaking and Listening	Female	34	1.97	0.94	0.24	58	NS
	Male	26	2.04	1.22			
Personal and Social Development	Female	34	2.65	1.15	1.58	58	NS
	Male	26	2.15	1.26			
Total Score	Female	34	19.35	6.36	1.93	58	NS
	Male	26	15.81	7.87			

Analysis comparing 'white' children to all other children revealed a significant difference on the QCA total score, with white children scoring higher than other children. One of the five QCA scales (writing) showed significant ethnic differences (see Table 53) in favour of white pupils.

Table 53: Ethnic differences on the five QCA Scales

	Ethnicity	N	Mean	SD	t	df	Sig.
Reading	White	55	6.58	2.84	1.96	58	NS
	Other	5	4.00	2.65			
Writing	White	55	2.89	1.17	2.74	58	p<0.01
	Other	5	1.40	1.14			
Mathematics	White	55	4.47	2.26	1.80	58	NS
	Other	5	2.60	1.82			
Speaking and Listening	White	55	2.02	1.08	0.44	58	NS
	Other	5	1.80	0.84			
Personal and Social Development	White	55	2.51	1.20	1.63	58	NS
	Other	5	1.60	1.14			
Total Score	White	55	18.40	7.05	2.14	58	p<0.05
	Other	5	11.40	6.27			

6.9.3 Reliability

Correlations between the first test scores and retest scores were calculated, for all four modules and the overall scores, using Pearson Product Moment Correlation Coefficient.

The results (see Table 54) indicate that all four modules and the overall scores have highly significant test-retest correlations.

Table 54: CoPS Baseline test-retest reliability coefficients

	N	r	Significance
Literacy	45	0.70	p<0.001
Mathematics	45	0.58	p<0.001
Communication	45	0.83	p<0.001
Personal and Social Development	45	0.99	p<0.001
Overall (mean) Score	45	0.91	p<0.001

6.9.4 *Validity*

Correlations between the CoPS Baseline scores and QCA Scale scores were calculated, for all four modules and the overall scores, using Pearson Product Moment Correlation Coefficient. The results (see Table 55) indicate that the CoPS Baseline modules and the overall scores all show significant correlations with the QCA Scales.

Table 55: CoPS Baseline and QCA Scale validity coefficients

CoPS Baseline	QCA	N	r	Sig
Literacy	Reading and Writing	60	0.56	p<0.001
Mathematics	Mathematics	60	0.26	p<0.05
Communication	Speaking and Listening	60	0.40	p<0.01
Personal and Social Development	Personal and Social Development	60	0.54	p<0.001
Overall (mean) Score	Total Score	60	0.64	p<0.001

6.10 Questionnaire study - Autumn 1997 - Spring 2001

Frequency distributions are shown for each question for the entire sample, with a breakdown by year of response, in Tables 56 to 97. 10 teachers returned questionnaires in October 1997, 13 in December 1998, 2 in October 1999 and 4 in March 2001.

Table 56: How did you find the installation of CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	5	3	2	0	0
A little difficult	6	3	2	1	0
Satisfactory	10	2	6	1	1
Easy	4	2	1	0	1
Very easy	4	0	2	0	2

Table 57: How did you find registering children on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	1	1	0	0	0
A little difficult	2	0	1	0	1
Satisfactory	13	4	6	2	1
Easy	9	4	4	0	1
Very easy	3	1	1	0	1

Table 58: Rate the CoPS Baseline training video for content

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	0	0	0	0	0
Satisfactory	5	0	5	0	0
Good	7	0	5	1	1
Excellent	0	0	0	0	0

Table 59: Rate the CoPS Baseline training video for presentation

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	0	0	0	0	0
Satisfactory	4	0	4	0	0
Good	8	0	6	1	1
Excellent	0	0	0	0	0

Table 60: Rate the CoPS Baseline training video for organisation

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	0	0	0	0	0
Satisfactory	4	0	4	0	0
Good	6	0	5	0	1
Excellent	2	0	1	1	0

Table 61: Rate the CoPS Baseline manual for content

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	0	0	0	0	0
Satisfactory	3	0	1	1	1
Good	11	0	7	1	3
Excellent	4	0	4	0	0

Table 62: Rate the CoPS Baseline manual for presentation

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	0	0	0	0	0
Satisfactory	4	0	2	1	1
Good	11	0	7	1	3
Excellent	3	0	3	0	0

Table 63: Rate the CoPS Baseline manual for organisation

	Frequency	1997	1998	1999	2001
Very poor	0	0	0	0	0
Poor	1	0	0	0	1
Satisfactory	5	0	4	1	0
Good	9	0	5	1	3
Excellent	3	0	3	0	0

Table 64: How did you find running the communication module on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	0	0	0	0	0
A little difficult	2	0	0	1	1
Satisfactory	14	6	5	0	3
Easy	9	3	5	1	0
Very easy	2	0	2	0	0

Table 65: Rate the children's response to the communication module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Greatly Disliked	0	0	0	0	0
Slightly disliked	0	0	0	0	0
Indifferent	5	1	3	1	0
Slightly enjoyed	15	5	6	0	4
Greatly enjoyed	7	3	3	1	0

Table 66: Rate the difficulty of the communication module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Far too easy	0	0	0	0	0
A little easy	1	1	0	0	0
About right	8	3	4	1	0
A little difficult	14	5	6	0	3
Far too difficult	3	0	1	1	1

Table 67: How useful are the stories for assessing expressive language?

	Frequency	1997	1998	1999	2001
Totally unsuitable	0	0	0	0	0
Not very suitable	9	1	4	1	3
Adequate	6	5	1	0	0
Fairly suitable	10	3	5	1	1
Highly suitable	2	0	2	0	0

Table 68: Rate the scoring system for the communication module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very difficult	2	0	2	0	0
A little difficult	8	2	3	1	2
Satisfactory	13	7	4	1	1
Fairly Easy	2	0	1	0	1
Very easy	2	0	2	0	0

Table 69: Rate the usefulness of the communication report on CoPS Baseline

	Frequency	1997	1998	1999	2001
No use at all	0	0	0	0	0
Needs improvement	4	1	2	0	1
Satisfactory	12	4	5	1	2
Fairly useful	5	2	2	1	0
Very useful	6	2	3	0	1

Table 70: Rate the clarity of the information in the communication report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very confused	0	0	0	0	0
A little confused	3	0	2	0	1
Satisfactory	16	8	4	2	2
Clear	6	1	4	0	1
Very clear	1	0	1	0	0

Table 71: Rate the content of the information in the communication report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Completely inadequate	0	0	0	0	0
A little inadequate	3	1	1	1	0
Satisfactory	15	6	6	0	3
Fairly comprehensive	8	2	4	1	1
Very comprehensive	1	0	1	0	0

Table 72: How did you find running the literacy module on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	0	0	0	0	0
A little difficult	1	0	1	0	0
Satisfactory	7	4	2	0	1
Easy	11	2	5	1	3
Very easy	8	4	3	1	0

Table 73: Rate the children's response to the literacy module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Greatly Disliked	0	0	0	0	0
Slightly disliked	0	0	0	0	0
Indifferent	1	0	0	1	0
Slightly enjoyed	17	6	7	1	3
Greatly enjoyed	9	4	4	0	1

Table 74: Rate the difficulty of the literacy module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Far too easy	0	0	0	0	0
A little easy	2	1	1	0	0
About right	19	5	9	2	3
A little difficult	6	4	1	0	1
Far too difficult	0	0	0	0	0

Table 75: Rate the educational relevance of the content of the items in the literacy module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Not at all relevant	0	0	0	0	0
Mostly irrelevant	0	0	0	0	0
Half and half	5	2	1	0	2
Mostly relevant	14	5	5	2	2
Very relevant	8	3	5	0	0

Table 76: Rate the usefulness of the literacy report on CoPS Baseline

	Frequency	1997	1998	1999	2001
No use at all	1	0	0	1	0
Needs improvement	4	2	1	0	1
Satisfactory	6	1	3	1	1
Fairly useful	10	2	7	0	1
Very useful	6	4	1	0	1

Table 77: Rate the clarity of the information in the literacy report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very confused	0	0	0	0	0
A little confused	4	0	3	0	1
Satisfactory	10	4	4	1	1
Clear	10	4	3	1	2
Very clear	3	1	2	0	0

Table 78: Rate the content of the information in the literacy report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Completely inadequate	1	0	0	1	0
A little inadequate	3	1	1	0	1
Satisfactory	9	3	3	1	2
Fairly comprehensive	9	2	6	0	1
Very comprehensive	5	3	2	0	0

Table 79: How did you find running the mathematics module on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	0	0	0	0	0
A little difficult	0	0	0	0	0
Satisfactory	8	4	3	1	0
Easy	13	2	6	1	4
Very easy	7	4	3	0	0

Table 80: Rate the children's response to the mathematics module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Greatly Disliked	0	0	0	0	0
Slightly disliked	0	0	0	0	0
Indifferent	2	0	1	1	0
Slightly enjoyed	18	6	7	1	4
Greatly enjoyed	8	4	4	0	0

Table 81: Rate the difficulty of the mathematics module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Far too easy	0	0	0	0	0
A little easy	1	0	1	0	0
About right	17	6	6	2	3
A little difficult	9	4	4	0	1
Far too difficult	1	0	1	0	0

Table 82: Rate the educational relevance of the content of the items in the mathematics module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Not at all relevant	0	0	0	0	0
Mostly irrelevant	0	0	0	0	0
Half and half	6	3	1	0	2
Mostly relevant	13	4	5	2	2
Very relevant	8	3	5	0	0

Table 83: Rate the usefulness of the mathematics report on CoPS Baseline

	Frequency	1997	1998	1999	2001
No use at all	0	0	0	0	0
Needs improvement	2	0	0	1	1
Satisfactory	8	4	2	1	1
Fairly useful	10	1	7	0	2
Very useful	6	4	2	0	0

Table 84: Rate the clarity of the information in the mathematics report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very confused	0	0	0	0	0
A little confused	2	1	1	0	0
Satisfactory	9	3	2	2	2
Clear	9	4	3	0	2
Very clear	5	1	4	0	0

Table 85: Rate the content of the information in the mathematics report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Completely inadequate	0	0	0	0	0
A little inadequate	3	1	0	1	1
Satisfactory	8	3	3	1	1
Fairly comprehensive	9	3	4	0	2
Very comprehensive	5	2	3	0	0

Table 86: How did you find running the personal and social development module on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	2	0	2	0	0
A little difficult	4	1	3	0	0
Satisfactory	9	3	3	1	2
Easy	6	2	1	1	2
Very easy	5	3	2	0	0

Table 87: Rate the educational relevance of the content of the items in the personal and social development module on CoPS Baseline

	Frequency	1997	1998	1999	2001
Not at all relevant	0	0	0	0	0
Mostly irrelevant	1	0	1	0	0
Half and half	6	2	2	0	2
Mostly relevant	16	5	7	2	2
Very relevant	5	3	2	0	0

Table 88: Rate the usefulness of the personal and social development report on CoPS Baseline

	Frequency	1997	1998	1999	2001
No use at all	1	0	1	0	0
Needs improvement	5	0	4	0	1
Satisfactory	6	3	0	2	1
Fairly useful	11	4	5	0	2
Very useful	3	2	1	0	0

Table 89: Rate the clarity of the information in the personal and social development report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very confused	0	0	0	0	0
A little confused	2	0	1	0	1
Satisfactory	7	1	4	1	1
Clear	14	8	3	1	2
Very clear	2	0	2	0	0

Table 90: Rate the content of the information in the personal and social development report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Completely inadequate	0	0	0	0	0
A little inadequate	5	0	4	0	1
Satisfactory	8	2	2	2	2
Fairly comprehensive	9	5	3	0	1
Very comprehensive	3	1	2	0	0

Table 91: How long did it take to administer each module (in minutes)?

	Mean	SD	1997	1998	1999	2001
Literacy	6.70	3.46	6.00	7.40	5.00	6.75
Mathematics	6.91	3.55	6.00	7.60	8.00	6.75
Communication	8.63	4.52	6.75	9.64	12.00	8.75
Personal and social development	6.43	4.85	4.67	7.10	6.00	7.50

Table 92: How did you find using the reports on CoPS Baseline?

	Frequency	1997	1998	1999	2001
Very difficult	2	1	1	0	0
A little difficult	0	0	0	0	0
Satisfactory	9	2	5	0	2
Easy	11	4	5	2	0
Very easy	6	3	2	0	1

Table 93: Rate the usefulness of the parent report on CoPS Baseline

	Frequency	1997	1998	1999	2001
No use at all	1	0	0	1	0
Needs improvement	6	2	3	0	1
Satisfactory	8	3	2	1	2
Fairly useful	6	1	4	0	1
Very useful	4	3	1	0	0

Table 94: Rate the clarity of the information in the parent report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Very confused	1	0	0	0	1
A little confused	5	2	3	0	0
Satisfactory	8	2	3	2	1
Clear	7	1	4	0	2
Very clear	6	4	2	0	0

Table 95: Rate the content of the information in the parent report on CoPS Baseline

	Frequency	1997	1998	1999	2001
Completely inadequate	1	0	0	1	0
A little inadequate	2	0	2	0	0
Satisfactory	7	4	1	1	1
Fairly comprehensive	9	2	5	0	2
Very comprehensive	7	3	3	0	1

Table 96: Rate CoPS Baseline in comparison with other baseline assessment systems

	Frequency	1997	1998	1999	2001
Much worse	2	1	1	0	0
A little worse	0	0	0	0	0
About the same	4	1	0	1	2
A little better	8	3	2	1	2
Much better	5	2	3	0	0

Table 97: Will you be using CoPS Baseline in the future?

	Frequency	1997	1998	1999	2001
Definitely not	2	0	1	1	0
Probably not	2	2	0	0	0
Maybe	4	0	2	0	2
Probably	13	5	7	0	1
Definitely	7	3	2	1	1

For details about the content of the CoPS Baseline manual, CoPS Baseline video and the CoPS Baseline reports, please see Appendix 8.

7 Discussion for the baseline assessment studies

Tymms (1999) states that:

> ...baseline assessments taking just 20 minutes can only hope to identify a general starting point. 'Official' baseline assessments must therefore be treated tentatively. They can act as an initial screen and may identify children with general learning abilities / difficulties. Such pupils may need further assessments to diagnose specific difficulties. Further, baseline assessments cannot be specific about the academic future of individuals. They should, however, be able to help schools spot pupils who fail to thrive. These will be an unusual, but important, minority of pupils (page 34).

7.1 Overview

The results from the baseline assessment research are discussed with a view to:

- the development of the tests;
- the literacy follow-up studies (after 12 months and after 32 months);
- the mathematics follow-up studies (after 20 months and after 32 months);
- group differences (in relation to gender and ethnicity);
- technical issues (including internal consistency, reliability and validity);
- the questionnaire study.

7.2 Test development

The results from the test development stage indicate that the CoPS Baseline literacy and mathematics modules provide satisfactory distributions of scores across the intended age range, without undesirable ceiling or floor effects. On the literacy module (see section 6.2.1), no child scored at or near chance level, even in the inner city school, which contained a number of children from deprived homes. However, on the mathematics module (see section 6.2.2), two children (one from the inner city school and one from a suburban school) scored below chance level. Nevertheless, even on the mathematics module, the lowest score is 12, which suggests that both modules have adequate range to meet the requirements of all schools. This avoids the difficulty that can occur when interpreting and reporting to parents, results obtained from certain teacher rating scales where some children can score zero in all areas (for example, the QCA Baseline Assessment Scales).

Intercorrelations between most of the eight skill / concept areas in the literacy module were significant (see section 6.2.1). However, rhyming ability did not significantly correlate with aural comprehension and verbal concepts did not significantly correlate with rhyming, alliteration or simple reading. The aural comprehension (e.g. 'show me the correct picture for ... the man is up the ladder') and verbal concepts (e.g. 'show me something we can clean our teeth with') skill / concept areas relate to the child's listening comprehension, whilst the rhyming and alliteration skill / concept areas are tests of phonological awareness. These are quite distinct psycholinguistic skills and so a strong correlation would not necessarily be expected. Intercorrelations between all of the mathematics skill / concept areas were significant (see section 6.2.2) with the exception of one (shape and relative size / quantity). However, these are fairly distinct mathematical skills. The shape skill / concept area is a strongly visual task which requires the pupil to select a shape that fits into a picture that has a piece missing. On the other hand, the relative size and quantity skill / concept area, although also requiring a visual judgement, involves understanding of the mathematical terms 'smallest', 'longest', 'shortest' and 'same'.

The principal components analysis for the CoPS Baseline literacy module showed that 19 factors had eigenvalues over 1, which together accounted for approximately 70% of the total variance (see section 6.2.1). All of the letter recognition items loaded highly onto Factor 1, as did some items from the knowledge about print and simple spelling skill / concept areas. This factor alone accounted for approximately 9% of the total variance. Six of the seven rhyme items loaded highly onto Factor 2, which accounted for a further 6% of the total variance. None of the remaining factors appear to be clearly related to any individual or group of skill / concept areas. It is clear that the factors extracted from the principal components analysis are not consistent with the eight skill / concept areas and it can also be seen that the items within the CoPS Baseline literacy module do not comprise a unitary concept. However, the exact nature of the extracted components is not clear, even using a rotated factor analysis.

The principal components analysis for the CoPS Baseline mathematics module showed that 21 factors had eigenvalues over 1, which together accounted for approximately 71% of the total variance (see section 6.2.2). Four of the five number recognition items loaded highly onto Factor 1, as did some items from the counting and addition skill / concept areas. This factor alone accounted for approximately 10% of the total variance. Four of the five sets items loaded highly onto Factor 2, which accounted for a further 5% of the total variance. None of the remaining factors appear to be clearly related to any individual or group of skill / concept areas. It is clear that the factors extracted from the principal components analysis are not consistent with the nine skill / concept areas and it can also be seen that the items within the CoPS Baseline mathematics module do not comprise a unitary concept. However, again, the exact nature of the extracted components is not clear, even using a rotated factor analysis.

A possible reason for the poor result of the principal components analysis is that the items within the literacy and mathematics modules are dichotomous, thus producing poorer intercorrelations. It is also likely that a principal components analysis, using only the items retained in the modules after the item elimination, would produce clearer results.

The adaptive forms of the CoPS Baseline literacy (see section 6.2.1) and mathematics (see section 6.2.2) modules correlated highly with the full forms (r=0.81 and r=0.80 respectively), justifying the use of the much shorter adaptive forms. The time taken to complete the adaptive versions of the literacy and mathematics modules is only approximately 35% of the time taken to complete the full versions. These results support the assertion of Singleton (1997a) that, by using a computerised adaptive test, it is possible to give an accurate test in a practical amount of time.

7.3 Literacy studies

The CoPS Baseline literacy module scores give a satisfactory overall correlation with reading (r = 0.73) and spelling ability (r = 0.69) 12 months later, which indicates that the composition of the module and its scoring procedure is sound (see section 6.3). These correlations compare favourably with those found by Strand (1997 and 1999), using a checklist and the LARR Test of Emergent Literacy (r=0.60), and by Lindsay (1998) using the Infant Rating Scale (r=0.54).

The correlations between the eight skill / concept areas and reading and spelling ability 12 months later are fairly consistent with other findings reported in the literature. For example, Stuart (1995) reported on a study with 30 children in which a battery of tests of phonological awareness and basic phonic knowledge (grapheme-phoneme correspondence) was compared with the Diagnostic Survey, an early literacy assessment scheme developed by Clay (1979), both instruments being administered on school entry. The two different screening instruments were found to correlate significantly (r = 0.78), and both correlated significantly with reading age on the British Ability Scales Word Reading Test one year later: r = 0.80 for the Diagnostic Survey; r = 0.73 for the phonological / phonics battery. Bryant et al (1990) looked at the relationships between early phonological skills and later reading and spelling abilities. They found a number of early phonological skills to be highly correlated with later reading, including alliteration oddity (r=0.78), rhyme / alliteration choice (r=0.74), rhyme oddity (r=0.67), phoneme deletion of the first sound (r=0.67), phoneme tapping (r=0.59) and phoneme deletion of the end sound (r=0.58). Bryant et al (1990) also found these skills to be highly correlated with later spelling: alliteration oddity (r=0.73), rhyme / alliteration choice (r=0.71), rhyme oddity (r=0.65), phoneme deletion of the first sound (r=0.64), phoneme tapping (r=0.63) and phoneme deletion of the end sound (r=0.54). The CoPS Baseline literacy skill / concept areas showed lower correlations with later reading and spelling than the previous studies, with an average correlation of 0.69 between early literacy measures and later literacy ability and an average correlation of 0.46 between early phonological skills and later literacy (see section 6.3).

Several studies in the literature report that measures of early literacy give better prediction of reading than do tests of phonological ability, and this was partly the case in the present study. Tymms (1999) reports that the best baseline predictors for reading ability in Year 2 are letter identification (r=0.61), number identification (r=0.60), writing (r=0.52), counting (r=0.50), simple sums (r=0.48) and then phonic skills (r=0.48). Ellis (1990) reports that, in children who have not yet begun to read, phonological skills promote the acquisition of letter knowledge and that these two abilities, together with visual short term memory, underpin the development of reading. However, once reading acquisition begins, reading promotes further growth of phonological skills and auditory short-term memory, and these phonological skills in turn lead to the development of visual short-term memory. Ellis (1990) found reading skill to contribute more to later proficiency in auditory short-term memory and phonological processing than the reverse. According to Francis (1994), reading, spelling, short-term memory and phonological skills are best predicted by earlier measures of the same ability. Francis (1994) found that reading was more implicated in the prediction of phonological skill than vice versa although, unlike Ellis, she found auditory short-term memory to predict reading ability. In the present study, measures of letter and print knowledge, and of simple reading and spelling all produced correlations with reading (see section 6.3.1) and spelling ability (see section 6.3.2) 12 months later that were in excess of 0.6. The same was true of alliterative skills for reading but not spelling, but the correlations for rhyming skills (although still significant) were lower at 0.36 with reading ability and 0.30 with spelling ability. This finding is supported by the regression analysis which reveals that simple spelling, letter recognition and simple reading are the only significant predictors of reading ability one year later, whilst simple spelling, knowledge about print and letter recognition are the only significant predictors of spelling ability over the same period.

It is apparent that there is a strong relationship between letter recognition and later reading (r=0.66) and spelling (r=0.62) abilities. Gavel (1958) found that the highest correlation between pre-reading tests and reading achievement was between the latter and letter-name knowledge. Horn and O'Donnell (1984) obtained a correlation of 0.51 between IQ at the start of the first grade and end-of-first-grade reading score, and a correlation of 0.59 between letter and number recognition at the start of the first grade and end-of-first-grade reading score. Blatchford et al (1987) found pre-school letter identification to be the best predictor of reading at age seven (r=0.61) followed by handwriting skills (0.49), vocabulary (r=0.36), word matching (r=0.31) and concepts about print (r=0.27). Letter identification, handwriting and vocabulary were the only significant predictors, accounting for 40% of the variance.

Ellis and Cataldo (1990) found that spelling predicted reading, but not vice versa, until the second year in school when it ceased to do so. Nevertheless, Francis (1994) argues that spelling can be predicted by earlier reading ability. The present study supports these findings (see section 6.3), with early spelling correlating highly with reading 12 months later (r=0.76) as well as early reading correlating well with spelling (r=0.73) after the same period of time. However, the regression analysis shows that simple spelling is the best predictor of reading ability 12 months later (see section 6.3.1) whilst simple reading is not a significant predictor of later spelling ability (see section 6.3.2). Indeed, Mommers et al (1986) suggest that there is an influence of spelling on reading that is unique to the early stage of literacy development. The results of the present study also support the finding by Goulandris (1991) that early spelling is predictive of both reading and spelling ability one year later.

The present study found strong correlations (see section 6.3) between knowledge about print (understanding and use of terms such as 'letter', 'word' and 'sentence') and later reading (r=0.62) and spelling (r=0.64). A number of other researchers (e.g. Reid, 1966; Downing, 1970; Francis, 1973) have also reported relationships between reading development and knowledge about print. Stuart (1995) found a correlation of 0.66 between concepts about print and later reading ability. Blatchford et al (1987) reported a correlation of 0.31 between knowledge about print and later reading and Tymms (1999) obtained a similar correlation of 0.33. However, Francis (1994) found knowledge about print to be more predictive of later writing and spelling abilities than reading ability.

For the discriminant function analysis, the rates of false positives and false negatives have been calculated as percentages of the total sample. A stricter method of calculating false positives and false negatives is to express misclassifications as percentages of the predicted group and the target group respectively. This would result in much higher numbers of false positives and false negatives than have been shown in the discriminant function analysis in this study. For example, Table 12 (section 6.3.1) shows that seven of the 19 children predicted to be 'poor readers', turned out to be 'not poor', giving a strict false positive rate of 37%. Whereas six of the readers later found to be 'poor' had been predicted to be 'not poor', giving a strict false negative rate of 33%. However, the majority of researchers do use the method used in the present study.

The results of the discriminant function analysis suggest that the CoPS Baseline literacy module is not a particularly good screening device for predicting reading (see section 6.3.1) and spelling difficulty (see section 6.3.2), since there were high levels of false positives and false negatives. However, the CoPS Baseline literacy module was never designed to fulfil this function but to give a balanced picture of children's early literacy abilities consistent with the Desirable Outcomes for Children's Learning on Entering Compulsory Education (SCAA, 1996b) and within the rules laid down in the National Framework for Baseline Assessment (SCAA, 1997b).

Furthermore, early screening devices are notoriously inaccurate when it comes to predicting reading difficulty. A number of researchers have reported that non-learning disabled children are more accurately identified than learning disabled children (Rubin et al, 1978; Satz et al, 1978; Horn and O'Donnell, 1984). Several researchers have shown other early screening instruments to have similar levels of false positives and false negatives. For example, Evans et al (1978) indicate that the Swansea Evaluation Profiles give 7% false positives and 10% false negatives. Also the Aston Index has been shown to give 19% false positives and 6% false negatives (Newton, Thomson and Richards, 1979), whilst the Infant Rating Scale gives 9% false positives and 11% false negatives (Lindsay, 1980). According to Tymms (1999), "In terms of value-added measures, controlling for about 50% of the pupil level variance is quite acceptable — the point is to make fair comparisons and that this is best accomplished by controlling for the dominant predictor of later success" (page 34). The multiple regression analysis shows that, in both reading and spelling, one predictor variable accounts for over 50% of the variance in later literacy scores.

In addition, it should be noted that the variance in BAS-II reading scores was relatively high. Inspection of results for individual children illustrates this: 26 of the children scored two or fewer correct items (ability score 25 or less), of whom nine failed to score at all. On the other hand, eight children obtained very high scores, with ability scores in excess of 106 (raw score 45) and one child scored 59 correct items (ability score 127, equivalent to a reading age of 8 years 9 months). On the BAS-II spelling test, 32 children obtained ability scores of 25 or less whilst one child obtained an ability score of 110. Such findings of fairly wide variability in attainment are not uncommon in studies of early literacy development. For example, Francis (1982) illustrated how 5 to 7-year-old children exhibited marked differences of quality and gradient in their progress in reading. Lindsay and Wedell (1982) also found various school entry measures to be poor predictors of reading at age seven and suggested that examination of progress curves over the early years would be useful. Furthermore, Francis (1992) carried out an 18-month longitudinal study of the reading development of 50 children in a suburban city primary school. She found that when first tested at age 5 years 9 months (this was the same age at which the children in the present study were assessed in reading and spelling), 26 children were scoring zero. The remaining 24 were at various stages of reading attainment, with the most advanced having a reading age of 8.6 years. All those children that were already above the mean at first assessment continued to make smooth, steady progress. However, those children that were below the mean on first assessment that were observed to 'take off' in reading during the period of the study (i.e. to progress to reasonable independence in being able to cope with simple text) did so at different points, some earlier, and some later. Such findings give some clues as to why it is so difficult to obtain satisfactory predictions of reading difficulty.

Nevertheless, despite the apparent poor predictive value of CoPS Baseline, there are highly significant differences between children classed at baseline as being poor, average and good at literacy on the follow-up reading (see section 6.3.1) and spelling tests (see section 6.3.2).

It can be seen that the CoPS Baseline literacy module scores correlate satisfactorily with vocabulary (r=0.59) 20 months later (see section 6.5). The baseline literacy skill / concept area that correlates most highly with later vocabulary is knowledge about print (r=0.60), whilst the lowest correlations are with the aural comprehension (r=0.21) and verbal concepts (r=0.25) skill / concept areas. This result is surprising as the verbal concepts skill / concept area appears to be the most similar to the BPVS. However, the pass rates for the verbal concepts and aural comprehension items are quite high. Even the youngest age group (4 years 0 months to 4 years 5 months) had a pass rate of 0.77 for the hardest aural comprehension item and 0.52 for the hardest verbal concepts item. Nevertheless, the National Framework for Baseline Assessment (SCAA, 1997b) requires that baseline assessment schemes must cover aspects of literacy as specified in the Desirable Outcomes for Children's Learning on Entering Compulsory Education (SCAA, 1996b), which states that by the time children enter compulsory education they should 'know that words and pictures carry meaning' (page 3). Furthermore, aural comprehension and verbal concepts are the first two skill / concept areas within the CoPS Baseline module and so the items have been selected to be fairly easy in order to motivate young children with little or no previous experience of testing.

The CoPS Baseline literacy module scores also give satisfactory correlations with reading (r=0.65), spelling (r=0.59) and writing ability (r=0.56) 32 months later (see section 6.6). These coefficients are slightly below the 0.70 limit to the predictability of reading quoted by Tymms (1999). The correlations between the eight skill / concept areas and reading and spelling ability 32 months later are fairly consistent with the previous findings from the 12 month follow-up study.

Again, the baseline literacy skill / concept area that correlates most highly with later reading (see section 6.6.1) is simple spelling (r=0.60), although knowledge about print (r=0.54) and alliteration (r=0.52) are now seen to be more predictive of reading ability than letter recognition (r=0.51) and simple reading (r=0.45). The finding that simple spelling is the most significant predictor of reading after 32 months (accounting for 36% of the variance) is contradictory to the assertion by Ellis and Cataldo (1990) that spelling ceases to be a predictor of reading from the second year in school. The regression analysis indicates that simple spelling, aural comprehension and rhymes are the only significant predictors of reading ability 32 months later. Ellis (1990) found phonological skill to be a better predictor of second year reading ability than early literacy ability. This finding is partially supported by the results of the present study, in that alliteration is more predictive of reading at this stage than are letter recognition and simple reading, although simple spelling is still the best predictor.

Alliteration (r=0.50) and knowledge about print (r=0.50) correlate most highly with spelling 32 months later (see section 6.6.2), followed by letter recognition (r=0.46), simple spelling (r=0.45), which had earlier been the best predictor of spelling, and simple reading (r=0.43). The regression analysis indicates that knowledge about print, alliteration and aural comprehension are the only significant predictors of spelling ability 32 months later. Ellis (1990) and Francis (1994) both found that from the second year in school, spelling ability was predicted by phonological skill more than by earlier literacy ability. These findings are partially supported by the results of the present study, in that alliteration is a better predictor of later spelling ability than earlier literacy ability.

The baseline literacy skill / concept areas that correlate most highly with later writing ability (see section 6.6.3) are letter recognition (r=0.41), knowledge about print (r=0.40) and rhymes (r=0.40). The regression analysis indicates that letter recognition and rhymes are the only significant predictors of writing ability 32 months later.

Again, the results of the discriminant function analyses show high levels of false positives and false negatives (see section 6.6). However, CoPS Baseline would not be expected to be a good predictor of literacy skills after 32 months given the poor prediction after 12 months. Nevertheless, there are still highly significant differences between children classed at baseline as being poor, average and good at literacy on the 32-month follow-up reading, spelling and writing tests.

7.4 Mathematics studies

The CoPS Baseline mathematics module scores also give a satisfactory overall correlation (r=0.66) with mathematical ability 20 months later, which indicates that the composition of the module and its scoring procedure is sound (see section 6.4).

The correlations between the baseline skill / concept areas and mathematics ability 20 months later are consistent with other findings reported in the literature. In the present study, measures of addition, subtraction, number recognition and counting all produced correlations with mathematics ability 20 months later that were in excess of 0.5. This finding is supported by the regression analysis, which reveals that addition, subtraction and number recognition are the only significant predictors of mathematics ability 20 months later. However, it should be noted that the mathematics test used in this follow-up study (WOND Numerical Operations Test, Wechsler, 1996b) assesses the ability to write dictated numerals and solve calculations involving addition, subtraction, multiplication and division. Nevertheless, Horn and O'Donnell (1984) obtained a correlation of 0.52 between IQ at the start of the first grade and end-of-first-grade mathematics score and a correlation of 0.51 between letter and number recognition at the start of the first grade and end-of-first-grade mathematics score. Tymms (1999) reports that the best baseline predictors for mathematics ability in Year 2 are number identification (r=0.55), letter identification (r=0.54), counting (r=0.50), writing (r=0.49), simple sums (r=0.49), phonic skills (r=0.45) and then matching shapes (r=0.45). In the present study a correlation of 0.54 was found between early number recognition and later mathematics ability, which is comparable to those obtained by Horn and O'Donnell (1984) and Tymms (1999). The correlations with counting (r=0.52) and addition (r=0.57) are also similar to those obtained by Tymms (1999), although the correlation with the shapes skill / concept area is somewhat lower (r=0.29).

Bryant et al (1990) found a number of early phonological skills to be predictive of later arithmetic ability (at age seven). They report correlations between later arithmetic and rhyme / alliteration choice (r=0.58), phoneme tapping (r=0.54), rhyme oddity (r=0.53), alliteration oddity (r=0.48), phoneme deletion of first sound (r=0.46) and phoneme deletion of end sound (r=0.33). The phoneme tapping test involves counting and gives a very similar correlation to that found in the present study between early counting and later mathematical ability (r=0.52).

Dimitrovsky and Almy (1975) suggest that early number conservation (i.e. recognising that a set of blocks of a certain number consists of the same number whether grouped together or spread out) is predictive of later arithmetic performance. They found that children who conserved number in kindergarten were significantly more likely to perform at or above grade level in arithmetic 18 months later than those children who did not. Early conservation of number was also found to be predictive of arithmetic achievement after 30 months for girls, but not for boys (Dimitrovsky and Almy, 1980). Number conservation is not actually tested by the CoPS Baseline mathematics module. However, three of the items in the relative size and quantity skill / concept area involves recognising that two pictures have the same number of things in (these may be identical or non-identical items). This skill / concept area shows a correlation of 0.48 with later mathematical ability but it was not shown by the regression analysis to be a significant predictor of mathematical ability after 20 months.

The results of the discriminant function analysis suggests that the CoPS Baseline mathematics module is not a particularly good screening device for predicting mathematical difficulty, since there were high levels of false positives and false negatives (see section 6.4). Again, CoPS Baseline assessment was never designed to fulfil this function, but to give a balanced picture of children's early mathematics abilities. Nevertheless, the analysis of variance results show that there are highly significant differences between children classed at baseline as being poor, average and good at mathematics on the follow-up mathematics tests 20 months later.

It can be seen that the CoPS Baseline mathematics module scores correlate satisfactorily with vocabulary (r=0.56) 20 months later (see section 6.5). The baseline mathematics skill / concept area that correlates most highly with later vocabulary is relative size and quantity (r=0.57), whilst the lowest correlation is with the shape (r=0.14) skill / concept area. This result is not surprising as the relative size and quantity skill / concept area requires an understanding of the terms 'smallest', 'longest', 'shortest' and 'same'.

The CoPS Baseline mathematics module scores also give satisfactory correlations with mathematics ability 32 months later (r=0.60). This correlation is below the 0.70 limit to the predictability of mathematics quoted by Tymms (1999). The correlation between the CoPS Baseline mathematics module and later mathematics ability is 0.34 when vocabulary is partialled out (see section 6.7). This indicates that the children's vocabulary does affect their performance on the mathematics test. This is unsurprising, since the Mathematics 8 test involves relatively high use of language, in contrast with tests of basic calculation, e.g. WOND Numerical Operations Test (which was employed in the first follow-up study).

The correlations between the skill / concept areas and mathematics ability 32 months later are fairly consistent with the previous findings from the 20 month follow-up study (see section 6.7). The highest correlations with mathematics after 32 months are with subtraction (r=0.47), number recognition (r=0.43) and addition (r=0.40). These findings are supported by the regression analysis, which reveals that subtraction and number recognition are the only significant predictors of mathematics ability 32 months later. The main difference between these findings and the earlier results is that addition is no longer the best predictor of mathematics ability. It may be that subtraction is more predictive of mathematics at this stage because the later maths test involves more subtraction items than the earlier test.

Again, the results of the discriminant function analysis show relatively high levels of false positives and false negatives. However, CoPS Baseline would not be expected to be a good predictor after 32 months given the poor prediction after one year. Nevertheless, there are still highly significant differences between children classed at baseline as being poor, average and good at mathematics on the 32-month follow-up mathematics tests.

7.5 Group differences

7.5.1 Gender

During the test development stage, no significant gender differences were found on the literacy module (see section 6.2.1) and only one of the mathematics skill / concept areas (sets) showed a significant gender difference (see section 6.2.2). This is counter to the reports by a number of researchers that girls are outperforming boys in baseline assessment (e.g. Lindsay, 1980; Strand, 1997; Lindsay and Desforges, 1999). However, the research literature is not entirely consistent on this point. Tizard et al (1988) found no significant difference between boys and girls in pre-literacy attainment at the end of the nursery stage. On the other hand, Tymms, Merrell and Henderson (1997) reported that during their time in the reception class, girls tend to make greater progress in literacy than boys. In a large scale study of baseline assessment and follow-up, Strand (1997) found that, at the age of four, girls had significantly higher attainment than boys on a 10-item rated checklist of early literacy skills and on the LARR Test of Emergent Literacy (Ayres, Downing and Schaefer, 1982). By age seven the gap between boys and girls had widened considerably. It should be noted that the scale of the present study is small (n = 153 for literacy; n = 140 for mathematics) in comparison with many of the studies that have reported gender differences in baseline assessment. For example, in Strand's 1997 study, the analysis of gender differences was based on 1,699 children.

It is possible that the failure to find a significant gender difference in this study is due to the relatively small sample size. However, examination of the checklist used by Strand (1997) shows a prevalence of items that assess what might be described as 'enthusiastic interaction with the learning environment' (e.g. 'responds to instructions', 'listens and responds to stories', 'looks at books for pleasure' and 'gives explanations'). Arguably, such items are more susceptible to subjectivity effects than are those that concern more concrete skills, such as 'writes own name' and 'uses some letter symbols'. The type of items included in Strand's checklist are not dissimilar to those to be found in a large number of accredited baseline assessment systems already in use across the UK. By contrast, the CoPS Baseline modules assess more specific early literacy skills (such as phonological awareness and simple spelling) and early mathematics skills (such as shape and numbers) in an objective manner.

During the relaibility and validity stage, an overall significant gender difference was found using the CoPS Baseline test, with girls obtaining a higher overall score than boys (see section 6.9.1). However, further analysis reveals that there are no significant differences on the two fully computerised modules of CoPS Baseline (literacy and mathematics) but there are significant gender differences on the two more subjective modules (communication and personal and social development), with girls obtaining higher scores on both modules. Significant gender differences were also found on two of the five QCA scales (reading and writing), with girls again obtaining higher scores. These findings may support the hypothesis that fully computerised tests are more objective than teacher ratings. However, the interaction with a computer in the CoPS Baseline tasks might have enhanced the motivation of the boys more than that of the girls and this could have compensated for (or masked) gender differences in performance. There is therefore the possibility that it is not the content of certain items in baseline assessment instruments that is to be preferred over others but, rather, that the manner of the assessment makes a difference to the results obtained. When the overall approach and item content are such that subjectivity is ruled out and data are based solely on the child's performance in standardised decision-making tasks, it is likely that gender differences will be less marked than in conventional baseline assessment systems that rely on teacher ratings. It remains to be seen whether larger-scale studies using more objective approaches to baseline assessment (such as computerised administration) will lead to a modification of the accepted wisdom regarding the gender issue.

7.5.2 Ethnicity

No significant ethnic differences were found in the results obtained from using the CoPS Baseline system, although the scores of the white group are consistently higher than the scores of the other children (see section 6.9.1). However, a significant difference was found on the QCA total score, with white children obtaining a higher overall score than children from other ethnic backgrounds (see section 6.9.2). Further analysis reveals that there is a significant difference between these groups on only one of the five QCA scales (writing). These results are contrary to the findings of many researchers who have found white children to score significantly higher than children from other ethnic backgrounds on baseline assessment (e.g. Lindsay, 1988; Strand, 1995 and 1999; Everett, Farnsworth and Mitchell, 1997; Lindsay and Desforges, 1999). However, Strand (1999) argues that where English is the main home language, Indian, Pakistani and Chinese pupils do as well as white children on baseline assessment. In the present study, the majority of children from non-white ethnic backgrounds spoke English as a first language and so these results support Strand's finding. The lack of any significant difference may also be due to the very small number of children in the non-white sample. Nevertheless, the fact that there is a significant difference on the basis of ethnic group on the QCA scales suggests, again, that teacher rating scales, as opposed to objective computerised tests, are subject to bias.

7.6 Technical issues

7.6.1 Reliability

7.6.1.1 Internal Consistency

The literacy (see section 6.2.1) and mathematics modules (see section 6.2.2) of CoPS Baseline show high levels of overall consistency (0.91 and 0.88 respectively). These are comparable to other reported estimates of internal consistency in other standardised baseline schemes. The Performance Indicators in Primary Schools (PIPS) system shows internal consistency coefficients of 0.95 for reading and 0.88 for mathematics (Tymms, 1999). Lindsay and Desforges (1999) report that the internal consistency of the Infant Index / Baseline-PLUS is 0.92. Strand (1999) reports internal consistency coefficients of 0.85 for the LARR (Linguistic Awareness in Reading Readiness) Test of Emergent Literacy and 0.90 for a baseline checklist. However, some of the CoPS

Baseline literacy and mathematics skill / concept areas showed rather poor internal consistency (verbal concepts, aural comprehension, simple reading, shape and seriation).

7.6.1.2 Test-retest reliability

CoPS Baseline appears to be reliable over time, with an overall test-retest reliability of 0.91 (see section 6.9.3). This is very similar to the reported test-retest reliability of 0.89 for the Infant Index (Desforges and Lindsay; 1995b). Test-retest reliability coefficients reported by Tymms (1999) for sub-tests of PIPS range from 0.32 (picture identification) to 0.95 (identifying two digit numbers), with an overall test-retest reliability of 0.93. The lowest test-retest reliability found in CoPS Baseline was for the mathematics module (r=0.58), although this was still very highly significant.

7.6.2 Validity

For details of the predictive validity of CoPS Baseline, refer to the sections covering the literacy follow-up studies (section 7.3) and the mathematics follow-up studies (section 7.4).

CoPS Baseline has an overall concurrent validity of 0.64 with the QCA Scales (see section 6.9.4). The correlation between the CoPS Baseline literacy modules and the QCA literacy total is moderate (r=0.56). When this is examined in more detail, the correlations between the CoPS Baseline literacy module and the QCA reading total (r=0.52) and the QCA writing score (r=0.50) are very close, despite the fact that CoPS Baseline does not assess children's writing. The QCA reading total is made up of three reading scores which all correlate significantly with the CoPS Baseline literacy score: reading for meaning and enjoyment (r=0.41); letter knowledge (r=0.52); and, phonological awareness (r=0.43). The lowest of these correlations is with reading for meaning and enjoyment. Nevertheless, this is still a significant moderate correlation, despite the fact that the CoPS Baseline literacy module does not directly assess these abilities. However, as noted by Nutbrown (1999), the QCA Scale does not test the use and knowledge of environmental print, which is part of the CoPS Baseline literacy modules. The QCA Scale also does not test listening comprehension or spelling.

There is a low, but nonetheless significant, correlation (r=0.26) between the CoPS Baseline mathematics module and the QCA maths total score. However, the correlations between the CoPS Baseline mathematics module and the individual QCA maths scales are not significant (number, r=0.24; mathematical language, r=0.23). The number sub-test involves addition and subtraction, whilst the mathematical language sub-test involves size, position, numbers 1-10 and explaining addition. These are all tested on the CoPS Baseline mathematics module, with the exception of the explanation of addition. The CoPS Baseline mathematics module is more diverse than the QCA Scale, in that it also assesses shape, pattern, sets and counting. This may, in part, explain the low correlation between the two tests.

The correlation between the CoPS Baseline communication module and the QCA speaking and listening scale is moderate and significant (r=0.40). It should be noted that the QCA scale involves the child making up a story and telling it as well as recounting a story, which is the sole requirement of the CoPS Baseline communication module.

There is a moderate significant correlation (r=0.54) between the CoPS Baseline personal and social development module and the QCA personal and social development scale. However, the QCA scale is based on only four (yes / no) items, as opposed to 10 (3-point) items in the CoPS Baseline test.

7.7 Questionnaire study

The majority of CoPS Baseline users had no difficulties with installing the system (62%) or registering pupils (89%). Indeed, those that did experience any difficulty tended to be users of early versions of the software (1997 and 1998). Both the CoPS Baseline training video and the manual were rated as 'good' for content (58% and 83% respectively), presentation (67% and 78%) and organisation (67% and 67%) by the majority of users (see section 6.10).

Most users found running the communication module satisfactory or easy (93%) and 81% reported that the children enjoyed the module. Despite the fact that many respondents (65%) felt that the communication module was too difficult for the children, the majority (67%) saw it as being adequate or suitable as a test of expressive language. The scoring system for the communication modules was found by 37% of users to be difficult whilst the report was generally highly rated for usefulness (85% satisfactory or useful), clarity (88% satisfactory or clear) and content (89% satisfactory or comprehensive).

Most users found running the literacy module satisfactory or easy (96%) and 96% reported that the children enjoyed the module. The majority of respondents felt that the difficulty level of the literacy module was about right for the children (70%), although 22% thought that it was a little too difficult, and most regarded it as educationally relevant (81%). The literacy report was generally highly rated for usefulness (81% satisfactory or useful), clarity (85% satisfactory or clear) and content (85% satisfactory or comprehensive).

All users found running the mathematics module satisfactory or easy and 93% reported that the children enjoyed the module. The majority of respondents felt that the difficulty level of the mathematics module was about right for the children (61%), whilst 36% reported that it was too difficult. However, some teachers were not aware that the test was adaptive and that children answering items correctly were being given more difficult items. Most users regarded the mathematics module as educationally relevant (78%). The mathematics report was generally highly rated for usefulness (92% satisfactory or useful), clarity (92% satisfactory or clear) and content (88% satisfactory or comprehensive).

Most respondents found running the personal and social module satisfactory or easy (77%) and 96% regarded it as educationally relevant. The personal and social report was generally highly rated for usefulness (77% satisfactory or useful), clarity (92% satisfactory or clear) and content (80% satisfactory or comprehensive).

The average reported time taken to complete all four modules is 28.67 minutes. However, as the personal and social module does not involve any child testing time, the overall reported time spent testing each child is, on average, 22.24 minutes, which is close to the SCAA 'target' of 20 minutes (SCAA, 1997b) for baseline assessment.

Most respondents found using the reports on CoPS Baseline satisfactory or easy (93%). The parent report was generally highly rated for usefulness (72% satisfactory or useful), clarity (78% satisfactory or clear) and content (88% satisfactory or comprehensive).

68% of respondents rated CoPS Baseline as better than other baseline assessment schemes, whilst 21% rated it as about the same and 10% rated it as worse than other baseline assessment schemes. 71% of users stated that they would be using CoPS Baseline in the future, 14% were unsure and 14% said that they would not use the system. Most of those reporting that they would not be using the system in future were users of an early version of the software (1997 and 1998), which was subject to some technical difficulties.

7.8 Summary and conclusions

CoPS Baseline fits the requirements of the National Framework for Baseline Assessment, in that it:

- covers aspects of language, literacy, mathematics and personal and social development
- includes guidance to teachers on how the outcomes of the assessment can be used to inform planning for the class and for individual children
- provides numerical outcomes capable of being used for value-added analysis
- specifies the period after the child has started school within which the assessment should be completed (seven weeks).

Furthermore, there is evidence of the technical quality (internal consistency, test-retest reliability, predictive validity and concurrent validity) of CoPS Baseline. In this respect, the system compares well to other baseline systems providing such information, although there are many baseline assessment schemes that do not provide evidence of technical quality and, indeed, are not required to. The use of two adaptive modules within the CoPS Baseline system drastically cuts the assessment time whilst retaining the accuracy of the test. In addition, CoPS Baseline contains standardised norms and is completely objective in its testing procedures. The CoPS Baseline system therefore addresses many of the challenges facing baseline assessment and has been found to be an acceptable on-entry assessment scheme by teachers.

Several researchers have commented on the wide variation in baseline assessment schemes that have been accredited in the UK and the difficulties that this creates for accountability, monitoring of standards and consistent calculation of value added (e.g. Lindsay, 1998; Murphy, 1998; Lindsay and Desforges, 1998). The use of norm-based tests, such as CoPS Baseline, would make baseline testing more consistent and aid the calculation of value-added.

The research involving CoPS Baseline has revealed no significant gender differences, a finding that is contrary to much other baseline assessment research. A difficulty with conventional baseline assessment systems is that they are seen to produce results that suggest that girls are 'more advanced' than boys on school entry. It is then possible that some teachers may, consciously or subconsciously, make allowances in their ratings for boys or interpret baseline assessment results to imply that boys are potential 'failures' when they have only just entered education, thus creating self-fulfilling prophecies (Ghouri, 1998).

Many conventional methods of baseline assessment that rely on subjective ratings, including checklists and teacher observation schedules, have been accredited for use in primary schools in England and Wales. The use of such varied tests makes it impossible to conduct any valid analysis of baseline data on either a national or local basis. Nevertheless, such analysis is still attempted and the results used to make important decisions about educational policy and differential resourcing (Singleton, Horne and Thomas, 1999). As Lindsay (1993) has pointed out, "We must acknowledge the wider educational, and increasingly, the socio-political contexts for baseline assessment if it is to be developed in a useful manner" (page 58). More objective methods of baseline assessment, including the use of computerised approaches, can help to prevent these difficulties and are therefore worth the investment in research and development and the expenditure of teacher time. As Wolfendale (1993b) states '...it behoves us all to secure baseline assessment practice that is sound and just on behalf of all children' (page 45).

8 Dyslexia assessment

8.1 Aims of the chapter

This chapter aims to:

- provide a definition of dyslexia
- state the purposes of dyslexia assessment
- provide an overview of the ability versus attainment discrepancy model of dyslexia
- provide an overview of alternative discrepancy models
- provide an overview of the phonological deficit hypothesis
- provide evidence of memory difficulties in individuals with dyslexia
- identify the challenges facing dyslexia assessment.

8.2 Definitions of dyslexia and purposes of assessment

There are two forms of dyslexia – developmental dyslexia and acquired dyslexia. Developmental dyslexia is a specific learning difficulty which is inherited in 70% to 80% of cases, with birth difficulties being the other main cause. Acquired dyslexia is an impairment of literacy skills caused by neurological trauma, illness or brain disease. The majority of people with dyslexia have developmental dyslexia and this is the form of dyslexia that this chapter will focus on.

Vellutino (1979) differentiated between 'extrinsic' and 'intrinsic' causes of reading difficulty. Extrinsic causes include social background, lack of opportunity, frequent changes of school and bad teaching; whilst intrinsic causes include emotional and behavioural problems such as hyperactivity, difficulties in attention, low intelligence, brain damage and sensory deficits. He asserts that developmental dyslexia is a severe and persistent difficulty with written language in the absence of any intrinsic or extrinsic cause.

A degree of confusion exists within this area as there is no one agreed definition of dyslexia. However, there are some well-accepted definitions in use at present. The definition presented by the Orton Dyslexia Association (1994) expands a great deal on that previously given by Vellutino, stating that:

> dyslexia is one of several distinct learning disabilities. It is a specific language-based disorder of constitutional origin characterised by difficulties in single word coding, usually reflecting insufficient phonological processing abilities. These difficulties in single word decoding are often unexpected in relation to age and other cognitive and academic abilities; they are not the result of generalized developmental disability or sensory impairment. Dyslexia is manifested by variable difficulty with different forms of language, often including, in addition to problems of reading, a conspicuous problem with acquiring proficiency in writing and spelling (page 5).

The British Dyslexia Association (1998) gives a similar definition but adds that "notational skills (music), motor function and organisational skills may also be involved" (page 48). Memory difficulties are not specifically stated in either of these definitions. However, the National Working Party on Dyslexia in Higher Education (1999) describes dyslexia as 'a complex neurological condition that occurs in approximately 4% of the population, and which primarily affects acquisition and use of written language, memory and organisational skills' (page 1).

Dyslexia is therefore best described as a syndrome rather than a single disorder. Not all people with dyslexia will display the same range or severity of difficulties. According to the National Working Party on Dyslexia in Higher Education (1999), about 4% of the population are affected to a significant extent. This figure is based on the incidence of pupils who have received normal schooling and who do not have significant emotional, social or medical aetiology, but whose literacy development by the end of the primary school is more than two years behind levels which would be expected on the basis of chronological age and intelligence. This estimate corresponds roughly to incidence figures derived from epidemiological studies of specific reading retardation (Rutter and Yule, 1975; Miles, 1991) and other authoritative estimates in the literature (British Dyslexia Association, 1998; Thomson, 1993; Turner, 1997).

Wright and Groner (1993) suggest that dyslexia may consist of a number of subtypes which could explain the range of deficits on a number of cognitive tasks which have been found when individuals with dyslexia have been compared to controls. There is little agreement on the number of subtypes but the most commonly reported are an auditory linguistic subtype (when the individual displays difficulties in the discrimination of speech sounds, in 'sound blending' and in naming visual stimuli) and a visual perceptual subtype (when the individual displays problems in visual perception and visual discrimination). In almost all studies the language subtype is reported more frequently than the visual subtype. However, the visual subtype does exist and should not be ignored. A mixed subtype is also often described, comprising individuals with both auditory and visual-perceptual difficulties.

Johnson and Myklebust (1967) first made the distinction between auditory and visual dyslexia, arguing that deficiencies might be auditory and / or visual in nature. Boder (1973) classified individuals with dyslexia into three subtypes. The 'dysphonetic' group have a "primary deficit in symbol-sound integration, resulting in inability to develop phonetic word analysis-synthesis skills" (Boder, 1973, page 667). The 'dyseidetic' group show a "deficit in the ability to perceive letters and whole words as configurations or visual gestalts" (Boder, 1973, page 667). She argues that a third subtype of dyslexia is the 'mixed' group.

Bakker (1990) found that at an early age there is a greater amount of activity in the right hemisphere when reading a word but at the age of seven or eight there is a switch to predominant activity in the left hemisphere. He argues that some children do not switch to left-brain learning at this age and he groups these as 'P-types' ('perceptual'). Bakker suggests that other children switch to left-brain learning too early and he groups these as 'L-types' ('linguistic'). However, Robertson (1997) found it difficult to classify children with dyslexia using these subtypes.

There is a need for a clear definition of dyslexia in order for it to be identified. Rispens and Yperen (1990) looked at five different methods for identifying dyslexia and found that the number of children identified as dyslexic depended on which measurement method was employed, ranging from between 4% and 18.3%. Furthermore, the definition of dyslexia used influences the relative numbers of males and females identified as having dyslexia. Traditionally, more boys than girls are identified, with ratios ranging from 2:1 to 5:1 in the UK (Critchley, 1970). Support for this finding comes from Snowling (1987), who argues that dyslexia affects three to four boys for every girl, and Zabell and Everatt (2000) who state that "many studies concerning the incidence of dyslexia have identified a four to one ratio of males to females" (page 83).

However, Shaywitz et al (1990) argue that reports of increased prevalence of dyslexia in boys may simply reflect referral biases and that the ratio is markedly lower in a research identified population (1.2:1). Other studies have also reported gender ratios closer to 1:1 (Wadsworth et al, 1992; Lubs et al, 1993). Indeed, Miles, Haslum and Wheeler (1998) reanalysed national sample data which gave a ratio of 4.5:1 using a definition of specific developmental dyslexia (Critchley, 1970) and found that using a discrepancy only definition, the ratio was 1.69:1. According to Share, Silva and Adler (1987), current definitions of dyslexia do not take account of the finding that reading expectancies are higher for girls than for boys. A girl may be underachieving relative to peers of her own sex but not relative to members of the opposite sex. Scarborough (1984) computed separate regression equations for males and females and virtually eliminated the traditional sex ratios.

A working group of the Division of Educational and Child Psychology of the British Psychological Society (1999) provides a 'working definition' of dyslexia. This definition states that "dyslexia is evident when accurate and fluent word reading and/or spelling develops very incompletely or with great difficulty" (page 11). Accordingly, three aspects are required to be evaluated through the assessment process. First, that the pupil has not learnt to read or spell accurately and fluently. Second, that appropriate learning opportunities have been provided. Finally, that progress has been made only as a result of much additional effort or instruction and that difficulties have, nevertheless, persisted. There are some difficulties in applying this definition as it is not clear what 'very incompletely' means, what learning opportunities can be deemed 'appropriate' and how much 'additional effort' is required. A diagnosis of dylsexia could not be made until additional instruction has been given and not found to have brought about significant improvements. However, dyslexic children's literacy skills can be improved by specialist tuition. The report also states that culture-fair assessment demands that dyslexia is identified across languages, cultures, socio-economic status, race and gender.

Reid (1998) proposes that the aims of assessment should be:

- to identify the student's general strengths and weaknesses
- to indicate the student's current attainments
- to explain the student's lack of progress
- to identify aspects of the student's performance in reading, writing and spelling, which may typify a 'pattern of errors'
- to identify specific areas of competence
- to give an understanding of the student's learning style
- to indicate aspects of the curriculum which may interest and motivate the learner.

He continues that assessment for specific learning difficulties should consider three aspects – difficulties, discrepancies and differences. The 'difficulties' may include decoding or encoding of print, phonological processing, memory problems, organisational and sequencing difficulties, motor co-ordination, language problems or auditory or visual perceptual difficulties. 'Discrepancies' may be between decoding and reading or listening comprehension, between oral and written responses, or between different subject areas. The 'differences' being looked at are between individual learners in their learning and cognitive styles.

8.3 The ability versus attainment discrepancy model

The traditional method used to identify dyslexia has been to look for a significant discrepancy between intelligence and attainment in reading, spelling and writing. According to the National Working Party on Dyslexia in Higher Education (1999), significant means "so much poorer than the level which would be expected on the basis of the person's age and intelligence that the discrepancy is unlikely to be due to normal variation within the population or to chance" (page 99). The accepted method of looking at the discrepancy between performance on different tests uses standardised scores such as centiles or standard deviation units. Snyderman and Rothman (1988) argue that intelligence is a significant predictor of academic success and so Kline (1993) suggests that the cause of academic failure may be low intelligence, thus the first step in the assessment process should be the intelligence test. Until the intelligence of a student is known, it is not possible to know how he or she should be performing in literacy.

When the discrepancy is based on a simple standard score difference, an individual is identified as having dyslexia if the attainment standard score is significantly lower than the ability standard score. A problem with this method is that comparison of two correlated test scores is associated with regression to the mean – the more extreme a score on one measure is, the higher the probability that the score obtained by the same individual on another measure is closer to the mean than that of the first measure. Wright and Groner (1993) argue that if a direct comparison is made between measures then there is a likelihood that very intelligent children will be identified as underachievers. Furthermore, Turner (1997) claims that the method is sometimes unfair to pupils of lesser ability. A preferable method uses linear regression. This method takes into account the correlation between the attainment and ability scores and predicts attainment scores avoiding the problems associated with regression to the mean (Turner, 1997).

There are three intelligence tests commonly used in the assessment of children for dyslexia. First, the Wechsler Intelligence Scale for Children (WISC) - III (Wechsler, 1992) which provides an overall IQ score as well as verbal and performance IQ scores. The scale is for children aged 6 to 16:11 and shows high levels of validity and reliability. Indeed, Cooper (1995) asserts that the WISC tests are probably the best validated and most widely accepted measures of children's intellectual functioning. Second, the British Ability Scales (BAS) II (Elliott, Smith and McCulloch, 1996) is a commonly used alternative to the WISC. The test is designed for 2 to 17 year olds and comprises two batteries: the Early Years Battery, consisting of cognitive scales only; and the School Age Battery consisting of cognitive and achievement scales (including single oral word reading and spelling). The test was standardised in 1995 and also displays high standards of validity and reliability. As the test includes attainment scales, it is capable of showing any discrepancy between literacy achievement and ability. Third is the less well known Woodcock-Johnson test (Revised) (Woodcock and Johnson, 1989-1990). The first edition in 1977 was the first fully developed test which included co-normed tests of ability and achievement with the means of evaluating discrepancies between the two. The revised version has 35 tests (21 tests of cognitive ability and 14 tests of scholastic achievement) and provides norms for 2 to 79 year olds. The cognitive tests fall into six categories: verbal, visuo-spatial, non-verbal reasoning, speed, verbal short-term memory and verbal strategy. The first three categories give a general ability measure and the second three categories give a measure of information skills. Any disparity between the two may be an indication of learning difficulty.

Some researchers have doubted the relevance of intelligence tests in diagnosing dyslexia for a number of reasons (McDougall and Ellis, 1994; Siegel, 1988; Stanovich, 1991a, 1991b, 1991c; Reid, 1998). Probably the most important reason is that poor readers with average or high IQs do not necessarily constitute a different group from 'garden variety' readers (Siegel, 1989). 'Garden variety' is the term used for individuals whose poor reading is thought to be a consequence of low intelligence rather than dyslexia (Wolf, 1991). Stanovich and Siegel (1994) found no difference between disabled readers defined by means of a discrepancy and 'garden variety' disabled readers with regard to the nature of the reading disability.

Second, children who may have missed school through long term illness, who have not been properly taught in literacy or who speak English as a second language would be identified on a test looking for a discrepancy between ability and attainment as underachieving in literacy relative to ability. Although some of these children could have dyslexia, it is likely that the other factors are the cause of their literacy difficulties and so further evidence is necessary for a diagnosis of dyslexia, such as difficulties in memory or phonological skills.

Third, although a discrepancy between ability and poor performance should signal a possibility of dyslexia, the problem is often not identified at an early age, when a child starts learning to read and write. Assessment of cognitive skills that are precursors of dyslexia is more effective for children starting school.

Fourth, intelligence tests are often restricted to use by psychologists rather than teachers. Beech and Singleton (1997) argue that if an education authority was to adopt a discrepancy approach then mass screening by a psychologist would be ruled out on the grounds of cost. Instead, only children already displaying significant difficulties in literacy are likely to be assessed.

Fifth, children with a high IQ and slightly above average reading attainment could be described as reading disabled when using this type of discrepancy criterion (Siegel, 1989). She argues that "calculating a discrepancy and requiring that achievement tests and an IQ score be sufficiently discrepant seems an illogical way of calculating whether or not there is a learning disability" (page 472). Miles and Miles (1999) support this argument, stating that it is not the discrepancy but the child's actual needs that should be the key factor in an assessment.

Finally, some conventional IQ measures include tests of working memory which individuals with dyslexia have a tendency to score poorly on. This has the effect of depressing their overall IQ measure significantly. Siegel (1989) claims that IQ tests do not necessarily measure intelligence, but in fact measure factual knowledge, expressive language ability, short-term memory and other skills related to learning. Hence, individuals with dyslexia, because of the nature of their difficulties, are likely to get artificially low IQ scores. According to Beech (1997) there are three tests within the WISC that overtly test short-term memory. These are digit span, mental arithmetic and coding. The digit span and arithmetic tests require the individual to retain verbal information and the coding test involves visual information. Many psychologists exclude these sub-tests when calculating the overall IQ due to the association between dyslexia and performance on memory tests. Thomson (1982) administered sub-tests from the BAS to children aged 8 to 16 with specific reading difficulties. They scored significantly less well on Speed of Information Processing, Immediate and Delayed Visual Recall, Recall of Digits, Basic Arithmetic and Word Reading compared to all other abilities. Similarly, Hooper and Hynd (1986) investigated the use of the Kaufman Assessment Battery for Children in differentiating between normal readers and matched readers with dyslexia aged 8 to 12. They found significant differences on hand movements, number recall, word order, matrix analogies, all the achievement sub-tests, sequential and achievement factors and mental processing in favour of normal readers. Whitehouse (1983) examined the relationship between performance on the WISC-R and reading ability. A group of disabled readers scored significantly lower than a matched group of normal readers on Arithmetic, Digit Span and Coding. However, not all poor coders had dyslexia, nor were all individuals with dyslexia poor coders.

Spafford (1989) found relatively low scores on Arithmetic, Coding, Information and Digit Span on the WISC test for individuals with dyslexia. This pattern is commonly known as the ACID profile. However, when boys and girls with dyslexia were compared, the boys tended to display the ACID profile whilst the girls were more likely to display an AVID profile with low scores on Arithmetic, Vocabulary, Information and Digit Span and a relatively strong performance on Coding. Digit Span sub-test scores were found to be consistently low for boys and girls with dyslexia. Turner (1997) found that girls perform better than boys on WISC Coding and Digit Span. Boys are therefore more likely to have their scores deemed exceptionally low and so to be identified as having dyslexia, girls must perform at a lower level than boys on these two tests. A solution to this problem would be the provision of separate norms for the two groups.

It is suggested that intelligence tests do, however, have a role to play in dyslexia assessment (Miles, 1996; Turner, 1997). They are useful in the identification of strengths and weaknesses, since these differences are a feature of a dyslexic profile, and in summarising the overall ability of an individual. But, the use of a global IQ score may be misleading since many of the items are not suitable for individuals with dyslexia and so do not provide an accurate measure of their abilities. An overall IQ score will also mask relevant intra-individual differences, for example, verbal and non-verbal abilities and low scores on the digit span and coding sub-tests.

According to the National Working Party on Dyslexia in Higher Education (1999), the reading difficulties evident in students with dyslexia are extracting sense from written material without substantial re-reading, slow reading speed and inaccurate reading. Padget, Knight and Sawyer (1996) argue that difficulty reading individual words and spelling are the primary symptoms of dyslexia and that the standard scores on these are expected to be at least one standard deviation lower than the IQ for a diagnosis of dyslexia. Attainment tests used in dyslexia assessment should be standardised and have proven validity and reliability. Diagnostic attainment tests are not usually restricted and so can be used by teachers.

Turner (1997) asserts that three types of reading test are important in the assessment of dyslexia. Context-free word recognition is, he argues, the main difficulty in dyslexia and so a word reading test is essential in the assessment of reading attainment. Second, the child must match written to spoken language in terms of sentence structure and must make some sense of what is read. This can be tested using cloze tests of sentence completion with multiple-choice answers or passage reading tests. Finally, reading comprehension tests on which individuals with dyslexia tend to perform better than they do on reading accuracy or reading rate. Turner continues that testing for the purpose of measuring progress following specialist teaching is desirable but difficult. He proposes that reading should be retested using the same instrument, that more than one measure of reading should be used (such as word recognition and passage reading) and that progress should be estimated termly. Tests with parallel forms and adaptive tests allow for repeated testing.

8.4 Alternative discrepancy models

A number of researchers argue that other discrepancy models are more useful in dyslexia assessment than ability versus attainment. Reid (1998) suggests that it may be possible to supplement the traditional IQ-reading discrepancy with other types of discrepancies involving aspects of phonological skills, single word reading and listening comprehension being compared to other measures, such as chronological age and reading comprehension. According to Turner (1997), alternative discrepancies often found in dyslexia assessment include those between verbal and non-verbal abilities, non-verbal and spatial abilities, performance on visuo-spatial tasks with and without a motor component, reading and spelling, decoding and comprehension, literacy and number skills, and word and non-word reading.

Using an attainment versus age discrepancy model, a child is identified as having dyslexia if his or her reading score is significantly below that which would be expected on the basis of his or her chronological age. However, according to Wright and Groner (1993) using this type of discrepancy has the tendency to identify 'garden variety' poor readers rather than children whose performance is significantly different from their ability.

Aaron and Joshi (1992) suggest that individuals with dyslexia have listening comprehension scores that are significantly higher than their reading comprehension scores. Furthermore, Reid and Weedon (1997) argue that a discrepancy between reading and listening comprehension is one of the best predictors in identifying dyslexia. Stanovich (1991a; 1991b; 1991c) proposes a discrepancy model between reading comprehension and listening comprehension as a child who does not understand spoken language could not be expected to read well. However, Turner (1997) argues that when using listening comprehension to show a discrepancy with reading comprehension, it requires 'greater failure' as the separation of reading ability and linguistic ability takes longer to establish developmentally and is more evident in children who are older.

8.5 The phonological deficit hypothesis

The report by the Working Party of the Division of Educational and Child Psychology of the British Psychological Society (1999) states that:

> phonological processing is broadly defined as the ability to process sounds in spoken language. Phonology is that part of language that concerns the sounds of words, rather than their meanings or grammatical structures... A weakness in phonological processing ... will have an impact on aspects of speaking and, through difficulties in the establishment of grapheme-phoneme (letter sound) links, on the acquisition of literacy skills (page 30).

The National Working Party on Dyslexia in Higher Education (1999) adds that inadequate phonological processing abilities affect the acquisition of phonic skills in reading and spelling so that unfamiliar words are frequently misread, which may in turn affect comprehension.

A number of researchers have examined phonological skills across a wide range of individuals with dyslexia and found a stable phonological deficit (Bruck, 1990, 1992; Frith, 1995; Padget, Knight and Sawyer, 1996; Snowling et al , 1997; Gottardo, Siegel and Stanovich, 1997). According to the British Psychological Society (1999), even when the learner's home language is not English, research has shown that phonological difficulties, as one important determinant of literacy, can be identified in the language of tuition. Phonological skills include rhyming, alliteration, rapid naming, non-word reading and phoneme deletion.

Ellis and Large (1987) found that rhyming and alliteration abilities reliably differentiated children with specific reading retardation from their better-reading peers when the groups were matched for intelligence. Singleton, Thomas and Horne (2000) suggest that studies in which groups are matched for intelligence may be criticised on the grounds that the participants are not equivalent in the amount of reading they have experienced, and this factor could account for the findings. Good readers will have had more practice in reading, which should have facilitated greater fluency in memory and phonological processes. Stanovich (1986) calls this the 'Matthew effect'. Other researchers have also identified difficulties with rhyme and / or alliteration in children with dyslexia (Snowling, 1995; Nicolson and Fawcett, 1995a; Frith, 1997; Snowling and Nation, 1997; Reid, 1998).

Furthermore, Bowers, Steffy and Tate (1988) found that when non-verbal intelligence was controlled for in a comparison between individuals with and without dyslexia, the groups were found to differ significantly on rapid naming ability. It could be argued that rapid naming ability is a function of both phonological processing and memory. Other researchers have also found that children with dyslexia frequently perform poorly in tasks requiring the rapid naming of words (Denckla and Rudel, 1976; Wolf, 1991, 1996; Nicolson and Fawcett, 1995a, 1995b; Wolf and Obregon, 1992). However, Nicolson and Fawcett's (1995a) data show a trend for the difference in naming speed between children with and without dyslexia to reduce with age.

In addition, Aaron and Joshi (1992) suggest that individuals with dyslexia have difficulty reading pronounceable non-words. Several other researchers have also suggested that non-word reading differentiates between individuals with and without dyslexia (Rack, Snowling and Olson, 1992; Turner, 1997; Snowling et al, 1997). Everatt (1997) found that adults with dyslexia perform significantly lower than adults without dyslexia on measures of single non-word naming time. He argues that rapid non-word reading may therefore be a more effective diagnostic indicator of dyslexia than naming of familiar items, the effect of which reduces with age.

Fletcher et al (1994) confirmed that poor phoneme deletion and word finding skills characterised reading disabled children. However, Turner (1997) argues that Fletcher made no distinction between low-achieving and under-achieving poor readers. Nevertheless, Snowling et al (1997) found that students with dyslexia, compared to a group of students without dyslexia, did have difficulties with tasks involving phoneme deletion and phoneme fluency. Padget, Knight and Sawyer (1996) found that when compared to typical readers, matched for age or reading level, readers with dyslexia evidence severe deficits in word analysis as well as in awareness and manipulation of phonemes. They argue that difficulty with phonological coding is a primary symptom of dyslexia and it is expected that skills on phonological awareness tasks will be well below age level.

Hanley (1997) found that students with dyslexia performed significantly lower on non-word reading and rhyming tests than students without dyslexia. 90% of the individuals with dyslexia displayed evidence of a phonological processing deficit. However, it is important that tests of dyslexia also identify those individuals who differ from the typical phonological dyslexic profile. Indeed, Rack (1997) examined a sample of adults with dyslexia and found that two-thirds showed difficulties in working memory and phonological processing but that about 20% appeared to have a different pattern of specific difficulties relating to weaknesses in visual-motor co-ordination, and half of these had additional phonological and memory problems. The National Working Party on Dyslexia in Higher Education (1999) contends that dyslexia is not simply a phonological processing difficulty and tests of phonological processing can not replace all the other diagnostic evidence that is necessary. Beaton, McDougall and Singleton (1997) and Nicolson and Fawcett (1995a) also argue that dyslexia is more than a phonological processing difficulty.

8.6 Memory difficulties

One of the most frequently cited features of dyslexia is a working or short term memory difficulty (Rack, 1994; McLoughlin, Fitzgibbon and Young, 1994; Beech, 1997; Rack, 1997). According to Gathercole and Baddeley (1993), short-term memory is the ability to retain information in short-term storage, while working memory implies that some additional processing is being carried out on the information when it is being held in short-term store. Singleton, Thomas and Horne (2000) point to an established association between reading and memory (Baddeley, 1986; Beech, 1997; Brady, 1986; Jorm, 1983; Wagner and Torgesen, 1987). It is generally suggested that phonological processes underpin the development of a phonological recoding strategy in reading and that working memory plays an important part in this, allowing constituent sounds and / or phonological codes to be stored in the short-term until they can be recognised as a word and it's meaning accessed in long-term memory (Gathercole and Baddeley, 1993; Wagner et al, 1993).

Swanson (1994) found that although short-term memory and working memory were both significant predictors of reading comprehension, working memory was the most predictive of word recognition and, in good readers, working memory was found to be more predictive than short-term memory of both word recognition and comprehension. Several studies have shown that when groups of good and poor readers that have been matched on intelligence are compared on different psychological measures, significant differences are observed in short-term memory (Ellis and Large, 1987; Bowers, Steffy and Tate, 1988; Gottardo, Siegel and Stanovich, 1997).

Thomson (1982) administered sub-tests from the BAS to children with specific reading difficulties and found that they scored significantly less well on Immediate and Delayed Visual Recall and Recall of Digits compared to other abilities. Gathercole (1995) argues that non-word repetition is more highly associated with current and later reading development than digit span.

Fein, Davenport, Yingling and Galin (1988) examined verbal and non-verbal short-term memory in children with or without dyslexia and found that those with dyslexia may have deficits in verbal and visual memory. Half of those with dyslexia were below normal on both visual and verbal memory performance, 41% were below normal in only one type of memory, 9% had normal visual and verbal memory and 16% had normal verbal memory. Indeed, the British Psychological Society (1999) indicate that it is significant that measures of visual sequential memory play no apparent part in current research literature. However, Turner (1997) purports that when a visual memory difficulty is evident in dyslexia, it is usually the linguistic component which is at fault and that the purer the measure of visual memory, the better the individual with dyslexia is at doing it. He continues that the BAS visual recall test may involve both visual and verbal memory. Nevertheless, Horne and Singleton (1997) found that university students with dyslexia scored below average on the BAS Recall of Designs, which is a much purer measure of visual memory than the Visual Recall test. The Visual Recall test involves studying a card with 20 pictures on for two minutes and then verbally recalling as many as possible. The Recall of Designs test involves studying a series of simple abstract line drawings for five seconds each and then drawing each one from memory immediately after it has been displayed.

Beech (1997) reports on the difficulty of administering digit span tests as there is such a variability between testers in the rate and quality of articulation. A solution to this problem would be the use of a computerised digit span test which would administer the test in exactly the same way to each individual.

8.7 The challenges facing dyslexia assessment

Sutherland and Smith (1991) examined three established dyslexia screening tests with first year secondary school students, namely, the Boder Test of Reading and Spelling Patterns, the Aston Index and the Bangor Dyslexia Test. The Boder Test of Reading and Spelling Patterns (Boder and Jarrico, 1982) aims to differentiate dyslexia from non-specific reading disorders and classifies readers with dyslexia into one of three subtypes. The Aston Index, which was developed at Aston University, claims to identify children with potential language-associated problems early in their education and to analyse and diagnose reading and language difficulties displayed by children who are experiencing difficulties in coping with basic attainments. The Aston Index has been criticised for weaknesses in its construction and standardisation and Pumfrey and Reason (1991) doubt its potential to discriminate between groups of children experiencing specific learning difficulties. The Bangor Dyslexia Test (Miles, 1982 / 1997) consists of ten sub-tests including left / right confusion, repeating polysyllabic words, subtraction, multiplication tables, saying the months of the year forwards and in reverse order, digits forward and reversed, b-d confusion and familial incidence.

However, it does not give a definitive diagnosis and is not fully standardised. Sutherland and Smith found that a number of sub-tests on the Aston Index had ceiling effects, the results were difficult to interpret and it was very time-consuming. The Bangor Dyslexia Test was found to be too general to benefit classroom teachers. There are many dyslexia checklists available for teachers to use but they cannot give a diagnosis of dyslexia and are of limited value. Sutherland and Smith have indicated that teachers face some difficulty in finding tests that are well standardised, that can discriminate between groups of children with specific learning difficulties, that are easy to interpret and give a definitive diagnosis, that are appropriate to the age range (do not show floor and ceiling effects) and that are not too time-consuming but still give enough detailed information.

As teachers are restricted in their use of many tests, educational psychologists are required to administer complete dyslexia assessments. However, many pupils with dyslexia are not referred to an educational psychologist due to financial reasons or because their difficulties have simply not been noticed. There is also concern that more boys than girls are referred to educational psychologists for assessment. A screening test that can be used by teachers to assess all pupils in a school should prevent individuals with dyslexia from being missed and would give teachers evidence that there is a need for a pupil to be referred for an assessment with an educational psychologist. It should also prohibit the possibility of referral bias.

Another challenge is that there is little agreement on the actual definition of dyslexia or on the composition of a dyslexia assessment. However, the National Working Party on Dyslexia in Higher Education (1999) suggests that, in practice, assessors should consider evidence of an ability versus attainment discrepancy as a necessary but not sufficient condition for a diagnosis, and should seek further evidence of dyslexia, for example, cognitive impairments in phonological processing and / or short-term or working memory.

8.8 Summary and conclusions

Although there is no one definition of developmental dyslexia, it is generally considered to be a syndrome that can affect written language, number skills, memory, phonological skills, organisational skills and motor function. However, not all individuals with dyslexia will exhibit the same range or severity of symptoms. The main purposes of dyslexia assessment are to identify students' difficulties and strengths and look for discrepancies between different abilities that may be indicative of dyslexia.

The traditional method of identifying dyslexia has been to look for a significant discrepancy between intelligence and attainment in literacy. However, a number of researchers have opposed the use of intelligence tests in the diagnosis of dyslexia for several reasons. Poor readers with average or high IQ do not necessarily constitute a different group from 'garden variety' readers. Children who have missed a lot of school, not been taught properly or speak English as a second language would be seen to underachieve in literacy in relation to their ability. Also, children with a high IQ and slightly above average reading skills could still be described as reading disabled. A significant discrepancy between ability and literacy would not be identified at an early age and intelligence tests are generally restricted to use by psychologists. Finally, some IQ measures include tests of working memory, on which individuals with dyslexia tend to score poorly; this can depress the IQ score and result in misleading findings. However, other researchers defend the role of intelligence tests in dyslexia assessment as they identify strengths and weaknesses, which may indicate a dyslexic profile, and summarise the overall ability of an individual.

A number of alternative discrepancy models have been suggested, including the comparison of listening and reading comprehension, verbal and non-verbal abilities, non-verbal and spatial abilities, reading and spelling, word and non-word reading and age and attainment.

The phonological deficit hypothesis has been put forward by a number of researchers who have found that individuals with dyslexia display a stable phonological deficit. This may involve difficulties in rhyming, alliteration, rapid naming ability, non-word reading, phoneme deletion and phoneme fluency. However, it has been shown that not all individuals with dyslexia have a phonological processing deficit and that evidence of such a deficit is not, by itself, sufficient for a diagnosis of dyslexia.

One of the most common features of dyslexia is a difficulty in working or short term memory. This is most often evident in auditory memory, although a number of individuals with dyslexia display difficulties in visual memory or in both auditory and visual memory.

There is therefore a need for a test that can be used by teachers, rather than just psychologists, to screen pupils for dyslexia. This test should not only show any discrepancy between ability and literacy attainment but should also give an indication of students abilities in phonological processing and verbal and visual memory. The test should be set at a level such that there are no floor or ceiling effects for the age group. If there are significant gender differences then separate norm tables for both sexes should be given. The test must be standardised and meet satisfactory psychometric criteria for validity and reliability.

9 Methodology for the secondary assessment studies

9.1 Overview

The secondary assessment research encompasses a number of phases. First, the test development phase, during which test items were developed and tested nationally and the results analysed. Second, the standardisation stage that involved administering the final computerised forms of the tests to a different national sample of children. Finally, reliability and validity tests were conducted using two more samples of children on a national basis.

9.2 Test development - Spring 1998

9.2.1 Participants

The 2366 participants (1302 boys and 1064 girls) in this study attended 28 different secondary schools in England, Scotland and Northern Ireland. These students' ages ranged from 11 years 0 months to 15 years 11 months, with a mean of 13 years and 3 months and a standard deviation of 21.6 months.

The numbers of participants in each test are shown in Table 98.

Table 98: Number of participants for each test

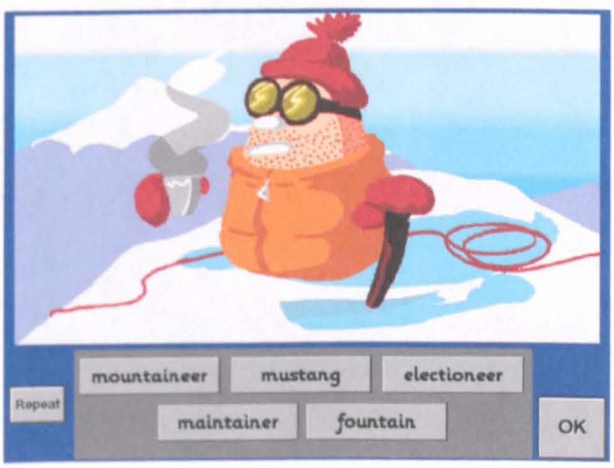
Test	N	Boys	Girls
Single word reading	2100	1166	934
Sentence reading	2234	1232	1002
Spelling	2210	1218	992
Reasoning	730	389	341
Visual memory	79	46	33
Auditory memory	80	47	33
Non-word reading	80	47	33
Syllable segmentation	80	47	33

The students were selected at random from Years 7 to 11. All students were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

9.2.2 Procedure

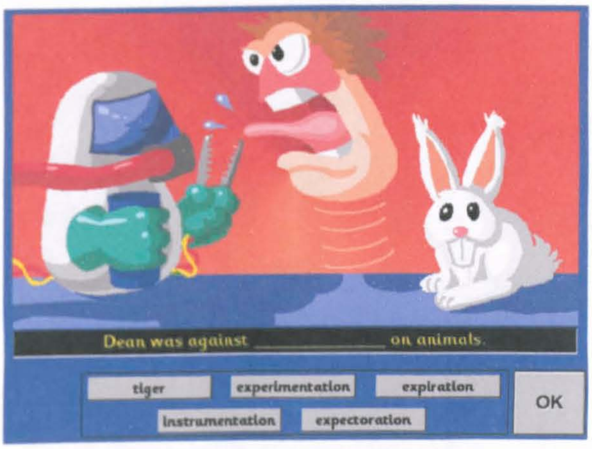
LASS Secondary comprises eight modules: single word reading; sentence reading; spelling; reasoning; visual memory; auditory memory; non-word reading; and, syllable segmentation. For each of these modules, various numbers of items were devised which were administered to the sample. Examples of the items are given below (See Appendix 9 for a list of all items).

Figure 23: LASS Secondary - Single word reading test



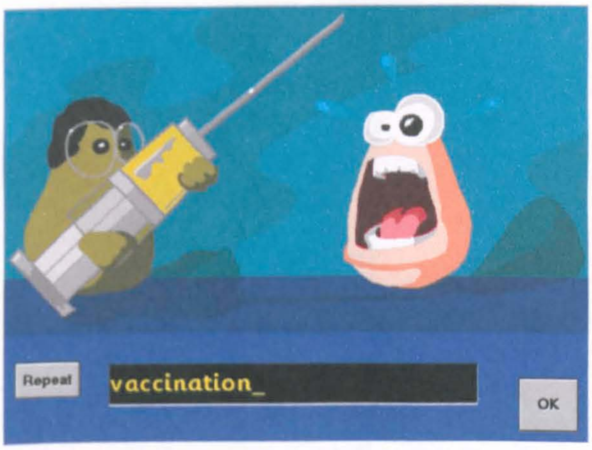
Click on the word which says 'mountaineer'. Then click on OK to move on.

Figure 24: LASS Secondary - Sentence reading test



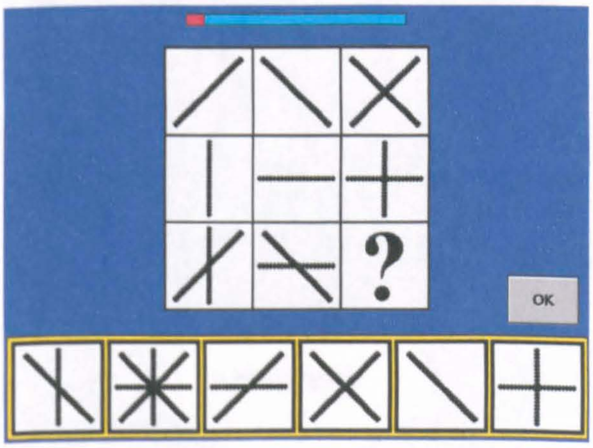
Click on the correct word to complete the sentence. Then click on OK to move on.

Figure 25: LASS Secondary - Spelling test



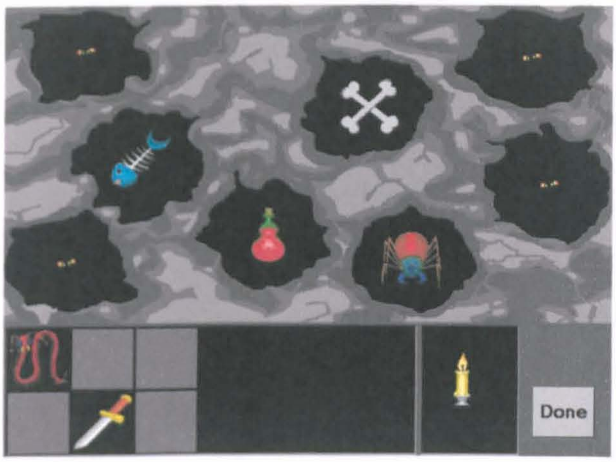
Tim needed a vaccination before his holiday. Spell the word vaccination.

Figure 26: LASS Secondary - Reasoning test



Look carefully. One picture is missing from the grid. Click on the correct picture which fits in the grid. Choose from the six pictures below. If you change your mind, just click on a different one. When you've finished, click on OK to move on.

Figure 27: LASS Secondary - Visual memory test

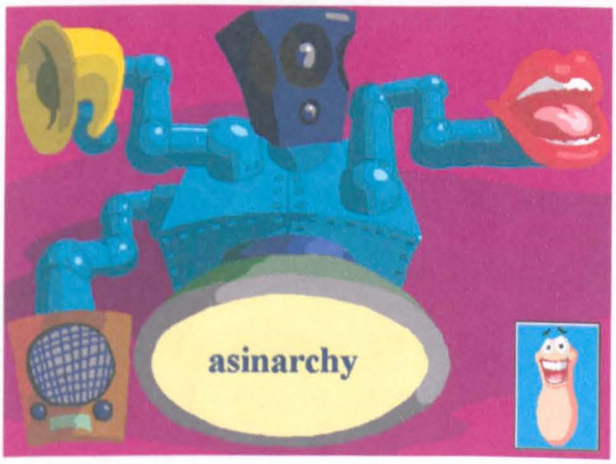


See the phantoms appear...now move the phantoms back to where you saw them.

Figure 28: LASS Secondary - Auditory memory test

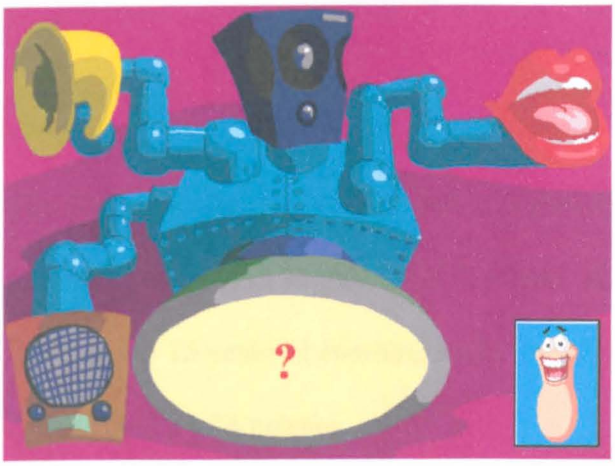
Remember this telephone number and enter the number you hear...53729.

Figure 29: LASS Secondary - Non-word reading test



Click on the speaker which you think has spoken the word correctly. Then click on Nak to move on.

Figure 30: LASS Secondary - Syllable segmentation test



Dictionary without 'tion' is?

Single word reading comprised 30 items; Sentence reading comprised 70 items; Spelling comprised 100 items; and, Reasoning comprised 60 items. These four modules were administered to pupils in a pen and paper format.

Visual memory and Auditory memory were administered as computerised quasi-adaptive tests. The Visual memory module involves remembering where different phantoms appear in a cave, starting with three phantoms and moving up one level after each correct answer until the student has completed 12 iterations. The Auditory memory module involves remembering a phone number, starting with two attempts at two digits, two attempts at three digits and so on until the student fails both attempts at one level.

Non-word reading comprised 20 items; and Syllable segmentation comprised 33 items. These two modules were administered to pupils in a computerised format and pupils were required to attempt all items in these tests.

9.3 Standardisation project - Autumn 1998

9.3.1 Participants

The 505 participants (300 boys and 205 girls) in this study attended 14 different secondary schools in England and Scotland. These students' ages ranged from 11 years 0 months to 15 years 11 months, with a mean of 13 years and 2 months and a standard deviation of 14.34 months.

The numbers of participants in each test are shown in Table 99.

Table 99: Number of participants for each test

Test	N	Boys	Girls
Sentence reading	487	285	202
Spelling	487	285	202
Reasoning	476	279	197
Visual memory	498	296	202
Auditory memory	489	292	197
Non-word reading	485	288	197
Syllable segmentation	476	280	196

9.3.2 Instruments

LASS Secondary (Horne, Singleton and Thomas, 1999) is a fully computerised test consisting of eight modules: single word reading; sentence reading; spelling; reasoning; visual memory; auditory memory; non-word reading; and, syllable segmentation. Each module takes five to ten minutes to complete.

In the single word reading test, students are presented with a picture of an object on the screen and hear the word spoken by the computer. Students select the correct word from five words at the bottom of the screen. The items are in order of difficulty with the easiest items first. Students complete all 30 items.

The sentence reading test is an adaptive test that involves finding the missing word in a sentence. Students are presented with a sentence that has one word missing and a picture to go with the sentence. Students select the correct word from five words at the bottom of the screen. Items are selected from a bank of 70 items. This involves administering five probe items of increasing difficulty in order to start the test at an appropriate level. As soon as a student fails a probe item the test begins at the item following the last correctly answered probe. Progress through the test is dependent on the students' responses. When a student answers three consecutive items correctly, they jump forward three items, and when a student answers three consecutive items incorrectly, they jump back nine items. The tests are discontinued when a student fails four out of six items.

Spelling is an adaptive test that involves spelling single words. Students are presented with a picture on the screen and hear a word and a sentence putting the word into context. Students spell the word using keyboard entry. Items are selected, from a bank of 98 items, on the basis of the adaptive algorithm described in the sentence reading test.

Reasoning is an adaptive test involving matrix puzzles that can be solved by application of logical reasoning, using both visual and verbal strategies. Students are shown a 3 x 3 matrix with the bottom right hand square empty. Students choose which of six squares at the bottom of the screen completes the pattern. Items are selected, from a bank of 58 items, on the basis of the adaptive algorithm described in the sentence reading test.

The visual spatial memory test is set in a cave with eight hollows in the wall. Different 'phantoms' (e.g. ghost, gremlin, bat etc) appear in different hollows one at a time and then disappear. The student must remember which phantom went in which hollow. After the phantoms have disappeared these are shown on the bottom of the screen along with two distracters. The students must select the phantoms that were originally presented and place them in the correct hollow. Students start with a presentation of three phantoms and complete twelve trials in total. When a student has correctly placed three phantoms they move on to four phantoms and so on. The maximum number of phantoms that can be presented is eight.

The auditory memory test involves digit span. Students are given a telephone number to remember which they then enter onto a representation of a mobile phone using the mouse. Students must get both practice items (three digit numbers) correct before moving on to the test items. Students start with trials of three digit numbers and if they answer one or both correctly then they move on to two trials of four digit numbers and so on up to nine digits. If a student fails both trials on a level then the test is discontinued.

The non-word reading test is a test of phonic decoding skills. Students are presented with 25 non-words in order of item difficulty. The non-word is presented visually on the screen of 'Nak's sound machine'. The machine plays four different versions of the word, which the student can hear as many times as they want to before selecting the version that they think is correct.

The syllable segmentation test is a test of syllable and phoneme deletion that identifies poor phonological processing ability. Students are presented with 32 words in order of item difficulty and asked what the word would sound like if part of the word is removed. The 'Nak's sound machine' plays four different answers that the students can hear as many times as they want to before selecting the answer that they think is correct.

9.3.3 Procedure

The eight LASS Secondary modules were administered individually in a random order to each pupil in a quiet environment.

9.4 Reliability project - Spring 2001

9.4.1 Participants

The participants in this study attended seven different secondary schools in England and Scotland. One of the schools was located in an inner city area, five were located in a suburban area and one school was in a rural area. The inner city school had a high proportion of students of families of the unemployed. One of the suburban schools had mostly students of manual workers and the unemployed, two suburban schools had a high proportion of students from skilled and professional families, and two consisted mostly of students from professional families. The rural school had a high proportion of students from professional families.

The inner city school had 1315 pupils in the whole school, 195 per year group and 144 on the SEN register. The suburban schools had an average of 792 pupils in the whole school, 95 per year group and 101 on the SEN register. The rural school had 766 in the whole school, 30 per year group and 30 on the SEN register.

101 students (55 boys and 46 girls) participated in the study. Their ages ranged from 11 years 6 months to 15 years 11 months, with a mean of 13 years 8 months and a standard deviation of 16.49 months.

15 students from each school were selected randomly from the Year 7 to Year 11 registers. All students were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

9.4.2 Instruments

LASS Secondary (Horne, Singleton and Thomas, 1999) is a fully computerised test consisting of eight modules: single word reading; sentence reading; spelling; reasoning; visual memory; auditory memory; non-word reading; and, syllable segmentation. Each module takes five to ten minutes to complete. For further information, see section 9.3.2.

9.4.3 Procedure

The eight LASS Secondary modules were administered in a specified order, with each school completing the tests in a different order. Pupils were re-tested on the LASS Secondary modules, in the same order, four weeks later.

9.5 Validity project - Spring 2001

9.5.1 Participants

The participants in this study attended five different secondary schools in England and Wales. The schools in this study are not the same schools that were involved in the reliability study. Two of the schools were located in an inner city area, two were located in a suburban area and one school was in a rural area. One of the inner city schools had a high proportion of students of families of the unemployed and manual workers whilst the other inner city school consisted mostly of students from families of manual and skilled workers. Both of the suburban schools had a high proportion of students from skilled and professional families. The rural school consisted of students from a very mixed range of backgrounds.

The inner city schools had an average of 865 pupils in the whole school, 165 per year group and 222 on the SEN register. The suburban schools had an average of 982 pupils in the whole school, 176 per year group and 102 on the SEN register. The rural schools had an average of 687 in the whole school, 115 per year group and 161 on the SEN register.

75 students (47 boys and 28 girls) participated in the study. Their ages ranged from 11 years 6 months to 15 years 11 months, with a mean of 13 years 6 months and a standard deviation of 17.01 months.

15 students from each school were selected randomly from the Year 7 to Year 11 registers. All students were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

9.5.2 Instruments

LASS Secondary (Horne, Singleton and Thomas, 1999) is a fully computerised test consisting of eight modules: single word reading; sentence reading; spelling; reasoning; visual memory; auditory memory; non-word reading; and, syllable segmentation. Each module takes five to ten minutes to complete. For further information, see section 9.3.2.

The British Ability Scales (Second Edition) Word Reading Test (Elliott, 1996) is a conventional, standardised, single-word, out-of-context oral reading test that comprises 60 items in increasing order of difficulty. The age range of the test is 5 years 0 months to 17 years 11 months.

The NFER-Nelson Group Reading Sentence Completion Test (Macmillan Test Unit, 1990) is a standardised, contextual reading test that comprises 45 items in increasing order of difficulty. The age range of the test is 8 years 2 months to 15 years 4 months.

The British Spelling Test Series Level 3 (Vincent and Crumpler, 1997) is a 68 item standardised, contextual spelling test involving single word spelling, cloze passages, dictation passages and correction of errors. The age range of the test is 9 years 11 months to 15 years 1 month.

The Matrix Analogies Test - Short Form (Naglieri, 1985) is a test of non-verbal ability involving 34 colour matrix problems in increasing order of difficulty. The age range of the test is 5 years 0 months to 17 years 11 months.

The Wechsler Memory Scale-III Spatial Span Test (Wechsler, 1997) involves the examiner pointing to a series of blocks and the child repeating the sequence in the same order or the reverse order. The test starts with a sequence of two blocks and goes up to a maximum of nine blocks, with two trials at each level.

The Wechsler Intelligence Scale for Children-III Digit Span Test (Wechsler, 1991) requires the child to repeat a sequence of numbers in either the same order or the reverse order. The test starts with a sequence of two digits and goes up to a maximum of nine digits forward and eight digits backward, with two trials at each level. The age range of the test is 6 years 0 months to 16 years 11 months.

The Phonological Assessment Battery (Frederickson, Frith and Reason, 1997) is designed to assess phonological processing and comprises nine tests. The Alliteration Test assesses ability to isolate initial sounds in single syllable words and comprises 10 items. In the Picture Naming Test, the child is shown a card with 50 randomly presented pictures of five objects and given 30 seconds to name as many as they can in sequence. In the Digit Naming Test, the child is shown a card with 50 randomly presented digits and given 30 seconds to name as many as they can in sequence. The Rhyme Test assesses ability to identify rhyme in single syllable words and consists of 21 items. The Spoonerisms Test, comprising 20 items, assesses whether students can segment single syllable words and then synthesize the segments to provide new words. The Semantic Fluency Test requires the child to name as many words within a particular category (e.g. names of animals) within 30 seconds. The Alliteration Fluency Test requires the child to name as many words within a particular category (e.g. words beginning with 'm') within 30 seconds. The Rhyme Fluency Test requires the child to name as many words within a particular category (e.g. words that rhyme with 'bat') within 30 seconds. The Non-Word Reading Test assesses the decoding of letter strings and comprises 20 items.

9.5.3 Procedure

Students completed the eight LASS Secondary modules in a specified order, with each school completing the tests in a different order, and the equivalent validation tests, in the same order, with a four week gap in between. Half the schools completed the LASS Secondary tests first and half completed the validation tests first.

9.6 Dyslexia project - Spring 2001

9.6.1 Participants

The participants in this study attended the twelve different secondary schools in England, Wales and Scotland that were involved in the reliability and validity projects. Three of the schools were located in an inner city area, seven were located in a suburban area and two schools were in a rural area. Two of the inner city schools had a high proportion of students of families of manual workers and the unemployed whilst the other inner city school consisted mostly of students from families of manual and skilled workers. One of the suburban schools had mostly students of manual workers and the unemployed, four suburban schools had a high proportion of students from skilled and professional families, and two schools consisted mostly of students from professional families. One of the rural schools had a high proportion of students from professional families and the other rural school consisted of students from a very mixed range of backgrounds.

The inner city schools had an average of 1015 pupils in the whole school, 175 per year group and 196 on the SEN register. The suburban schools had an average of 858 pupils in the whole school, 92 per year group and 87 on the SEN register. The rural schools had an average of 1013 in the whole school, 164 per year group and 249 on the SEN register.

176 students (102 boys and 74 girls) participated in the study. Their ages ranged from 11 years 6 months to 15 years 11 months, with a mean of 13 years 7 months and a standard deviation of 16.70 months.

The group of students with dyslexia consisted of 30 students (21 boys and 9 girls). Their ages ranged from 11 years 6 months to 15 years 11 months, with a mean of 13 years 7 months and a standard deviation of 17.36 months.

The group of students with other special educational needs consisted of 17 students (11 boys and 6 girls). Their ages ranged from 11 years 7 months to 15 years 8 months, with a mean of 13 years 8 months and a standard deviation of 17.36 months.

The group of students not identified as having any special educational needs consisted of 129 students (76 boys and 59 girls). Their ages ranged from 11 years 6 months to 15 years 11 months, with a mean of 13 years 7 months and a standard deviation of 16.58 months.

All students were given the option of declining to participate and of withdrawing from the study at any time they wished, but none did so.

9.6.2 Instruments

The following tests were used (for further information see section 9.5.2):

- LASS Secondary (Horne, Singleton and Thomas, 1999)
- NFER-Nelson Group Reading Sentence Completion Test (Macmillan Test Unit, 1990)
- British Spelling Test Series Level 3 (Vincent and Crumpler, 1997)
- Matrix Analogies Test - Short Form (Naglieri, 1985)
- Wechsler Memory Scale-III Spatial Span Test (Wechsler, 1997)
- Wechsler Intelligence Scale for Children-III Digit Span Test (Wechsler, 1991)
- Phonological Assessment Battery (Frederickson, Frith and Reason, 1997)

9.6.3 Procedure

The eight LASS Secondary modules were administered individually in a random order to each pupil in a quiet environment. The six conventional tests were administered to a subset of pupils either individually or as a group.

9.7 Questionnaire study - Autumn 1998 - Spring 2001

9.7.1 Participants

Questionnaires were sent to schools involved in the LASS Secondary standardisation in October 1998 and the reliability and validity projects in March 2001.

26 teachers using LASS Secondary, in England, Wales and Scotland, returned questionnaires (14 in October 1998 and 12 in March 2001).

9.7.2 Instruments

The LASS Secondary Questionnaire (October 1998) consisted of 18 questions covering installation, registration, the tests, the reports and the system overall.

The questionnaire was revised in March 2001 to include separate questions on each of the individual modules. It now consists of 23 questions (the questionnaire is shown in full in Appendix 10).

9.7.3 Procedure

Questionnaires were sent to two samples of teachers, using LASS Secondary, on different occasions (October 1998 and March 2001).

10 Results for the secondary assessment studies

10.1 Test development - Spring 1998

The means, standard deviations and ranges for each of the eight LASS Secondary modules are given in Table 100.

Table 100: Descriptive statistics for the LASS Secondary modules

	N	Minimum	Maximum	Mean	Standard Deviation
Single word reading	2100	6	30	29.46	1.79
Sentence reading	2234	1	70	55.03	14.53
Spelling	2210	11	100	62.76	19.25
Reasoning	730	3	56	34.96	10.05
Visual memory	79	0	4	2.41	0.93
Auditory memory	80	2	13	6.89	2.07
Non-word reading	80	2	20	13.81	4.46
Syllable segmentation	80	11	31	22.93	4.86

Analysis by gender revealed significant differences (see Table 101) on three of the LASS Secondary modules (single word reading, sentence reading and spelling).

Table 101: Gender differences on the LASS Secondary modules

	Gender	N	Mean	SD	t	df	Significance
Single word reading	Female	934	29.55	1.56	2.05	2098	$p<0.05$
	Male	1166	29.39	1.94			
Sentence reading	Female	1002	55.92	13.93	2.60	2232	$p<0.01$
	Male	1232	54.31	14.96			
Spelling	Female	992	65.14	18.78	5.28	2208	$p<0.001$
	Male	1218	60.82	19.42			
Reasoning	Female	341	35.45	9.76	1.24	728	NS
	Male	389	34.53	10.29			
Visual memory	Female	33	2.18	1.01	1.84	77	NS
	Male	46	2.57	0.83			
Auditory memory	Female	33	7.42	2.24	1.98	78	NS
	Male	47	6.51	1.88			
Non-word reading	Female	33	14.39	4.87	0.98	78	NS
	Male	47	13.40	4.15			
Syllable segmentation	Female	33	24.06	5.00	1.78	78	NS
	Male	47	22.13	4.65			

Internal consistency was calculated for seven of the eight modules using the Kuder-Richardson formula. The coefficients range from 0.71 to 0.97 (see Table 102). The visual memory module was not included in this analysis as the items administered varied between children depending on their level of ability.

Table 102: Kuder-Richardson coefficients for the LASS Secondary modules

Module	Kuder-Richardson
Single word reading	0.87
Sentence reading	0.97
Spelling	0.97
Reasoning	0.91
Auditory memory	0.71
Non-word reading	0.84
Syllable segmentation	0.81

Pass rates for each item were calculated. Those items with satisfactory reliability and appropriate difficulty for the age ranges were included within the final suite. Five items were removed (Spelling, 2 items; Reasoning, 2 items; and, Syllable Segmentation, 1 item). Five more difficult items were added to the Non-word Reading test.

In order to reduce the time taken for assessment in three of the modules (Sentence Reading, Spelling and Reasoning), experimentation with various adaptive models was carried out, checking the output of each model against the real data that had been obtained from the students. The most suitable algorithm (i.e. one that administered a smaller number of items but correlated highly with the full form) involved administering five probe items of sharply increasing difficulty, in order to start the test at an appropriate level. As soon as a student fails a probe item the test begins at the item following the last correctly answered probe. Progress through the test is dependent on the students' responses. When a student answers three consecutive items correctly, they jump forward three items, and when a student answers three consecutive items incorrectly, they jump back nine items. The tests are discontinued when a student fails four items out of any block of six consecutive items.

Other algorithms that were tested but rejected involved the use of: age-related start points rather than probe items; jumps forward of 4 or 5 items after answering 4 or 5 items consecutively; jumps back of 7, 8 or 10 after answering 4 or 5 items incorrectly; and, discontinuation after 4, 5 or 6 consecutive failures or 5 out of 6 or 6 out of 8 failures.

The validity of the adaptive algorithm was investigated by simulating the adaptive form of the three LASS modules with data that had been collected from the students in this study using the full form of the test. The two measures (original raw scores and adaptive form score) were correlated using the Pearson Product Moment Correlation, with significant positive correlations being obtained (see Table 103).

Table 103: Correlations between full forms and adaptive versions of the three adaptive LASS Secondary modules

	Correlation	Significance	N
Sentence Reading	0.92	$p<0.001$	2234
Spelling	0.93	$p<0.001$	2210
Reasoning	0.80	$p<0.001$	730

10.2 Standardisation project - Autumn 1998

Single word reading was not included in the standardisation project due to the ceiling effect found during the test development stage. However, after consultation with teachers involved in the test development and standardisation stages, single word reading was included in the final version of LASS Secondary. It was argued, particularly by teachers in special schools, that the single word reading test was useful in assessing the reading abilities of students who performed poorly on the sentence reading test. Data collected during the test development stage were used to produce norms for the single word reading test. The means, standard deviations and ranges for the scores of the other seven LASS Secondary modules are given in Table 104. It should be noted that the scores for the sentence reading, spelling and reasoning tests are an average of the pass rates of the hardest three items that the student passes in that test. Therefore, a lower average pass rate indicates harder items and therefore is a higher score.

Table 104: Descriptive statistics for the LASS Secondary modules

	Age	N	Minimum	Maximum	Mean	SD
Sentence reading	11	96	0.345	0.971	0.75	0.19
Sentence reading	12	150	0.345	0.972	0.70	0.22
Sentence reading	13	115	0.345	0.971	0.68	0.22
Sentence reading	14	85	0.345	0.974	0.64	0.22
Sentence reading	15	38	0.345	0.968	0.62	0.22
Sentence reading	All	487	0.345	0.974	0.68	0.22
Spelling	11	96	0.111	0.998	0.74	0.24
Spelling	12	150	0.071	0.998	0.68	0.27
Spelling	13	115	0.097	0.994	0.66	0.27
Spelling	14	85	0.093	0.990	0.59	0.26
Spelling	15	38	0.136	0.992	0.57	0.28
Spelling	All	487	0.071	0.998	0.65	0.27
Reasoning	11	91	0.332	0.959	0.71	0.17
Reasoning	12	147	0.209	0.959	0.69	0.19
Reasoning	13	114	0.219	0.959	0.68	0.19
Reasoning	14	84	0.209	0.966	0.64	0.20
Reasoning	15	38	0.206	0.959	0.62	0.21
Reasoning	All	476	0.206	0.966	0.66	0.19
Visual memory	11	101	11	43	27.75	6.78
Visual memory	12	148	5	55	29.22	8.13
Visual memory	13	117	10	54	29.25	8.86
Visual memory	14	88	1	53	29.27	9.60
Visual memory	15	39	12	44	30.54	8.35
Visual memory	All	498	1	55	29.06	8.36
Auditory memory	11	97	1	9	4.71	1.68
Auditory memory	12	149	1	10	5.40	2.02
Auditory memory	13	114	1	11	5.45	2.11
Auditory memory	14	86	2	10	5.72	1.97
Auditory memory	15	39	2	9	5.73	1.67
Auditory memory	All	489	1	13	5.39	2.01

	Age	N	Minimum	Maximum	Mean	SD
Non-word reading	11	97	2	24	11.28	5.34
Non-word reading	12	145	2	24	12.41	5.61
Non-word reading	13	115	2	23	12.63	5.83
Non-word reading	14	86	3	24	13.57	5.08
Non-word reading	15	38	4	23	14.08	5.47
Non-word reading	All	485	2	24	12.90	5.57
Syllable segmentation	11	95	6	31	17.39	6.26
Syllable segmentation	12	146	1	30	18.86	6.59
Syllable segmentation	13	112	5	32	18.95	6.93
Syllable segmentation	14	83	3	31	19.90	7.07
Syllable segmentation	15	37	4	29	19.98	6.75
Syllable segmentation	All	476	1	32	18.88	6.73

Analysis by gender revealed significant differences in two of the modules, with girls scoring higher than boys in spelling and boys scoring higher than girls in the visual memory module (see Table 105).

Table 105: Gender differences on the LASS Secondary modules

	Gender	N	Mean	SD	t	df	Significance
Sentence reading	Female	202	0.6641	0.21	1.04	485	NS
	Male	285	0.6850	0.22			
Spelling	Female	202	0.6019	0.27	3.57	485	p<0.001
	Male	285	0.6883	0.26			
Reasoning	Female	197	0.6801	0.18	1.47	474	NS
	Male	279	0.6539	0.20			
Visual memory	Female	202	27.82	8.54	2.74	496	p<0.01
	Male	296	29.90	8.15			
Auditory memory	Female	197	5.37	2.15	0.21	487	NS
	Male	292	5.40	1.91			
Non-word reading	Female	197	13.28	5.46	1.25	483	NS
	Male	288	12.64	5.63			
Syllable segmentation	Female	196	18.72	7.08	0.43	474	NS
	Male	280	18.99	6.48			

Percentile scores were calculated for each age group on each module. These scores are shown in Appendix 11.

10.3 Reliability project - Spring 2001

The median centile scores and ranges for seven of the eight LASS Secondary modules are given in Table 106. Only pupils scoring below the 20th centile on LASS sentence reading were required to complete the single word reading test. As only one pupil fulfilled this criteria, the results of the single word reading test have not been included in the following analyses.

Table 106: Descriptive statistics (centile scores) for the LASS Secondary modules

	Session	N	Minimum	Maximum	Median
Sentence Reading	1	100	15	99	80
	2	98	17	99	78
Spelling	1	99	22	99	80
	2	99	21	99	83
Reasoning	1	101	8	99	78
	2	99	2	99	81
Visual Memory	1	99	1	99	70
	2	99	1	99	66
Auditory Memory	1	100	1	99	80
	2	100	1	99	73
Non-word Reading	1	98	1	99	83
	2	97	1	99	87
Syllable Segmentation	1	98	1	99	76
	2	97	3	99	81

Analysis by gender revealed no significant differences on any of the eight modules for the first testing session (see Table 107).

Table 107: Gender differences on the eight LASS Secondary modules (first testing session z-scores)

	Gender	N	Mean	SD	t	df	Significance
Sentence Reading	Female	45	1.00	0.94	0.12	98	NS
	Male	55	0.98	0.92			
Spelling	Female	45	1.01	0.83	0.83	97	NS
	Male	54	0.87	0.86			
Reasoning	Female	46	0.69	0.71	0.16	99	NS
	Male	55	0.71	0.76			
Visual Memory	Female	45	0.42	0.78	0.35	97	NS
	Male	54	0.49	0.96			
Auditory Memory	Female	46	0.74	1.03	1.20	98	NS
	Male	54	0.48	1.09			
Non-word Reading	Female	44	0.95	0.90	1.55	96	NS
	Male	54	0.64	1.07			
Syllable Segmentation	Female	44	0.61	0.92	0.31	96	NS
	Male	54	0.66	0.83			

Analysis by gender also revealed no significant differences on any of the eight modules for the second testing session (see Table 108).

Table 108: Gender differences on the eight LASS Secondary modules (second testing session z-scores)

	Gender	N	Mean	SD	t	df	Significance
Sentence Reading	Female	43	0.99	0.93	0.21	96	NS
	Male	55	1.03	0.93			
Spelling	Female	44	1.01	0.78	0.97	97	NS
	Male	55	0.86	0.79			
Reasoning	Female	44	0.74	0.81	0.12	97	NS
	Male	55	0.76	0.82			
Visual Memory	Female	44	0.58	0.90	1.31	97	NS
	Male	55	0.29	1.24			
Auditory Memory	Female	45	0.64	1.19	0.66	98	NS
	Male	55	0.49	1.06			
Non-word Reading	Female	42	1.16	0.84	1.50	95	NS
	Male	55	0.87	1.02			
Syllable Segmentation	Female	42	0.81	0.92	0.33	95	NS
	Male	55	0.75	0.78			

The 101 pupils in the sample came from five ethnic backgrounds (see Table 109). However, as the majority of pupils were categorised as 'white', with only one or two students in each of the other categories, analysis by ethnic background was not feasible.

Table 109: Ethnic background of children

Ethnic background	N	Percent
White	96	95
Black Caribbean	2	2
Indian	1	1
Chinese	1	1
Other	1	1

Analysis comparing 'white' children to all other children revealed no significant differences on the eight modules for the first testing session. One of the eight modules (reasoning) showed a significant ethnic difference on the second testing session (see Table 110), with white students scoring less well than students from other ethnic backgrounds.

Table 110: Ethnic differences on the eight LASS Secondary modules (z-scores)

	Ethnicity	N	Mean	SD	t	df	Significance
Sentence reading 1	White	95	1.01	0.93	1.11	98	NS
	Other	5	0.55	0.49			
Spelling 1	White	94	0.92	0.84	0.89	97	NS
	Other	5	1.26	1.02			
Reasoning 1	White	96	0.68	0.74	1.44	99	NS
	Other	5	1.16	0.48			
Visual memory 1	White	94	0.44	0.89	1.12	97	NS
	Other	5	0.89	0.54			
Auditory memory 1	White	95	0.60	1.09	0.03	98	NS
	Other	5	0.61	0.57			
Non-word reading 1	White	93	0.79	1.03	0.50	96	NS
	Other	5	0.56	0.37			
Syllable segmentation 1	White	93	0.65	0.89	0.41	96	NS
	Other	5	0.48	0.53			
Sentence reading 2	White	94	1.02	0.93	0.17	96	NS
	Other	4	0.93	1.01			
Spelling 2	White	94	0.91	0.79	0.92	97	NS
	Other	5	1.24	0.78			
Reasoning 2	White	95	0.71	0.81	2.35	97	$p<0.05$
	Other	4	1.67	0.26			
Visual memory 2	White	94	0.39	1.12	0.98	97	NS
	Other	5	0.89	0.70			
Auditory memory 2	White	95	0.55	1.11	0.13	98	NS
	Other	5	0.62	1.43			
Non-word reading 2	White	93	1.00	0.96	0.38	95	NS
	Other	4	0.82	1.00			
Syllable segmentation 2	White	93	0.79	0.86	0.68	95	NS
	Other	4	0.49	0.34			

Analysis comparing children with English as a first language with children with English as a second language revealed no significant differences on the eight modules on either of the testing sessions (see Table 111).

Table 111: Language differences on the eight LASS Secondary modules (z-scores)

	Language	N	Mean	SD	t	df	Significance
Sentence reading 1	E1L	96	1.01	0.93	1.14	98	NS
	E2L	4	0.48	0.54			
Spelling 1	E1L	95	0.93	0.84	0.15	97	NS
	E2L	4	1.00	0.96			
Reasoning 1	E1L	97	0.69	0.74	0.94	99	NS
	E2L	4	1.04	0.46			
Visual memory 1	E1L	95	0.44	0.89	0.93	97	NS
	E2L	4	0.86	0.62			
Auditory memory 1	E1L	96	0.61	1.08	0.33	98	NS
	E2L	4	0.43	0.46			
Non-word reading 1	E1L	94	0.80	1.02	0.76	96	NS
	E2L	4	0.40	0.16			
Syllable segmentation 1	E1L	94	0.65	0.88	0.83	96	NS
	E2L	4	0.28	0.32			
Sentence reading 2	E1L	95	1.01	0.92	0.05	96	NS
	E2L	3	0.99	1.23			
Spelling 2	E1L	95	0.92	0.79	0.49	97	NS
	E2L	4	1.12	0.84			
Reasoning 2	E1L	96	0.72	0.81	1.84	97	NS
	E2L	3	1.59	0.27			
Visual memory 2	E1L	95	0.41	1.13	0.48	97	NS
	E2L	4	0.68	0.59			
Auditory memory 2	E1L	96	0.56	1.10	0.18	98	NS
	E2L	4	0.46	1.60			
Non-word reading 2	E1L	94	1.02	0.96	1.26	95	NS
	E2L	3	0.32	0.01			
Syllable segmentation 2	E1L	94	0.78	0.85	0.74	95	NS
	E2L	3	0.42	0.37			

Correlations between the first test scores and retest scores were shown to be highly significant for all eight modules, using Pearson Product Moment Correlation Coefficient (see Table 112).

Table 112: LASS Secondary test-retest reliability coefficients

	N	r	Significance
Sentence Reading	98	0.85	p<0.001
Spelling	98	0.93	p<0.001
Reasoning	99	0.51	p<0.001
Visual memory	98	0.53	p<0.001
Auditory memory	99	0.58	p<0.001
Non-word Reading	96	0.77	p<0.001
Syllable Segmentation	96	0.74	p<0.001

10.4 Validity project - Spring 2001

The median centile scores and ranges for each of the eight LASS Secondary modules are given in Table 113. Again, only pupils scoring below the 20th centile on LASS sentence reading were required to complete the single word reading test. As only four pupils fulfilled this criteria, the results of the single word reading test have not been included in the following analyses.

Table 113: Descriptive statistics (centile scores) for the LASS Secondary modules

	N	Minimum	Maximum	Median
LASS Sentence Reading	71	4	99	65
LASS Spelling	70	15	99	65
LASS Reasoning	71	9	98	66
LASS Visual Memory	71	1	99	55
LASS Auditory Memory	71	1	99	69
LASS Non-word Reading	71	1	99	69
LASS Syllable Segmentation	71	1	99	67

The median centile scores and ranges for each of the validation measures are given in Table 114. Only pupils scoring below the 20th centile on LASS sentence reading or on the NFER Group Reading Test were required to complete the BAS word reading test. As only four pupils fulfilled this criteria, the results of the BAS word reading test have not been included in the following analyses.

Table 114: Descriptive statistics (centile scores) for the validation measures

	N	Minimum	Maximum	Median
NFER Group Reading Test	75	1	99	32
BSTS 3	75	1	99	34
MAT	75	1	98	38
WMS-III Spatial Span (forward)	74	1	98	37
WMS-III Spatial Span (backwards)	74	1	98	50
WMS-III Spatial Span (total)	74	1	95	50
WISC-III Digit Span (forward)	75	1	99	69
WISC-III Digit Span (backwards)	75	1	99	31
WISC-III Digit Span (total)	75	1	99	37
PhAB Alliteration	74	7	50	50
PhAB Rhyme	74	1	96	34
PhAB Spoonerisms	74	1	98	37
PhAB Non-word Reading	73	1	96	42
PhAB Picture naming speed	74	1	98	46
PhAB Digit naming speed	73	1	98	39
PhAB Fluency (alliteration)	74	1	92	37
PhAB Fluency (rhyme)	73	1	98	42
PhAB Fluency (semantic)	74	1	98	28

Analysis by gender (see Table 115) revealed no significant differences on any of the LASS modules.

Table 115: Gender differences on the eight LASS Secondary modules (z-scores)

	Gender	N	Mean	SD	t	df	Significance
LASS Sentence Reading	Female	27	0.64	0.70	1.25	69	NS
	Male	44	0.36	0.99			
LASS Spelling	Female	26	0.69	0.76	1.93	68	NS
	Male	44	0.25	1.00			
LASS Reasoning	Female	27	0.51	0.70	0.93	69	NS
	Male	44	0.33	0.80			
LASS Visual Memory	Female	27	0.02	0.93	0.44	69	NS
	Male	44	0.13	1.03			
LASS Auditory Memory	Female	27	0.54	0.74	0.96	69	NS
	Male	44	0.29	1.17			
LASS Non-word Reading	Female	27	0.49	0.97	0.55	69	NS
	Male	44	0.35	1.10			
LASS Syllable Segmentation	Female	27	0.49	0.86	1.23	69	NS
	Male	44	0.23	0.84			

Analysis by gender revealed a significant difference on only one of the validation measures (BSTS 3 spelling test), with girls outperforming boys on this test (see Table 116).

Table 116: Gender differences on the validation measures (z-scores)

	Gender	N	Mean	SD	t	df	Significance
NFER Group Reading Test	Female	28	-0.29	0.83	1.37	73	NS
	Male	47	-0.63	1.16			
BSTS 3	Female	28	-0.01	0.88	2.23	73	p<0.05
	Male	47	-0.62	1.27			
MAT	Female	28	-0.24	0.83	0.83	73	NS
	Male	47	-0.44	1.08			
WMS-III Spatial Span (forward)	Female	28	-0.13	0.86	1.64	72	NS
	Male	46	-0.48	0.93			
WMS-III Spatial Span (backwards)	Female	28	-0.05	1.30	1.04	72	NS
	Male	46	-0.34	1.05			
WMS-III Spatial Span (total)	Female	28	-0.08	1.09	1.45	72	NS
	Male	46	-0.43	0.97			
WISC-III Digit Span (forward)	Female	28	0.42	0.98	1.40	73	NS
	Male	47	0.08	1.04			
WISC-III Digit Span (backwards)	Female	28	-0.15	1.07	0.91	73	NS
	Male	47	-0.38	1.10			
WISC-III Digit Span (total)	Female	28	0.11	0.96	1.83	73	NS
	Male	47	-0.34	1.08			
PhAB Alliteration	Female	28	-0.23	0.50	0.54	72	NS
	Male	46	-0.17	0.44			
PhAB Rhyme	Female	28	-0.18	0.93	1.56	72	NS
	Male	46	-0.58	1.16			
PhAB Spoonerisms	Female	28	-0.15	0.88	0.82	72	NS
	Male	46	-0.34	1.02			
PhAB Non-word Reading	Female	28	0.12	0.81	0.78	71	NS
	Male	45	-0.07	1.15			
PhAB Picture naming speed	Female	28	-0.11	0.79	0.09	72	NS
	Male	46	-0.13	1.08			
PhAB Digit naming speed	Female	28	-0.25	0.68	0.31	71	NS
	Male	45	-0.33	1.19			
PhAB Fluency (alliteration)	Female	28	-0.43	0.85	0.68	72	NS
	Male	46	-0.28	1.00			
PhAB Fluency (rhyme)	Female	28	-0.14	0.96	0.56	71	NS
	Male	45	-0.27	0.93			
PhAB Fluency (semantic)	Female	28	-0.60	1.02	1.06	72	NS
	Male	46	-0.31	1.24			

The 75 pupils in the sample came from four ethnic backgrounds (see Table 117). However, as the majority of children were categorised as 'white', with only one or two students in each of the other categories, analysis by ethnic background was not feasible.

Table 117: Ethnic background of children

Ethnic background	N	Percent
White	70	93.3
Black other	2	2.7
Pakistani	1	1.3
Other	2	2.7

Analysis comparing 'white' children to all other children revealed no significant differences on the eight LASS modules (see Table 118).

Table 118: Ethnic differences on the eight LASS Secondary modules (z-scores)

	Gender	N	Mean	SD	t	df	Significance
LASS Sentence Reading	White	66	0.49	0.92	0.70	69	NS
	Other	5	0.19	0.57			
LASS Spelling	White	65	0.42	0.94	0.30	68	NS
	Other	5	0.30	1.01			
LASS Reasoning	White	66	0.40	0.77	0.15	69	NS
	Other	5	0.35	0.77			
LASS Visual Memory	White	66	0.06	1.02	0.85	69	NS
	Other	5	0.45	0.36			
LASS Auditory Memory	White	66	0.41	0.99	0.58	69	NS
	Other	5	0.13	1.62			
LASS Non-word Reading	White	66	0.42	1.05	0.41	69	NS
	Other	5	0.22	1.02			
LASS Syllable Segmentation	White	66	0.34	0.84	0.29	69	NS
	Other	5	0.22	1.07			

Analysis comparing 'white' children to all other children also revealed no significant differences on the validation measures (see Table 119).

Table 119: Ethnic differences on the validation measures (z-scores)

	Gender	N	Mean	SD	t	df	Significance
NFER Group Reading Test	White	70	-0.51	1.05	0.17	73	NS
	Other	5	-0.42	1.21			
BSTS 3	White	70	-0.41	1.18	0.28	73	NS
	Other	5	-0.25	1.18			
MAT	White	70	-0.36	1.00	0.06	73	NS
	Other	5	-0.34	1.01			
WMS-III Spatial Span (forward)	White	69	-0.33	0.92	0.60	72	NS
	Other	5	-0.59	0.97			
WMS-III Spatial Span (backwards)	White	69	-0.24	1.16	0.33	72	NS
	Other	5	-0.06	1.04			
WMS-III Spatial Span (total)	White	69	-0.30	1.02	0.05	72	NS
	Other	5	-0.32	1.23			
WISC-III Digit Span (forward)	White	70	0.20	1.04	0.42	73	NS
	Other	5	0.40	0.91			
WISC-III Digit Span (backwards)	White	70	-0.30	1.08	0.13	73	NS
	Other	5	-0.24	1.29			
WISC-III Digit Span (total)	White	70	-0.16	1.04	0.34	73	NS
	Other	5	-0.33	1.39			
PhAB Alliteration	White	69	-0.21	0.48	0.96	72	NS
	Other	5	0.00	0.00			
PhAB Rhyme	White	69	-0.44	1.11	0.36	72	NS
	Other	5	-0.26	0.93			
PhAB Spoonerisms	White	69	-0.29	0.98	0.82	72	NS
	Other	5	0.08	0.77			
PhAB Non-word Reading	White	68	-0.01	1.00	0.36	71	NS
	Other	5	0.16	1.48			
PhAB Picture naming speed	White	69	-0.16	0.99	1.22	72	NS
	Other	5	0.39	0.37			
PhAB Digit naming speed	White	68	-0.30	1.04	0.11	71	NS
	Other	5	-0.35	0.93			
PhAB Fluency (alliteration)	White	69	-0.36	0.96	0.78	72	NS
	Other	5	-0.01	0.64			
PhAB Fluency (rhyme)	White	68	-0.24	0.97	0.76	71	NS
	Other	5	0.09	0.24			
PhAB Fluency (semantic)	White	69	-0.40	1.18	0.65	72	NS
	Other	5	-0.75	0.91			

All 75 pupils in this sample have English as their first language and so no language analysis has been conducted.

Correlations between the LASS Secondary scores and their equivalent validation tests were calculated, for all eight modules, using Pearson Product Moment Correlation

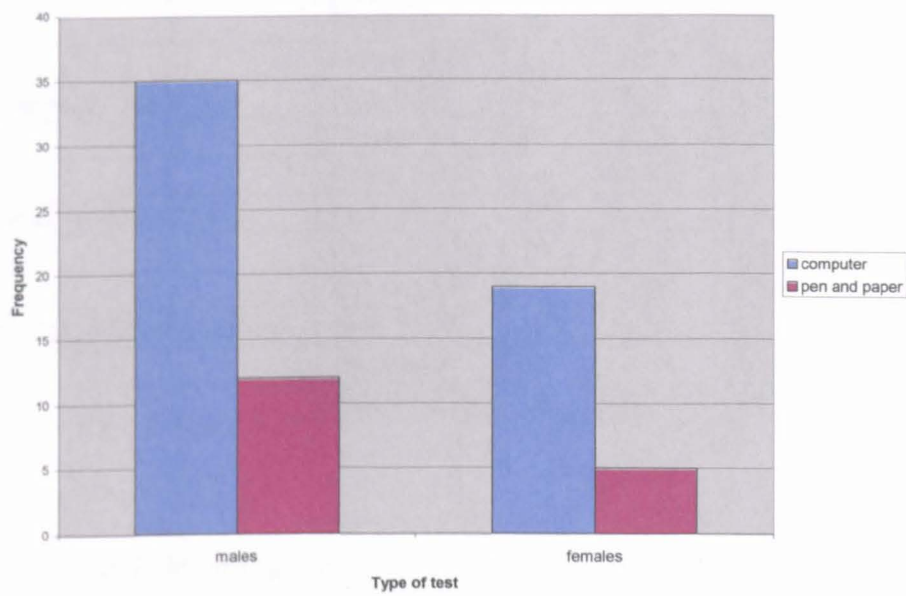
Coefficient. The results (see Table 120) indicate that all the LASS Secondary modules are significantly correlated with the validation measures.

Table 120: LASS Secondary and validity measure correlation coefficients

LASS test	Validation test	N	r	Significance
Sentence Reading	NFER Sentence Completion	71	0.75	p<0.001
Spelling	BSTS 3	70	0.88	p<0.001
Reasoning	MAT	71	0.52	p<0.001
Visual Memory	WMS-III Spatial Span (total)	70	0.37	p<0.01
Auditory Memory	WISC-III Digit Span (total)	71	0.55	p<0.001
Non-word Reading	PhAB Non-word Reading	69	0.43	p<0.001
Syllable Segmentation	PhAB Spoonerisms	70	0.45	p<0.001

54 of the 75 pupils preferred the computer tests, whilst 17 preferred the paper and pencil tests. There was no significant gender difference (see Figure 31) in the preference for computerised or paper and pencil tests (chi-square = 0.19, df = 1, not significant).

Figure 31: Gender differences in preferences for different test types



10.5 Dyslexia project - Spring 2001

Analysis of scores of students not identified as having dyslexia, students with dyslexia and students with other types of special educational need showed significant differences on all seven LASS modules (see Table 121). The placement of students into these categories was based on the information supplied by the teacher on the pupil information sheet (see Appendix 12).

Table 121: SEN group differences on the LASS Secondary modules z scores (ANOVA)

	Group	N	Mean	SD	F	df	Significance
LASS Sentence Reading	Non SEN	125	1.06	0.81	28.65	2	p<0.001
	Dyslexic	30	-0.03	0.86			
	Other SEN	16	0.03	0.80			
LASS Spelling	Non SEN	123	1.04	0.78	42.88	2	p<0.001
	Dyslexic	30	-0.22	0.69			
	Other SEN	16	-0.01	0.58			
LASS Reasoning	Non SEN	126	0.69	0.76	7.41	2	p<0.01
	Dyslexic	30	0.43	0.61			
	Other SEN	16	-0.02	0.74			
LASS Visual Memory	Non SEN	124	0.41	0.90	4.85	2	p<0.01
	Dyslexic	30	0.20	0.96			
	Other SEN	16	-0.33	1.00			
LASS Auditory Memory	Non SEN	125	0.77	0.94	17.27	2	p<0.001
	Dyslexic	30	-0.26	1.00			
	Other SEN	16	-0.07	1.07			
LASS Non-word Reading	Non SEN	123	0.91	0.95	22.96	2	p<0.001
	Dyslexic	30	-0.22	0.83			
	Other SEN	16	-0.03	0.88			
LASS Syllable Segmentation	Non SEN	123	0.74	0.83	18.67	2	p<0.001
	Dyslexic	30	-0.04	0.71			
	Other SEN	16	-0.19	0.65			

Further analysis, using Tukey's (HSD) test reveals where these significant differences actually lie (see Table 122).

Table 122: SEN group differences on the LASS Secondary modules z scores (Tukey)

	Group A	Mean	Group B	Mean	Diff.	Sig.
LASS Sentence Reading	Non SEN	1.06	Dyslexic	-0.03	1.10	p<0.001
	Non SEN	1.06	Other SEN	0.03	1.02	p<0.001
	Dyslexic	-0.03	Other SEN	0.03	0.07	NS
LASS Spelling	Non SEN	1.04	Dyslexic	-0.22	1.26	p<0.001
	Non SEN	1.04	Other SEN	-0.01	1.06	p<0.001
	Dyslexic	-0.22	Other SEN	-0.01	0.20	NS
LASS Reasoning	Non SEN	0.69	Dyslexic	0.43	0.25	NS
	Non SEN	0.69	Other SEN	-0.02	0.71	p<0.001
	Dyslexic	0.43	Other SEN	-0.02	0.46	NS
LASS Visual Memory	Non SEN	0.41	Dyslexic	0.20	0.21	NS
	Non SEN	0.41	Other SEN	-0.33	0.74	p<0.01
	Dyslexic	0.20	Other SEN	-0.33	0.53	NS
LASS Auditory Memory	Non SEN	0.77	Dyslexic	-0.26	1.04	p<0.001
	Non SEN	0.77	Other SEN	-0.07	0.85	p<0.01
	Dyslexic	-0.26	Other SEN	-0.07	0.19	NS
LASS Non-word Reading	Non SEN	0.91	Dyslexic	-0.22	1.14	p<0.001
	Non SEN	0.91	Other SEN	-0.03	0.95	p<0.001
	Dyslexic	-0.22	Other SEN	-0.03	0.19	NS
LASS Syllable Segmentation	Non SEN	0.74	Dyslexic	-0.04	0.78	p<0.001
	Non SEN	0.74	Other SEN	-0.19	0.93	p<0.001
	Dyslexic	-0.04	Other SEN	-0.19	0.14	NS

Analysis of scores of students not identified as having dyslexia, students with dyslexia and students with other types of special educational need showed significant differences on most of the validation measures (see Table 123).

Table 123: SEN group differences on the validation measures z scores (ANOVA)

	Group	N	Mean	SD	F	df	Significance
NFER Group Reading Test	Non SEN	49	-0.02	0.83	23.85	2	p<0.001
	Dyslexic	19	-1.39	0.89			
	Other SEN	7	-1.46	0.60			
BSTS 3	Non SEN	49	0.27	0.78	58.57	2	p<0.001
	Dyslexic	19	-1.69	0.63			
	Other SEN	7	-1.55	0.61			
MAT	Non SEN	49	-0.16	0.89	4.55	2	p<0.05
	Dyslexic	19	-0.58	1.16			
	Other SEN	7	-1.22	0.65			
WMS-III Spatial Span (forward)	Non SEN	48	-0.19	0.91	2.21	2	NS
	Dyslexic	19	-0.64	0.88			
	Other SEN	7	-0.66	0.88			

	Group	N	Mean	SD	F	df	Sig.
WMS-III Spatial Span (backwards)	Non SEN	48	0.06	1.04	5.64	2	p<0.01
	Dyslexic	19	-0.64	1.11			
	Other SEN	7	-1.13	1.32			
WMS-III Spatial Span (total)	Non SEN	48	-0.05	0.93	4.61	2	p<0.05
	Dyslexic	19	-0.66	1.04			
	Other SEN	7	-0.99	1.13			
WISC-III Digit Span (forward)	Non SEN	49	0.50	0.85	6.95	2	p<0.01
	Dyslexic	19	-0.26	1.13			
	Other SEN	7	-0.56	1.11			
WISC-III Digit Span (backwards)	Non SEN	49	0.11	1.07	12.71	2	p<0.001
	Dyslexic	19	-1.08	0.64			
	Other SEN	7	-0.99	0.66			
WISC-III Digit Span (total)	Non SEN	49	0.31	0.86	25.47	2	p<0.001
	Dyslexic	19	-1.16	0.75			
	Other SEN	7	-0.93	0.61			
PhAB Alliteration	Non SEN	48	-0.10	0.34	2.79	2	NS
	Dyslexic	19	-0.35	0.61			
	Other SEN	7	-0.38	0.65			
PhAB Rhyme	Non SEN	48	-0.07	0.92	9.31	2	p<0.001
	Dyslexic	19	-1.19	0.96			
	Other SEN	7	-0.80	1.45			
PhAB Spoonerisms	Non SEN	48	0.02	0.93	7.21	2	p<0.01
	Dyslexic	19	-0.85	0.73			
	Other SEN	7	-0.66	1.05			
PhAB Non-word Reading	Non SEN	47	0.44	0.89	18.04	2	p<0.001
	Dyslexic	19	-0.79	0.75			
	Other SEN	7	-0.83	0.78			
PhAB Picture naming speed	Non SEN	48	0.08	0.92	3.24	2	p<0.05
	Dyslexic	19	-0.44	1.10			
	Other SEN	7	-0.64	0.44			
PhAB Digit naming speed	Non SEN	48	0.06	0.94	11.46	2	p<0.001
	Dyslexic	19	-1.01	0.82			
	Other SEN	6	-0.98	0.82			
PhAB Fluency (alliteration)	Non SEN	48	-0.23	0.91	1.25	2	NS
	Dyslexic	19	-0.41	0.89			
	Other SEN	7	-0.81	1.28			
PhAB Fluency (rhyme)	Non SEN	47	-0.04	0.76	3.33	2	p<0.05
	Dyslexic	19	-0.68	0.90			
	Other SEN	7	-0.13	1.67			
PhAB Fluency (semantic)	Non SEN	48	-0.28	1.18	1.19	2	NS
	Dyslexic	19	-0.76	1.11			
	Other SEN	7	-0.45	1.14			
PhAB Highlighted low scores	Non SEN	47	0.85	1.20	10.98	2	p<0.001
	Dyslexic	19	2.89	2.21			
	Other SEN	7	2.86	3.34			

Further analysis, using Tukey's (HSD) test reveals where these significant differences actually lie (see Table 124).

Table 124: SEN group differences on the validation measures z scores (Tukey)

	Group A	Mean	Group B	Mean	Diff.	Sig.
NFER Group Reading Test	Non SEN	-0.02	Dyslexic	-1.39	1.37	p<0.001
	Non SEN	-0.02	Other SEN	-1.46	1.43	p<0.001
	Dyslexic	-1.39	Other SEN	-1.46	0.06	NS
BSTS 3	Non SEN	0.27	Dyslexic	-1.69	1.96	p<0.001
	Non SEN	0.27	Other SEN	-1.55	1.82	p<0.001
	Dyslexic	-1.69	Other SEN	-1.55	0.14	NS
MAT	Non SEN	-0.16	Dyslexic	-0.58	0.43	NS
	Non SEN	-0.16	Other SEN	-1.22	1.06	p<0.05
	Dyslexic	-0.58	Other SEN	-1.22	0.64	NS
WMS-III Spatial Span (forward)	Non SEN	-0.19	Dyslexic	-0.64	0.46	NS
	Non SEN	-0.19	Other SEN	-0.66	0.48	NS
	Dyslexic	-0.64	Other SEN	-0.66	0.02	NS
WMS-III Spatial Span (backwards)	Non SEN	0.06	Dyslexic	-0.64	0.71	p<0.05
	Non SEN	0.06	Other SEN	-1.13	1.20	p<0.05
	Dyslexic	-0.64	Other SEN	-1.13	0.49	NS
WMS-III Spatial Span (total)	Non SEN	-0.05	Dyslexic	-0.66	0.61	NS
	Non SEN	-0.05	Other SEN	-0.99	0.94	NS
	Dyslexic	-0.66	Other SEN	-0.99	0.33	NS
WISC-III Digit Span (forward)	Non SEN	0.50	Dyslexic	-0.26	0.76	p<0.05
	Non SEN	0.50	Other SEN	-0.56	1.06	p<0.05
	Dyslexic	-0.26	Other SEN	-0.56	0.30	NS
WISC-III Digit Span (backwards)	Non SEN	0.11	Dyslexic	-1.08	1.18	p<0.001
	Non SEN	0.11	Other SEN	-0.99	1.09	p<0.05
	Dyslexic	-1.08	Other SEN	-0.99	0.09	NS
WISC-III Digit Span (total)	Non SEN	0.31	Dyslexic	-1.16	1.47	p<0.001
	Non SEN	0.31	Other SEN	-0.93	1.24	p<0.01
	Dyslexic	-1.16	Other SEN	-0.93	0.23	NS
PhAB Alliteration	Non SEN	-0.10	Dyslexic	-0.35	0.25	NS
	Non SEN	-0.10	Other SEN	-0.38	0.28	NS
	Dyslexic	-0.35	Other SEN	-0.38	0.03	NS
PhAB Rhyme	Non SEN	-0.07	Dyslexic	-1.19	1.12	p<0.001
	Non SEN	-0.07	Other SEN	-0.80	0.73	NS
	Dyslexic	-1.19	Other SEN	-0.80	0.39	NS
PhAB Spoonerisms	Non SEN	0.02	Dyslexic	-0.85	0.87	p<0.01
	Non SEN	0.02	Other SEN	-0.66	0.68	NS
	Dyslexic	-0.85	Other SEN	-0.66	0.19	NS
PhAB Non-word Reading	Non SEN	0.44	Dyslexic	-0.79	1.23	p<0.001
	Non SEN	0.44	Other SEN	-0.83	1.28	p<0.01
	Dyslexic	-0.79	Other SEN	-0.83	0.04	NS
PhAB Picture naming speed	Non SEN	0.08	Dyslexic	-0.44	0.52	NS
	Non SEN	0.08	Other SEN	-0.64	0.72	NS
	Dyslexic	-0.44	Other SEN	-0.64	0.21	NS

	Group A	Mean	Group B	Mean	Diff.	Sig.
PhAB Digit naming speed	Non SEN	0.06	Dyslexic	-1.01	1.07	p<0.001
	Non SEN	0.06	Other SEN	-0.98	1.04	p<0.05
	Dyslexic	-1.01	Other SEN	-0.98	0.02	NS
PhAB Fluency (alliteration)	Non SEN	-0.23	Dyslexic	-0.41	0.18	NS
	Non SEN	-0.23	Other SEN	-0.81	0.58	NS
	Dyslexic	-0.41	Other SEN	-0.81	0.40	NS
PhAB Fluency (rhyme)	Non SEN	-0.04	Dyslexic	-0.68	0.63	p<0.05
	Non SEN	-0.04	Other SEN	-0.13	0.08	NS
	Dyslexic	-0.68	Other SEN	-0.13	0.55	NS
PhAB Fluency (semantic)	Non SEN	-0.28	Dyslexic	-0.76	0.48	NS
	Non SEN	-0.28	Other SEN	-0.45	0.18	NS
	Dyslexic	-0.76	Other SEN	-0.45	0.31	NS
PhAB Highlighted low scores	Non SEN	0.85	Dyslexic	2.89	2.04	p<0.001
	Non SEN	0.85	Other SEN	2.86	2.01	p<0.05
	Dyslexic	2.89	Other SEN	2.86	0.04	NS

Students were classified by LASS as having difficulties if they had a LASS reasoning score at least 0.66 standard deviations higher than one or more of the LASS literacy tests and at least 0.66 standard deviations higher than one or more of the LASS cognitive tests. A discrepancy of 0.66 or more standard deviations is statistically significant ($p<0.05$). LASS Reasoning scores for each pupil were then adjusted to allow for regression to the mean. Discrepancies were then calculated using the previous method.

Students were classified as having difficulties using the validation measures if they had a MAT reasoning score at least 0.66 standard deviations higher than one or more of the literacy validation measures and at least 0.66 standard deviations higher than one or more of the cognitive validation measures. MAT reasoning scores for each pupil were then adjusted to allow for regression to the mean. Discrepancies were then calculated using the previous method.

Students were identified by the PhAB as having difficulties if they had three or more low scores (below standard score 85) on any of the eight PhAB subtests (excluding the semantic fluency test).

Of the 19 students with dyslexia, 15 were identified by LASS as having difficulties, compared to 12 identified using the validation measures and 11 using the PhAB alone (see Table 125). Allowing for regression to the mean, 14 students were identified by LASS as having difficulties, compared to 16 identified using the validation measures (see Appendix 13 for scores and calculations for each pupil).

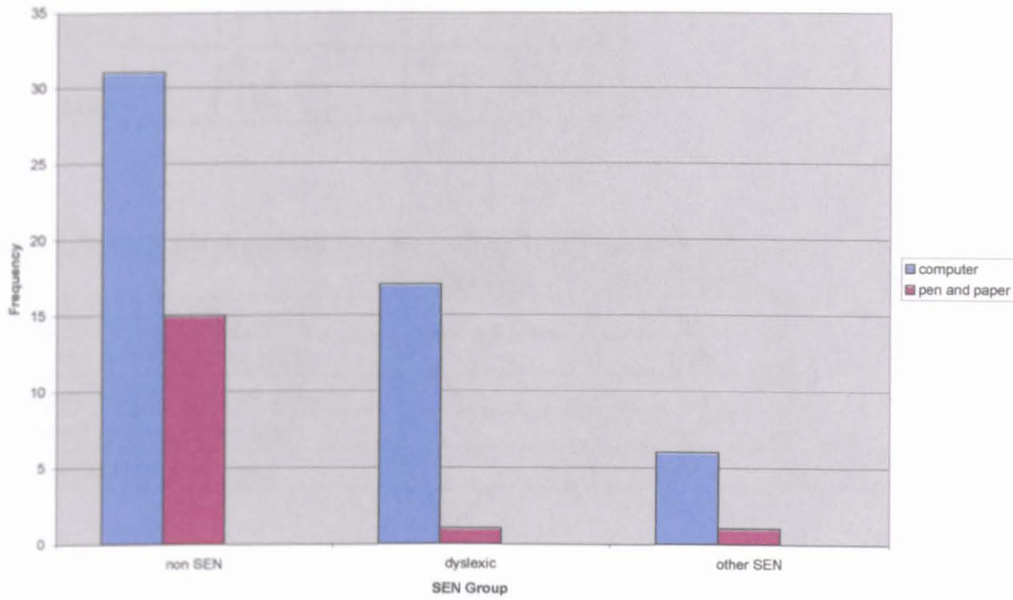
Table 125: Students with dyslexia identified as having difficulties

Student	Gender	LASS	LASS*	Validation	Validation *	PhAB
1	F	✓	✓	✓	✓	
2	M	✓	✓	✓	✓	✓
3	F	✓	✓		✓	
4	M	✓	✓	✓	✓	✓
5	M	✓	✓	✓	✓	✓
6	M	✓		✓	✓	✓
7	M	✓	✓			✓
8	M		✓		✓	✓
9	M	✓	✓	✓	✓	
10	M	✓		✓	✓	
11	M	✓	✓	✓	✓	✓
12	M		✓	✓	✓	✓
13	M					✓
14	M				✓	✓
15	M	✓	✓	✓	✓	
16	F	✓	✓		✓	✓
17	M	✓	✓	✓	✓	
18	M	✓				
19	F	✓	✓	✓	✓	
Total		15	14	12	16	11

* allowing for regression to the mean.

There was no significant SEN group difference (see Figure 32) in the preference for computerised or paper and pencil tests (chi-square = 5.60, df = 2, not significant).

Figure 32: SEN group differences in preferences for different test types



10.6 Questionnaire study - Autumn 1998 - Spring 2001

Frequency distributions are shown for each question for the entire sample, with a breakdown by year of response, in Tables 126 to 150. 14 teachers returned questionnaires in October 1998 and 12 returned questionnaires in March 2001.

Table 126: How did you find the installation of LASS Secondary?

	Frequency	1998	2001
Very difficult	0	0	0
A little difficult	1	0	1
Satisfactory	5	5	0
Easy	10	5	5
Very easy	10	4	6

Table 127: How did you find registering pupils on LASS Secondary?

	Frequency	1998	2001
Very difficult	0	0	0
A little difficult	1	0	1
Satisfactory	5	2	3
Easy	10	7	3
Very easy	10	5	5

Table 128: How did you find running the tests on LASS Secondary?

	Frequency	1998	2001
Very difficult	2	1	1
A little difficult	1	0	1
Satisfactory	3	1	2
Easy	9	6	3
Very easy	11	6	5

Table 129: How did the pupils find the tests on LASS Secondary?

	Frequency	1998	2001
Enjoyed none of the tests	0	0	0
Enjoyed a few of the tests	0	0	0
Enjoyed some but not others	6	2	4
Enjoyed most of the tests	16	9	7
Enjoyed all of the tests	4	3	1

Table 130: Which tests did the pupils enjoy most on LASS Secondary?

	Frequency	1998	2001
Single word reading	1	0	1
Sentence reading	3	0	3
Spelling	3	1	2
Reasoning	4	2	2
Visual memory	9	6	3
Auditory memory	14	7	7
Non-word reading	3	2	1
Syllable segmentation	3	2	1

Table 131: Which tests did the pupils enjoy least on LASS Secondary?

	Frequency	1998	2001
Single word reading	0	0	0
Sentence reading	4	2	2
Spelling	3	1	2
Reasoning	8	4	4
Visual memory	7	4	3
Auditory memory	0	0	0
Non-word reading	4	2	2
Syllable segmentation	4	2	2

Table 132: Rate the suitability of 'Cave' as an assessment of visual memory?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	0
Adequate	3
Fairly suitable	4
Highly suitable	4

Table 133: Rate the suitability of 'Mobile' as an assessment of auditory memory?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	0
Adequate	3
Fairly suitable	3
Highly suitable	6

Table 134: Rate the suitability of Non-word reading as an assessment of phonic skills?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	0
Adequate	2
Fairly suitable	7
Highly suitable	3

Table 135: Rate the suitability of Syllable segmentation as an assessment of phonological processing?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	1
Adequate	2
Fairly suitable	3
Highly suitable	6

Table 136: Rate the suitability of Sentence reading as an assessment of reading ability?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	1
Adequate	3
Fairly suitable	2
Highly suitable	6

Table 137: Rate the suitability of Spelling as an assessment of spelling ability?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	0
Adequate	3
Fairly suitable	3
Highly suitable	6

Table 138: Rate the suitability of Reasoning as an assessment of non-verbal intelligence?

	Frequency (2001 only)
Totally unsuitable	0
Not very suitable	0
Adequate	2
Fairly suitable	4
Highly suitable	6

Table 139: How did you find using the reports on LASS Secondary?

	Frequency (1998 only)
Very difficult	0
A little difficult	1
Satisfactory	1
Easy	7
Very easy	5

Table 140: Rate the usefulness of the results on LASS Secondary?

	Frequency	1998	2001
No use at all	0	0	0
Needs improvement	2	2	0
Satisfactory	2	1	1
Fairly useful	9	7	2
Very useful	12	4	8

Table 141: How would you rate LASS Secondary overall?

	Frequency	1998	2001
Poor	0	0	0
Not very good	0	0	0
Average	2	0	2
Good	12	8	4
Very good	12	6	6

Table 142: Will you be using LASS Secondary in the future?

	Frequency	1998	2001
Definitely not	0	0	0
Probably not	0	0	0
Maybe	1	0	1
Probably	5	3	2
Definitely	20	11	9

Table 143: How did you find running the following pen and paper tests with pupils? (2001 only)

	Spatial span	Digit span	PhAB	NFER SCT	BSTS 3	MAT
Very difficult	0	0	0	0	1	0
A little difficult	2	1	0	1	0	0
Satisfactory	1	2	1	1	1	1
Easy	1	0	0	1	1	2
Very easy	0	1	2	1	1	1

Table 144: Which of the following tests did pupils enjoy most? (2001 only)

LASS Visual memory	WMS-III Spatial span
2	2

Table 145: Which of the following tests did pupils enjoy most? (2001 only)

LASS Auditory memory	WISC-III Digit span
4	0

Table 146: Which of the following tests did pupils enjoy most? (2001 only)

LASS Non-word reading	PhAB Non-word reading
4	0

Table 147: Which of the following tests did pupils enjoy most? (2001 only)

LASS Syllable segmentation	PhAB Spoonerisms
2	2

Table 148: Which of the following tests did pupils enjoy most? (2001 only)

LASS Sentence reading	NFER Sentence completion
3	1

Table 149: Which of the following tests did pupils enjoy most? (2001 only)

LASS Spelling	BSTS 3
4	0

Table 150: Which of the following tests did pupils enjoy most? (2001 only)

LASS Reasoning	MAT
2	2

11 Discussion for the secondary assessment studies

11.1 Overview

The results from the secondary assessment research are discussed with a view to:

- the development and standardisation of the tests
- group differences (in relation to gender, ethnicity and language)
- technical issues (including internal consistency, reliability and validity)
- the identification of pupils with dyslexia
- the questionnaire study.

11.2 Test development and standardisation

The results from the test development stage indicate that seven of the LASS Secondary modules provide satisfactory distributions of scores across the intended age range, without undesirable ceiling or floor effects (see section 10.1). These results were confirmed by the later standardisation project (see section 10.2). However, the single word reading module had a marked ceiling effect with a mean of 29.46 out of a total of 30. Despite this, the test was found to be useful for special needs schools and was therefore retained in the final test suite. The students within the special school system comprise a different population to that in the mainstream school system and many struggled with the LASS Secondary sentence reading test. Scoring zero on any test is discouraging and it is important to include a reading test suitable to the abilities of students in special education. However, it is recommended that the single word reading test is administered only to pupils scoring below the 20th centile on the LASS Secondary sentence reading test.

The adaptive forms of the LASS Secondary sentence reading, spelling and reasoning modules correlated highly with the full forms, with correlation coefficients ranging from 0.80 (reasoning) to 0.93 (spelling), justifying the use of the much shorter adaptive forms (see section 10.1).

The sample size in the standardisation stage of the project (505) is smaller than the samples used in some similar standardisation studies. For example, the British Ability Scales-II were standardised on 1689 children in 245 schools (Elliott, 1996) and the NFER Sentence Completion Test (Macmillan Test Unit, 1990) standardisation involved 3917 children in 130 schools. In addition, the British Spelling Test Series - Level 3 was standardised using 3936 children (Vincent and Crumpler, 1997) and the Matrix Analogies Test (Naglieri, 1985) standardisation included 4468 students. However, the Wechsler Intelligence Scale for Children-III was standardised on a smaller sample of 824 children (aged 6 to 16) in the UK (Wechsler, 1991) and the Phonological Assessment Battery (Frederickson, Frith and Reason, 1997) standardisation involved just 629 children (aged 6 to 14). It should be noted that the LASS Secondary modules were standardised on the age range 11 to 15, with an average of 101 per year group. This is, therefore, a larger sample per year than the average per year in both the WISC-III (75 per year group) and the PhAB standardisations (70 per year group) and is comparable in size to the BAS-II standardisation sample (105 per year group). Despite the relatively small size of the UK standardisation sample, the WISC-III is reported to be the best validated and most widely accepted measure of children's intellectual functioning (Cooper, 1995). However, the re-standardisation of a test in the UK when it has already been developed and standardised in the USA is different to the introduction of a new test (e.g. LASS Secondary) in the UK.

The percentile scores calculated for each age group (see Appendix 11) show that, in all eight modules, the scores increased with age, indicating that there are no age-related anomalies on any of the LASS Secondary modules.

11.3 Group differences

11.3.1 Gender

During the test development stage, significant gender differences were found on three of the LASS Secondary modules (single word reading, sentence reading and spelling) with females outperforming males in all of these cases (see section 10.1). This supports the finding by Fergusson and Horwood (1997) that males achieve less well than females from school entry to age 18, despite similar IQ scores (in the present study, there was no significant gender difference on the reasoning module). Furthermore, Blatchford (1997) reports that females obtain significantly higher scores than males in reading at 7, 11 and 16 years of age. Hare (1985) also found that black boys score lower than black girls in reading achievement (the term 'black' was undefined by the author). The findings in the present study are consistent with those reported by Heath (2001) that females aged 10 to 12 score significantly higher than males in reading. Lee (2000) also found that males obtain lower scores than females on language tests and this deficiency increases over the school years. However, Lee also suggests that males have an advantage over females in non-verbal reasoning and that this strengthens over time. This finding is not supported by the present study.

It should be noted that the sample sizes for the modules that showed significant gender differences during the test development stage ranged from 2098 to 2232, compared to sample sizes of 77 to 728 for the other modules. This may account for significant differences only being shown in the first three modules. However, an alternative explanation is that the three tests in which significant gender differences were found were administered as pen and paper tests during the test development stage and are arguably less objective and more subject to bias (this topic is followed up in Chapter 12). Four of the other tests which showed no significant gender difference (visual memory, auditory memory, non-word reading and syllable segmentation) were administered as computerised tests at this stage.

The LASS Secondary standardisation revealed significant gender differences on only the spelling and visual memory modules (see section 10.2). Once again, the females outperformed the males on the spelling test, whilst the males scored significantly higher on the visual memory test. This latter finding is contrary to that reported by Lee (2000), that girls' memory skills overshadowed boys and strengthened over time. The significant differences found during the test development stage were based on a much larger sample than the standardisation stage, which may explain why more significant gender differences were found during the test development stage than during the standardisation. However, during the standardisation, all of the LASS Secondary tests were administered in their computerised form. It may be the case that in using more objective testing methods, gender differences are less likely to be observed.

No significant gender differences were found on either testing session during the reliability project (see section 10.3) or the validity project (see section 10.4). The validation measures show a significant gender difference on spelling, with females outperforming males. The validation measures used were traditional pen and paper tests which, as previously noted, are less objective than computerised assessment methods. The gender difference in spelling ability on the BSTS 3 spelling test is concordant with a finding by Coren (1989) that even at an undergraduate level a gender difference, favouring females, exists in spelling.

The results of the validity study suggest that there is no significant effect of gender in students' preference for computerised or pen and paper tests (see section 10.4).

11.3.2 Ethnicity

One significant ethnic group difference was found in the reliability project, with white students scoring less well than students from other ethnic backgrounds on the second testing of the LASS reasoning module (see section 10.3). This finding is contrary to those reported by a number of American researchers who argue that individuals from an African ethnic background score lower on cognitive ability tests than white individuals (Reynolds et al, 1987; Schmidt, 1988; Hartigan and Wigdor, 1989; and, Neisser et al, 1996). Jensen and Inouye (1980) report findings that suggest that Asian children score significantly higher than white children, who score significantly higher than black children, on memory span whilst Asian and white children score significantly higher than black children on non-verbal intelligence.

However, Vincent (1991) suggests that, although the IQ difference between black and white adults has remained constant, IQ differences between black and white children are declining in the United States. Jensen (1973) found that 11-year old black children obtained scores on Raven's Coloured Progressive Matrices 14 points below white children. Whereas, Raven et al (1986) found a difference of only seven points for 12-year olds on Raven's Standard Progressive Matrices. Furthermore, Vance and Gaynor (1976) found no significant differences between black and white 10 and 11- year old children on the performance scale and full scale of the WISC. Skuy et al (2000) and Salois (1999) have also reported that black and Indian children score significantly lower than white children on the WISC-R verbal scale and full scale but not on the performance scale. Indeed, Salois notes that two of the performance tests (picture completion and mazes) showed significantly higher scores for Indian children compared to the national norm. Furthermore, Gonzales and Roll (1985) found Mexican children to score the same as white children in non-verbal intellectual ability but less well in vocabulary. The use of a non-verbal reasoning test such as the one in LASS Secondary may be a more culture-fair assessment of intelligence. Tate and Gibson (1980) found black children (aged 10-18) to score significantly higher than white children on intelligence, when matched for socio-economic status and educational level. This finding is partially supported by the present study although the results of many other studies are not consistent with this result.

No significant differences on the basis of ethnic background were found in any of the LASS Secondary modules or the validation measures during the validity project (see section 10.4). These findings are contrary to those reported by Plewis (1991) who found educational differences between ethnic groups, with African Caribbean pupils performing less well than white pupils in language. The present findings are also inconsistent with those of Herring (1989) and Heath (2001) whose results indicated that white students at middle and high schools score significantly higher in reading than other ethnic groups. Kenney and Anderson (1984) obtained similar results at university level. Holland (1996) reports that school failure is a characteristic of African American boys in urban America.

Blatchford (1997) found no significant differences between black and white pupils in reading at age 7 and 11, although at age 16, black pupils obtained significantly higher scores than white pupils in English. The findings of the present study do, at least in part, support Blatchford's results. Sammons (1994) also found a marked improvement by African Caribbean and Asian groups over white during secondary education. Such results were not apparent in the current research. However, Michael Brooks (The Independent, December 4 1997) reports that black boys perform worse than other groups in technical subjects, mathematics and science at GCSE level. He asserts that, as they proceed through the education system, African Caribbean boys slide from the top to the bottom of the achievement scales. Additionally, in The Times (O'Leary and Betts, March 11 1999), it is reported that the Office for Standards in Education expressed concern about the progress made by Pakistani, Bangladeshi and Black Caribbean children. They suggest that Bangladeshi and Pakistani children perform poorly in primary school but do better as their English improves, whereas Black Caribbean children start school well but have the worst GCSE results of all ethnic groups. Nevertheless, the results of the validation study are consistent with the finding by Nichols and McKinney (1977) that there are no significant differences between black and white students in reading and language tests. However, the lack of significant differences, in the present study, may be due to the small number of students in the non-white sample.

11.3.3 Language

No significant differences on the basis of language were found in any of the LASS Secondary modules during the reliability project (see section 10.3). Miller (1979) suggests that by the end of secondary education, students learning English as a second language have no problem with different forms of written English. However, the lack of significant differences may, again, be due to the small number of students in the sample speaking English as a second language. There were no children speaking English as a second language involved in the validity project.

11.4 Technical issues

11.4.1 Reliability

11.4.1.1 Internal Consistency

All of the LASS Secondary modules show high levels of internal consistency (see section 10.1), ranging from 0.71 (auditory memory) to 0.97 (sentence reading and spelling), with a median of 0.87 (single word reading). These are comparable to levels of internal consistency found in other similar tests. For example, the NFER Sentence Completion Test shows internal consistency coefficients of 0.90 (n=3917) for form X and 0.89 (n=1269) for form Y (Macmillan Test Unit, 1990), whilst the British Spelling Test Series - Level 3 (Vincent and Crumpler, 1997) has an internal consistency of 0.97. The Matrix Analogies Test (Naglieri, 1985) reports an internal consistency coefficient of 0.83, which is slightly lower than that found in the equivalent LASS reasoning test (0.91). The Wechsler Memory Scale-III has internal consistency coefficients ranging from 0.74 to 0.93 (Wechsler, 1997) and the PhAB (Frederickson, Frith and Reason, 1997) reports coefficients ranging from 0.19 to 0.96, although most are above 0.80.

11.4.1.2 Test-retest reliability

LASS Secondary appears to be fairly reliable over time (see section 10.3) with test-retest reliability coefficients ranging from 0.51 (reasoning) to 0.93 (spelling), with a median of 0.74 (syllable segmentation). These coefficients are comparable to other test-retest reliability coefficients quoted in the literature. For example, the BAS-II (Elliott, 1996) test-retest reliability coefficients range from 0.38 (delayed recall of objects, age 3:6 to 4:5, n=100) to 0.98 (word reading, age 5:0 to 14:0, n=40). Furthermore, the MAT (Naglieri, 1985) has a test-retest reliability of 0.75 (n=163) and the WMS-III (Wechsler, 1997) shows coefficients ranging from 0.62 (Faces-II) to 0.82 (Verbal paired associates-I). The test-retest reliability coefficient for the WMS-III spatial span total is 0.71, which is higher than that found by the equivalent LASS visual memory test (r=0.53). The WISC-III (Wechsler, 1991) shows test-retest reliability coefficients ranging from 0.57 (mazes) to 0.89 (vocabulary). The test-retest reliability coefficient for the WISC-III digit span test is 0.73, which is slightly higher then that for the LASS Secondary auditory memory test (r=0.58). It should be noted that three of the LASS Secondary modules are adaptive tests (sentence reading, spelling and reasoning) and so the actual items administered to a particular pupil may vary from one testing session to another. This type of test administration differs from the validation measures in which pupils are administered the same items on each testing session. Nevertheless, the test-retest reliability coefficients for the LASS spelling (r=0.93) and sentence reading (r=0.85) tests are very high.

11.4.2 Validity

LASS Secondary shows concurrent validity coefficients (see section 10.4) ranging from 0.37 (visual memory) to 0.88 (spelling), with a median of 0.52 (reasoning). These correlations are comparable to those reported for other similar tests. For example, the BAS-II (Elliott, 1996) shows correlations (n=38) with the WISC-III (Wechsler, 1991) ranging from 0.34 (BAS-II speed of information processing and WISC-III coding) to 0.74 (BAS-II word definitions and WISC-III vocabulary).

The NFER Sentence Completion Test - form X (Macmillan Test Unit, 1990) shows a correlation of 0.73 (n=218) with a previous NFER reading test, which is very similar to the correlation found between LASS sentence reading and the NFER Sentence Completion test (r=0.75). The LASS sentence reading test and NFER Group Reading sentence completion test are very similar in terms of their item format.

The BSTS 3 (Vincent and Crumpler, 1997) correlates highly with the Schonell spelling test (Schonell, 1971), with coefficients of 0.93 (n=29) for form X and 0.92 (n=30) for form Y. Again these correlations are very close to the correlation between LASS spelling and the BSTS 3 (r=0.88). The LASS Secondary spelling test involves contextual word spelling whilst the BSTS 3 involves a wider range of spelling items, including single word spelling, cloze passages, dictation passages and correction of errors. Despite the differences between the two tests in their item format, the correlation between them is high.

The MAT (Naglieri, 1985) test has been shown to correlate well (r=0.68, n=54) with the WISC-R (Wechsler, 1974), which is higher than the correlation found between the LASS Reasoning module and the MAT (r=0.52). This correlation is fairly low, considering that the LASS Secondary reasoning and MAT items are very similar in their format and both are timed tests.

The WMS-III (Wechsler, 1997) shows correlations with the Children's Memory Scale (Cohen, 1997) ranging from 0.26 to 0.74. The correlation between the WMS-III visual immediate memory and the Children's Memory Scale visual immediate memory is 0.55, which is somewhat higher than the correlation between LASS visual memory and WMS-III spatial span (r=0.37). The WMS-III spatial span test is a measure of spatial-sequential memory. However, the LASS Secondary visual memory test is a measure of spatial-associative memory. The fact that these tests measure slightly different aspects of visual memory may explain the low correlation between them.

The correlation between the WMS-III (Wechsler, 1997) auditory immediate memory and the Children's Memory Scale (Cohen, 1997) verbal immediate memory (r=0.74) is higher than that found between the LASS auditory memory test and the WISC-III (Wechsler, 1991) digit span test (r=0.55). The WISC-III digit span test consists of two parts: forward digit span and backward digit span. The LASS Secondary auditory memory test is purely a measure of forward digit span. The LASS auditory memory test correlates similarly with the WISC-III backward digit span (r=0.42) as with the WISC-III forward digit span (r=0.40).

The PhAB (Frederickson, Frith and Reason, 1997) has correlations with the Neale Analysis of Reading Ability (Neale, 1989) ranging from 0.24 (PhAB semantic fluency and Neale accuracy) to 0.72 (PhAB non-word reading and Neale accuracy). The PhAB also reports correlations with the BAS-II ranging from 0.05 (PhAB digit naming speed and BAS-II recall of designs) to 0.44 (PhAB spoonerisms and BAS-II quantative reasoning). This latter correlation is comparable to those obtained between the LASS Secondary non-word reading test and the PhAB non-word reading test (r=0.43) and between the LASS Secondary syllable segmentation test and the PhAB spoonerisms test (r=0.45). The LASS non-word reading test and the PhAB non-word reading test are similar in their item format. In the LASS Secondary syllable segmentation test, students are presented with a word and asked what the word would sound like if part of the word is removed. They hear four different answers and select the one they think is correct. Whereas, in the PhAB spoonerisms test, students are asked to segment words and synthesise the segments to provide new words. Despite the differences between these two tests, there is still a moderate correlation between them.

The results of the validity study (see section 10.4) suggest that the majority of the students preferred the computerised tests (72%) to the validation measures (28%). This is consistent with other findings reported in the literature that indicate that children often prefer computer-based assessment to traditional assessment by a human assessor (Skinner and Allen, 1983; Moore, Summer and Bloor, 1984; French, 1986; Singleton, Horne and Vincent, 1995; Singleton, 1994 and 1997a).

11.5 Identification of pupils with dyslexia

The children involved in the dyslexia project were classified as being non-SEN, dyslexic or other SEN using the information provided by their teachers on the pupil information sheet (see Appendix 12).

The dyslexic group scored significantly lower than the non-SEN group (see section 10.5) on five of the seven LASS tests (sentence reading, spelling, auditory memory, non-word reading and syllable segmentation). There were no significant differences between the dyslexic group and the non-SEN group on reasoning or visual memory. There were no significant differences between the dyslexic group and the other SEN group. The other SEN group scored significantly lower than the non-SEN group on all seven LASS Secondary tests. These results fit well with the definitions of dyslexia provided by the Orton Dyslexia Association (1994), the British Dyslexia Association (1998) and the National Working Party on Dyslexia in Higher Education (1999). The dyslexic group are significantly behind the non-SEN group in literacy, phonological skills and auditory memory and these difficulties are not due to low intelligence as the dyslexic group show similar levels of intelligence to the non-SEN group. The dyslexic group also show similar visual memory ability to the non-SEN group, which is contrary to findings reported by Thomson (1982) and Fein et al (1988). However, Turner (1997) argues that when a visual memory difficulty is evident in dyslexia, it is usually the linguistic component that is at fault and that the purer the measure of visual memory, the better the individual with dyslexia is at doing it. The LASS Secondary visual memory test is a much purer (i.e. less susceptible to verbal encoding support strategies) measure of visual memory than many other commonly used visual memory tests (for example, the BAS-II Immediate Recall test). This may explain the lack of a significant difference between the dyslexic group and the non-SEN group on the LASS Secondary visual memory test. The finding that the other SEN group scored significantly lower than the non-SEN group on all of the LASS Secondary tests was as expected.

The dyslexic group scored significantly lower than the non-SEN group (see section 10.5) on several of the validation measures (NFER group reading test, BSTS 3 spelling test, all digit span measures, backward spatial span, PhAB rhyme, PhAB spoonerisms, PhAB non-word reading, PhAB digit naming speed and PhAB rhyme fluency). The dyslexic group had significantly more low PhAB scores than the non-SEN group. There were no significant differences between the dyslexic group and the non-SEN group on the remaining validation measures (MAT, forward and total spatial span, PhAB alliteration, PhAB picture naming speed, PhAB alliteration fluency and PhAB semantic fluency). Again, these results fit reasonably well with the previous definitions of dyslexia, with the dyslexic group obtaining significantly lower scores than the non-SEN group in literacy, auditory memory and some phonological skills, whilst displaying a similar level of intelligence. The WMS-III spatial span test is also a pure measure of visual memory which may explain why there is only a significant difference between the dyslexic group and the non-SEN group on the backward spatial span. Surprisingly, the dyslexic group did not score significantly lower than the non-SEN group on the

PhAB alliteration, picture naming speed and alliteration fluency tests as would be expected. These results are contrary to the finding by Ellis and Large (1987) that alliteration ability reliably differentiates children with dyslexia from their better-reading peers. However the Ellis and Large study involved much younger children (aged five to eight years). Bowers, Steffy and Tate (1988) suggest that individuals with and without dyslexia differ significantly on rapid naming ability. This finding is not fully supported by the present study. However, Nicolson and Fawcett (1995a) argue that this difference reduces with age, which may explain these results.

There were no significant differences between the dyslexic group and the other SEN group on any of the validation measures (see section 10.5). The other SEN group scored significantly lower than the non-SEN group on several of the validation measures (NFER group reading test, BSTS 3 spelling test, MAT, all digit span measures, backward spatial span, PhAB non-word reading and PhAB digit naming speed). The other SEN group also had significantly more low PhAB scores than the non-SEN group. There were no significant differences between the other SEN group and the non-SEN group on the remaining validation measures (forward and total spatial span, PhAB alliteration, PhAB rhyme, PhAB spoonerisms, PhAB picture naming speed, PhAB alliteration fluency, PhAB rhyme fluency and PhAB semantic fluency).

15 of the 19 dyslexic students were identified as having dyslexia by LASS Secondary, compared to 12 identified using the validation measures and 11 identified using the PhAB on its own (see section 10.5). A problem with the discrepancy model is that it is associated with regression to the mean. According to Wright and Groner (1993), very intelligent children will be identified as underachievers and Turner (1997) argues that this method may also be unfair to lower ability pupils. Allowing for regression to the mean involves taking into account the correlation between the attainment and ability

scores. When allowing for regression to the mean, 14 of the 19 dyslexic students were identified as having dyslexia by LASS Secondary, compared to 16 identified using the validation measures. All of the students are identified as having difficulties by at least one of the three methods. Three of the four students in the dyslexic group who were not initially identified by LASS as having dyslexia had low reasoning scores and would usually be seen as being of low ability. Allowing for regression to the mean, two of these students were identified by LASS as having dyslexia. It is possible that having a measure of verbal intelligence within the assessment system would give a better overall picture of student's intelligence.

The criteria used for identifying a student as having dyslexia on LASS and the validation measures is that there is a significant discrepancy between the student's intelligence and their literacy and memory scores. Using this discrepancy model, 63% to 84% of the dyslexic students were correctly identified. Whereas, identification of students as having dyslexia using the PhAB (i.e. three or more standard scores below 85) is based on the phonological deficit hypothesis. This method only correctly identified 58% of the dyslexic students as having difficulties.

Further analysis was carried out involving all of the students in the dyslexia project (including the non-SEN and other SEN students). Decisions were made for each student as to whether they would be identified as having difficulties by the three screening methods. Not all students were administered the validation measures and the PhAB. For the LASS system and the validation measures, no calculations were carried out but a decision was made according to the profile of scores, with obvious discrepancies being looked for. This is the method that would be used by teachers using these tools as screening devices. Students were identified by the PhAB as having difficulties if they had three or more low scores (i.e. below standard score 85).

Using the LASS tests, only 3 out of 126 non-SEN students (2%) were identified as having difficulties and 5 out of 16 other SEN students (31%) were displaying a dyslexic profile. The validation measures indicated that 9 out of 49 non-SEN students (18%) were having difficulties and 2 out of 7 other SEN students (29%) were similarly identified. Using the PhAB, 7 out of 48 non-SEN students (15%) were identified as having phonological difficulties and 3 out of 7 other SEN students (43%) were seen to have a phonological deficit. These results will be discussed in more detail in Chapter 12.

The ratio of males to females in the dyslexic group in this study is 4:1 (see section 10.5). This is comparable with ratios found by several researchers (Critchley, 1970; Snowling, 1987; and Zabell and Everatt, 2000). However, among those identified as having difficulties, the ratio of males to females varies according to the screening test used. The LASS Secondary system identifies three males to one female and this ratio changes to 2.5:1 when regression to the mean is allowed for. Whereas, the validation measures identify five males to one female and this ratio changes to 3:1 when regression to the mean is allowed for. When the PhAB is used on its own, the ratio of males to females, identified as having difficulties, is extremely high (10:1). These findings support the conclusion given by Miles, Haslum and Wheeler (1998) that the apparent differences in gender ratio reported in the literature have arisen because different criteria for diagnosing dyslexia have been used.

The results of the validity study suggest that there is no significant effect of SEN group in students' preference for computerised or pen and paper tests (see section 10.5). Within all three groups, the majority of students show a preference for computerised tests. However, although the effect is not significant, 94% of the dyslexic group and 86% of the other SEN group prefer the computerised assessment over the traditional pen and paper tests, compared to 67% of the non-SEN group. This finding supports the conclusion by Singleton, Horne and Vincent (1995) that children with special educational needs, who feel that they are more likely to perform poorly on a particular test, show a greater preference for computerised assessment.

11.6 Questionnaire study

The majority of LASS Secondary users had no difficulties (see section 10.6) with installing the system (96%), registering pupils (96%) or running the tests (88%).

77% of users reported that the pupils enjoyed most or all of the tests, with the auditory memory test being the most popular test and the reasoning test being the least popular.

All of the tests were rated as being suitable tests by all of the teachers with the exception of syllable segmentation and sentence reading which were rated by 8% of users as being 'not very suitable'.

Most respondents found using the reports on LASS Secondary satisfactory or easy (93%) and the results were rated as being satisfactory or useful by 92% of users.

Overall, LASS Secondary was rated as good or very good by the majority of respondents (92%), whilst the other 8% reported that it was 'average'. 96% of users stated that they would be using LASS Secondary in the future and 4% were unsure.

The teachers of the schools involved in the LASS Secondary validity project were asked some further questions concerning the pen and paper tests. Some respondents reported experiencing difficulties running some of the pen and paper tests, including WMS-III spatial span (50%), WISC-III digit span (25%), NFER sentence completion test (25%) and the BSTS 3 spelling test (25%).

Three of the LASS Secondary modules were reported by all of the users to be preferred by pupils over the equivalent pen and paper tests (auditory memory, non-word reading and spelling). The LASS Secondary sentence reading test was reported by 75% of users to be preferred by pupils over the NFER sentence completion test. The remaining LASS Secondary modules (visual memory, syllable segmentation and reasoning) were reported by 50% of the respondents to be preferred by pupils over the equivalent pen and paper tests. None of the pen and paper tests were reported by the majority of users to be preferred by pupils over the computerised tests.

11.7 Summary and conclusions

LASS Secondary has been shown to be an appropriate test for the age range of 11 to 15 years, with items suitable for a range of ability levels. The use of the three adaptive modules reduces the assessment time whilst retaining the accuracy of the test. LASS Secondary contains standardised norms and is completely objective in its test administration. The majority of students show a preference for the computerised LASS Secondary assessments over the traditional pen and paper tests. Furthermore, the system was highly rated by teachers for its ease of use, suitability of tests and usefulness of reports.

There is evidence of a gender difference, favouring girls, on the LASS spelling test but not on the other LASS modules. There are no apparent differences on the LASS tests on the basis of ethnic background or language, although white students did score lower than students from other ethnic backgrounds on the LASS reasoning retest. The pen and paper tests, used as validation measures, showed more evidence of group differences on the basis of gender and ethnicity. These results suggest that the computerised LASS Secondary modules are relatively fair assessment tools compared to more traditional methods.

There is evidence of the technical quality (internal consistency, test-retest reliability and concurrent validity) of LASS Secondary. The modules show high levels of internal consistency that compare favourably to those of similar tests. LASS Secondary is also fairly reliable over time, although the coefficients vary somewhat between modules. Nevertheless, these are comparable to reliability coefficients reported for several well-known conventional tests. The results also show that the LASS test scores correlate well with the validation measures and, again, these are similar to other validity coefficients reported in the literature.

Obviously, the most important function of a dyslexia screening system is that it can differentiate between students with and without dyslexia and students with other special educational needs. The results of the dyslexia project suggest that LASS Secondary does indeed fulfil this function. Non-SEN students score significantly higher than dyslexic students on all tests except reasoning and visual memory and significantly higher than other SEN students on all tests. 79% of the dyslexic sample were correctly identified by LASS Secondary as having difficulties using a discrepancy model. This is comparable to the validation measures and preferable to using a test of phonological ability (e.g. PhAB). Only 2% of the non-SEN students were incorrectly identified as having dyslexia and, again, this result compares favourably to the validation measures and the PhAB used alone.

The results of these studies suggest that the LASS Secondary system generally fulfils the requirements of dyslexia assessment put forward by Reid (1998). The system identifies students' general strengths and weaknesses, indicates the student's current attainments in reading and spelling, indicates if the student's lack of progress is due to low intelligence or a specific learning difficulty and identifies particular errors in students' reading and spelling. Furthermore, LASS Secondary identifies difficulties (in decoding and encoding of print, phonological processing and auditory and visual memory), discrepancies (between intelligence and literacy skills) and differences (between individual students). These three aspects (difficulties, discrepancies and differences) are proposed by Reid (1998) to be paramount to the assessment process.

The LASS Secondary system encompasses the elements put forward by Turner (1997) as being important in dyslexia assessment. It includes tests of single word reading, which Turner argues is essential in the assessment of dyslexia, as well as a cloze test of sentence reading. Furthermore, as the sentence reading test is adaptive, the test can be administered every term to monitor students' progress. LASS Secondary also addresses some of the problems highlighted by Sutherland and Smith (1991). They suggest that there are difficulties in finding tests that are standardised, that can discriminate between groups of children with specific learning difficulties, that are easy to interpret, are appropriate to the age range and are not too time-consuming but give enough detailed information. Furthermore, LASS Secondary is designed for use by teachers as well as educational psychologists, which should avoid students with dyslexia remaining unidentified and, where necessary, provides teachers with evidence for the need for a referral to an educational psychologist for a full assessment.

Although the LASS Secondary system is based upon the ability versus attainment discrepancy model, there is scope for making use of other discrepancy models when looking at students' results. For example, Reid (1998) suggests that comparisons can be made between scores on phonological skills or single word reading and chronological age. The norms for LASS Secondary are age-related and it is therefore easy to see a discrepancy between a student's scores on single word reading or non-word reading and the average for their age group (i.e. percentile 50). The difficulty with this type of discrepancy model is that there is a tendency to identify 'garden variety' poor readers, rather than students whose performance is significantly different from their ability (Wright and Groner, 1993). However, use of the diagnostic tests in LASS Secondary (i.e. visual and auditory memory and phonological processing) helps to avoid this.

Researchers who subscribe to the phonological deficit hypothesis have suggested that tests of non-word reading (Rack, Snowling and Olson, 1992; Turner, 1997; Snowling et al, 1997) and phoneme deletion (Snowling et al, 1997; Padget, Knight and Sawyer, 1996) can discriminate between students with and without dyslexia. Again, using LASS Secondary, it is possible to make such a decision,. However, it is also argued that dyslexia is not simply a phonological processing difficulty and other diagnostic information is necessary (Nicolson and Fawcett, 1995a; Beaton, McDougall and Singleton, 1997; National Working Party on Dyslexia in Higher Education, 1999).

Fawcett, Pickering and Nicholson (1993) suggest that the computer-based approach to dyslexia assessment provides "the opportunity for constructing a new generation of psychometric tests, more sensitive than traditional tests and more easily administered, thus de-skilling the administration requirement and enabling low-cost screening for dyslexia (and other problems)" (Fawcett, Pickering and Nicholson; 1993; pages 489-490). The present study with LASS Secondary has borne out their predictions.

12 General discussion

12.1 Overview

In this chapter, the findings from both Study 1 (baseline assessment) and Study 2 (dyslexia screening at secondary school level) are discussed with specific reference to:

- gender bias
- how the use of computers addresses the challenges of educational assessment
- the advantages of computerised educational assessment
- the disadvantages of computerised educational assessment
- changes in baseline assessment
- changes in dyslexia screening
- outcomes of assessment

12.2 Gender bias

The findings of the present studies, that there are no significant gender differences in the computerised baseline literacy (see section 6.9.1) or in the secondary reading and spelling tests (see section 10.4.1), are contrary to the results of previous studies, which suggested that girls outperform boys in conventional pen and paper literacy tests (see section 2.5). It appears likely that the use of a computerised test eliminates bias in the administration and scoring of tests. Indeed, Friedman and Davenport (1998) report that teachers' attitudes toward their students has been suggested as an explanation for gender differences in tests. The result of the present study is not likely to be due to small sample sizes as significant differences between males and females were found within the same samples when administered conventional pen and paper literacy tests.

Another explanation for boys achieving similar scores to girls on the computerised tests is that boys may have more previous experience with computers and this could boost their scores. A number of researchers have reported that girls do not view computers as positively as boys do (e.g. Fife-Shaw et al, 1986; Hughes, Brackenridge and Macleod, 1987; Hoyles, 1988; Durndell, 1991). Crook (1996) suggests that girls' attitudes towards technology may become more negative as they go through school. This is consistent with the finding of Newton and Beck (1993) that, in the early 1990s, the percentage of women studying for computer science degrees was falling. Martinez and Mead (1988) found that, in general, males demonstrate a slightly higher level of computer competence than females. Scott, Cole and Engel (1992) report that "at the student level, many studies show considerable differences between the computing experiences of boys and girls. Boys habitually have more access, whether at school, home, or recreational (arcade game) computers. Where computer programming is offered, more boys take the subject than girls (the girls take word-processing courses). Parents are more likely to buy computers for their sons than their daughters, and boys are more likely to attend after-school computer club meetings" (pages 227-228). However, Crook and Steele (1987) found no significant gender differences amongst reception class pupils choosing to engage in computer activities in school and Essa (1987) obtained similar results among pre-schoolers. Crook (1996) concludes that 'gender-based attitude differences are not convincingly present at the start of schooling: they must somehow be cultivated within the early school years' (page 25).

Nevertheless, if gender differences in computer interest do exist, then these could bias the results of a computerised assessment. However, in the present studies, no significant gender differences were evident using the computerised tests for either secondary school pupils (aged 11 to 15) or reception class pupils (aged four or five). At the early age of four, it is unlikely that boys have more experience of computer use than girls do. Additionally, the CoPS Baseline literacy module and LASS Secondary reading and spelling tests are not speeded and therefore more adept use of a mouse is unlikely to be an advantage. Taylor et al (1999) also found no relationship between computer familiarity and level of performance on a computerised test of English as a foreign language, after controlling for English language ability.

Alternatively, Maccoby and Jacklin (1974) suggest that males are more competitive than females and so, according to Hayes and Waller (1994), this may interfere with accurate performance under situations such as group testing. In the baseline validity study, both the computerised CoPS Baseline tests and the comparative QCA Scales were administered individually and therefore the explanation given by Hayes and Waller (1994) for the gender difference would not apply. Nevertheless, the conventional reading and spelling tests used in the LASS Secondary validity study were group tests (see section 9.5.2), whilst the computerised tests were administered individually in most cases. It is therefore possible that the male students attempted to finish the conventional tests quickly and so made more mistakes, whereas during the individual computerised tests boys were under less pressure to complete the test quickly. One school had the LASS Secondary program installed on several machines within a computer room so that students could be tested simultaneously (wearing headphones). In such situations, it may be a necessary precaution to advise that students being tested at the same time be administered the tests in different orders.

12.3 Addressing the challenges of education assessment

Ebel (1972) reported that the increase in the use of educational tests has been accompanied by an increase in the criticism of such tests. One criticism is than tests vary in quality with some tests used in schools being particularly poor. However, the results of both the baseline and the secondary studies show that CoPS Baseline and LASS Secondary have high levels of internal consistency (see sections 6.2 and 10.1), test-retest reliability (see sections 6.9.3 and 10.3) and validity (see sections 6.9.4 and 10.4).

Another criticism of educational testing is that it may damage a child's self-esteem or decrease motivation. The pupils in both the baseline and the secondary studies expressed an overwhelming preference for the computerised assessment over the conventional tests (see section 10.4). There were no significant gender differences in preferences for one or the other type of test. These findings suggest that computerised tests have considerable potential in the field of educational assessment, particularly if the pupils find the tasks to be enjoyable. Additionally, through the use of adaptive testing, pupils do not have the experience of failing lots of items and are therefore less likely to have negative feelings about the assessment itself. With both CoPS Baseline and LASS Secondary, the results are not directly available to the pupil and therefore competition between pupils is minimised. Moreover, LASS Secondary may identify students who have been struggling in school for many years as having dyslexia. The fact that pupils are struggling in school is likely to have a severe negative effect on their self-esteem. Hence, identifying a particular reason for their difficulties (e.g. dyslexia) may actually raise their self-esteem. A number of researchers have found links between academic performance and self-esteem (Burns, 1982; Chapman, Lambourne and Silva, 1990) and many have reported that poor readers have low self-esteem compared to other children (Lawrence, 1971, 1987; Butkowsky and Willows, 1980; Gjessing and Karlsen, 1989; Saracoglu, Minden and Wilchesky, 1989; Huntington and Bender, 1993). Sadovnik (2000) suggests that children with learning disabilities have low levels of self-esteem and higher levels of aggression (which may be a cover for low self-esteem). A child who thinks that he or she has failed may vent his or her frustration and anger on others. Riddick et al (1999) also report evidence that dyslexic adults have significantly lower self-esteem than a matched control group. The dyslexic group reported themselves as feeling more anxious and less competent in their written work at school and less competent in their written work and academic achievements at university. Additionally, Riddick (1996) found that mothers reported particularly low self-esteem in their dyslexic children before their problems were identified and specific support was offered. Furthermore, LaFrance (1997) argues that children who are gifted and dyslexic have lower self-esteem and poor motivation. Arguably, it is these gifted dyslexics who are most likely to proceed through the education system without being identified as having dyslexia because they are more adept at developing strategies to compensate for (or, perhaps conceal) their difficulties. As Ebel (1972) suggests, educational tests should be evaluated in terms of how much they increase achievement by motivating and directing the efforts of students and teachers. Both CoPS Baseline and LASS Secondary allow teachers to see the strengths of pupils as well as their weaknesses.

Educational assessment has also been criticised for encouraging the development of a single ability and thus reducing the diversity of talent within society. However, most subjects included in the school curriculum have a reliance on good literacy skills and so the testing of such skills with LASS Secondary, in order to identify difficulties and aid in the planning of teaching, will have the effect of widening access to other parts of the curriculum. A further criticism is that assessment which is intended to support the curriculum can come to dominate it, as what is assessed is likely to be taken as an indication of what is important (Hambleton and Murphy, 1992; Mehrens, 1992; Brady, 1997). Again, good literacy skills are an important prerequisite to many curriculum subjects and the importance of literacy is unlikely to increase just because it is widely assessed or because computers allow for easier assessment of literacy skills.

According to Singleton (1997a), equivalence of computerised assessment with conventional test forms is key to the acceptance of computerised assessment in education. Any computerised assessment, as with any traditional assessment, must be shown to be both valid and reliable. In the early years of computerised assessment, the issue of equivalence was generally focussed on the translation of existing tests to a computerised format. Equivalence is particularly important where speed of response is critical to the test, because clicking with a mouse is different from ticking a box or producing an oral response. Furthermore, reading text on a computer screen may be more difficult than conventional reading. Indeed, Dillon (1992) reports that reading from a computer screen has been found to be 20 to 30% slower than reading paper-based text. Demonstrating equivalence of conventional assessments and computerised assessment generally requires evidence of a high correlation between the two formats. Singleton (2001) found that, in a study using computerised and conventional tests of verbal and non-verbal (mental rotation) intelligence, the different versions correlated well, indicating that the format used does not significantly affect the ability being assessed. Both the CoPS Baseline and LASS Secondary systems were originally designed as computerised assessments rather than being computerised versions of conventional pen and paper tests. Evidence of equivalence between forms is therefore irrelevant. However, there is evidence from both the baseline and secondary studies that the computerised tests are highly correlated with other conventional pen and paper tests that measure similar abilities and attainments (see sections 6.9.4 and 10.4).

12.4 Advantages of computerised assessment

There appear to be a number of advantages of computerised assessment over conventional assessment in education. Since the computer does most of the work of assessment, including administering items, recording responses and scoring results, there are significant savings in labour and cost in using computerised assessment compared with conventional assessments (French, 1986; Woodward and Rieth, 1997). The time taken to administer CoPS Baseline is 20 minutes, which is the same as the time taken to administer the QCA scales (or, in theory, any other baseline assessment scheme as all systems are required to adhere to the QCA accreditation criteria of an average testing time of up to 20 minutes). However, the CoPS Baseline system saves time as there is no preparation involved (e.g. the teacher does not have to select suitable texts, as they would using the QCA scales) and the scoring is carried out by the computer. It takes each pupil an average of 41 minutes to complete all eight modules in the LASS Secondary system, compared to an average of 94 minutes to complete the equivalent eight tests used in the validation study. LASS Secondary is also time-saving as pupils aged 11 to 15 can carry out the tests by themselves whilst the supervising teacher continues with other work. Again, the system requires no time in preparation or scoring. Several researchers have shown that teachers prefer computerised assessments because the results are immediately available, saving time in scoring responses and calculating standard scores (Wesson et al, 1986; Fuchs et al, 1987; Fuchs, Fuchs and Hamlett; 1993; Woodward and Rieth, 1997). Time savings are maximised with the use of adaptive computerised assessment. The adaptive forms of the tests used in both the baseline and the secondary studies correlate highly with the full forms (see secions 6.2.1, 6.2.2 and 10.1) and the baseline study shows that the adaptive literacy and mathematics tests take approximately 35% of the time taken to administer the full forms (see section 7.2).

Computerised assessment allows for greater precision in the timing and delivery of test items and measurement of responses. This is particularly important where timing is critical, for example in the assessment of short-term memory (Singleton and Thomas, 1994). Computerised assessment ensures the standardised presentation of digits on the LASS Secondary auditory memory tests, whereas with human administered digit-span tasks some variation is inevitable. Arguably, this helps to improve the reliability of measurement (Singleton, 1997a). Computers also allow the use of tests that would be very difficult, or even impossible, to be administered by a human. An example of such a test is the LASS Secondary visual memory test which involves remembering the positions of different phantoms in a cave. Replicating this task without the use of a computer would be practically impossible. Additionally, response time data, which is difficult to measure using traditional tests, enables distinctions to be drawn between children who are accurate and fast, and those who are accurate but much slower in their responses (Singleton, Thomas and Leedale, 1996).

Computerised assessment may also increase the motivation of individuals being tested. Students with special educational needs often respond negatively to traditional pen-and-paper tests (Wade and Moore, 1993). However, a number of researchers have reported that children and adults, particularly those of low ability, feel less threatened by computerised assessment than by conventional assessment (Skinner and Allen, 1983; Moore, Summer and Bloor, 1984; French, 1986; Watkins and Kush, 1988; Singleton, Horne and Vincent, 1995; Singleton, 1997a). Furthermore, in an adaptive test, pupils are not administered items that are far too easy or far too difficult for them, which also enhances test motivation (Singleton, 1997a). In the present study, all groups of children displayed a clear preference for the computerised assessment over the conventional assessment (see section 10.4).

Computerised assessment allows the assessment of children with severe physical disabilities or profound sensory impairments. A wide variety of applications of computerised assessment in special education have been noted (for review see Woodward and Rieth, 1997). Certainly children with co-ordination difficulties or physical disabilities can be easily tested on the majority of the CoPS Baseline and LASS Secondary tests by using a touch screen or by having the teacher use the mouse on the child's behalf. Pupils with moderate or severe learning difficulties have also been successfully tested using CoPS Baseline and LASS Secondary. Furthermore, it is possible, to a considerable extent, to test children with a hearing impairment or mild visual impairment using these tests.

Computerised assessment also allows for wider access to educational assessments, although the potential of the internet for delivery of up-to-date versions of tests anywhere in the world has not been exploited. This should prevent the use of old or obsolete tests that have outdated norms, as any program changes (either for fixing bugs or updating norms) need only be carried out on the server. There has been a 'phenomenal' growth in electronic communication networks, in particular the internet (Buchanan and Smith, 1999) and internet connection now comes as standard with new computers. Buchanan and Smith (1999) report that a possible difficulty with internet-based testing is that different users will be using different browser software packages, configured differently, running on different computer platforms with different displays.

This may affect the speed at which the test items are administered, which is particularly important in timed or speeded tests. For example, a digit-span test usually involves administration of digits at a rate of one per second. Any divergence from this rate will have a major impact on the test results of a particular student. However, Bartram (1999) reports that we are moving towards common standards for browsers. Furthermore, this problem can be overcome by downloading time critical material as an applet (Bartram, 2000).

Bartram (2000) argues that there are a number of issues arising from the use of distance assessment. These include: ensuring that all test-takers have had sufficient practice without being over-exposed to specific test content; ensuring the person taking the test is actually who they say they are; ensuring that test-takers are taking the test unaided; ensuring the test-takers results are held confidentially; and, protecting the test publisher's intellectual property rights. However, if an internet-based educational test, e.g. for screening for dyslexia, can only be accessed by registered users, who are qualified teachers or psychologists, then most of the potential difficulties listed above would not apply as the teacher or psychologist would have control over which individuals are administered the test. As Bartram (2000) states, "Just as stand-alone computer-based tests require the presence of a qualified test administrator, so will distance-assessment techniques" (page 269). Furthermore, only the test items and report would be publicly available over the internet, whilst the scoring process would remain secure on the server.

12.5 Disadvantages of computerised assessment

Due to programming and other technical requirements, the development, validation and standardisation of a computerised test usually takes longer and is more expensive to develop than an equivalent conventional test. Thus, compared with conventional tests, computerised tests are often more expensive for schools and psychologists to purchase initially, although this expense is likely to be offset by savings in personnel time and costs of administration. For example, CoPS Baseline costs £85 for one year and LASS Secondary costs £125 for one year, with both of these licences being for an unlimited number of pupils. However, the costs of the equivalent pen and paper tests used as validation measures for the secondary studies are considerably higher. For example: the NFER group reading sentence completion test costs £10.85 for a manual and £56 per 100 record forms; the British Spelling Test Series 3 costs £9.35 for a manual and £93.50 per 100 record forms; the Matrix Analogies test costs £42 for a manual and £188 per 100 record forms; and, the Phonological Assessment Battery costs £79.50 for a manual and £108.50 per 100 record forms. Therefore, to test 100 pupils on each of these assessments costs £5.88 per pupil, whereas to test them on LASS Secondary costs £1.25 per pupil.

There are also limitations in the type of assessment that are suitable for computer administration. Computerised assessment is best suited to measurement of cognitive and intellectual abilities, fact knowledge, basic skills and curriculum content. A number of important aspects of behaviour are currently impossible to measure using computerised assessment, including expressive language, social and emotional behaviour and any assessment involving the reading of large amounts of text. Another disadvantage of computerised assessment is the risk of technology failure, although the reliability of hardware and software is improving.

There is a further danger that computerised tests may be used by untrained or inexperienced users, although this is also the case for conventional tests. However, both the CoPS Baseline and LASS Secondary programs are sold only to qualified teachers and psychologists. As recognised by the British Psychological Society, "...the ease of computerised assessment construction does pose a serious threat to effective and fair practice in the use of psychological tests" (BPS, 1999). To counter this threat, the BPS formulated guidelines for the development and use of computerised assessment (BPS, 1999). These guidelines assert that all computerised assessments should be supported by clear documentation on rationale, validity, appropriateness, administration and interpretation, and that users need to be aware of what constitutes best practice in computerised assessment, so that they can make informed evaluations and choices between available systems. The test manuals for both CoPS Baseline and LASS Secondary contain information concerning all these key areas specified in the BPS guidelines.

12.6 Changes in baseline assessment

The current arrangements for baseline assessment are set to change in the future with the introduction of a new statutory assessment at the end of the reception year. However, if schools wish to continue to assess children on entry to the reception class in addition to the new statutory assessment at the end of reception, progress over that period of time will be tracked (Merrell, 2001).

Lindsay, Lewis and Phillips (2000) report that although it is more common for schools to use a single national assessment schemes, such as the end of Key Stage 1 assessment, this is not universal (for example, GCSE and 'A' level examinations are set by several different boards). They argue that when different schemes are approved for a single purpose, the comparability of schemes becomes more important. However, baseline assessment schemes have been used for a number of different purposes, including guiding teaching in the early weeks of reception and the measurement of value-added, and 91 different schemes have been accredited by QCA. Lindsay, Lewis and Phillips (2000) conducted a survey evaluating accredited baseline assessment schemes. Data were obtained from 982 schools and 102 LEAs, and follow-up interviews were conducted in 16 LEAs. One head teacher stated that neighbouring schools appeared to be "making lower assessments because of the value-added element" (page 47). They also found that most teachers and head teachers wanted a single national baseline assessment scheme to be introduced in order to make transfer of information between schools easier (in the case of children changing schools) and to tie the assessment in with the National Curriculum. A few teachers and head teachers were against a single national baseline assessment scheme because it would waste the data that had already been built up and because the work put in by their education authority in devising a scheme showed how much the authority valued baseline assessment. Lindsay, Lewis and Phillips (2000) add that 'with this came an implicit trust that the authority had produced a scheme most suitable for their own children and situation' (page 49). They also report that teachers and head teachers assume that baseline assessment data is reliable and valid. One conclusion that may be drawn from this is that QCA should have required evidence of reliability and validity as a criteria for accreditation of baseline schemes and then schools would be able to rely on the data from those accredited schemes.

Lindsay, Lewis and Phillips (2000) also comment on the effect on baseline assessment results of the age at which children start school. Some LEAs admit children at the start of each term whilst others have one intake in the September of the school year in which the child will be five, in which case many children being assessed are likely to be developmentally immature compared with other children in the class. They recognise that this is not a problem if baseline assessment results are used for planning teaching but it is problematic if the assessment information is used for value added purposes. However, normative baseline assessment schemes, such as CoPS Baseline, give centile scores for different age bands, thus overcoming this problem.

Staff in special schools were concerned that many baseline assessment schemes did not usefully include children with special needs. As Lindsay, Lewis and Phillips (2000) state, "given the current focus on inclusion it was ironic and unfortunate if children with special needs were marginalized or overlooked in baseline assessment schemes" (page 54). With CoPS Baseline, children with known special educational needs (or where an assessment of special needs is pending) can still be given some assessment. If a child has a profound hearing impairment, the text spoken by the computer can be signed to the child using British Sign Language or Signed Supported English. CoPS Baseline is not suitable for children who are blind, although children with a moderate degree of visual impairment may be tested. Children with physical disabilities or co-ordination difficulties can be assessed on CoPS Baseline with either the teacher using the mouse on behalf of the child or the child using a touch screen. Children with moderate or severe learning difficulties may be tested at a later stage and age equivalent scores calculated using the tables in the Teacher's manual.

Lindsay and Lewis (2001) argue that baseline assessment at the end of the foundation stage, using a single scheme, could improve accountability and aid the SEN system, regarding comparability of children's developmental levels for resourcing. However, they suggest that early identification requires action at school entry rather than at the end of the reception year. Teachers appear to be positive about the timing of school entry assessment for getting to know pupils, target setting within the class and as a basis for discussion with parents (Lindsay and Lewis, 2001). However, they suggest that the removal of the statutory requirement could be either negative (in diminishing the appeal of baseline assessment) or positive (in re-accentuating its child-centred nature). Some scheme providers may wish to alter on-entry tests to improve their reliability as they will not be constrained by specific accreditation criteria. For example, the CoPS Baseline system could be changed to include more items in the literacy and mathematics modules (previously this would have been impossible if the whole assessment was to conform to the 20 minutes administration time specified by QCA). This should increase validity and reliability, and could make it a better predictor of special educational needs. Lindsay and Lewis (2001) conclude that the new baseline assessment at the end of the foundation stage should be a part of a continuous system of child-focussed monitoring, drawing upon the collective support of the LEA and schools, respecting differences in needs (e.g. special schools) and providing information to optimise children's learning by appropriate teaching.

12.7 Changes in dyslexia screening

The discrepancy approach to identifying dyslexia has been the main method used for a number of years. Using this method, individuals are identified as having dyslexia if there is a significant difference between their reading or spelling scores and that predicted by a general cognitive ability measure. There are two main approaches to the discrepancy method. The first is a cut-off approach, e.g. IQ must be at least 90 and reading age must be 18 months behind chronological age. The second is a regression approach, e.g. reading age must be at least 1.5 standard deviations below that expected on the basis of the individual's IQ. As Nicolson (2001) states, "The key difference between the two is that the regression method explicitly takes account of the child's IQ - so a child with an IQ of 140 and reading standard score of 100 (exactly normal) may turn out to be dyslexic. By contrast, the cut-off method might classify a child of IQ 92 and reading standard score 90 as dyslexic, whereas the regression method would not" (page 16). For this reason, use of the regression method is preferred over the cut-off approach.

Kirk, McLoughlin and Reid (2001) state that 'The principle of 'specificity', that is, the notion that dyslexic people have an underlying neurological inefficiency, is central to the screening and assessment process. Consequently, the accurate diagnosis of dyslexia requires the measurement of general ability and working memory' (page 295). They suggest that an assessment which does not include these cognitive tests is likely to produce both false positives (the incorrect identification of a person who has low intelligence as dyslexic) and false negatives (the failure to identify a dyslexic person as such because they have developed strategies that compensate for their dyslexia). Kirk, McLoughlin and Reid (2001) conclude that 'the reliance on the assessment of literacy skills alone as the sole or main diagnostic procedure is inappropriate and uninformed' (page 295).

However, Lindsay (2001) argues that the discrepancy method has often been misused or poorly used, as reading tests and tests of general cognitive abilities do not correlate perfectly and are not perfectly reliable, so affecting the predicted literacy score and the discrepancy needed. He reports that 'on the basis of the statistical properties of the tests, discrepancies need to be substantial before they can be considered significant' (page 259). Thus, the discrepancy approach has fallen out of favour with educational psychologists. Another difficulty with the discrepancy approach is that older children and adults with dyslexia may, to some extent, have overcome their difficulties in reading, thus reducing the discrepancy between their reading and their general ability. Nevertheless, spelling or other skills may be taken into account (Snowling, 2000) and Stein, Talcott and Witton (2001) suggest that, with such individuals, it is appropriate to identify their phonological difficulties. A longitudinal study conducted by Shaywitz et al (1992) shows that, using a regression approach with a cut-off of 1.5 standard deviations, diagnosis of dyslexia is not stable over time. Only 28% of the children diagnosed as dyslexic at the end of the first grade were given the same diagnosis during the third grade and only 47% of these were diagnosed as dyslexic in the fifth grade. Additionally, Snowling (2000) asserts that a child who does no reading is unlikely to have literacy levels in line with expectation, but the reason for the discrepancy should not be assumed to be dyslexia. Nevertheless, she also suggests that it may be necessary to continue to use IQ in diagnosing dyslexia, as intelligence may mitigate the effects of the phonological deficit on reading skills. Furthermore, as stated by Torgesen (1989), "saying that phonological deficits are the key does not necessarily deny that low IQ can be a cause of poor reading skill" (page 485). Indeed, 16% of the variance in reading is accounted for by IQ (Snowling, 2000). Snowling (2000) states that 'the discrepancy approach needs to be supplemented by positive diagnostic markers that will allow practitioners to identify children who show early or residual signs of dyslexia that require intervention, and do not depend solely on the extent of the child's reading problem' (page 25).

The relationship between phonological awareness and the development of literacy is well established (Liberman et al, 1974; Bradley and Bryant, 1978; Vellutino, 1979; Lundberg, Olofsson and Wall, 1980; Snowling, 1981; Bradley and Bryant, 1983;

Goswami, 1986; Stanovich, 1986; Wagner, 1986; Bryant and Goswami, 1987; Snowling, 1987; Stanovich, 1988; Snowling, 2000). Indeed, the British Psychological Society (1999) views the identification of phonological difficulties as the key to identifying dyslexia and Stanovich (1996) argues that such difficulties are evident in the full range of IQ scores. Furthermore, Lundberg and Høien (2001) state that the association between phonological awareness and reading acquisition is "one of the most robust findings in developmental cognitive psychology and it has been replicated over and over again across several languages, ages, and tasks used to assess phonological awareness" (page 112). A number of researchers report that individuals with poor phonological skills will have difficulties reading non-words (Snowling, 1980; Rack, Snowling and Olson, 1992; Lindsay, 2001). According to Lundberg and Høien (2001) the most characteristic phonological difficulties in individuals with reading problems are segmenting words into phonemes, keeping strings of letters in short term memory, repeating long non-words, reading and writing non-words, rapid naming, speech and manipulating phonemes. There is also evidence of phonological difficulties in teenage and adult dyslexics (Høien and Lundberg, 1989; Pennington et al, 1990; Bruck, 1992; Elbro, Nielsen and Petersen, 1994; Gallagher et al, 1996), which suggests that older dyslexics do not overcome these problems. However, Wolf and O'Brien (2001) report that 'there have been no significant differences found for phonological awareness measures between discrepant and non-discrepant readers' (page 127). Stanovich and Siegel (1994) report that all poor readers differ from normal readers in phonological awareness but dyslexic readers differ from other poor readers in working memory and listening comprehension.

Several researchers have suggested that, as well as the phonological deficit in developmental dyslexia, there is a second core deficit of rapid processing, which is necessary for fluent reading and reading comprehension (Denckla and Rudel, 1976; Wolf, 1986; Bowers and Wolf, 1993; Van Daal and Van der Leij, 1999; Wolf and O'Brien, 2001). Wolf and O'Brien (2001) report that this problem is seen in dyslexics from kindergarten through to adulthood and that dyslexic individuals fall into three subtypes (those with phonological deficits only, those with naming-speed deficits only, and those with both deficits). It is suggested that dyslexics with a naming-speed deficit, but no phonological deficit, would be missed by most diagnostic batteries. Additionally, Wolf and O'Brien have found that there are differences between dyslexics and garden-variety poor readers in how they name letters and numbers. Dyslexic adults have also been found to show rapid naming deficits (Felton and Wood, 1989; Pennington et al, 1990). According to Wolf and Bowers (1999), some dyslexics will have serious reading problems due to the 'double deficit' in phonological skill and naming speed.

There is a view, particularly in the United States, that 'garden-variety' poor readers (i.e. those without a discrepancy) exhibit the same phonological difficulties as dyslexics and there is therefore no need for a distinction to be made between the two groups (i.e. there is no need to identify 'dyslexia' but merely 'poor reading'). Snowling (2000) argues against the discrepancy definition, suggesting that the only difference between dyslexic readers and generally backward readers is in IQ and this does not necessitate differences in remediation. However Nicolson (1996) contends that, although the symptoms may appear similar, the underlying causes and the best approach to treatment may be different and so the two groups should be distinguished. Indeed, Hatcher (2000) found that dyslexic children benefited more from an intervention linking reading with phonology than did children with moderate learning difficulties. Additionally, Olson et al (1999) found a higher genetic susceptibility for dyslexia than for garden-variety poor reading. Share et al (1987) suggest that the high proportion of boys within the dyslexic population is due to the fact that reading scores are lower in boys and so when a cut-off point is used in diagnosing a reading difficulty, it will identify more male poor readers than females. However, the LASS Secondary reading test does not appear to show this bias in favour of girls.

In the last few years, it has been proposed that reading difficulties are caused genetically by the impaired development of magnocells in the brain, which control the timing of sensory and motor events (Stein and Walsh, 1997; Stein, Talcott and Witton, 2001). According to Stein, Talcott and Witton (2001), these visual and auditory impairments can be identified early and remediated before children begin to fail at reading. Deficits in the cerebellum are being seen as increasingly important in reading development (Nicolson and Fawcett, 1995a; Fawcett, Nicolson and Dean, 1996). Fawcett and Nicolson (2001) also suggest that abnormalities in the cerebellum, which is known to be involved in skill learning and automatisation, particularly in language-related cognitive tasks, are characteristic in dyslexic children. According to Fisher and Smith (2001), "a large amount of data from many families will be required in order to pinpoint the specific gene variants that predispose to developmental dyslexia in the general population" (page 59). However, even if the genes that predispose individuals to dyslexia are identified, and this appears to be some way off, there is still a need for effective tests to measure levels of difficulty and monitor progress.

There appears to be a growing acceptance of the use of computers in dyslexia screening with the development of programs such as Lucid CoPS (Singleton, Thomas and Leedale, 1996), LASS Secondary (Horne, Singleton and Thomas, 1999), LASS Junior (Thomas, Singleton and Horne, 2001) and LADS (Singleton, Horne and Thomas, 2002).

Indeed, Nicolson, Fawcett and Miles (1993), having carried out an international survey of researchers and practitioners in adult dyslexia, report that:

> there was consensus that new developments in computer technology made it feasible to introduce computer-supported testing procedures that do not require the direction of a trained clinician / diagnostician, and could therefore be carried out cost-effectively in centres such as adult literacy centres, units for young offenders or job centres, subject to the provisos that a follow-up second stage testing procedure was available and that the screening was integrated within a support framework (page 8).

12.8 Outcomes of assessment

12.8.1 CoPS Baseline

The value of baseline assessment is that it helps the teacher to establish the strengths and possible weaknesses of the individual child from the beginning of formal education. Careful use of that information can help to ensure that planning for both the class and the individual pupil is as effective as possible. The creation of a secure, stimulating learning environment is a continuing challenge to every teacher. This entails providing surroundings which maximise pupils' concentration and time on task, and ensuring that the work is set at an appropriate level so that all children achieve their potential. Children begin school having come from a variety of backgrounds. Some will have had pre-school experience in play groups or nurseries. Some pupils will come from homes where there is little conversation and few (if any) story books and where life is governed by the television. Others will have a rich experience of conversation, stories and other meaningful activities. It is important to remember that children develop in different ways and at different rates. Every experience they have helps to mould their perceptions and outlook on learning and life. Consequently, successful and fulfilling early school experiences are vital in ensuring that a real desire to learn is maintained.

The importance of providing suitable teaching styles to meet the variety of individual learning styles within a class cannot be over-emphasised. All individuals use a variety of senses but each of us has a preferred sensory modality, that is a sense that we prefer for receiving and processing information. That preference may change depending on what we are doing, but usually one will be used more frequently than others. The 'visual learner' prefers to learn by seeing and so finds visually presented material easier to understand and remember. He may have difficulty with auditory skills such as recalling sounds and names that will affect phonic ability. He prefers a visual approach to reading and spelling. It can be difficult for him to sit and listen for very long, and he may be easily distracted by movement nearby. The 'auditory learner' finds it easier to remember information that he hears and tends to use a phonic approach to reading and spelling. He may have difficulty with visual skills and so may find the early stages of reading difficult if the school uses a 'Look and Say' approach. He may even learn a reading book off by heart. He usually thinks in sounds and is easily distracted by noise. The 'tactile/kinaesthetic learner' learns best by having direct involvement in what he is learning, and especially benefits from 'hands-on' experience. He likes to write words down to get the feel of them and generally prefers to learn through action. He can seem distractible and have difficulty attending at times, whether the information is visually or orally presented. Being aware that children learn in different ways enables the class teacher to use a variety of teaching strategies in order to accommodate the preferred learning styles within a class. A multi-sensory approach, which involves simultaneous input from different sensory channels and which can be observed in many infant classrooms, greatly assists the learning process.

CoPS Baseline does not provide direct assessment of special educational needs. However, when children score in the 'low' grade (i.e. lowest 5%) on one or more modules, this should raise serious concerns and further investigation of the child's needs should be carried out. If a child has two or more of the attainment modules (i.e. Communication, Literacy and Mathematics) scoring in the 'low' grade, this should be regarded as an indication of risk as such cases will fall in the bottom 2% of pupils in the population. In this case, further investigation of the child's needs should always be carried out.

Fundamental expressive language abilities are required for good communication in all areas of life and attention should be paid to all the important elements as assessed by CoPS Baseline. Speech is an important social skill and is a vital part of making relationships, expressing needs, beliefs and opinions. For each member of society, speech should provide a means of questioning, challenging and evaluating. We use language to clarify our own thinking, and it is often in verbalising that we learn to understand or arrive at a conclusion. Young children are still learning to manage language and remember what they learn. Labels are powerful tools for them and each one enlarges their personal vocabulary and bank of experience. Without that existing resource, decoding of language is difficult and reading text is meaningless. Children need to be taught to use language effectively and appropriately in a variety of situations, and to be allowed to express their opinions without fear of criticism. Communication is, however, a two-way process and it is as important to be able to listen as well as talk. Knowledge is usually imparted to children in the classroom by talk, at least initially. Pupils can then internalise that knowledge by a process of simplification and analysis, establishing patterns and relationships, using language that is familiar to them. Without the development of effective listening skills, children will almost certainly miss out on vital information, and may even foster inappropriate behaviour strategies for coping in the classroom.

In a society that expects a high level of literacy, it is essential that all pupils make as good a start as possible. Children should not be allowed to fail in their introduction to literacy, so early targets should be limited, achievable and differentiated depending on pre-school experience. Pupils should be allowed to proceed at their own pace, and repetition and practice should be built into programmes to ensure that learning is consolidated.

As with literacy, early learning in mathematics is vital. The sequential nature of the subject makes it imperative that the foundations on which knowledge is built are as firm as possible. Any gaps in the first stages can affect learning later when new concepts are being taught. Early experiences with the subject can make or mar future progress.

Education should encourage the development of suitable personal and social skills as well as moral values and academic progress. As children start school with a variety of different experiences, it is important for the class teacher to establish ground rules, so that pupils quickly learn what constitutes acceptable behaviour. Rules should be explained and discussed with the children so that they understand the reasons for them. Children should be taught as early as possible that they are responsible for their own actions. There should be praise for good behaviour and any threatened punishment should be carried out, so that the children learn that the teacher means what she says. Children often need to be taught how to learn and which strategies are possible in a given situation. Some will have had little experience of listening and need help to attend and concentrate. Others need to be shown how to develop personal skills and complete simple tasks because they do not learn incidentally. All class members should be taught to become active, creative, independent learners. All children need the friendship of others. Some will need to be taught social skills such as how to share and take turns. Each child within the class must feel valued as an individual if future problems are to be pre-empted. Counselling should be an intrinsic component of teaching sessions; humour should be kept alive and the importance of laughter remembered. The classroom environment should be one in which children are prepared to take risks. All pupils should be given opportunities to experience success.

12.8.2 LASS Secondary

Many secondary schools routinely assess the general abilities of all students (especially in verbal and non-verbal abilities) but sometimes in literacy attainment as well as mathematics and quantitative reasoning skills. In many cases, this is carried out at the point of entry to secondary education. LASS Secondary can fulfil several of these functions, including the non-verbal ability and literacy attainment components. When used for this purpose, it would not normally be necessary to administer the modules assessing memory (Cave and Mobile) or phonological skills (Nonwords and Segments), because these are essentially diagnostic tests.

LASS Secondary also provides schools with a straightforward screening system for special educational needs. When used for this purpose, students who gain low scores on any of the routine profiling modules (Reasoning, Single Word Reading, Sentence Reading and Spelling) or who display a significant discrepancy between their scores on Reasoning compared with their score(s) on Single Word Reading, Sentence Reading or Spelling, would automatically be administered the diagnostic modules.

The two main literacy modules in LASS (Sentence Reading and Spelling) are both adaptive tests that can be used at regular intervals to monitor progress. The minimum interval between administration of the same module on a second or subsequent occasion should be about four months. When a particular problem (e.g. specific learning difficulty or dyslexia) has been identified and intervention, such as specialist teaching, has been implemented, teachers will naturally wish to evaluate the student's response to that intervention.

Any individual LASS Secondary module result which falls below the 20th centile is significantly below average and thus indicates an area of weakness. This is a fairly conventional cut-off point in identifying special needs or moderate educational weaknesses. Sometimes a weakness is identified which can be remedied by appropriate training. In some cases the problem is more pervasive and requires a differentiated approach to teaching in basic skills. Where there is strong confirmation (e.g. a number of related tests at or below the 20th centile) then the assessor can be convinced that concern is appropriate. Where a student is scoring below the 5th centile on any particular module, this generally indicates a serious difficulty and should always be treated as diagnostically significant, and usually this will be a strong indication that a student requires intervention. Again, where there is strong confirmation (e.g. a number of related tests at or below the 5th centile) then the assessor can be even more confident about the diagnosis. However, it should not be forgotten that LASS Secondary is also a profiling system, so when making interpretations of results it is important to consider the student's overall profile. For example, a centile score of 30 for reading or spelling would not normally give particular cause for concern because it does not fall below the 20th centile threshold. But if the student in question had a centile score of 85+ on the reasoning module, there would be a significant discrepancy between ability and attainment, which would give cause for concern.

When specific areas of learning difficulty have been identified by LASS Secondary, there are a wide range of teaching strategies that can be used to build on the student's strengths to mitigate or remediate the weaknesses. Most schools will already have a range of reading and spelling activities, worksheets, prompt cards, teaching schemes and devices, which can now be selected and used in a more focused way. In general, strategies for addressing the learning problems of students in this age range will focus more on support than on remediation. The latter, particularly if it involves withdrawal from ordinary classes can often be embarrassing and stigmatising for an older student. The most important thing for dyslexic students at the secondary education stage is to be enabled to access the curriculum, despite their difficulties. This can be achieved by various strategies, including use of assistive technology and support assistants. However, some students may still need to improve their basic skills, particularly in phonic decoding, word recognition and spelling. In such cases, suitable computer software designed to provide stimulating practice in the appropriate areas, can often be the most acceptable and effective solution.

Dyslexic students who continue to experience persistent phonological difficulties into secondary age are likely to require particularly careful literacy teaching. In such cases, a well-structured multisensory approach incorporating plenty of practice in phonic skills (over-learning) is strongly recommended. Without adequate training in applying phonics, students with such weaknesses are liable to develop an over-reliance on visual (whole word) and contextual strategies in reading (especially if they are bright). This, in turn, will have a deleterious effect on their text comprehension, especially in dealing with more complex curriculum-related material.

Most students enjoy using a computer, so they tend to be well disposed to technology suggestions. There is often a reluctance to produce written material on paper, when it has to be re-written after spelling errors are corrected and/or great efforts still produce unattractive handwriting. Using any electronic keyboard removes much of the hassle of editing spellings and punctuation and the final printed product looks smart. For students who need more support, a computer with sound and a Windows environment, facilitates the writing process even more, as spoken prompts of errors come instantly and on-screen word banks provide access to a richer range of vocabulary.

12.9 Summary

The advantages and potential benefits of computerised assessment in educational assessment appear to outweigh their disadvantages and potential risks. As with conventional assessment, the data derived from computerised assessments are useful in monitoring the progress of pupils and identifying and assessing children with special educational needs. However, computerised assessments are beneficial as they enable teachers to carry out assessments that would otherwise require large amounts of time in learning test administration procedures, and in delivering and scoring tests. Furthermore, computerised assessments can allow young children (and disaffected pupils) to be assessed in a way that is motivating and enjoyable for them. They may also be more objective than conventional tests and reduce the traditional gender bias, favouring girls, on literacy tests. The most significant disadvantage of computerised assessments is the risk of misuse, although this risk applies equally to conventional assessment where untrained or inexperienced users are concerned. The British Psychological Society is in the process of instigating a programme of training and accreditation for educational personnel in the use of psychometric tests. It is to be hoped that this initiative, together with the adoption of guidelines on the development and use of computerised tests (BPS, 1999) will enable teachers to take full advantage of computerised educational assessments. Meier (1994) suggests that the increased efficiency offered by computerised tests will ensure their increasing use in test administration. Furthermore, Bartram (2000) reports that "there are clear signs that attitudes to computer-based assessment are changing as people come to appreciate the real benefits of technology for assessment, and as the technological infrastructure needed to support these applications becomes increasingly ubiquitous" (page 262).

It is clear that compulsory baseline assessment is facing a major change in the near future. However, the CoPS Baseline program will remain available for the assessment of 4 to 5½-year-olds, and its use in other countries is increasing. It is hoped that any system adopted as a national scheme in the UK for the assessment of children at the end of the foundation stage is standardised, reliable and valid.

The assessment of dyslexia appears to have shifted from the traditional discrepancy approach to the identification of a phonological deficit (and, more recently, a rapid naming deficit) and research is being carried out into the identification of magnocellular impairment and cerebellum abnormalities. The diagnosis of a predisposition to dyslexia through the identification of specific genes is some way off. Whichever method of identifying dyslexia is used, Salvia and Ysseldyke (1988) argue that there is an ethical responsibility to ensure that assessment decisions are '...based on both objective information and professional interpretation of that information' (page 56). The LASS Secondary program is capable of identifying a discrepancy between cognitive ability and reading or spelling skills as well as a providing evidence of deficits in phonological processing and memory using objective, standardised measures that have been shown to be reliable and valid.

13 Conclusions and proposals for future research

"The computer has been acknowledged as an integral part of all aspects of modern life at least since Time magazine celebrated it in 1983 as 'man of the year" (Reinking, Labbo and McKenna, 2000, page 114).

13.1 Conclusions

13.1.1 CoPS Baseline has been shown to be reliable and valid assessment of pupils' skills in literacy, mathematics, communication and personal and social development on entry to school at age four or five

The modules show no ceiling or floor effects, indicating that they are appropriate assessments for this age group. CoPS Baseline shows significant test-retest correlations and also correlates significantly with the QCA Scales. CoPS Baseline has been shown to be predictive of children's later reading, spelling, writing and mathematics ability up to three years after the initial testing.

13.1.2 LASS Secondary has been shown to be a reliable and valid assessment of students' reading, spelling, reasoning, auditory memory, visual memory, phonological processing and phonic skills from the ages of 11 to 15

The modules show no ceiling or floor effects, indicating that they are appropriate assessments for this age group. LASS Secondary shows significant test-retest correlations and also correlates significantly with the conventional validation measures.

13.1.3 LASS Secondary has been shown to be a good indicator of dyslexia in the 11 to 15 year age group

There are significant differences between scores of dyslexic students and non-SEN students on the sentence reading, spelling, auditory memory, non-word reading and syllable segmentation tests. LASS Secondary correctly identified 79% of the dyslexic students as having dyslexia.

13.1.4 Computerised assessment is more objective than assessment administered by a human and test administration and scoring is completely standardised

Gender differences favouring females that are evident in conventional literacy tests are not evident in either the CoPS Baseline literacy module or the LASS Secondary reading and spelling tests. This may be due to computerised tests being less subjective than tests administered by human assessors.

13.1.5 Computerised educational assessment saves time in test administration and scoring

CoPS Baseline and LASS Secondary can be administered by less highly trained personnel, scoring is automatic and results are available immediately. The adaptive modules within these assessments allows tests to be tailored to individuals and reduces time taken for adaptive modules to approximately 35% of the time taken for the full versions of the modules. The adaptive modules correlate highly with the full versions of the modules.

13.1.6 Computerised assessment is enjoyable and motivating for children, particularly children who have specific difficulties

Teachers using CoPS Baseline and LASS Secondary reported that pupils enjoyed the tests. Pupils tested on LASS Secondary as well as on conventional assessments showed an overwhelming preference for the computerised tests. 92% of the dyslexic and other-SEN students preferred the computerised assessment.

13.2 Suggestions for future research

A large-scale study looking at gender differences on conventional and computerised tests is necessary to establish if computerised tests actually are less prone to bias. This would involve testing pupils on various conventional measures and computerised versions of the exact same tests. Equivalency of the two formats would have to be established first. Half of the pupils would need to be administered the conventional assessments first and the others the computerised assessments first.

Due to the small samples of non-white children and severe lack of children speaking English as a second language in both the secondary and the baseline studies, more extensive trials involving these groups are necessary.

It would appear to be reasonable in the future for further tests to be added to the LASS Secondary system, including tests of reading comprehension, verbal intelligence, mathematics and motor skills. Turner (1997) argues that reading comprehension tests are important in the assessment of dyslexia as dyslexic individuals tend to perform better on such tests than they do on tests of reading accuracy or reading rate. Indeed, dyslexic children use context to compensate for poor decoding skills, whilst children with poor reading comprehension skills benefit less from context (Nation and Snowling, 1998). However, there are difficulties with administering reading comprehension tests using a computer as such tests involve reading large amounts of text which students must be able to refer back to when answering the questions (in order to avoid them becoming memory tests). A number of researchers have reported lower verbal intelligence scores than non-verbal intelligence scores in dyslexic individuals (Patakfalvine and Kiss, 1971; Farrag et al, 1995; Shalev et al, 1995). However, other research has shown that verbal intelligence is higher than non-verbal intelligence in children with specific spelling difficulties (Newman, Fields and Wright, 1993; Warrington and Langdon, 1994) and specific arithmetic difficulties (Rourke and Finlayson, 1978; Shalev et al, 1995; Badian, 1999). It would therefore be advantageous to include tests of both verbal and non-verbal intelligence in the LASS Secondary system. Several studies have shown that some individuals with dyslexia exhibit difficulties in mathematics (Beauvais, 1973; Joffe, 1983; Smith, 1992; Miles, 1995) and motor skills (Stein, 1993; Nicolson and Fawcett, 1994; Fawcett and Nicolson, 1995). It is also recognised that individuals without reading difficulties may suffer from specific arithmetic disabilities or dyscalculia (Klauer, 1992; Temple, 1992; Lyytinen, Ahonen and Raesenen, 1994; Padget, 1998; Geary and Hoard, 2001). It is therefore useful for tests of mathematics and motor skills to be included within a screening test.

Follow-up tests of students assessed on LASS Secondary would provide information about a number of teaching outcomes. Students shown to have particular difficulties on LASS Secondary could be given different teaching programs and then re-tested to establish the effect of these programs. Furthermore, the effects of age and cognitive profile can also be measured.

There is also the possibility that reading tests in the future could make use of speech recognition technology. In such a test, students could be shown a word on a computer screen that they read to the computer, which determines if it has been read correctly or not. However, although these systems of dictating to a computer have greatly improved over recent years, they are not yet at reliable enough for use in assessment (Singleton, 1997a).

Administering CoPS Baseline and LASS Secondary over the internet is also a possibility. However, equivalency between the current versions of the tests and any internet version must be established. It is also critical that necessary security precautions are taken to avoid the test being mis-used. CoPS Baseline and LASS Secondary are intended for use by teachers and psychologists, who are able to interpret the results given.

It is recognised that there is a need for an effective screening test for dyslexia in adults (National Joint Committee on Learning Disabilities, 1987; National Working Party on Dyslexia in Higher Education, 1999). Since early 2001, Singleton, Horne and Thomas have been carrying out research on computerised screening for dyslexia in adults. The 'LADS' ('Lucid Adult Dyslexia Screening') system that has been produced comprises tests of: word recognition (speeded recognition of real words from nonwords); word construction (speeded encoding of nonwords from syllables); and short-term memory (backwards digit span). These tasks place heavy demands on phonological processing, working memory, lexical access and speed of processing. It is well established in the research literature that all these tasks are difficult for most adult dyslexics, even bright well-compensated ones (Felton, Naylor and Wood, 1990; Everatt, 1997; Gottardo, Siegel and Stanovich, 1997; Rack, 1997; Brunswick et al, 1999). The tests in LADS are adaptive and the whole screening can usually be completed in about 15 minutes.

Preliminary trials of LADS in a number of universities, FE colleges and adult literacy centres have shown it to give a highly significant and reliable discrimination between known dyslexic and known non-dyslexic individuals, with very low incidence of false negatives and false positives.

It has been noted that literacy difficulties are common amongst individuals who are deaf / hearing-impaired (Favez-Boutonier, 1967; Engel-Eldar and Rosenhouse, 2000) and these difficulties are generally attributed to their hearing problems. It is likely that in some cases these literacy difficulties are a result of dyslexia rather than just due to a hearing impairment. It is necessary to have tests suitable for use with deaf / hearing-impaired individuals in order to assess and monitor literacy skills and identify possible cases of dyslexia. For many of the tests in CoPS Baseline and LASS Secondary (e.g. CoPS Baseline mathematics, CoPS Baseline literacy - except rhymes and alliteration, LASS single word reading, LASS sentence reading, LASS spelling, LASS reasoning and LASS visual memory) there would be little change except to have the speech files shown in British Sign Language using digital video graphics. A digit span test could also be used but would no longer be a test of auditory memory but of visual memory.

References

Aaron, P. G., & Joshi, R. M. (1992). *Reading Problems: Consultation and Remediation.* New York / London: Guilford Press.

Anastasi, A. (1972). Culture-fair testing. In G. H. Bracht, K. D. Hopkins, & J. C. Stanley (Eds.), *Perspectives in Educational and Psychological Measurement.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Ayres, D., Downing, J., & Schaefer, B. (1982). *The Linguistic Awareness in Reading Readiness (LARR) - Revised version.* Windsor, Berks: NFER-Nelson.

Baddeley, A. D. (1986). Working memory, reading and dyslexia. In E. Hjelmquist & L. Nillsson (Eds.), *Communication and Handicap: Aspects of Psychological Compensation and Technical Aids.* Amsterdam: Elsevier Science.

Badian, N. A. (1999). Persistent arithmetic, reading, or arithmetic and reading disability. *Annals of Dyslexia, 49*, 45-70.

Bakker, D. J. (1990). *Neuropsychological Treatment of Dyslexia.* Oxford: Oxford University Press.

Bartram, D. (1994). Computer-based assessment. *International Review of Industrial and Organizational Psychology, 9,* 31-69.

Bartram, D. (1999). *Testing and the Internet: Current realities, issues and future possibilities.* Paper presented at the Test User Conference.

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment, 8*(4), 261-274.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology, 57*, 221-237.

Basden, A. (1984). On the application of expert systems. In M. J. Coombs (Ed.), *Developments in Expert Systems* (pp. 59-75). London: Academic Press.

Beaton, A., McDougall, S., & Singleton, C. H. (1997). Humpty Dumpty grows up? - Diagnosing dyslexia in adulthood. *Journal of Research in Reading, 20*(1), 1-6.

Beaumont, J. G. (1981). Microcomputer aided assessment and its implications for clinical practice. *Orchard Lodge Studies of Deviancy, 1*, 21-42.

Beaumont, J. G. (1985). The effect of microcomputer presentation and response medium on digit span performance. *International Journal of Man-Machine Studies, 22*, 11-18.

Beaumont, J. G., & French, C. C. (1984). *Microcomputer aided assessment in psychiatric and neurological states: Phase II (September 1982 - August 1984)*. London: Department of Health and Social Security.

Beauvais, J. (1973). The contribution of modern mathematics (to the study of dyslexia). *Pratique des Mots, 13-23.*

Beck, L. R., & St George, R. (1983). The alleged cultural bias of PAT: reading comprehension and reading vocabulary tests. *New Zealand Journal of Educational Studies, 18*(1), 32-47.

Beech, J. R. (1997). Assessment of memory and reading. In J. R. Beech & C. H. Singleton (Eds.), *The Psychological Assessment of Reading*. London: Routledge.

Beech, J. R., & Singleton, C. H. (1997). The psychological assessment of reading: Theoretical issues and professional solutions. In J. R. Beech & C. H. Singleton (Eds.), *The Psychological Assessment of Reading*. New York: Routledge.

Bennett, R. E. (1999). Computer-based testing for examinees with disabilities: On the road to generalized accommodation. In S. J. Messick (Ed.), *Assessment in higher education: Issues of access, quality, student development and public policy*. Mahwah, New Jersey: Lawrence Erlbaum.

Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., & Zack, J. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Bahavior, 15*(3-4), 283-294.

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behaviour perceptions and gender on teachers' judgements of students' academic skills. *Journal of Educational Psychology, 85*, 347-356.

Birmingham City Council. (1996). *Baseline Assessment for the Primary Phase: Autumn 1995, Analysis of Results*. Birmingham: Birmingham Local Education Department.

Black, P. (1987). *Task Group on Assessment and Testing*. London: DES.

Black, P. (1995). Can teachers use assessment to improve learning? *British Journal of Curriculum and Assessment, 5*(2), 7-11.

Blatchford, P. (1997). Pupils' self assessments of academic attainment at 7, 11 and 16 years: Effects of sex and ethnic group. *British Journal of Educational Psychology, 67*(2), 169-184.

Blatchford, P., Burke, J., Farquhar, C., Plewis, I., & Tizard, B. (1987). Associations between pre-school reading related skills and later reading achievement. *British Educational Research Journal, 13*(1), 15-23.

Blatchford, P., & Cline, T. (1992). Baseline assessment for school entrants. *Research Papers in Education, 7*(3), 247-269.

Bloxom, B. (1989). Adaptive Testing: A review of recent results. *Zeitschrift fur Differentielle und Diagnostiche Psychologie, 10*, 1-17.

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. *Developmental Medicine and Child Neurology, 15*, 663-687.

Boder, E., & Jarrico, S. (1982). *Boder Test of Reading-Spelling Patterns*. New York: Grune and Stratton.

Booth, T., & Statham, J. (1982). *The Nature of Special Education*. London: Croom Helm.

Bowers, P., Steffy, R., & Tate, E. (1988). Comparison of the effects of IQ control methods on memory and naming speed predictors of reading disability. *Reading Research Quarterly, 23*, 304-319.

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skills in dyslexia. *Reading and Writing: An Interdisciplinary Journal, 5*, 69-85.

BPS (1999). *Guidelines for the Development and Use of Computer-Based Assessments*. Leicester: British Psychological Society.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271*, 746-747.

Bradley, L. L., & Bryant, P. E. (1983). Categorising sounds and learning to read - a causal connection. *Nature, 301*, 419-521.

Brady, D. (1997). Assessment and the curriculum. In C. Cullingford (Ed.), *Assessment versus Evaluation.* London: Cassell.

Brady, S. (1986). Short-term memory, phonological processing and reading ability. *Annals of Dyslexia, 36*, 138-153.

British Dyslexia Association (1998). *The Dyslexia Handbook.* Reading, England: BDA.

British Psychological Society (1999). *Dyslexia, Literacy and Psychological Assessment.* Leicester: The British Psychological Society.

Brooks, M. (1997). Education: Closer inspection could narrow the ethnic gap. *The Independent,* December 4, 1997.

Bruck, M. (1990). Word-recognition skills in adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26*, 439-454.

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Developmental Psychology, 26*, 874-886.

Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., & Frith, U. (1999). Explicitly and implicit processing of words and pseudowords by adult developmental dyslexics. *Brain, 122*(10), 1901-1917.

Bryans, T. (1992). Educational psychologists working in a biased society. In S. Wolfendale, T. Bryans, M. Fox, A. Labram, & A. Sigston (Eds.), *The Profession and Practice of Educational Psychology: Future Directions*. London: Cassell Educational Limited.

Bryant, P. E., & Goswami, U. (1987). Phonological awareness and learning to read. In J. Beech & A. Colley (Eds.), *Cognitive Approaches to Reading*. Chichester: Wiley.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhymes and alliteration, phoneme detection and learning to read. *Developmental Psychology, 26*(3), 429-438.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125-144.

Bullock, A. (1975). *A Language for Life*. London: HMSO.

Burns, R. (1982). *Self-concept Development and Education*. London: Holt, Rinehart and Winston.

Burt, C. (1921). *Mental and Scholastic Tests*. London: Staples Press.

Burt, C. (1957). *The Causes and Treatment of Backwardness*. London: University of London Press.

Butkowsky, T. S., & Willows, D. M. (1980). Cognitive motivation and characteristics of children varying in reading ability: Evidence of learned helplessness in poor readers. *Journal of Educational Psychology, 72*, 408-422.

Butterfield, S. (1995). *Educational Objectives and National Assessment.* Buckingham: Open University Press.

Byers, A. P. (1981). Psychological evaluation by means of an on-line computer. *Behaviour Research Methods and Instrumentation, 13*, 585-587.

Calvert, E. J., & Waterfall, R. C. (1982). A comparison of conventional and automated administration of Raven's Standard Progressive Matrices. *International Journal of Man-Machine Studies, 17*, 305-310.

Cameron, R. (1991). Curriculum related assessment: The importance of educationally relevant data. In L. Harding & J. R. Beech (Eds.), *Educational Assessment of the Primary School Child*. Windsor, Berkshire, UK: NFER-Nelson.

Carr, A. C., Wilson, S. L., Ghorb, A., Angell, R. J., & Mums, R. T. (1982). Automated testing of geriatric patients using a microcomputer based system. *International Journal of Man-Machine Studies, 17*(3), 297-300.

Chapman, J. W., Lambourne, R., & Silva, P. A. (1990). Some antecedents of academic self-concept: A longitudinal study. *British Journal of Educational Psychology, 60*(2), 142-152.

Cisero, C. A., Royer, J. M., Marchant, H. G., & Jackson, S. J. (1997). Can the Computer-based Academic Assessment System (CAAS) be used to diagnose reading disability in college students? *Journal of Educational Psychology, 89*(4), 599-620.

Clarke, A. M., & Clarke, A. D. B. (1976). *Early Experience: Myth and Evidence.* London: Open Books.

Clay, M. (1975). *What Did I Write?* Auckland, New Zealand: Heinemann Education Books.

Clay, M. (1979). *The Early Detection of Reading Difficulties*. London: Heinemann.

Cline, T., & Shamsi, T. (2000). *Language Needs or Special Needs? The assessment of learning difficulties in literacy among children learning English as an additional language: A literature review*. London: Department for Education and Employment.

Cohen, M. (1997). *Children's Memory Scale*. San Antonio, Texas: The Psychological Corporation.

Cooper, C. (1995). Inside the WISC-III UK. *Educational Psychology in Practice, 10*(4), 215-219.

Coren, S. (1989). The Spelling Component Test: Psychometric evaluation and norms. *Educational and Psychological Measurement, 49*(4), 961-971.

Critchley, M. (1970). *The Dyslexic Child*. London: Heinemann Medical Books.

Crook, C. (1996). *Computers and the Collaborative Experience of Learning: A Psychological Perspective*. London: Routledge.

Crook, C., & Steele, J. (1987). Self-selection of simple computer activities by infant school pupils. *Educational Psychology, 7*, 23-32.

Cummins, J. (1984). *Bilingualism and Special Education Issues in Assessment and Pedagogy*. Clevedon: Multi-lingual Matters.

Davis, E. G., & Swezey, R. W. (1983). Human factors guidelines in computer graphics: A case study. *International Journal of Man-Machine Studies, 18*, 113-133.

Day, C. L. (1999). A predictive validity study of computer adaptive placement tests for Tennessee higher education institutions. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 59.*

Delap, M. R. (1995). Teachers' estimates of candidates' performance in public examinations. *Assessment in Education, 2*, 75-92.

Denckla, M. B., & Rudel, R. (1976). Rapid 'Automatized' Naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*, 471-479.

Department for Education (1994). *Code of Practice on the Identification and Assessment of Special Educational Needs*. London: Department for Education.

Department for Education and Employment (1998). *Baseline Assessment Circular.* London: Department for Education and Employment.

Department of Education and Science (1988). *National Curriculum Task Group on Assessment and Testing Report.* London: Department for Education and Science.

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

Desforges, M., & Lindsay, G. (1995a). Baseline assessment. *Educational and Child Psychology, 12*, 42-51.

Desforges, M., & Lindsay, G. (1995b). *Infant Index.* London: Hodder and Stoughton.

Desforges, M., & Lindsay, G. (1998). *Baseline-PLUS.* London: Edexcel.

Dessent, T. (1992). Educational psychologists and 'the case for individual casework'. In S. Wolfendale, T. Bryans, M. Fox, A. Labram, & A. Sigston (Eds.), *The Profession and Practice of Educational Psychology: Future Directions*. London: Cassell Educational Limited.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326.

Dimitrovsky, L., & Almy, M. (1975). Early conservation as a predictor of arithmetic achievement. *The Journal of Psychology, 91*, 65-70.

Dimitrovsky, L., & Almy, M. (1980). Kindergarten conservation as a predictor of second-grade arithmetic achievement. *The Journal of Psychology, 106*, 219-225.

Downing, J. (1970). Children's concepts of language in learning to read. *Educational Research, 12*, 106-112.

Druckman, L. E. (1997). An investigation of the utility of using local curriculum-based measurement norms versus group standardized norms to predict students at risk for academic failure in reading and mathematics. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 57*(12A).

Dunn, L. M., & Dunn, L. M. (1997). *The British Picture Vocabulary Scale - Second Edition*. Windsor, Berkshire: NFER-Nelson.

Durndell, A. (1991). The persistence of the gender gap in computing. *Computers and Education, 16*, 283-287.

Dyer, H. S. (1972). Needed changes to sweeten the impact of testing. In G. H. Bracht, K. D. Hopkins, & J. C. Stanley (Eds.), *Perspectives in Educational and Psychological Measurement* . Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Eaton, L. (1991). The law relating to assessment. In L. Harding & J. R. Beech (Eds.), *Educational Assessment of the Primary School Child.* Windsor, Berkshire, UK: NFER-Nelson.

Ebel, R. L. (1972). The social consequences of educational testing. In G. H. Bracht, K. D. Hopkins, & J. C. Stanley (Eds.), *Perspectives in Educational and Psychological Measurement.* Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia, 44*, 205-226.

Elithorn, A., Mornington, S., & Stavrou, A. (1982). Automated psychological testing: Some principles and practice. *International Journal of Man-Machine Studies, 17*, 247-263.

Elliot, S. N. (1996). Assessing learning of all students: Becoming an essential service provider once again. In R. C. Talley & T. Kubiszyn (Eds.), *Making psychologists in schools indispensable: Critical questions and emerging perspectives.* Washington: American Psychological Association.

Elliott, C. D. (1979). *British Ability Scales.* Windsor, Berkshire: NFER-Nelson.

Elliott, C. (1996). *British Ability Scales - Second Edition.* Windsor, Berkshire: NFER-Nelson.

Ellis, N. (1990). Reading, phonological skills and short-term memory: Interactive tributaries of development. *Journal of Research in Reading, 13*(2), 107-122.

Ellis, N., & Cataldo, S. (1990). The role of spelling in learning to read. *Language and Education, 4*, 1-28.

Ellis, N. C., & Large, B. (1987). The development of reading: As you seek so shall you find. *British Journal of Psychology, 78*(1), 1-28.

Elwood, D. L. (1970). Automation of testing procedures: Elwood replies to Moses. *American Psychologist, 25*(8), 764-765.

Elwood, D. L. (1972). Test-retest reliability and cost analyses of automated and face-to-face intelligence testing. *International Journal of Man-Machine Studies, 4*, 1-22.

Engel-Eldar, R., & Rosenhouse, J. (2000). Reading difficulty characteristics in dyslexia and hearing-impaired students. *Educational Psychology, 20*(4), 459-482.

Essa, E. L. (1987). The effect of a computer on pre-school children's activities. *Early Childhood Research Quarterly, 2*, 377-382.

Evans, L. D., Tannehill, R., & Martin, S. (1995). Children's reading skills: A comparison of traditional and computerised assessment. *Behavior Research Methods, Instruments and Computers, 27*(2), 162-165.

Evans, R., Davies, P., Ferguson, N., & Williams, P. (1978). *Swansea Evaluation Profiles for School Entrants: Research Edition.* Windsor, Berkshire: NFER.

Everatt, J. (1997). The abilities and disabilities associated with adult developmental dyslexia. *Journal of Research in Reading, 20*(1), 13-21.

Everett, S., Farnsworth, S., & Mitchell, R. (1997). *The Nottinghamshire Framework of Achievement: The First Report of the Nottinghamshire Entry and Baseline Scheme and the Key Stage 2 to 3 Tracking Project*. Nottingham: Nottinghamshire County Council.

Eyde, L. D. (1987). Computerised psychological testing: An introduction. *Applied Psychology: An International Review, 36*(3/4), 223-235.

Eyde, L. D., & Kowal, D. M. (1985). Psychological decision support software for the public: Pros, cons and guidelines. *Computers in Human Behaviour, 1*, 321-336.

Farrag, A. F., Shaker, H., Hamdy, N. A., & Wafaa, M. A. (1995). Clinical characteristics of a poulation of dyslexic children in Assiut, Egypt. *Neuroepidemiology, 14*(2), 92-99.

Farrell, A. D. (1989). Impact of standards for computer-based tests on practice: Consequences of the information gap. *Computers in Human Behavior, 5*, 1-11.

Favez-Boutonier, J. (1967). Language. *Bulletin de Psychologie, 20*(23-24), 1378-1381.

Fawcett, A. J., & Nicolson, R. I. (1995). Persistent deficits in motor skill of children with dyslexia. *Journal of Motor Behavior, 27*(3), 235-240.

Fawcett, A. J., & Nicolson, R. I. (2001). Dyslexia: The role of the cerebellum. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Fawcett, A. J., Nicolson, R. I., & Dean, P. (1996). Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia, 46*, 259-283.

Fawcett, A. J., Pickering, S., & Nicolson, R. I. (1993). Development of the DEST test for the early screening of dyslexia. In S. F. Wright & R. Groner (Eds.), *Facets of Dyslexia and its Remediation*. Amsterdam: Elsevier Science Publishers BV.

Federico, P.-A. (1991). Measuring recognition performance using computer-based and paper-based methods. *Behavior Research Methods, Instruments and Computers, 23*, 341-347.

Fein, G., Davenport, L., Yingling, C. D., & Galin, D. (1988). Verbal and non-verbal memory deficits in pure dyslexia. *Developmental Neuropsychology, 4*, 181-197.

Feldman, S. E., & Sullivan, D. S. (1971). Factors mediating the effects of enhanced rapport on children's performance. *Journal of Consulting and Clinical Psychology, 36*, 302.

Felton, R. H., Naylor, C. E., & Wood, F. B. (1990). Neuropsychological profile of adult dyslexics. *Brain and Language, 39*(4), 485-497.

Felton, R. H., & Wood, P. B. (1989). Cognitive deficits in reading disability and attention deficit disorder. *Journal of Learning Disabilities, 22*, 3-22.

Fergusson, D. M., & Horwood, L. J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies, 32*(1), 83-96.

Fife-Shaw, C., Breakwell, G., Lee, T., & Spencer, J. (1986). Patterns of teenage computer usage. *Journal of Computer-Assisted Learning, 2*, 152-161.

Fisher, D. (1995). *Baseline Assessment*. Slough, Berkshire: National Foundation for Educational Research.

Fisher, S. E., & Smith, S. D. (2001). Progress towards the identification of genes influencing developmental dyslexia. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Flaugher, R. L. (1978). The many definitions of test bias. *American Psychologist, 33*, 671-679.

Fletcher, J. M., & Satz, P. (1984). Test-based versus teacher-based predictions of academic achievement: A three-year longitudinal follow-up. *Journal of Pediatric Psychology, 9*(2), 193-203.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D., Katz, L., Liberman, I. Y., Stuebling, K. K., Francis, J. M., Fouler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparison of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6-24.

Francis, H. (1973). Children's experience of reading and notions of units in language. *British Journal of Educational Psychology, 43*, 17-23.

Francis, H. (1982). *Learning to Read.* London: Allen and Unwin.

Francis, H. (1992). Patterns of reading development in the first school. *British Journal of Educational Psychology, 62*(2), 225-232.

Francis, H. (1994). Literacy development in the first school - what advice? *British Journal of Educational Psychology, 64*(1), 29-44.

Frederickson, N. (1992). Curriculum-based assessment: Broadening the base. In T. Cline (Ed.), *The Assessment of Special Educational Needs: International Perspectives* (pp. 147-169). London: Routledge.

Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological Assessment Battery.* Windsor, Berkshire: NFER-Nelson.

French, C. C. (1986). Microcomputers and psychometric assessment. *British Journal of Guidance and Counselling, 14*(1), 33-45.

French, C. C., & Beaumont, G. B. (1992). Microcomputer version of a digit span test in clinical use. *Interacting with Computers, 4*(2), 163-178.

French, C. C., & Beaumont, J. G. (1990). A clinical study of the automated assessment of intelligence by the Mill Hill Vocabulary test and the Standard Progressive Matrices test. *Journal of Clinical Psychology, 46*(2), 129-140.

French, C. C., & Hemmings, R. K. (1984). Software review of 'Cattell IQ Test'. *Bulletin of the British Psychological Society, 37*, 383-384.

Friedman, L., & Davenport, E. (1998). Literacy gender gaps: Evidence from the National Adult Literacy Survey. In M. C. Smith (Ed.), *Literacy for the Twenty-First Century: Research, Policy, Practices and the National Adult Literacy Survey*. London: Praeger.

Frith, U. (1995). Dyslexia: Can we have a shared theoretical framework? *Educational and Child Psychology, 12*(1), 6-17.

Frith, U. (1997). Brain, mind and behaviour in dyslexia. In C. Hulme & M. Snowling (Eds.), *Dyslexia: Biology, Cognition and Intervention* (pp. 1-19). London: Whurr.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488-500.

Fuchs, L. S., & Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education? *RASE: Remedial and Special Education, 5*(4), 29-32.

Fuchs, L., Fuchs, D., & Hamlett, C. (1993). Technological advances linking the assessment of students' academic proficiency to instructional planning. *Journal of Special Education Technology, 12*(1), 49-62.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Hasselbring, T. S. (1987). Using computers with curriculum-based monitoring: Effects on teacher efficiency and satisfaction. *Journal of Special Education Technology, 8*(4), 14-27.

Gallagher, A., Laxon, V., Armstrong, E., & Frith, U. (1996). Phonological difficulties in high functioning dyslexics. *Reading and Writing: An Interdisciplinary Journal, 8*, 499-509.

Gathercole, S. E. (1995). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology, 65*(2), 155-164.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working Memory and Language*. Hove, Sussex: Erlbaum.

Gavel, S. R. (1958). June reading achievement of first-grade children. *Journal of Education, Boston University, 140*, 37-43.

Geary, D. C., & Hoard, M. K. (2001). Numerical and arithmetical deficits in learning-disabled children: Relation to dyscalculia and dyslexia. *Aphasiology, 15*(7), 635-647.

Ghouri, N. (1998). Tests may brand boys as failures. *Times Educational Supplement*, June 19, 1998.

Gilberstadt, H., Lushene, R., & Buegel, B. (1976). Automated assessment of intelligence: The TAPAC test battery and computerised report writing. *Perceptual and Motor Skills, 43*, 627-635.

Gillborn, D., & Gipps, C. (1996). *Recent Research on the Achievements of Ethnic Minority Pupils*. London: HMSO.

Gipps, C. (1990). *Assessment: A Teacher's Guide to the Issues*. London: Hodder and Stoughton.

Gipps, C., Steadman, S., Blackstone, T., & Stierer, B. (1983). *Testing Children: Standardised Testing in Local Education Authorities and Schools*. London: Heinemann Educational.

Gjessing, H. J., & Karlsen, B. (1989). *A Longitudinal Study of Dyslexia*. New York: Springer.

Godfrey Thompson Unit (1993). *Edinburgh Reading Tests*. London: Hodder and Stoughton.

Gonzales, R. R., & Roll, S. (1985). Relationship between acculturation, cognitive style and intelligence: A cross-sectional study. *Journal of Cross Cultural Psychology, 16*(2), 190-205.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42*, 73-83.

Gottardo, A., Siegel, L. S., & Stanovich, K. E. (1997). The assessment of adults with reading disabilities: What can we learn from experimental tasks? *Journal of Research in Reading, 20*(1), 42-54.

Goulandris, N. K. (1991). Alphabetic spelling: Predicting eventual literacy attainment. In C. M. Sterling & C. Robson (Eds.), *Psychology, Spelling and Education*. Clevedon: Multilingual Matters.

Greenwood, C., & Rieth, H. (1994). Current dimensions of technology-based assessment in special education. *Exceptional Children, 61*(2), 105-113.

Groth-Marnet, G., & Schumaker, J. (1989). Computer-based psychological testing: Issues and guidelines. *American Journal of Orthopsychiatry, 59*, 257-263.

Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education, 5*(1), 1-16.

Hanley, J. R. (1997). Reading and spelling impairments in undergraduate students with developmental dyslexia. *Journal of Research in Reading, 20*(1), 22-30.

Harding, L., & Beech, J. R. (1991). Introduction to educational assessment. In L. Harding & J. R. Beech (Eds.), *Educational Assessment of the Primary School Child.* Windsor, Berkshire, UK: NFER-Nelson.

Hare, B. R. (1985). Re-examining the achievement central tendency: Sex differences within race and race differences within sex. In H. P. McAdoo & J. L. McAdoo (Eds.), *Black Children: Social, Educational and Parental Environments* (pp. 139-155). Thousand Oaks, California: Sage Publications.

Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalisation, minority issues and the general aptitude test battery.* Washington, DC: National Academy Press.

Hatcher, P. J. (2000). Sound links in reading and spelling with discrepancy-defined dyslexics and children with moderate learning difficulties. *Reading and Writing, 13*, 257-272.

Hayes, Z. L., & Waller, T. G. (1994). Gender differences in adult readers: A process perspective. *Canadian Journal of Behavioural Science, 26*(3), 421-437.

Heath, W. J. (2001). Learning style differences between gifted and non-gifted sixth and seventh grade students on Texas Assessment of Academic Skills test. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 61*, 2591.

Helms, J. E. (1995). Why is there no study of cultural equivalence in standardised cognitive ability testing? In N. R. Goldberger & J. B. Veroff (Eds.), *The Culture and Psychology Reader* (pp. 674-719). New York: New York University Press.

Herring, R. D. (1989). Minimum competency testing: Implications for ethnicity. *Journal of Multicultural Counseling and Development, 17*(3), 98-104.

Holland, S. H. (1996). PROJECT 2000: An educational mentoring and academic support model for inner-city African American boys. *Journal of Negro Education, 65*(3), 315-321.

Hooper, S. R., & Hynd, G. W. (1986). Performance of normal and dyslexic readers on the Kaufman Assessment Battery for Children: A discriminant analysis. *Journal of Learning Disabilities, 19*(4), 206-210.

Horn, W. F., & O'Donnell, J. P. (1984). Early identification of learning disabilities: A comparison of two methods. *Journal of Educational Psychology, 76*(6), 1106-1118.

Horne, J. K., & Singleton, C. H. (1997). *Assessment of Dyslexia in Higher Education.* Paper presented at the Fourth International Conference of the British Dyslexia Association: Biological Bases, Identification and Intervention, York.

Horne, J. K., Singleton, C. H., & Thomas, K. V. (1999). *Lucid Assessment System for Schools (Secondary)*. Beverley, East Yorkshire: Lucid Creative Limited.

Hoyles, C. (1988). Girls and computers. *Bedford Way Papers 34*.

Høien, T., & Lundberg, I. (1989). A strategy for assessing problems in word recognition among dyslexics. *Scandinavian Journal of Educational Research, 33*, 185-201.

Huebner, E. S., & Cummings, J. A. (1986). A naturalistic study of school psychologists' use of diagnostic test data in making decisions. *Journal of Psychoeducational Assessment, 4*(3), 183-190.

Hughes, M., Brackenridge, A., & Macleod, H. (1987). Children's ideas about computers. In J. Rutkowska & C. Crook (Eds.), *Computers, Cognition and Development*. Chichester: Wiley.

Hunt, E., & Hertzog, C. (1981). *Age related changes in cognition during the working years*. Washington: University of Washington.

Hunt, E., & Pellegrino, J. (1985). Using interactive computing to expand intelligence testing: A critique and prospectus. *Intelligence, 9*, 207-236.

Huntington, D. D., & Bender, W. D. (1993). Adolescents with learning disabilities at risk? Emotional well-being, depression and suicide. *Journal of Learning Disabilities, 26*, 159-166.

Ingenkamp, K. (1972). *Zur Problematik der Jahrgangsklasse*. Weinheim: Beltz.

Ingenkamp, K. (1977). *Educational Assessment*. Windsor, Berkshire, UK: NFER-Nelson.

Jensen, A. R. (1973). *Educability and Group Differences*. New York: Harper and Row.

Jensen, A. R., & Inouye, A. R. (1980). Level I and level II abilities in asian, white and black children. *Intelligence, 4*(1), 41-49.

Joffe, L. (1983). School mathematics and dyslexia: A matter of verbal labelling, generalisation, horses and carts. *Cambridge Journal of Education, 13*(3), 22-27.

Johnson, D. J., & Myklebust, H. R. (1967). *Learning Disabilities: Educational Principles and Practices*. New York: Grune and Stratton.

Johnson, J. H., & Johnson, K. N. (1981). Psychological considerations related to the development of computerized testing stations. *Behaviour Research Methods and Instrumentation, 13*, 421-424.

Jorm, A. F. (1983). Specific reading retardation and working memory: A review. *British Journal of Psychology, 74*, 311-342.

Joyce, J. (1988). The development of an anti-racist policy in Leeds. In S. Wolfendale, I. Lunt, & T. Carroll (Eds.), *Educational Psychologists Working in Multi-cultural Communities, DECP* (Vol. 5 (2)).

Kelly, V. (1992). Concepts of assessment: An overview. In G. M. Blenkin & A. V. Kelly (Eds.), *Assessment in Early Childhood Education*. London: Paul Chapman.

Kenney, J. B., & Anderson, H. E. (1984). A comparative study of selected black and white university sophomore students on a required reading test. *Journal of Experimental Education, 53*(1), 20-28.

Kibby, M. W. (1995). *Practical Steps for Informing Literacy Instruction: A Diagnostic Decision-making Model*. Newark, Delaware: International Reading Association.

Kirk, J., McLoughlin, D., & Reid, G. (2001). Identification and intervention in adults. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Klauer, K. J. (1992). Do more girls have problems in mathematics and more boys in reading and writing? On the assessment of specific learning problems. *Zeitschrift fuer Entwicklungspsychologie und Paedagogische Psychologie, 24*(1), 48-65.

Kline, P. (1993). *The Handbook of Psychological Testing*. London: Routledge.

LaFrance, E. B. (1997). The gifted / dyslexic child: Characterizing and addressing strengths and weaknesses. *Annals of Dyslexia, 47*, 163-182.

Lam, T. C. M. (1993). Testability: A critical issue in testing language minority students with standardized achievement tests. *Measurement and Evaluation in Counselling and Development, 26*, 179-191.

Lawrence, D. (1971). The effects of couselling on retarded readers. *Educational Research, 13*, 119-124.

Lawrence, D. (1987). *Enhancing Self-esteem in the Classroom*. London: Paul Chapman.

Lee, D. S. (2000). Gender differences in academic achievement and ability: A longitudinal study. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(4-B), 2244.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Lindquist, E. F., & Hieronymus, A. N. (1972). Use and interpretation of test results. In G. H. Bracht, K. D. Hopkins, & J. C. Stanley (Eds.), *Perspectives in Educational and Psychological Measurement*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Lindsay, G. (1980). The infant rating scale. *British Journal of Educational Psychology, 50*(2), 97-104.

Lindsay, G. (1981). *The Infant Rating Scale*. Sevenoaks: Hodder and Stoughton.

Lindsay, G. (1984). *Screening for Children with Special Needs*. Beckenham: Croom Helm.

Lindsay, G. (1988). Early identification of learning difficulties: Screening and beyond. *School Psychology International, 9*, 61-68.

Lindsay, G. (1991). The assessment of cognitive abilities. In L. Harding & J. R. Beech (Eds.), *Educational Assessment of the Primary School Child.* Windsor, Berkshire, UK: NFER-Nelson.

Lindsay, G. (1993). Baseline assessment and special educational needs. In S. Wolfendale (Ed.), *Assessing Special Educational Needs* (pp. 57-71). London: Cassell.

Lindsay, G. (1995). Early identification of special educational needs. In I. Lunt, B. Norwich, & V. Varma (Eds.), *Psychology and Education for Special Needs*. London: Arena.

Lindsay, G. (1998). Baseline assessment: A positive or malign initiative? In B. Norwich & G. Lindsay (Eds.), *Baseline Assessment: Benefits and Pitfalls*. Tamworth: National Association for Special Educational Needs.

Lindsay, G. (2001). Identification and intervention in the primary school. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Lindsay, G., & Desforges, M. (1998). *Baseline Assessment: Practice, Problems and Possibilities*. London: David Fulton.

Lindsay, G., & Desforges, M. (1999). The use of the Infant Index / Baseline-PLUS as a baseline assessment measure of literacy. *Journal of Research in Reading, 22*(1), 55-66.

Lindsay, G., & Lewis, A. (2001). *Evaluation of the National Baseline Assessment Scheme in England.* Paper presented at the British Psychological Society Developmental and Education Sections' Joint Annual Conference, University College Worcester.

Lindsay, G., Lewis, A., & Phillips, E. (2000). *Evaluation of Accredited Baseline Assessment Schemes 1999/2000*: Qualifications and Curriculum Authority.

Lindsay, G. A., & Wedell, K. (1982). The early identification of educationally 'at risk' children revisited. *Journal of Learning Disabilities, 15*(4), 212-217.

Loftus, G., & Loftus, E. F. (1983). *The Mind at Play.* New York: Basic Books.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Lubs, H. A., Rabin, M., Feldman, E., Jallad, B. J., Kushch, A., & Gross-Glenn, K. (1993). Familial dyslexia: Genetic and medical findings in eleven three-generation families. *Annals of Dyslexia, 43*, 44-60.

Lundberg, I., & Høien, T. (2001). Dyslexia and phonology. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Lyytinen, H., Ahonen, T., & Raesenen, P. (1994). Dyslexia and dyscalculia in children: Risks, early precursers, bottlenecks and cognitive mechanisms. *International Journal of Child and Adolescent Psychiatry, 56*(3), 179-192.

Maccoby, E. E., & Jacklin, C. N. (1974). *The Psychology of Sex Differences.* Stanford, CA: Stanford University Press.

MacKenzie, C. G. (1989). The eleven-plus examination in developing countries: A case study. *Educational Studies, 15*(3), 281-300.

Macmillan Test Unit. (1990). *Group Reading Test.* Windsor, Berkshire: NFER-Nelson.

Male, D. B. (1994). Time taken to complete assessments and make statements: Does the nature of SEN matter and will new legislation make a difference? *British Journal of Special Education, 21*(4), 147-151.

Martinez, M. E., & Mead, N. A. (1988). *Computer competence: The first national assessment* (ETS 17-CC-01). Princeton, NJ: Educational Testing Service.

Matarazzo, J. D. (1983). Computerized psychological testing. *Science, 221*(22), 323.

McCullough, C. S., & Wenck, L. S. (1984). Current microcomputer applications in school psychology. *School Psychology Review, 13*(4), 429-439.

McDougall, S., & Ellis, A. W. (1994). On the problems of comparing dyslexics with other groups of readers (and the quest for the cognitive basis of dyslexia). *Early Childhood Development and Care, 101*, 33-49.

McLoughlin, D., Fitzgibbon, G., & Young, V. (1994). *Adult Dyslexia: Assessment, Counselling and Training*. London: Whurr.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11*(1), 3-9.

Meier, S. (1994). *The Chronic Crisis in Psychological Measurement and Assessment: A historical Survey*. San Diego: Academic Press.

Merrell, C. (2001). *The development of an assessment to measure the skills and progress of pre-school children.* Paper presented at the British Psychological Society Developmental and Education Sections' Joint Annual Conference, University College Worcester.

Miles, T. R. (1982). *The Bangor Dyslexia Test.* Wisbech, Cambridgeshire: Learning Development Aids.

Miles, T. R. (1991). Towards a prevalence figure for dyslexia. In M. Snowling & M. Thomson (Eds.), *Dyslexia: Integrating theory and practice*. London: Whurr.

Miles, T. R. (1995). Dyslexia: The current status of the term: II. *Child Language Teaching and Therapy, 11*(1), 23-33.

Miles, T. R. (1996). Do dyslexic children have IQs? *Dyslexia, 2*(3), 175-178.

Miles, T. R. (1997). *The Bangor Dyslexia Test.* Wisbech, Cambridgeshire: Learning Development Aids.

Miles, T. R., Haslum, M. N., & Wheeler, T. J. (1998). Gender ratio in dyslexia. *Annals of Dyslexia, 48*, 27-55.

Miles, T. R., & Miles, E. (1999). *Dyslexia: A Hundred Years On.* (2nd Ed.). Buckingham, UK: Open University Press.

Miller, G. (1979). The ABC's of English. *International Newsletter: Educational Evaluation and Research, 20*, 10-11.

Mommers, M. J., Van Leeuwe, J. F., Oud, J. H., & Janssens, J. M. (1986). Decoding skills, reading comprehension and spelling: A longitudinal investigation. *Tijdschrift voor Onderwijsresearch, 11*(2), 97-113.

Moore, N. C., Summer, K. R., & Bloor, R. N. (1984). Do patients like psychometric testing by computer? *Journal of Clinical Psychology, 40*, 875-877.

Moses, J. L. (1969). Automation of testing procedures. *American Psychologist, 24*, 1174.

Murphy, N. (1998). Baseline assessment at school entry. *The Psychologist, 11*, 274.

Murphy, R. (1991). National developments in primary school assessment. In L. Harding & J. R. Beech (Eds.), *Educational Assessment of the Primary School Child.* Windsor, Berkshire, UK: NFER-Nelson.

Naglieri, J. A. (1985). *Matrix Analogies Test - Short Form*: The Psychological Corporation.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69*(4), 996-1011.

National Joint Committee on Learning Disabilities. (1987). Adults with learning disabilities: A call to action. *Journal of Learning Disabilities, 20*(3), 172-175.

National Working Party on Dyslexia in Higher Education. (1999). *Dyslexia in Higher Education: Policy, Provision and Practice*. Hull: University of Hull.

NCET (1994). *Using IT for Assessment - Going Forward.* Coventry: National Council for Educational Technology.

NCET (1995). *Using IT for Assessment - Case Studies Reports.* Coventry: National Council for Educational Technology.

Neale, M. D. (1989). *Neale Analysis of Reading Ability.* Windsor, Berkshire: NFER-Nelson.

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Newman, S. P., Fields, H., & Wright, S. (1993). A developmental study of specific spelling disability. *British Journal of Educational Psychology, 63*(2), 287-296.

Newton, M. J., Thomson, M. E., & Richards, I. L. (1979). *Readings in Dyslexia.* Wisbech: LDA.

Newton, P., & Beck, E. (1993). Computing: An ideal occupation for women? In J. Benyon & H. Mackay (Eds.), *Computers into Classrooms*. London: The Falmer Press.

NFER-Nelson (1993). *LARR Test of Emergent Literacy.* Windsor, Berkshire: NFER-Nelson.

Nichols, N. J., & McKinney, A. W. (1977). Black or white socio-economically disadvantaged pupils - they aren't necessarily inferior. *Journal of Negro Education, 46*(4), 443-449.

Nicolson, R. I. (1996). Developmental dyslexia: Past, present and future. *Dyslexia: An International Journal of Research and Practice, 2*, 190-207.

Nicolson, R. I. (2001). Developmental dyslexia: Into the future. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Nicolson, R. I., & Fawcett, A. J. (1994). Comparison of deficits in cognitive and motor skills among children with dyslexia. *Annals of Dyslexia, 44*, 147-164.

Nicolson, R. I., & Fawcett, A. J. (1995a). Dyslexia is more than a phonological deficit. *Dyslexia: An International Journal of Research and Practice, 1*, 19-36.

Nicolson, R. I., & Fawcett, A. J. (1995b). Speed of processing, motor skill, automaticity and dyslexia. In A. Fawcett & R. Nicolson (Eds.), *Dyslexia in Children, Multidisciplinary Perspectives*. Hemel Hempstead: Harvester Wheatsheaf.

Nicolson, R. I., Fawcet, A. J., & Miles, T. R. (1993). *Feasibility study for the development of a computerised screening test for dyslexia in adults* (OL176). Sheffield: Employment Department.

Norwich, B. (1988). Educational psychology services in LEAs: What future? In N. Jones & J. Sayer (Eds.), *Management and the Psychology of Schooling*. London: The Falmer Press.

Nutbrown, C. (1999). Baseline assessment of writing: The need for reconsideration. *Journal of Research in Reading, 22*(1), 37-44.

O'Leary, J., & Betts, H. (1999). Inspectors accuse schools of racism. *The Times*, March 11, 1999.

Olsen, J. B. (1990). Applying computerized adaptive testing in schools. *Measurement and Evaluation in Counselling and Development, 23*, 31-38.

Olson, R. K., Datta, H., Gayan, J., & DeFries, J. C. (1999). A behavioural-genetic analysis of reading disabilities and component processes. In R. Klein & P. Macmullan (Eds.), *Converging Methods for Understanding Reading and Dyslexia* (pp. 133-152). Cambridge, MA: MIT Press.

Orton Dyslexia Society (1994). *Perspectives in Dyslexia.*

Padget, S. Y. (1998). Lessons from research on dyslexia: Implications for a classification system for learning disabilities. *Learning Disability Quarterly, 21*(2), 167-178.

Padget, S. Y., Knight, D. F., & Sawyer, D. J. (1996). Tennessee meets the challenge of dyslexia. *Annals of Dyslexia, 46*, 51-72.

Paitich, D. (1973). Computers in behavioural science: A comprehensive automated psychological examination and report. *Behavioural Science, 18*, 131-136.

Patakfalvine, C. R., & Kiss, Z. (1971). Etiological and psychological data bearing on the function disturbance of defective reading. *Magyar Pszichologiai Szemle, 28*(3), 407-416.

Patilla, P. (1994). *Mathematics 8.* Windsor, Berkshire: NFER-Nelson.

Pearson, L., & Quinn, J. (1986). *Bury Infant Check.* Windsor: NFER-Nelson.

Penney, C. G., & Blackwood, P. A. (1989). Recall mode and recency in immediate serial recall: Computer users beware! *Bulletin of the Psychonomic Society, 27*, 545-547.

Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development, 61*, 1753-1778.

Plewis, I. (1991). Pupils' progress in reading and mathematics during primary school: Associations with ethnic group and sex. *Educational Research, 33*, 133-140.

Plewis, I. (1997). Inferences about teacher expectations from national assessment at key stage one. *British Journal of Educational Psychology, 67*, 235-247.

Pole, C. (1998). Assessing, recording and testing achievement: Some issues in accountability, professionalism and education markets. In D. Shorrocks-Taylor (Ed.), *Directions in Educational Psychology*. London: Whurr Publishers Ltd.

Potton, A. (1983). *Guides to Assessment in Education Screening*. London: Macmillan Education Ltd.

Pumfrey, P. D., & Reason, R. (1991). *Specific Learning Difficulties: (Dyslexia) Challenges and Responses*. Windsor: NFER-Nelson.

QCA (1997). *Baseline Assessment Scales*. Hayes, Middlesex: Qualifications and Curriculum Authority.

QCA (1999a). *Baseline Assessment Schemes Accredited by QCA: Final List*. London: Qualifications and Curriculum Authority.

QCA (1999b). *Annual Report on Sample Data from the QCA Baseline Assessment Scales*. London: Qualifications and Curriculum Authority.

QCA (1999c). *Final Report of the Evaluation of the First Year of Statutory Baseline Assessment 1998/1999*. London: Qualifications and Curriculum Authority.

Rack, J. P. (1994). Dyslexia: The phonological deficit hypothesis. In A. Fawcett & R. Nicolson (Eds.), *Dyslexia in Children: Multidisciplinary Perspectives* (pp. 5-38). New York / London: Harvester Wheatsheaf.

Rack, J. (1997). Issues in the assessment of developmental dyslexia in adults: Theoretical and applied perspectives. *Journal of Research in Reading, 20*(1), 66-76.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The non-word reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29-53.

Raven, J., Summer, B., Birchfield, M., Brosier, G., Burciaga, L., & Byrkit, B. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Research Supplement no. 3*. London: Lewis.

Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and Practice, 8*, 11-16.

Reid, G. (1998). *Dyslexia: A Practitioner's Handbook*. Chichester: John Wiley and Sons.

Reid, G., & Weedon, C. (1997). *Insights gained from developing an instrument to profile the emerging literacy skills of six year olds*. Edinburgh: Moray House Centre for the Study of Dyslexia.

Reid, J. (1966). Learning to think about reading. *Educational Research, 9*, 56-62.

Reinking, D., Labbo, L. D., & McKenna, M. C. (2000). From assimilation to accommodation: A developmental framework for integrating digital technologies into literacy research and instruction. *Journal of Research in Reading, 23*(2), 110-122.

Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardisation sample as a function of the stratification variables. *Journal of School Psychology, 25*(4), 323-342.

Riddick, B. (1996). *Living with Dyslexia: The Social and Emotional Consequences of Specific Learning Difficulties*. London: Routledge.

Riddick, B., Sterling, C., Farmer, M., & Morgan, S. (1999). Self-esteem and anxiety in the educational histories of adult dyslexic students. *Dyslexia: An International Journal of Research and Practice, 5*(4), 227-248.

Rispens, J., & van Yperen, T. A. (1990). The identification of specific reading disorders: Measuring a severe discrepancy. In G. T. Pavlidis (Ed.), *Perspectives on Dyslexia Vol 2: Cognition, Language and Treatment*. Chichester: John Wiley and Sons.

Robertson, J. (1997). *Neuropsychological studies in intervention in dyslexia: Recent studies in differential treatment*. Paper presented at the Fourth World Congress on Dyslexia, York.

Rourke, B. P., & Finlayson, M. A. (1978). Neuropsychological significance of variations in patterns of academic performance: Verbal and visual-spatial abilities. *Journal of Abnormal Child Psychology, 6*(1), 121-133.

Rubin, R. A., Balow, B., Dorle, J., & Rosen, M. (1978). Preschool prediction of low achievement in basic school skills. *Journal of Learning Disabilities, 11*, 664-667.

Ruch, G. M. (1933). Recent developments in statistical procedures. *Review of Educational Research, 3*, 39-40.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181-197.

Sadovnik, J. (2000). A program design to improve self-esteem and aggression in dyslexic children. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(5-B), 2779.

Sainsbury, M. (1997). Test group slips under the baseline. *Times Educational Supplement 2*, February 7, 1997.

Sainsbury, M., Schagen, I., Whetton, C., & Caspall, L. (1999). An investigation of hierarchical relationships in children's literacy attainments at baseline. *Journal of Research in Reading, 22*(1), 45-54.

Saint-Laurent, L., Herbert, M., Royer, E., & Pierard, B. (1996). Identification of students with academic difficulties: Implications for research and practice. *Canadian Journal of School Psychology, 12*(2), 143-154.

Salmon-Cox, L. (1980). *Teachers and tests: What's really happening?* Paper presented at the American Educational Research Association, Boston.

Salois, K. A. N. (1999). A comparative study of the Wechsler Intelligence Scale for Children - Third Edition (WISC-III) test performance: Northern Cheyenne and Blackfeet Reservation Indian children with the standardisation sample. *Dissertation Abstracts International Section B: The Sciences and Engineering, 60 (4-B)*, 1909.

Salvia, J., & Ysseldyke, J. E. (1988). Using estimated true scores: A response to Sabers, Feldt and Reschly. *Journal of Special Education, 22*(3), 367-373.

Sammons, P. (1994). *Gender, ethnic and socio-economic differences in attainment and progress: A logitudinal analysis of student achievement over nine years.* Paper presented at the American Educational Research Association Annual Meeting, New Orleans, USA.

Saracoglu, B., Minden, H., & Wilchesky, M. (1989). The adjustment of students with learning disabilities to university and its relationship to self-esteem and self-efficacy. *Journal of Learning Disabilities, 22*, 590-592.

Satz, P., Taylor, G., Friel, J., & Fletcher, J. (1978). Some developmental and predictive precursors of reading disabilities. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An Appraisal of Current Knowledge*. New York: Oxford University Press.

SCAA (1996a). *Baseline Assessment: Draft Proposals*. London: School Curriculum and Assessment Authority.

SCAA (1996b). *Desirable Outcomes for Children's Learning on Entering Compulsory Education* . London: School Curriculum and Assessment Authority.

SCAA (1996c). *Trials of Baseline Assessment Schemes*. London: School Curriculum and Assessment Authority.

SCAA (1997a). *Baseline Assessment Scales*. London: School Curriculum and Assessment Authority.

SCAA (1997b). *The National Framework for Baseline Assessment: Criteria and Procedures for Accreditation of Baseline Assessment Schemes*. London: School Curriculum and Assessment Authority.

Scarborough, H. S. (1984). Continuity between childhood dyslexia and reading. *British Journal of Psychology, 75*, 329-348.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 33*, 272-292.

Schoenfeldt, L. F. (1989). Guidelines for computer-based psychological tests and interpretations. *Computers in Human Behaviour, 5*, 13-21.

Schonell, F. J. (1971). *Graded Word Spelling Test*. Edinburgh: Oliver and Boyd Ltd.

Scott, T., Cole, M., & Engel, M. (1992). Computers and education: A cultural constructivist perspective. *Review of Research in Education, 18*, 191-251.

Shalev, R. S., Manor, O., Amir, N., & Wertman, E. R. (1995). Developmental dyscalculia and brain laterality. *Cortex, 31*(2), 357-365.

Share, D. L., McGee, R., McKenzie, D., Williams, S., & Silva, P. A. (1987). Further evidence relating to the distinction between specific reading retardation and general reading backwardness. *British Journal of Developmental Psychology, 5*, 35-44.

Share, D. L., Silva, P. A., & Adler, C. J. (1987). Factors associated with reading plus spelling retardation and specific spelling retardation. *Developmental Medicine, 29*, 72-84.

Shaywitz, B. A., Shaywitz, M. D., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut longitudinal study. *Journal of the American Medical Association, 264*, 998-1002.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makugh, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326*, 145-150.

Siegel, L. S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42*, 201-215.

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*(8), 469-478.

Singleton, C. H. (1994). Computer applications in the identification and remediation of dyslexia. In D. Wray (Ed.), *Literacy and Computers: Insights from Research* (pp. 55-61). Widness: United Kingdom Reading Association.

Singleton, C. H. (1996). Computerised screening for dyslexia. In G. Reid (Ed.), *Dimensions of Dyslexia*. Edinburgh: Moray House Publications.

Singleton, C. H. (1997a). Computer-based assessment of reading. In J. R. Beech & C. H. Singleton (Eds.), *The Psychological Assessment of Reading*. London: Routledge.

Singleton, C. H. (1997b). Screening early literacy. In J. R. Beech & C. H. Singleton (Eds.), *Psychological Assessment of Reading*. London: Routledge.

Singleton, C. H. (2001). Computer-based assessment in education. *Educational and Child Psychology, 18*(3), 58-74.

Singleton, C. H., Horne, J. K., & Thomas, K. V. (1999). Computerised baseline assessment of literacy. *Journal of Research in Reading, 22*(1), 67-80.

Singleton, C. H., Horne, J. K., & Thomas, K. V. (2002). *Lucid Adult Dyslexia Screening*. Beverley, East Yorkshire: Lucid Research Limited.

Singleton, C. H., Horne, J. K., & Vincent, D. (1995). *Implementation of a Computer-based System for the Diagnostic Assessment of Reading.* Coventry: National Council for Educational Technology.

Singleton, C. H., & Thomas, K. V. (1994). Computerised screening for dyslexia. In C. H. Singleton (Ed.), *Computers and Dyslexia: Educational Applications of New Technology* (pp. 172-184). Hull: Dyslexia Computer Resource Centre, University of Hull.

Singleton, C. H., Thomas, K. V., & Horne, J. K. (1998). *CoPS Baseline Assessment System.* Beverley, East Yorkshire: Lucid Research Limited.

Singleton, C. H., Thomas, K. V., & Horne, J. K. (2000). Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading, 23*(2), 158-180.

Singleton, C. H., Thomas, K. V., & Leedale, R. C. (1996). *Lucid CoPS Cognitive Profiling System.* Beverley, East Yorkshire: Lucid Research Limited.

Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug and tobacco use. *Journal of Consulting and Clinical Psychology, 51*, 267-275.

Skuy, M., Taylor, M., O'Carroll, S., Fridjhon, P., & Rosenthal, L. (2000). Performance of black and white South African children on the Wechsler Intelligence Scale for Children - Revised and the Kaufman Assessment Battery. *Psychological Reports, 86*(3), 727-737.

Smith, S. (1992). Familial patterns of learning disabilities. *Annals of Dyslexia, 42*, 143-158.

Snowling, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology, 29*, 294-305.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43*, 219-234.

Snowling, M. (1987). *Dyslexia: A Cognitive Developmental Perspective*. Oxford: Basil Blackwell Ltd.

Snowling, M. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading, 18*, 132-138.

Snowling, M. J. (2000). *Dyslexia*. (Second ed.). Oxford: Blackwell Publishers Ltd.

Snyderman, M., & Rothman, S. (1988). *The IQ Controversy, the Media and Public Policy*. New Brunswick, NJ: Transaction Books.

Snowling, M., & Nation, K. (1997). Language, phonology and learning to read. In C. Hulme & M. Snowling (Eds.), *Dyslexia: Biology, Cognition and Intervention* (pp. 153-166). London: Whurr.

Snowling, M., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading, 20*(1), 31-41.

Space, L. G. (1981). The computer as psychometrician. *Behaviour Research Methods and Instrumentation, 13*, 595-606.

Spafford, C. S. (1989). Wechsler digit span subtest: Diagnostic usefulness with dyslexic children. *Perceptual and Motor Skills, 69*, 115-125.

Stake, R. E., & Hastings, J. T. (1972). Review of the Stanford Achievement Battery. In G. H. Bracht, K. D. Hopkins, & J. C. Stanley (Eds.), *Perspectives in Educational and Psychological Measurement*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590-612.

Stanovich, K. E. (1991a). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26* (1), 7-29.

Stanovich, K. E. (1991b). The theoretical and practical consequences of discrepancy definitions of dyslexia. In M. J. Snowling & M. E. Thomson (Eds.), *Dyslexia; Integrating Theory and Practice*. London: Whurr.

Stanovich, K. E. (1991c). Conceptual and empirical problems with discrepancy definitions of reading disability. *Learning Disability Quarterly, 14*(4), 269-280.

Stanovich, K. E. (1996). Towards a more inclusive definition of dyslexia. *Dyslexia, 2*(3), 154-166.

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

Stein, J. (1993). Dyslexia: Impaired temporal information processing? In P. Tallal & A. M. Galaburda (Eds.), *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia* (pp. 83-86). New York: New York Academy of Sciences.

Stein, J., Talcott, J., & Witton, C. (2001). The sensorimotor basis of developmental dyslexia. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Stein, J. F., & Walsh, V. (1997). To see but not to read: The magnocellular theory of dyslexia. *Trends in Neuroscience, 20*(4), 508-514.

Stocking, M. L. (1987). Two simulated feasibility studies in computerised adaptive testing. *Applied Psychology: An International Review, 36*(3/4), 263-277.

Stout, R. L. (1981). New approaches to the design of computerized interviewing and testing systems. *Behaviour Research Methods and Instrumentation, 13*, 436-442.

Strand, S. (1995). *Wandsworth Baseline Assessment 1994 / 95*. London: Wandsworth Education Department.

Strand, S. (1997). Pupil progress during key stage 1: A value added analysis of school effects. *British Educational Research Journal, 23*(4), 471-487.

Strand, S. (1999). Baseline assessment results at age 4: Associations with pupil background factors. *Journal of Research in Reading, 22*(1), 14-26.

Stuart, M. (1995). Prediction and qualitative assessment of 5 and 6-year old children's reading: A longitudinal study. *British Journal of Educational Psychology, 65*, 287-296.

Sutherland, M. J., & Smith, C. D. (1991). Assessing literacy problems in mainstream schooling: A critique of three literacy screening tests. *Educational Review, 43*(1), 39-48.

Sutton, R. (1992). *Assessment: A framework for teachers*. London: Routledge.

Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities. *Journal of Learning Disabilities, 27*, 34-50.

Tate, D., & Gibson, G. (1980). Socio-economic status and black and white intelligence revisited. *Social Behavior and Personality, 8*(2), 233-237.

Taylor, C., Kirsch, I., Eignor, D., & Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49*(2), 219-274.

Temple, C. M. (1992). Developmental dyscalculia. In S. J. Segalowitz & I. Rapin (Eds.), *Handbook of Neuropsychology* (Vol. 7,). New York: Elsevier Science.

Terman, L. M. (1916). *The Measurement of Intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1937). *Measuring Intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1972). *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin.

Thomas, K. V., Singleton, C. H., & Horne, J. K. (2001). *Lucid Assessment System for Schools (Junior)*. Beverley, East Yorkshire: Lucid Creative Limited.

Thomson, M. E. (1982). The assessment of children with specific reading difficulties (dyslexia) using the British Ability Scales. *British Journal of Psychology, 73*(4), 461-478.

Thomson, M. (1993). *Developmental Dyslexia*. (3rd ed.). London: Whurr.

Tizard, B., Blatchford, P., Burke, J., Farquhar, C., & Plewis, I. (1988). *Young Children at School in the Inner City*. Hove: Lawrence Erlbaum.

Topping, K. J. (1978). Consumer confusion and professional conflict in educational psychology. *Bulletin of the British Psychological Society, 31*, 265-267.

Torgesen, J. K. (1989). Why IQ is relevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 484-486.

Troman, G. (1989). Testing tensions: The politics of educational assessment. *British Educational Research Journal, 15*(3), 279-295.

Turner, M. (1997). *Psychological Assessment of Dyslexia*. London: Whurr.

Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading, 22*(1), 27-36.

Tymms, P. B., Merrell, C., & Henderson, B. (1997). The first year at school: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation, 3*, 101-118.

Vale, C. D. (1981). Design and implementation of a microcomputer-based adaptive testing system. *Behaviour Research Methods and Instrumentation, 13*, 399-406.

Van Daal, V. H. P., & Van der Leij, A. (1999). Developmental dyslexia: Related to specific or general deficits? *Annals of Dyslexia, 49*, 71-104.

Vance, H. B., & Gaynor, P. E. (1976). A note on cultural difference as reflected in the Wechsler Intelligence Scale for Children. *The Journal of Genetic Psychology, 129*(1), 171-172.

Vardill, R. (1996). Imbalance in the numbers of boys and girls identified for referral to educational psychologists: Some issues. *Support for Learning, 11*(3), 123-129.

Vellutino, F. R. (1979). *Dyslexia: Theory and Research.* Cambridge, MA: MIT Press.

Vincent, K. R. (1991). Black / white IQ differences: Does age make the difference? *Journal of Clinical Psychology, 47*(2), 266-270.

Vincent, D., & Crumpler, M. (1997). *British Spelling Test Series.* Windsor, Berkshire: NFER-Nelson.

Wade, B., & Moore, M. (1993). The test's the thing: Viewpoints of students with special educational needs. *Educational Studies, 19*(2), 181-191.

Wadsworth, S. J., DeFries, J. C., Stevenson, J., Gilger, J. W., & Pennington, B. F. (1992). Gender ratios among reading-disabled children and their siblings as a function of parental impairment. *Journal of Child Psychology and Psychiatry, 33*, 1229-1239.

Wagner, R. K. (1986). Phonological processing abilities and reading: Implications for disabled readers. *Journal of Learning Disabilities, 19*, 623-630.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83-103.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (1990). *Computerized Adaptive Testing: A Primer.* Hillsdale, NJ: Lawrence Erlbaum.

Warnock (1978). *Report of the Committee of Enquiry into the Education of Handicapped Children and Young People.* London: Department of Education and Science.

Warrington, E. K., & Langdon, D. (1994). Spelling dyslexia: A deficit of the visual word form. *Journal of Neurology, Neurosurgery and Psychiatry, 57*(2), 211-216.

Watkins, M. W., & Kush, J. C. (1988). Assessment of academic skills of learning disabled students with classroom microcomputers. *School Psychology Review, 17*(1), 81-88.

Watts, K., Baddeley, A., & Williams, M. (1982). Automated tailored testing using Raven's Matrices and the Mill Hill Vocabulary Tests: A comparison with manual administration. *International Journal of Man-Machine Studies, 17*, 331-344.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* New York: The Psychological Corporation.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children - Revised.* New York: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children - Third Edition.* San Antonio, Texas: The Psychological Corporation.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children - III.* (3rd ed.). New York: Harcourt Brace Jovanovitch.

Wechsler, D. (1996a). *Wechsler Objective Language Dimensions.* London: The Psychological Corporation.

Wechsler, D. (1996b). *Wechsler Objective Numerical Dimensions.* London: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Memory Scale - Third Edition.* London: The Psychological Corporation.

Wesson, C., Fuchs, L. S., Tindal, G., Mirkin, P., & Deno, S. (1986). Facilitating the efficiency of on-going curriculum-based measurement. *Teacher Education and Special Education, 9*, 166-172.

Whitehouse, C. C. (1983). Analysis of WISC-R coding performance of normal and dyslexic readers. *Perceptual and Motor Skills, 57*, 951-960.

Wilson, S. L. (1987). The development of an automated test of immediate memory and its evaluation on severely physically disabled adults. *Applied Psychology: An International Review, 36*(3/4), 311-327.

Wilson, S. L., Thompson, J. A., & Wylie, G. (1982). Automated psychological testing for the severely physically handicapped. *International Journal of Man-Machine Studies, 17*, 291-296.

Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language, 27*, 360-379.

Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26*, 123-160.

Wolf, M. (1996). *The Double-Deficit Hypothesis for the Developmental Dyslexics.* Paper presented at the 47th Annual Conference of the Orton Dyslexia Society, Boston.

Wolf, M., & Bowers, P. G. (1999). The double deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415-438.

Wolf, M., & Obregon, M. (1992). Early naming deficits, developmental dyslexia, and a specific deficit hypothesis. *Brain and Language, 42*, 219-247.

Wolf, M., & O'Brien, B. (2001). On issues of time, fluency and intervention. In A. Fawcett (Ed.), *Dyslexia: Theory and Good Practice*. London: Whurr.

Wolfendale, S. (1993a). *Baseline Assessment for Children Starting School: Issues and Challenges.* Stoke-on-Trent: Trentham Books.

Wolfendale, S. (1993b). *Baseline Assessment: A Review of Current Practice, Issues and Strategies for Effective Implementation.* Stoke-on-Trent: Trentham Books.

Wolfendale, S., & Bryans, T. (1972). Identification and follow-up of slow readers in the primary school. *Journal of the Association of Educational Psychologists, 3*(2).

Wolfendale, S., & Bryans, T. (1979). *Identification of Learning Difficulties: A Model for Intervention.* Stafford: National Association for Remedial Education.

Wolfendale, S., & Lindsay, G. (1999). Issues in baseline assessment. *Journal of Research in Reading, 22*(1), 1-13.

Wood, R. (1991). *Assessment and Testing: A survey of research commissioned by the University of Cambridge Local Examinations Syndicate*. Cambridge: University of Cambridge Press.

Woodcock, R. W., & Johnson, M. B. (1989-1990). *Woodcock-Johnson Psycho-Educational Battery - Revised*. Allen, TX: DLM.

Woodings, V. (1997). How primary schools deal with assessment in the National Curriculum. In C. Cullingford (Ed.), *Assessment versus Evaluation*. London: Cassell.

Woods, M. (1973). Caught in the Act. *Association of Educational Psychologists Journal and Newsletter, 3*(3), 33-36.

Woodward, J., & Rieth, H. (1997). A historical review of technology research in special education. *Review of Educational Research, 67*(4), 503-536.

Wright, H. J., & Payne, T. A. N. (1979). *An Evaluation of a School Psychology Service: The Portsmouth Pattern*. Winchester: Hampshire County Council.

Wright, S. F., & Groner, R. (1993). Dyslexia: Issues of definition and subtyping. In S. F. Wright & R. Groner (Eds.), *Facets of Dyslexia and its Remediation* (pp. 437-453). Amsterdam: Elsevier Science.

Yeh, J. (1978). *Test Use in the Schools*. Los Angeles: Centre for the study of evaluation, University of California.

Young, D. (1987). *SPAR Spelling and Reading Tests - Second Edition*. London: Hodder and Stoughton.

Zabell, C., & Everatt, J. (2000). Gender differences in dyslexia, *The Dyslexia Handbook* (pp. 83-87). Reading, England: British Dyslexia Association.

Appendix 1: CoPS Baseline Literacy items

Set A – Aural comprehension

Four pictures are displayed simultaneously on the screen: an open door; a closed door; an open window; a closed window. *Show me the correct picture for 'The door is open'.*

Four pictures are displayed simultaneously on the screen: a man up a ladder; a man next to a ladder; a man carrying a ladder; a man standing on a chair. *Show me the correct picture for 'The man is up the ladder'.*

Four pictures are displayed simultaneously on the screen: a baby asleep; a baby playing; a baby being fed; a man. *Show me the correct picture for 'The baby is asleep'.*

Four pictures are displayed simultaneously on the screen: a rabbit outside a cage; a rabbit in a cage with the door shut; a rabbit in a cage with the door open; a rabbit in a cage with a cat outside the cage. *Show me the correct picture for 'The rabbit is out of the cage'.*

Four pictures are displayed simultaneously on the screen: a boy with a girl; a boy alone; a boy with a cat; a girl with a dog. *Show me the correct picture for 'boy is with the girl'.*

Four pictures are displayed simultaneously on the screen: a spoon in a jar; a spoon next to a jar; a spoon on a plate; a knife on a plate. *Show me the correct picture for 'The spoon is in the jar'.*

Four pictures are displayed simultaneously on the screen: a cat on a table; a cat under a table; a cat on a chair; a vase of flowers on a table. *Show me the correct picture for 'The cat is on the table'.*

Set B – Verbal concepts

Four pictures are displayed simultaneously on the screen: a toothbrush; a spoon; a trumpet; an umbrella. *Show me something we can clean our teeth with.*

Four pictures are displayed simultaneously on the screen: a pair of scissors; a fork; a broom; a comb. *Show me something we can cut paper with.*

Four pictures are displayed simultaneously on the screen: a pen; a comb; a pair of scissors; a spoon. *Show me something we can write with.*

Five pictures are displayed simultaneously on the screen: a hand holding a duster; a duck; a saw; a television; a pair of shoes. *Show me the picture which goes best with the word cleaning.*

Five pictures are displayed simultaneously on the screen: a tree; a ball; a pair of shoes; a pen; a car. *Show me the picture which goes best with the word growing.*

Five pictures are displayed simultaneously on the screen: an oven; a vacuum cleaner; a bed; a paint brush; a hammer. *Show me the picture which goes best with the word cooking.*

Five pictures are displayed simultaneously on the screen: a shoe; a pair of roller skates; a chair; a cup; a spade. *Show me the picture which goes best with the word walking.*

Set C – Knowledge about print

Four pictures are displayed simultaneously on the screen: a book; a picture; a pen; a radio. *Show me something we can read.*

An animation appears on the top half of the screen with a story underneath which is read by the computer. *The frog hopped on to a leaf. He could see a fish in the water. Show me a capital letter.*

Five children's names appear on the screen: the child's own name and four other names randomly selected from the pupils database. *Show me your name.*

Four pictures are displayed simultaneously on the screen: a page with writing on; a page with a pattern on; a blank sheet of lined paper; a picture. *Show me some writing.*

An animation appears on the top half of the screen with a story underneath which is read by the computer. *The frog hopped on to a leaf. He could see a fish in the water. Show me a word.*

An animation appears on the top half of the screen with a story underneath which is read by the computer. *The frog hopped on to a leaf. He could see a fish in the water. Show me where we start reading*

Set D – Letter recognition

An animation appears on the top half of the screen with a story underneath which is read by the computer. *The frog hopped on to a leaf. He could see a fish in the water. Show me where we finish reading.*

Six letters appear simultaneously on the screen: s; a; f; k; m; w. *Show me the letter s.*

Six letters appear simultaneously on the screen: W; E; T; O; B; Y. *Show me the letter W.*

Six letters appear simultaneously on the screen: N; X; R; O; H; E. *Show me the letter N.*

Six letters appear simultaneously on the screen: m; a; f; k; s; w. *Show me the letter m.*

Six letters appear simultaneously on the screen: R; X; O; N; E; H. *Show me the letter R.*

Six letters appear simultaneously on the screen: g; p; d; y; h; b. *Show me the letter g.*

Six letters appear simultaneously on the screen: a; d; i; n; h; t. *Show me the letter a.*

Set E – Phonological Awareness (Rhyming)

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully. House…cat…drawer…chair…mouse. Which one sounds like mouse?*

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully. Bike…man…sun…book…van. Which one sounds like van?*

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully. Cake…spoon…fish…bird…dish. Which one sounds like dish?*

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully. Dog…cup…bat…glove…hat. Which one sounds like hat?*

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully. Pen…door…chair…egg…hen. Which one sounds like hen?*

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully.*

River...coat...frog...lamp...boat. Which one sounds like boat?

Five pictures appear one by one on the screen and are named by the computer. The rhyme object is displayed in the middle of the screen. *Listen carefully.*

Tree...ring...zip...queen...king. Which one sounds like king?

Set F – Phonological Awareness (Alliteration)

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully.*

Knife...mouth...apple...balloon...bottle. Which one begins with the same sound as bottle?

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully.*

Umbrella...door...cat...shoe...dog. Which one begins with the same sound as dog?

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully.*

Chimney...flower...train...glasses...tree. Which one begins with the same sound as tree?

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully.*

Carrot...eye...fish...mouse...five. Which one begins with the same sound as five?

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully. Egg cup...hand...tee shirt...picture...house. Which one begins with the same sound as house?*

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully. Stars...moon...bus...woman...window. Which one begins with the same sound as window?*

Five pictures appear one by one on the screen and are named by the computer. The alliteration object is displayed in the middle of the screen. *Listen carefully. Table...candle...leaf...bucket...ladder. Which one begins with the same sound as ladder?*

Set G – Reading

Four pictures appear one by one on the screen and are named by the computer. The letter c is then displayed in the middle of the screen. *Cat...hat...sun...goat. This is the letter c. It makes the sound k. Which of these pictures begins with the letter c?*

Four pictures appear one by one on the screen and are named by the computer. The letter f is then displayed in the middle of the screen. *Shop...fox...box...train. This is the letter f. It makes the sound fffff. Which of these pictures begins with the letter f?*

Four pictures appear one by one on the screen and are named by the computer. The letter b is then displayed in the middle of the screen. *Bird...dog...hand...cup. This is the letter b. It makes the sound b. Which of these pictures begins with the letter b?*

A picture of a fish appears at the top of the screen with five words underneath: fish; fun; dish; hush; ship. *This is a fish. Show me the word that says fish.*

Four pictures appear one by one on the screen and are named by the computer. The letter d is then displayed in the middle of the screen. *Bag...pen...bread...brick. This is the letter d. It makes the sound d. Which of these pictures ends with the letter d?*

Four pictures appear one by one on the screen and are named by the computer. The letter n is then displayed in the middle of the screen. *Ship...broom...table...spoon. This is the letter n. It makes the sound nnn. Which of these pictures ends with the letter n?*

A picture of a girl appears at the top of the screen with five words underneath: girl; good; gale; fill; rule. *This is a girl. Show me the word that says girl.*

Set H – Spelling

Four letters appear on the screen (s; t; o; w) with a picture of the sun in the middle. *This is the sun. Which letter does the word sun begin with?*

Four letters appear on the screen (t; d; a; g) with a picture of a teddy in the middle. *This is a teddy. Which letter does the word teddy begin with?*

Four letters appear on the screen (w; p; s; e) with a picture of some water in the middle. *This is some water. Which letter does the word water begin with?*

Four letters appear on the screen (d; n; u; g) with a picture of some dinner in the middle. *This is some dinner. Which letter does the word dinner begin with?*

Four letters appear on the screen (p; f; i; b) with a picture of a penguin in the middle. *This is a penguin. Which letter does the word penguin begin with?*

Ten letters appear on the screen (s; b; a; k; c; d; h; m; e; t) with a picture of a bed in the middle. *This is a bed. Can you spell the word bed? Press the smiley face when you've finished.*

Ten letters appear on the screen (s; b; a; k; c; d; h; m; e; t) with a picture of a hat in the middle. *This is a hat. Can you spell the word hat? Press the smiley face when you've finished.*

Appendix 2: CoPS Baseline Mathematics items

Set A – Number recognition

Four boats with different numbers on them (3; 1; 8; 7) appear simultaneously on the screen. *Here are some boats. Show me boat number 3.*

Five busses with different numbers on them (5; 9; 2; 10; 8) appear simultaneously on the screen. *Here are some busses. Show me bus number 5.*

A clock appears on the screen. *Show me number 7.*

Ten balloons with different numbers on them (1 to 10) appear simultaneously on the screen. *Here are some balloons. Show me balloon number 8.*

Six coins (1p; 20p; £1; 10p; 2p; 5p) appear simultaneously on the screen. *Here are some coins. Show me the 10p.*

Six coins (1p; 20p; £1; 10p; 2p; 5p) appear simultaneously on the screen. *Here are some coins. Show me the 20p.*

Six coins (1p; 20p; £1; 10p; 2p; 5p) appear simultaneously on the screen. *Here are some coins. Show me the £1.*

Set B – Counting

Five of Zoid's Friends appear on the screen holding different numbers of balloons (1; 2; 3; 4; 5). *Here are some of Zoid's friends with balloons. Show me the friend who has 4 balloons.*

Four flower pots appear on the screen with different numbers of flowers in (3; 5; 6; 8). *Here are some flower pots. Show me the pot that has 8 flowers.*

Four rabbit holes appear with different numbers of rabbits in (1; 2; 4; 5) and the number 5 appears in the middle of the screen. *What number is this? Show me the hole with this number of rabbits in.*

Four dogs appear with different numbers of spots on them (4; 6; 7; 9) and the number 9 appears in the middle of the screen. *What number is this? Show me the dog with this number of spots.*

Six kittens appear on the screen with a number line underneath (1 to 10). Jo has some kittens. How many kittens can you see? Show me the number on the number line.

A tree with seven apples on appears on the screen with a number line underneath (1 to 10). This is Zoid's apple tree. How many apples can you see? Show me the number on the number line.

Six dominoes appear on the screen with different numbers of spots (5/2; 3/4; 4/2; 5/5; 1/6; 1/2). *Here are some dominoes. Show me which domino has 10 spots.*

Set C – Relations

Five birds of various sizes appear on the screen (two are the same size). *Here are some birds. Show me the smallest bird.*

Four rabbit holes with different numbers of rabbits in (3; 4; 5; 7) appear on the screen. *Here are some rabbit holes. Show me the hole with the most rabbits in.*

Seven balloons with varying length strings appear on the screen (two are the same length). *Here are some balloons. Show me the balloon with the longest string.*

Five trees of varying heights appear on the screen (two are the same height). *Here are some trees. Show me the shortest tree.*

Five of Zoid's friends appear on the screen holding varying numbers of balloons (1; 2; 3; 3; 5). Here are some of Zoid's friends. Two of them have the same number of balloons. Show me which two friends have the same number of balloons.

Five pictures appear on the screen with different numbers of objects in them (2 balloons; 3 balloons; 4 cakes; 4 ice-creams; 5 cans of drink). *Here are some pictures. Two of the pictures have the same number of things in. Show me which two pictures have the same number of things in.*

Seven stars appear on the screen with different numbers of points on them (3; 4; 6; 7; 7; 8; 9). Here are some stars. Two of the stars have the same number of points. Show me which two stars have the same number of points.

Set D – Sets

Four pictures appear on the screen with different objects in (3 dogs; car, dog, tree; 2 cats, bike; cat, dog, aeroplane). Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.

Four pictures appear on the screen with different objects in (3 kites with the same pattern; 3 kites of 3 different patterns; 3 kites of 2 different patterns; 3 kites of 2 different patterns). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Four pictures appear on the screen with different objects in (5 crosses; 3 crosses, square, circle; 3 circles, 2 crosses; 3 squares, circle, cross). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Four pictures appear on the screen with different objects in (4 numbers; 2 shapes, letter, number; 2 numbers, letter, shape; 2 letters, 2 shapes). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Four pictures appear on the screen with different objects in (4 balloons; 2 fish, 2 spoons; 3 cows, balloon; cow, fish, balloon, spoon). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Four pictures appear on the screen with different objects in (4 shapes without straight edges; the other three pictures contain some straight edged shapes). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Four pictures appear on the screen with different objects in (4 line shapes – no area; the other three pictures contain some shapes with area). *Here are some pictures with different things in. One picture has things which all go together. Show me the picture that has things which all go together.*

Set E – Seriation

An animation shows seven of Zoid's friends queuing up to buy ice-cream at the seaside. Zoid's friends are at the seaside. They are going to buy ice-cream. Can you see them waiting in a line? Which friend is first in the line?

An animation shows seven of Zoid's friends queuing up to buy ice-cream at the seaside. Zoid's friends are at the seaside. They are going to buy ice-cream. Can you see them waiting in a line? Which friend is last in the line?

An animation shows seven of Zoid's friends queuing up to buy ice-cream at the seaside. Zoid's friends are at the seaside. They are going to buy ice-cream. Can you see them waiting in a line? Which friend is fifth in the line?

A picture of five children finishing a race appears on the screen. *These children are having a race. Show me the child that is first.*

Three Russian dolls appear at the top of the screen one at a time, decreasing in size from left to right. There is a space at the end of the row for children to select from the four at the bottom of the screen the one that comes next. *Here are some toys. Show me the toy that comes next.*

Three steps made from building blocks appear at the top of the screen one at a time, increasing in size from left to right. There is a space at the end of the row for children to select from the four at the bottom of the screen the one that comes next. *Here are some building blocks. Show me the one that comes next.*

Three dominoes appear at the top of the screen one at a time, the number of spots increasing from left to right (2, -, 6, 8). There is a space after the first domino for children to select from the four at the bottom of the screen (2, 4, 6, 10) the one that is missing. *Here are some dominoes, but one domino is missing. Show me the domino which goes where one is missing.*

Set F – Shape

Four shapes appear on the screen (square; circle; triangle; diamond). *Here are some shapes. Show me the triangle.*

A circle with a piece missing appears at the top of the screen. Underneath are four shapes. *There's a piece missing from this circle. Show me which piece fits to make the circle.*

Four shapes appear on the screen one at a time from left to right (square; circle; square; circle) with a space at the end of the row. There are four shapes at the bottom of the screen (circle; square; triangle; rectangle). *Here are some shapes in the grey box. Show me the shape which comes next.*

A square with a piece missing appears at the top of the screen. Underneath are four shapes. *There's a piece missing from this square. Show me which piece fits to make the square.*

A picture appears at the top of the screen showing a worm beside an apple which has been bitten into. At the bottom of the screen are five pieces of apple. *Wendy the worm has taken a big bite out of the apple. Show me which piece she had.*

Five shapes appear on the screen one at a time from left to right (triangle; square; circle; triangle; square) with a space at the end of the row. There are five shapes at the bottom of the screen (triangle; square; inverted triangle; circle; diamond). *Here are some shapes in a row. Show me the shape which comes next.*

Five shapes appear on the screen one at a time from left to right (large triangle; small triangle; large square; small square; large circle) with a space at the end of the row. There are six shapes at the bottom of the screen (small triangle; large square; small circle; small square; large circle; large triangle). *Here are some shapes in a row. Show me the shape which comes next.*

A picture of a cake with three candles on it and another candle next to it appears at the top of the screen. There are three cakes with different numbers of candles (2; 3; 4) at the bottom of the screen. *Last year Emma had 3 candles on her birthday cake. This year she needs one more candle. Show me the cake Emma will have this year.*

Two rabbit holes appear on the screen, one with 5 rabbits in and the other with 3 rabbits. A number line (1 to 10) appears at the bottom of the screen. *Here are 2 rabbit holes. One has 5 rabbits in and the other has 3 rabbits in. How many rabbits are there altogether? Show me the answer on the number line.*

A picture of Zoid's friends looking through two windows appears at the top of the screen. A number line (1 to 10) appears at the bottom of the screen. *Zoid is baking some cakes for his friends. There are 4 friends in one window and 3 friends in the other window. How many cakes will Zoid need to bake for his friends? Show me the number on the number line.*

A picture of two groups of Zoid's friends appears at the top of the screen. At the bottom of the screen are four sets of different numbers of hats (7; 8; 9; 10). *Here are some of Zoid's friends. There are 5 friends in one group and 4 friends in another group. They all want to wear hats. How many hats do they need altogether? Show me the picture with the right number of hats for all of them.*

A picture of a boy and a girl appears at the top of the screen. At the bottom of the screen are four groups of coins (2p, 2p, 2p, 1p; 5p, 1p, 1p; 5p, 2p, 2p; 10p). *Here are 2 friends. Peter has 5p in his purse. Julie has 4p in her purse. How much money have they altogether? Show me which picture has the right amount of money.*

At the top of the screen a sum (4+2) appears in 'Zoid's number factory'. At the bottom of the screen there is a number line (1 to 10). *Here is a sum for you to try. 4 add 2 equals what? Show me the answer on the number line.*

At the top of the screen a sum (6+3) appears in 'Zoid's number factory'. At the bottom of the screen there is a number line (1 to 10). *Here is a sum for you to try. 6 add 3 equals what? Show me the answer on the number line.*

Set H – Subtraction

At the top of the screen a picture appears of a girl next to a shelf with 4 books on. At the bottom of the screen are 3 pictures of different numbers of books (2; 3; 4). *Nazreen is getting a book to read. How many books will be left on the shelf when Nazreen has taken one? Show me the answer.*

At the top of the screen is a picture of a pot with 4 pencils in and 3 children standing next to it. At the bottom of the screen are 3 pictures of pots with different numbers of pencils in (0; 1; 2). *These 3 children are taking 1 pencil each out of the pot. How many pencils will be left? Show me the answer.*

At the top of the screen is a picture of 5 bananas and 2 monkeys. At the bottom of the screen are 4 pictures of different numbers of bananas (1; 2; 3; 4). *These 2 monkeys are getting a banana each. How many bananas will be left? Show me the answer.*

At the top of the screen is a picture of a tree with 9 apples on it and 4 of Zoid's friends standing next to it. At the bottom of the screen is a number line (1 to 10). *There are 9 apples on this tree. Zoid's friends are going to pick one apple each. How many apples will be left on the tree? Show me the answer on the number line.*

At the top of the screen is a picture of a boy in a sweet shop. At the bottom of the screen are four coins (1p; 2p; 5p; 10p). *John has 7p to spend. He spends 5p on some sweets. How much money will he have left? Show me the answer.*

At the top of the screen is a picture of 10 chairs and 6 children. At the bottom of the screen is a number line (1 to 10). There are 10 chairs in the classroom and 6 friends. Each friend is going to sit on a chair. How many empty chairs will be left when each friend has sat down? Show me the answer on the number line.

At the top of the screen a sum (7-3) appears in 'Zoid's number factory'. At the bottom of the screen there is a number line (1 to 10). *Here is a sum for you to try. 7 take away 3 equals what? Show me the answer on the number line.*

Appendix 3: CoPS Baseline Communication items

Dimension:	1. Interest
Definition:	The degree of interest in and enthusiasm for the story and its components.
Dimension:	2. Description
Definition:	The accuracy of the child's perception and observation as indicated by his/her descriptions of the scene and the events.
Dimension:	3. Vocabulary
Definition:	The appropriateness, specificity, range and maturity of the vocabulary which the child uses to describe the scene and the events.
Dimension:	4. Grammar
Definition:	The maturity of grammatical structures used by the child.
Dimension:	5. Phonology
Definition:	The maturity of the child's pronunciation.
Dimension:	6. Fluency
Definition:	The overall confidence, spontaneity and effectiveness of the child's expressive communication skills.

Appendix 4: CoPS Baseline Personal and Social Development items

Dimension	**General description**
Relationship to adults	Is the child generally positive in his/her relationships with familiar adults, or negative (i.e. difficult, uncooperative, very shy, withdrawn, unconfident, disobedient, hostile, aggressive)?
Relationship to other children	In work and play activities is the child generally co-operative in his/her relationships with other children of a similar age, or uncooperative (i.e. negative, very shy, withdrawn, unconfident, hostile, aggressive)?
Social-emotional maturity	Ability to cope with and display emotions in an age-appropriate manner, to share with other children, taking turns when appropriate, and to handle everyday disappointments. Ability to defer gratification, waiting until the appropriate time for rewards.
Self-confidence and perseverance	The child's degree of confidence in his/her own capabilities to learn, to persevere and accomplish tasks, and to achieve goals, being neither inappropriately *over-confident* nor *under-confident*.
Attention and concentration	Ability to attend to the teacher when required, to follow instructions and to concentrate on the task in hand without excessive susceptibility to distractions from other things going on in the classroom.
Motivation for learning	Degree of interest shown in the activities and new experiences encountered within school, and the degree of spontaneous interest shown in activities and events happening outside the school. Amount of enthusiasm displayed in exploration of materials and in spontaneous questioning and in response to questions.
Independence and self-reliance	Degree of independence and self-reliance shown by the child in dressing, in looking after their possessions, in using materials confidently in the classroom, and in operating in the physical and social environment of the school.
Imagination and creativity	Degree of imagination shown in play, in generation of ideas, in response to stories and artistic experiences, and creativity with materials.
Co-ordination (gross motor skills)	General co-ordination of movement of large muscles, including walking, running, and jumping.
Manipulation (fine motor skills)	Accuracy of fine-motor skills, including use of pens, brushes, scissors, manipulation of materials with hands and fingers, and eye-hand co-ordination skills.

Appendix 5: CoPS Baseline Questionnaire (March 2001)

Baseline Project Questionnaire

Name of School: ....

School Address: ....

....

....Postcode: ....

Telephone number: ....

Name of person completing questionnaire: ....

1) How did you find the installation of CoPS Baseline?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found it difficult to install CoPS Baseline, please briefly outline the difficulty you encountered: ....

....

....

2) Is there any way you think the installation should be improved? ....

....

....

3) How did you find registering children on CoPS Baseline?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found registering children difficult, please briefly outline the difficulty you encountered: ....

....

....

4) Is there any way you think the registration procedure should be improved?

....

....

....

5) Did you use the training video? Yes / No

If 'No' please explain why: ....

....

6) If 'Yes', how would you rate the Training Video:

For content

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

For presentation

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

For organisation

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

7) What did you find to be the most useful part of the Training Video?

....

....

....

8) What did you find to be the least useful part of the Training Video?

....

....

....

9) How would you rate the Teachers Manual?

For content

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

For presentation

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

For organisation

Very poor	Poor	Satisfactory	Good	Excellent
❑	❑	❑	❑	❑

10) What did you find to be the most useful part of the Teachers Manual?

....

....

....

11) What did you find to be the least useful part of the Teachers Manual?

....

....

....

12) How did you find running the Communication module with children?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found running the module difficult, please briefly outline the difficulty you encountered: ....

....

....

13) Please rate the children's response (in general) to the Communication module?

Greatly disliked	Slightly disliked	Indifferent	Slightly enjoyed	Greatly enjoyed
❑	❑	❑	❑	❑

14) How would you rate the level of difficulty of the Communication module for the children you have assessed with the system?

Far too easy	A little easy	About right	A little difficult	Far too difficult
❑	❑	❑	❑	❑

15) How useful do you consider the Bumble Street and Bumble Park stories to be as a way of assessing children's expressive language?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

16) How do you rate the scoring system for the Communication module?

Very difficult	A little difficult	Satisfactory	Fairly easy	Very easy
❑	❑	❑	❑	❑

17) How do you rate the <u>usefulness</u> of the information in the Communication Report?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
❑	❑	❑	❑	❑

18) How do you rate the <u>clarity</u> of the information in the Communication Report?

Very confused	A little confused	Satisfactory	Clear	Very clear
❑	❑	❑	❑	❑

19) How do you rate the <u>content</u> of the information in the Communication Report?

Completely inadequate	A little inadequate	Satisfactory	Fairly comprehensive	Very comprehensive
❑	❑	❑	❑	❑

20) How did you find running the Literacy module with children?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

21) Please rate the children's response (in general) to the Literacy module?

Greatly disliked	Slightly disliked	Indifferent	Slightly enjoyed	Greatly enjoyed
❑	❑	❑	❑	❑

22) How would you rate the level of difficulty of the Literacy module for the children you have assessed with the system?

Far too easy	A little easy	About right	A little difficult	Far too difficult
❑	❑	❑	❑	❑

23) How would you rate the educational relevance of the <u>content</u> of the items in the Literacy module?

Not at all Relevant	Mostly irrelevant	Half and half	Mostly relevant	Very relevant
❑	❑	❑	❑	❑

24) How do you rate the <u>usefulness</u> of the information in the Literacy Report?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
❑	❑	❑	❑	❑

25) How do you rate the <u>clarity</u> of the information in the Literacy Report?

Very confused	A little confused	Satisfactory	Clear	Very clear
❑	❑	❑	❑	❑

26) How do you rate the <u>content</u> of the information in the Literacy Report?

Completely inadequate	A little inadequate	Satisfactory	Fairly comprehensive	Very comprehensive
☐	☐	☐	☐	☐

27) How did you find running the Mathematics module with children?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
☐	☐	☐	☐	☐

28) Please rate the children's response (in general) to the Mathematics module?

Greatly disliked	Slightly disliked	Indifferent	Slightly enjoyed	Greatly enjoyed
☐	☐	☐	☐	☐

29) How would you rate the level of difficulty of the Mathematics module for the children you have assessed with the system?

Far too easy	A little easy	About right	A little difficult	Far too difficult
☐	☐	☐	☐	☐

30) How would you rate the educational relevance of the <u>content</u> of the items in the Mathematics module?

Not at all relevant	Mostly irrelevant	Half and half	Mostly relevant	Very relevant
☐	☐	☐	☐	☐

31) How do you rate the <u>usefulness</u> of the information in the Mathematics Report?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
☐	☐	☐	☐	☐

32) How do you rate the <u>clarity</u> of the information in the Mathematics Report?

Very confused	A little confused	Satisfactory	Clear	Very clear
☐	☐	☐	☐	☐

33) How do you rate the <u>content</u> of the information in the Mathematics Report?

Completely inadequate	A little inadequate	Satisfactory	Fairly comprehensive	Very comprehensive
❑	❑	❑	❑	❑

34) How did you find using the Personal and Social module?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

35) How would you rate the educational relevance of the <u>content</u> of the items in the Personal and Social module?

Not at all Relevant	Mostly irrelevant	Half and half	Mostly relevant	Very relevant
❑	❑	❑	❑	❑

36) How do you rate the <u>usefulness</u> of the information in the Personal and Social Report?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
❑	❑	❑	❑	❑

37) How do you rate the <u>clarity</u> of the information in the Personal and Social Report?

Very confused	A little confused	Satisfactory	Clear	Very clear
❑	❑	❑	❑	❑

38) How do you rate the <u>content</u> of the information in the Personal and Social Report?

Completely inadequate	A little inadequate	Satisfactory	Fairly comprehensive	Very comprehensive
❑	❑	❑	❑	❑

39) How long (on average) did it take you to administer each of the CoPS Baseline modules with an individual child?

Communication	Literacy	Mathematics	Personal and Social
........mins	.........mins	........mins	........mins

40) Are there any ways you think the CoPS Baseline modules should be improved?

....

....

....

41) Overall, how did you find using the reports facilities?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found using the reports facilities difficult, please briefly outline the difficulty you encountered: ....

....

....

42) How do you rate the usefulness of the information in the Parent Report?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
❑	❑	❑	❑	❑

43) How do you rate the clarity of the information in the Parent Report?

Very confused	A little confused	Satisfactory	Clear	Very clear
❑	❑	❑	❑	❑

44) How do you rate the content of the information in the Parent Report?

Completely inadequate	A little inadequate	Satisfactory	Fairly comprehensive	Very comprehensive
❑	❑	❑	❑	❑

45) Is there any way you think the reports facilities in general could be improved?

....

....

....

46) Overall, how do you rate CoPS Baseline in comparison with the QCA Scales?

Much worse	A little worse	About the same	A little better	Much better
❑	❑	❑	❑	❑

Please explain your reasons: ....

....

....

47) In your view what are the main benefits or advantages of CoPS Baseline?

....

....

....

48) In your view what are the main <u>weaknesses or disadvantages</u> of CoPS Baseline?

....

....

....

49) How would you like to see CoPS Baseline improved?

....

....

....

50) Will your school be using CoPS Baseline in the future?

Definitely not	Probably not	Maybe	Yes, probably	Yes, definitely
❑	❑	❑	❑	❑

THANK YOU FOR YOUR CO-OPERATION

Appendix 6: CoPS Baseline Literacy Factor Analysis

Rotated Component Matrix[a]

	Component																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
SCORE A1	.113	.155						.132	.237	.106	.707		-.126						-.185
SCORE A2								.106	.822		.103					.108	.108		
SCORE A3	-.102			-.193	.116		.449	.136	.611		.234	.130	.191				.118		
SCORE A4	.108		.732						.280	-.197	.339			.146	.101			-.119	
SCORE A5	.108		.279		.219			-.111		.728	-.180		.169	-.103	.145				
SCORE A6			.774					.114					-.218		-.129				.142
SCORE A7														.781				.128	
SCORE A8		-.107		.139	.111			-.135		.446	.129	.469		.506		-.131			
SCORE A9		-.154	-.120	.174			.102						-.159	.237	.206			.744	
Score A10			.135				.154		-.115	-.153		.717		.144		.137			
SCORE B1	.224			.231	.189					.167					.238	.627			
SCORE B2					.119	.126									.802	.133		.150	
SCORE B3			.228				.123			-.122			-.715						-.114
SCORE B4	.146		.795							.347	-.111				.187				
SCORE B5	.119		.635				.392	-.108		.275						.348			
SCORE B6				-.122			.859		.135				-.136						
SCORE B7							.864										.168		
SCORE C1		.509						.216			-.523		-.126		.126	-.186			
SCORE C2	.124	.778								-.114					-.154				
SCORE C3		.679	-.174			.240				-.127			.186	-.120					
SCORE C4	.115	.236		.214			-.139				-.159		.149	.120					.683
SCORE C5		.743	.135									.137							.234
SCORE C6		.718	-.139	.244					.167	.180		-.104					-.112	-.107	
SCORE C7		.665		.298			-.178			.106							.119	-.218	
SCORE D1	.213	.398	-.139	.859	.151	.125		.118											
SCORE D2	.173	.205	.122	.705	.145	.104												.133	.213
SCORE D3	.139				.757									.152					
SCORE D4				.184	.573	.331	-.107		.169		-.125		-.196	-.123					
SCORE D5	.181			.347			-.132	.648	.156				-.192	-.115	.152	-.200			-.112
SCORE D6	.253	.308	.122	.225	.555	.156		.276			.143								.106
SCORE D7	.275					.101		.744											
SCORE D8	.181	.119		.184	.213	.108	-.223	.183		.119			.141				-.589		.209
SCORE E1	.298		.351						.224	.173		-.170	.110				.571		
SCORE E2					-.235					.710	.281					.155			
SCORE E3			.340	.255		.165			.186	.185		.409	.175	-.259	-.119		-.101	.136	-.279
SCORE E4	.296	.108	-.212				.114	.239	.165		-.164	.483		-.116		-.185		-.334	.119
SCORE E5	.539	.107		.278	.178	.156		.273		.106						.140	.104	-.268	-.182
SCORE E6	.535				.167	.166	.112	.324			.133		-.105	.139		.104			
SCORE E7	.322	.102	.125		.140	.200	.152	.239				-.244	.243	.201		-.289			-.133
SCORE E8	.198	.140	.345			-.138	-.108		.248	.174	-.115		.187	-.105			.212	.437	.101
SCORE F1	.622		.183	.182				.210			-.125		.159	.138	-.116				-.101
SCORE F2	.688			.332				.131			.162				.175				.144
SCORE F3	.574	.225		.131	-.106	.132	-.140	-.114			.181		.232		.100	-.281		.108	
SCORE F4	.725														.157		.173		
SCORE F5	.643	.110				.116				.157	-.102	.104	-.248		-.304		-.134	-.105	
SCORE F6	.694				.162				-.157						-.142			.133	.121
SCORE F7	.705			-.171		.121	-.104					.218	-.294	-.155					
SCORE G1	.309	.111		.282	.310				.135			.102	-.143	.196	-.214		.520	.143	.219
SCORE G2	.317	.259		.296	.457							.186	.346	-.122	.146		-.202		
SCORE G3	.219	.159		.361	.247				-.187		.138		.109			-.583	.112		
SCORE G4						.742		.110			-.138			.128	.122	.149			
SCORE G5			-.109		.225	.487			-.104	.176	.418	-.123		-.213				.114	
SCORE G6	.378			.124	.204	.242	.190	.301	.120					-.163	.105		-.244	-.209	.191
SCORE G7								.344	-.540		.119	.356	.190			.174	.197	-.118	
SCORE H1	.346	.258			.368		.162	.318				.162	.198	-.160	-.322				
SCORE H2	.331	.115		.343	.239	.140	.124	.195					.104	-.265	-.359		.229	.180	
SCORE H3	.259			.379	.200	.291	.109	.222			-.170		.169		-.192	.158			-.301
SCORE H4	.328			.234	.138	.182	.228	.345			.245	.137		-.151		-.184		.164	.197
SCORE H5	.525		.108	.336	.341	.149		.202	-.111						-.128				
SCORE H6	.289	.147			.259	.653							-.103						
SCORE H7	.270	.110		.179	.130	.702				-.140		.128	.114						.159

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 28 iterations.

Appendix 7: CoPS Baseline Mathematics Factor Analysis

Rotated Component Matrix(a)

	Component																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
SCORE A1									-.101			-.135					.864				
SCORE A2	.268		.146			.890			.100	-.113		.105	-.150		-.151					.235	.179
SCORE A3	.124	.115				.854										.129				-.132	
SCORE A4		.144	.122		-.180	.185		-.200		-.405	-.174	.152	.255	.206	.255		.222		-.285	.142	
SCORE A5	.135			-.153					.121	.732			.174		.107		.150				
SCORE A6		.138			.136			.111			.220		.728		-.181						
SCORE B1	.165	.127	-.106	.138		.181	-.513			.173	.188			.260	.211	.240		.274			
SCORE B2	.184	.133			.108	.110			.152			.111		.747					.132		.124
SCORE B3	.358	.262	-.109	-.169		.133	-.169	.112		-.242	.301		.200			.145	-.181	-.154	.116		-.238
SCORE B4	.239		.188	.192		.122				.175			.521		.166		-.176	-.135			
SCORE B5					.249		-.214	.604				.283		.108	-.183	.121	.170				.165
SCORE B6	.193	.102		.152	.138		-.105	.129				.720	-.130				-.175	-.116			
SCORE B7	.108	.317			.413	.144	.113		-.326		.206	.125		.120		-.189		-.167	.137		-.138
SCORE C1					.102															.823	
SCORE C2	.245	.161	.214							-.230								.656		.116	
SCORE C3	.208		-.121			-.159			.108	.166	.102				.716	.200		.133			
SCORE C4	.271	.294	.244		.158			.172	.183					.252	.525		.170		-.123		
SCORE C5	.204	.177	.126	-.118	.295			.412	.112	.143	-.139	.229		.208	.213		.224				-.157
SCORE C6	.142	.420	-.123						.312		.222	.160			-.101			.225			-.372
SCORE C7		.127					.116	.776													
SCORE D1		.724	.102		.113									.151				.221			
SCORE D2		.528		-.101		.103	.132	.335	.200	-.127			.170		.195	-.109	-.108		.288		
SCORE D3		.786	.109	.129							-.104		.100		.109						
SCORE D4		.538	-.181		-.164		.155	-.101		.120	.321	.124	-.168	.104	-.165	.165	.111		.137	-.106	.239
SCORE D5	.307				-.187	-.220				-.161		.137		.104						.455	-.142
Score D'1		.167	.802			.158	-.125		.130											.170	.148
Score D'2	.116		.796			-.122	-.130						.138					.204		-.147	
SCORED'3	.160	.189	.512			.167	.140			.390				.153	.127	.131	-.102		-.146	.125	-.220
SCORED'4	.167	.106	.156	.206	-.187			.237			.142			.509		.234					-.242
SCORED'5			-.187	-.192		.132	.197	.109			.237	.651	.250	.151						.113	
SCORED'6			.101				.115				.744	.125	.168							-.115	
SCORED'7					.472	-.172	.149		-.190	.154	.148	.108		.468	-.219						
SCORE E1	.399	.182	.235						.104				.102						.112		.626
SCORE E2	.683				.101	.127		-.104		-.165	.162	.136			.128		.163	.110	-.184		.218
SCORE E3	.621	-.108	.284								.258			.102	.110		-.125				.334
SCORE E4	.759				.115	.235		.202	.112											.108	
SCORE E5	.518		.242	.128	.133	.123		.189	-.282		.136			-.112		.201		-.232	-.190		-.242
SCORE F1	.496	.220		.144		.109	-.170				.114		.107	.141	.280					-.144	.221
SCORE F2						.157									.152	.817			.104		
SCORE F3	.487	.183				-.127			.262	.337	.140				.169	.112			-.130	.216	.292
SCORE F4	.231			.115	.159			.109			.467	.166	-.152			.107	.285		-.160	.385	
SCORE F5	.711		-.172						.124				.154			-.152		.150		.144	
SCORE F6	.747	.179		-.108					.106	.180			.102				.108		.104		
SCORE G1	.329		.246	-.136		.317				.219	.109						-.110	.609	-.106		
SCORE G2	.333	.328		.191				.139	-.148	.363	-.108	.265			.174	.213	-.152	.203			.105
SCORE G3	.673		.142			-.109				.300	-.184	.175		.179		.190			-.121		
SCORE G4	.636	.102					.107		.172	.257	-.143					.199	.142	.295	-.125		
SCORE G5				.881			-.114			-.139			-.117					-.110			
SCORE G6	.225	.197			.335			-.233	.348		-.156	.184	.311			.369	.164		.134		-.151
SCORE G7	.227	.147	.134	.394	.143		.388	.305	.165			.207	.219				.111	.158		-.122	-.127
SCORE G8				.838			.244						.135								
SCORE H1	.154	.117	.152			.169	.107		.647	.199			-.101	.131	.170		-.111				
SCORE H2	.178	-.185	.157	.156					.285		.209	.380	.115		.114				.362		
SCORE H3	.296	.198			.159			.179	.824		.186	-.141	.164			.106					
SCORE H4	.168				.757				.170				.135		.172			.120			
SCORE H5		.134							-.102							.102			.835		
SCORE H6		.204	.123		.536	-.232		.133	.139		.127					.251			-.240		
SCORE H7			-.191		.129		.806						.123	.130							
SCORE H8	.114	.188	-.102	.115			.601	-.109			.243		-.140			.214	-.325				

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 27 iterations.

Appendix 8: Contents of the CoPS Baseline manual, video and reports

Manual Contents:

Using the Pupil database

Select a pupil to assess

Useful tip for finding a pupil from a long list

Create a new pupil

Age calculation

Pupils with the same name

Additional pupil information

Re-testing a child

Editing pupil details

Deleting a pupil

Using the Assessor database

Program facilities during assessment

Pausing during an assessment (F2 key)

Repeating an assessment item (F3 key)

Premature abandonment of an assessment (F4 key)

Recording of scores

Advanced options

Backup CoPS Baseline database to floppy disk

Restore CoPS Baseline database from floppy disk

Export CoPS data to a tab delimited file

Import new pupils from a tab delimited file

Import procedures

Change password

The Report Generator

Communication report and Personal and Social Development report

Mathematics report and Literacy report

The Reports toolbar

Viewing a report

Printing

Exporting the reports

Opening and modifying a report using a word processor

Opening a report in other applications

Opening a report in a spreadsheet application

Opening a report in a database application

General

Copying CoPS Baseline reports screen to another application

Structure

Assessment procedure

Is the assessment environment satisfactory?

Is the equipment functioning correctly?

Is the child properly prepared for the task?

Is the assessment being conducted correctly?

Giving encouragement, prompts and feedback

5 Using the Communication module

Choice of story

Listening to the child's description

Scoring the child's description

Example protocols with scoring

A. Luke (Bumble Street)

B. Victoria (Bumble Street)

C. Joel (Bumble Street)

D. Leyla (Bumble Park)

E. Connor (Bumble Park)

F. Alastair (Bumble Park)

Fluency

Literacy Module

Aural comprehension and use of verbal concepts

Knowledge about print and letter recognition

Phonological awareness (rhyming and alliteration)

Simple Reading and Spelling

Mathematics Module

Number recognition and counting

Understanding sets and mathematical relationships

Understanding order, seriation, shape and pattern

Simple arithmetic (addition and subtraction)

Personal and Social Development Module

Relationship to adults

Relationship to other children

Social-emotional maturity

Self-confidence and perseverance

Attention and concentration

Motivation for learning

Independence and self-reliance

Imagination and creativity

Co-ordination (gross motor skills)

Manipulation (fine motor skills)

Case studies

Recommendations for the case study children

9 Technical and Troubleshooting Guide

How to use this guide

System check during installation

My computer failed one or more of the system checks

Screen resolution not 800 x 600

Date is not set to the British format

The Date or Time shown by my computer is wrong

Other computer problems

I can't print out

I can't hear any sounds

The screen saver interrupts CoPS Baseline tests

Information to supply to technical support

What version of CoPS Baseline do I have?

10 Appendices

Acknowledgements

References

Record of child's previous experience

Teacher rating form for the communication module

Teacher rating form for the personal and social development module

Item descriptions

Literacy

Set A – Aural comprehension

Set B – Verbal concepts

Set C – Knowledge about print

Set D – Letter recognition

Set E – Phonological Awareness (Rhyming)

Set F – Phonological Awareness (Alliteration)

Set G – Reading

Set H – Spelling

Mathematics

Set A – Number recognition

Set B – Counting

Set C – Relations

Set D – Sets

Set E – Seriation

Set F – Shape

Set G – Addition

Set H – Subtraction

Communication module – description of features and events

Bumble Street

Sequence of Events

Bumble Park

Sequence of Events

Pass rates for Literacy and Mathematics modules

Literacy pass rates

Literacy pass rates for age range 4 yrs 0 ms – 4 yrs 5 mos.

Literacy pass rates for age range 4 yrs 6 ms – 4 yrs 8 mos.

Literacy pass rates for age range 4 yrs 9 mos – 4 yrs 11 mos

Literacy pass rates for age range 5 yrs 0 mos – 5 yrs 2 mos

Literacy pass rates for age range 5 yrs 3 mos – 5 yrs 6

Mathematics pass rates

Mathematics pass rates for age range 4 yrs 0 ms – 4 yrs 5 mos.

Mathematics pass rates for age range 4 yrs 6 ms – 4 yrs 8 mos.

Mathematics pass rates for age range 4 yrs 9 ms – 4 yrs 11 mos.

Mathematics pass rates for age range 5 yrs 0 ms – 5 yrs 2 mos.

Mathematics pass rates for age range 5 yrs 3 ms – 5 yrs 6 mos .

Video contents:

Section titles and duration:

1. Introduction .4' 50"
2. Getting started .4' 40"
3. Using the literacy and mathematics modules .6' 15"
4. Using the communication module 16' 15"
5. Using the personal and social development module .3' 45"
6. Obtaining reports .3' 25"
7. Reporting to parents .4' 35"
8. Applying the outcomes of assessment 12' 40"
9. Special educational needs .3' 35"
10. Conclusions 11' 20"

(Total running time 72 minutes).

Reports:

CoPS BASELINE **CONFIDENTIAL**

Literacy Skills Report

Report for Ben Example Date of birth 16/07/93

Assessor's name	Date child assessed	Child's age at time of assessment
	13/09/97	5 years 6 months

	Pass	**Fail**	**Ability ascribed by teacher**	**Entry level**	**Entry mode**
Responses (out of 16)	11	5			
Average pass rates	0.48	0.41	Average	2	Child using mouse

Ben can write his name legibly

Raw scores

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Item	A2	A3	B3	B4	C4	C5	D5	D6	E6	E7	F6	F5	G5	G4	H5	H6
P/F	P	P	P	P	F	P	P	P	P	F	F	F	F	P	P	P
PR	0.90	0.90	0.90	0.90	0.65	0.26	0.23	0.16	0.26	0.19	0.42	0.45	0.35	0.29	0.35	0.10

T-Score	Centile score	z-Score	Grade
61.07	87.00	1.13	above average

Teaching recommendations

Building further on Ben's strengths in early-literacy related skills, which are in the areas of:

- Oral comprehension
- Ability to listen and understand
- Letter recognition

- Simple spelling

Providing a variety of activities and experiences which will address Ben's learning needs in early literacy-related skills, which are currently:

- Knowledge about print
- Phonological awareness (rhyming)
- Phonological awareness (alliteration)
- Simple reading

Teacher's comments

(Signed)

....

Class / Head Teacher

Baseline Assessment Report for Parents and Guardians

Report for Ben Example date of birth 16/07/93

All children in state schools in England and Wales are required to be assessed on school entry, using an accredited baseline assessment scheme. The purpose of this assessment is to find out children's strengths and learning needs when they first enter school, so that teachers can plan their education most effectively.

This school uses a computer-assisted baseline assessment system called CoPS BASELINE which we have chosen because of its efficiency and thoroughness and also because young children find it interesting and fun to do. CoPS BASELINE provides assessment in four key areas of early development: 1. Communication, 2. Literacy, 3. Mathematics, and 4. Personal and social development. The significance of these for your child's education is explained below.

In this report you will find a brief outline of how your child has been assessed and what his results are. We must stress that no system of assessment is perfect. Baseline assessment is like a 'snapshot' of each child's profile of capabilities at the time of the assessment - it does not necessarily show how well children will be able to learn. You should not become anxious about your child's results nor assume that they indicate how he or she is going to progress throughout schooling. We would welcome the opportunity to discuss results with all parents and guardians.

Baseline assessment is not about labelling children. All children are individuals - they have their learning needs as well as their strengths in learning. We need to know about children's strengths so that we can build on them, developing confidence and a sense of self-achievement. We also need to know about children's learning needs so that we do not expect too much at first and create unnecessary difficulty for the child; instead we want to provide the experiences which the child needs to meet their learning needs.

Module 1. Communication Skills

Good communication skills are essential for children to be able to learn. In this assessment Ben was shown a short animated cartoon and asked to talk about what happened in the story. If children are reluctant to talk at first, the teacher asked questions to find out what the child has noticed in the story and how the child describes the events.

We have found Ben's strengths in communication and expressive language to be:

- Ben showed great interest in the story.

- Ben was fairly observant and was able to describe the scene and the events in the story with reasonable accuracy.

- Ben's spoken grammar is reasonably developed for his age.

- Ben is easy to understand and speaks very clearly for his age.

- Overall, Ben expresses himself fairly well.

At the present time Ben's learning needs in communication and expressive language seem to be:

- Ben's vocabulary and range of words are rather below the average expected for his age.

Module 2. Literacy Skills

The development of effective reading and writing skills is founded on a rich variety of early experiences in language - such as listening and understanding, awareness of speech sounds, looking at books and hearing stories, and experience of visual observation, especially of the shapes of letters and words. In this assessment Ben was given a series of stimulating computer-delivered tasks involving these experiences. The computer automatically adjusts the difficulty of these tasks, so that the activities are interesting and at the right level for each individual.

We have found Ben's strengths in early literacy-related skills to be:

- Ben's ability to listen and understand is good for his age.

- Ben has some prior experience of books, and he is able to recognise a few letters of the alphabet.

- In simple reading and spelling activities, Ben showed average performance for his age.

At the present time Ben's learning needs in early literacy-related skills seem to be:

- Ben's awareness of language sounds (e.g. in rhymes) was not particularly good.

Module 3. Mathematics Skills

Mathematics development begins with an appreciation of numbers and counting, of mathematical relationships such as shape and size, of how things can be organised in to sets and series, and how useful calculations can be made by adding or subtracting things. In this assessment Ben was given a series of stimulating computer-delivered tasks involving these mathematical and number-related activities. The computer automatically adjusts the difficulty of these tasks, so that the activities are interesting and at the right level for each individual.

We have found Ben's strengths in early maths and number-related skills to be:

- Ben's awareness of numbers and ability to count is good for his age.

- Ben's understanding of mathematical relationships (e.g. size) is good for his age.

- Ben's awareness of series in number work and appreciation of shape is average for his age.

- In simple addition and subtraction, Ben showed above-average performance for his age.

Module 4. Personal and Social Skills

During Ben's time at school we have observed his interaction with adults and other children, how he responds to learning activities, his levels of attention and concentration, and his strengths and learning needs in various other aspects of social and personal development which are important for educational progress.

We have found Ben's strengths in personal and social development to be:

- Ben's relationships to adults in the school are fairly good, he is usually cooperative and positive in his manner.

- Socially and emotionally, Ben is very mature for his age.

- Ben is fairly confident in tackling tasks at school and shows some degree of perseverance in his work.

- Ben is able to follow instructions and has reasonable concentration skills in the classroom.
- Ben is very enthusiastic and shows a high degree of interest in learning.
- Ben displays some creativity and imagination in work and play.
- Ben has good overall physical coordination.
- Ben's manipulation skills and hand-eye coordination are average for his age.

At the present time Ben's learning needs in personal and social development seem to be:

- In general, Ben does not have particularly good relationships with other children.
- Ben is not very independent for his age, and is rather lacking in confidence.

Conclusions

(Signed) Class / Head Teacher

....

Appendix 9: LASS Secondary items

Single Word Reading

1. Ear
2. Fly
3. Key
4. Log
5. Van
6. Face
7. Game
8. Head
9. Leaf
10. Nail
11. Safe
12. Fence
13. Heart
14. Table
15. Jungle
16. Narrow

17. Sailor

18. Ticket

19. Valley

20. Weight

21. Balance

22. Ceiling

23. Licence

24. Receipt

25. Accident

26. Magazine

27. Laboratory

28. Rhinoceros

29. Mountaineer

30. Ventriloquist

Sentence Reading

Probe items:

There was a severe weather ____ on the news

warring warning earningraining warming

A new ____ of the book has been published

education library edition editing tuition

Vegetables provide a good form of ____

nonexistent abolishment nourishment fruit nonchalant

There is a huge ____ of water in the pool

quality quantities identity quantity liquefy

The builder gave a written ____ for the cost of the work

quotation question rotation quotient house

Test items:

Sonia was wearing ____ socks

hod odd owed ode add

Ellen enjoyed studying ____

salience science silence conscious scientist

Michael ____ go to the party

night stood mighty right might

There is a sign ____ the shop door

abode bough window glove above

Ian tried to ____ the biscuit tin

retch reach each react eats

The family are going ____ for their holiday

beach broader abroad abrupt broad

The actress had to ____ her lines

learn lean yearn leant lure

Anna does not eat meat because she is a ____

abattoir federation vegetarian vegetable proletarian

The campsite is in a ____ village

painful piece valley forceful peaceful

George reads a ____ paper

dairy dally gaily daily daises

Magda tried to ____ the noise outside

ignite gnaw loud snore ignore

Alex was very ____ with his money

generous onerous generals currency glorious

Carl heard an ____ in the cave

macho etch bats echo arch

The sisters are ____ twins

illogical identical political identity gender

The necklace was very ____

valuable variable viable arguable jewellery

The ____ came to sort out the problem

engender genuine engine dominion engineer

The programme showed the animals in their natural ____

habituate inhabit biting habitat acrobat

The surgeon was ready to begin the ____

oppression opposition doctor iteration operation

Sam had a good ____ of winning the race

change glance finish chance chant

The football team did not ____ for the World Cup

qualify quality amplify qualms qualified

Louise had a very vivid ____

immigration imagination emaciation imaginary pagination

Jo won the karate ____

tantamount tournament parliament judo torment

Hiroko made a ____ about her school subjects

decision incision deciding decisive deception

The ____ came to fix the cooker

elastication electrician kitchen electricity election

Alison had a good ____ of mathematics

kilometre neology knowledge addition acknowledge

The names in the telephone book are listed ____

apologetically alphabetically hypothetically dialling arithmetically

The experiment had a ____ effect

scientist patience negative derogative negation

Sunil chose the police force as his ____

occlusion inoculation truncheon occupation syncopation

Sharik had difficulty learning his ____ tables

misapplication mutilation implication multiplication eating

Mohammed is very ____ about sport

encyclopaedia infusion enthusiastic hockey ecclesiastic

Last summer the ____ reached 35°C

Winter temperate literature temperature temperament

The ____ reported on the fire

jauntiest engine journalist journalism interview

The ____ selected the correct medicine

pharmacist physics farmstead drug classicist

Abigail wanted to join a wildlife ____

urbanisation orchestration lion organisation origination

The builder positioned the shelves ____

horizontally horrify horizon incidentally wall

The family tried to be more ____ at the supermarket

ecology economical ecumenical anatomical shopping

The celebrity wrote her ____

geography famousautobiography typewriter accomplish

The house was built on a solid ____

formation inundation foundries foundation bricks

The new secretary was a very ____ worker

efficient effluent typist efficacy office

The company consulted a ____ advisor

fictional financed financier infinity financial

The singer had a pianist to ____ her

playing company accompany accomplice miscellany

There is a job ____ at the bookshop

valiancy vacancy infancy assistant vagrancy

The latest film is ____

secretarial sensibility sensation cinema sensational

Dean was against ____ on animals

experimentation instrumentation expiration tiger expectoration

A ____ shape has four sides

quadrennial triangle quadraphonic multilateral quadrilateral

Beth turned on the light to ____ the room

illiterate eliminate illuminate darken ruminate

The sailor fixed his position by measuring latitude and ____

locatable magnitude compass longitude longing

The doctor wrote a ____ for the medicine

preservation tablet scripture subscription prescription

Margaret is in hospital with ____

panorama numeral pneumatic ambulance pneumonia

The supermarket is a large ____

estrangement establishment astonishment cereals accomplishment

The unicorn is a ____ creature

mathematical morphology mythology horticultural mythological

The spy tried to be ____

incongruity inauspicious weapon inconspicuous agent

The village is in an ____ area

agronomical painter horticulture heraldry agricultural

Brenda needed ____ for her holiday

accumulation recommendation accommodation aeroplane consolidation

The new plane is very ____

aviation agoraphobic aerodrome hydrodynamic aerodynamic

There was an unforeseen ____ during the performance

complication implication compilation building computation

The bird watcher looked up the ____ of the bird

communication classification clarification simplification binoculars

Madhu looked at the work ____ carefully

sociable nodule schedule scheming studied

Julie's ____ annoyed Rick

potent refractiveness possessiveness possessively possible

The car manual had a ____ section

angry compliant supplementary suppository supplicatory

The ____ of the new school rules was successful

impersonation imperforation lesson simple implementation

A ____ existed within the company

oligarchy higher hierarchical hierarchy hilarity

The video was in the ____ section

miscalculates instantaneous mistake miscellaneous mischievous

____ can create a lot of work

Buoyancy Aristocrat Job Bureaucracy Bureau

The newspaper has a large ____

circulation civilisation calculation speculation crucible

Spelling

Probe items:

Foot – Thomas hurt his foot

Knife – The knife is very sharp

Machine – The chocolate machine was out of order

Harbour – The boat approached the harbour

Anchor – The ship dropped anchor in the storm

Test items:

Good – The film was very good

Bed – It was time for Alice to go to bed

Cup – The cup is full of coffee

Name – Luke's bag has his name on it

Bag – James put his shopping in the bag

Top – Rachael climbed to the top of the mountain

Red – Emily likes her red trousers

Day – It is a very sunny day

Hat – Michael wore his hat to work

Man – The man was late for the train

Dog – The dog ran after the stick

Pan – The pan is on the cooker

Jam – Peter put jam on his toast

Egg – Sarah had an egg for breakfast

Hen – The hen laid three eggs

Ice – Katy had some ice in her drink

Leg – Lauren hopped on one leg

Sea – The children swam in the sea

Pack – Louise had to pack her bag for school

Off – Rebecca got off the horse

Kick – Jack tried to kick the ball into the goal

Lid – The jar has a lid on it

Girl – The girl went to the park

Salt – Jonathan has salt and vinegar on his chips

Goal – Mark scored a goal in the cup final

Join – Hannah wanted to join the gang

Nose – The dog has a wet nose

Web – The fly got caught in the spider's web

Factory – The factory was very noisy

Idea – Shaun had an idea for the art competition

Pair – David bought a new pair of shoes

Nature – The nature programme was about insects

Planet – The alien came from another planet

Ocean – The ship sailed across the ocean

Diagram – The teacher drew a diagram on the board

Juice – The fruit juice was very sweet

Badge – The speaker wore a badge with his name on

Garage – The car was parked in the garage

Infection – Karen had a chest infection

Engine – The mechanic fixed the engine

Ghost – Daniel pretended to be a ghost

Increase – There was an increase in the prices at the school canteen

Education – The school had an excellent standard of education

Field – The sheep grazed in the field

Historical – The cathedral is a historical building

Treasure – Ancient treasure was discovered on a building site

Captain – The captain gave the team some instructions

Acrobat – The acrobat had performed in many countries

Festival – The festival went on for two days

Scientist – The scientist had some exciting findings

Celebrate – A party was held to celebrate Jo's birthday

Language – French is Kathryn's first language

Height – The two friends were the same height

Crumb – There was only one crumb of the cake left

Scene – The last scene of the play was the best

League – The team are at the top of the league

Vocabulary – Lisa was studying German vocabulary

Physics – Sam had forgotten his physics homework

Theatre – Joanna enjoyed going to the theatre

Sphere – A ball is a sphere

Quarter – A quarter of the cake had been eaten

Referee – The referee sent the player off

Favourite – Emma's favourite food is chocolate

Vegetable – Carrot is Kirsten's favourite vegetable

Orchestra – The orchestra played to their biggest audience

Aquarium – The fish are kept in an aquarium

Envelope – The card arrived in a blue envelope

Translator – The visitor from abroad needed a translator

Government – The government introduced a new policy

Mysterious – The lights in the sky were very mysterious

Mosquito – Matthew had a mosquito bite

Binoculars – Russell looked through his binoculars

Rehearse – The actors had to rehearse every scene of the play

Observatory – The astronomers worked at the observatory

Responsibility – Owning a pet is a big responsibility

Carriage – The bride was sat in a horse drawn carriage

Optician – Julia had an appointment with the optician

Literature – Heather enjoyed studying literature

Photosynthesis – Photosynthesis occurs in green plants

Exhibition – The art exhibition was very interesting

Environment – Recycled goods help to protect the environment

Architecture – The architecture of the building was elaborate

Legislation – New legislation is passed by Parliament

Circumference – Robert measured the circumference of the circle

Necessary – It was necessary for Andrew to call an ambulance

Vaccination – Tim needed a vaccination before his holiday

Quarantine – Animals must go into quarantine when arriving in the country

Rhythm – The drummer keeps rhythm for the band

Stethoscope – The doctor listened through his stethoscope

Hygienic – Brushing your teeth is hygienic

Reservoir – The reservoir provides water for nearby towns

Catastrophe – The earthquake was a major catastrophe

Surveillance – The shop had set up surveillance cameras

Auditory Memory

582

694

7392

4837

96184

53729

825736

314952

1845932

7436825

27359614

81395276

548263179

924716385

Non-word reading

onk

ank	ork	on'kay

frue

flu	fru'ee	fe'ru

dirsk

dir'esk	disk	dir'skay

im'pew

im'a'pew impel im'poo

wib'bing

why'bing wibb'inge webbing

phassid (fass'id)

fass'ide passid per'hass'id

shreb'ray

shree'bray shreb'ar'ay shreb'ry

rackent

racket rac'ken'tee rac'kay'ent

hindrious

hin'der'i'ous in'dri'ous hin'drus

lastermine

la'ster'min'y last'mine lass'er'mine

ver'mo'graph

fer'mo'graf ver'mo'grap ver'moj'er'af

esturgent (es'tur'jent)

es'tur'gent es'sur'jent es'tur'jen'tee

non'flup'pest

non'flu'pest non'fer'lupp'est no'en'flupp'est

moll'end'ing

moll'end'ine moll'end'inj moll'el'end'ing

corvationous (cor'va'shun'us)

cor'vat'ee'on'ous cor'va'shun'oose cor'fash'un'us

fir'iss'able

fire'iz'able fir'eye'zable fir'iss'ably

tersiptioning (ter'sip'shun'ing)

ter'sip'tee'on'ing ter'ser'ip'shun'ing ter'sipe'shun'ing

asinarchy (a'sin'a'key)

ass'in'o'key a'sin'ay'ar'key eye'seen'arch'ee

quidulousness

quid'lus'ness quid'u'loose'knees qui'due'low'us'ness

binassicacious (bin'ass'i'cay'shus)

bin'ace'i'cay'shus bin'ass'i'cak'ee'ous bin'ass'i'shus

gristulapsichon (griss'chew'lap'sick'on)

griss'chew'lap'sitch'on grize'chew'lap'sick'on griss'tull'ap'sick'own troughilicancy (troff'ill'ick'an'see)

trow'ill'ick'an'see troff'ill'iss'an'see troff'ill'ick'an'kee

crepurliscience (crep'url'ish'ee'ance)

crep'url'isk'ee'ance crep'url'she'ance crep'url'ee'si'ance

sprinantlistation (sprin'ant'list'ay'shun)

sprin'ant'list'at'ee'on sprin'list'ay'shun sprin'ant'lie'stay'shun

scorsprestulia (scores'press'chew'lee'a)

scores'press'tee'oo'lee'a scores'prest'lee'a scores'prest'ull'ee'a

<u>Syllable segmentation</u>

door<u>way</u>

door way orway do

app<u>le</u>

ap le pull plea

baggage

age bag agga agg

ladd<u>er</u>

lad er der add

music

sick mu muse uzi

orbit

bit or orb it

valley

val ey ley alley

window

dow win wind indo

caravan

avan ara van ava

enigma

igma nig ma ig

factory

tory act ree actor

guarantee

guaran rant ant gara

holiday

holly oli olid hol

kangaroo

kanoo	garoo	kangar	angoo

tomato

tomto	mato	tom	too'toe

umbrella

uma	umbra	umla	ulla

position

posi	posit	city	ossit

telephone

tele	elef	fone	telon

computer

comter	copter	commuter	pute

roundabout

rounda	round	about	runout

radio

dio	adi	raid	do

photograph

phograph	faf	potra	toga

recorder

corder re'order reader cord

dictionary

dickary dinnary dicty tinary

intuition

inition in'shun tu'shun into

normality

normity normal mality norlity

perimeter

pereter permeter peter metre

responsible

ressible response onsib ribble

scientific

scienic sky'fik scen'tik entif

formulated

formul'ed formated formula formed

educating

edcating editing educting dating

radiator

diator dater adiate rater

Appendix 10: LASS Secondary Questionnaire (March 2001)

LASS Secondary Validity Project Questionnaire

Name of School: ....

School Address: ....

....

.... Postcode: ....

Telephone number: ....

Name of person completing questionnaire: ....

1) How did you find the installation of LASS Secondary?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found it difficult to install LASS Secondary, please briefly outline the difficulty you encountered: ....

....

....

2) Is there any way you think the installation should be improved?

....

....

....

3) How did you find registering pupils on LASS Secondary?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found registering pupils difficult, please briefly outline the difficulty you encountered: ....

....

....

4) Is there any way you think the registration procedure should be improved?

....

....

....

5) In general, how did you find running the LASS Secondary tests with pupils?

Very difficult	A little difficult	Satisfactory	Easy	Very easy
❑	❑	❑	❑	❑

If you found running the tests difficult, please briefly outline the difficulty you encountered: ....

....

....

6) In general, how did the pupils find LASS Secondary?

Enjoyed none of the tests	Enjoyed only a few of the tests	Enjoyed some but not others	Enjoyed most of the tests	Enjoyed all of the tests
❑	❑	❑	❑	❑

7) Which tests did the pupils enjoy most?

....

....

Please briefly outline what they enjoyed about them:

....

....

....

8) Which tests did the pupils enjoy least?

....

....

Please briefly outline what they disliked about them:

....

....

....

9) How do you rate the suitability of the Cave test as an assessment of visual memory?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

10) How do you rate the suitability of the Mobile test as an assessment of auditory memory?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

11) How do you rate the suitability of the Non-words test as an assessment of phonic skills?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

12) How do you rate the suitability of the Segments test as an assessment of phonological processing?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

13) How do you rate the suitability of the Sentence test as an assessment of reading ability?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

14) How do you rate the suitability of the Spelling test as an assessment of spelling ability?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

15) How do you rate the suitability of the Reasoning test as an assessment of non-verbal intelligence?

Totally Unsuitable	Not very suitable	Adequate	Fairly suitable	Highly suitable
❑	❑	❑	❑	❑

16) Is there any way you think the LASS Secondary tests should be improved?

....

....

....

17) How do you rate the usefulness of the Reports given in LASS Secondary?

No use at all	Needs improvement	Satisfactory	Fairly useful	Very useful
❑	❑	❑	❑	❑

18) Is there any way you think the LASS Secondary Results should be improved?

....

....

....

19) How would you rate LASS Secondary overall?

Poor	Not very good	Average	Good	Very good
❑	❑	❑	❑	❑

If you rate LASS Secondary as poor or not very good, please briefly outline your reasons:

....

....

....

20) In your view what are the main benefits or advantages of LASS Secondary

....

....

....

21) In your view what are the main weaknesses or disadvantages of LASS Secondary?

....

....

....

22) How would you like to see LASS Secondary improved?

....

....

....

23) Will you be using LASS Secondary in the future?

Definitely not	Probably not	Maybe	Yes, probably	Yes, definitely
❑	❑	❑	❑	❑

24) How did you find running the following pen and paper tests with pupils?

	Very difficult	A little difficult	Satisfactory	Easy	Very easy
Spatial Span					
Digit Span					
PhAB					
NFER Sentence Completion					
BSTS 3					
MAT					

25) In general, which one of each of the following pairs of tests did pupils enjoy most?

a)	LASS Cave		OR	Spatial Span	
b)	LASS Mobile		OR	Digit Span	
c)	LASS Non-words		OR	PhAB Non-word reading	
d)	LASS Segments			PhAB Spoonerisms	
e)	LASS Sentence Reading		OR	NFER Sentence Completion	
f)	LASS Spelling		OR	BSTS 3	

g)	LASS Reasoning		OR	MAT	

THANK YOU FOR YOUR CO-OPERATION

Appendix 11: LASS Secondary centile scores

Cave

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
0	1	1	1	1	1
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	2	2	1	1	1
14	5	2	2	1	1
15	5	2	2	2	2
16	6	3	2	2	2
17	8	3	3	2	2
18	10	4	3	3	3
19	11	5	4	3	3
20	12	8	6	5	4
21	14	12	10	8	6
22	17	17	15	14	13
23	19	18	18	17	14
24	24	23	23	22	15
25	30	30	29	28	18
26	35	35	34	34	23
27	41	41	40	36	28
28	42	41	40	38	33
29	47	46	45	45	38
30	55	54	51	51	41
31	60	58	54	53	48
32	67	63	58	55	51
33	76	66	63	62	56
34	78	70	67	66	61
35	87	74	71	70	64
36	91	77	74	73	67
37	94	79	77	77	71
38	96	81	80	78	75
39	96	84	84	83	80
40	97	90	89	86	84

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
41	97	92	89	89	87
42	98	94	93	93	92
43	99	96	96	96	94
44	99	97	96	96	95
45	99	97	96	96	95
46	99	98	96	96	96
47	99	98	97	97	96
48	99	98	97	97	96
49	99	98	97	97	97
50	99	98	98	98	97
51	99	99	98	98	98
52	99	99	98	98	98
53	99	99	99	99	98
54	99	99	99	99	99
55	99	99	99	99	99
56	99	99	99	99	99
57	99	99	99	99	99
58	99	99	99	99	99
59	99	99	99	99	99
60	99	99	99	99	99
61	99	99	99	99	99
62	99	99	99	99	99
63	99	99	99	99	99
64	99	99	99	99	99
65	99	99	99	99	99
66	99	99	99	99	99
67	99	99	99	99	99
68	99	99	99	99	99
69	99	99	99	99	99
70	99	99	99	99	99

Mobile

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
0	1	1	1	1	1
1	4	2	1	1	1
2	8	3	2	2	2
3	10	8	8	3	3
4	20	15	15	15	8
5	48	34	34	28	26
6	69	52	51	45	44
7	86	73	70	67	61
8	94	85	82	81	80
9	99	93	91	90	89
10	99	97	96	95	94

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
11	99	99	99	98	97
12	99	99	99	99	98
13	99	99	99	99	99
14	99	99	99	99	99

Non-words

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
0	1	1	1	1	1
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	5	3	3	2	2
5	8	5	4	2	2
6	15	10	10	7	5
7	25	14	14	14	13
8	29	18	17	16	16
9	38	21	21	20	20
10	44	27	27	26	25
11	48	34	34	33	33
12	51	45	43	43	42
13	58	50	49	49	45
14	63	53	52	52	46
15	69	57	57	56	47
16	74	63	63	62	62
17	82	70	69	69	68
18	88	85	85	83	74
19	91	87	83	84	83
20	94	89	88	87	87
21	95	89	89	88	87
22	99	97	96	96	95
23	99	99	99	99	98
24	99	99	99	99	99
25	99	99	99	99	99

Segments

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
0	1	1	1	1	1
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	2	1	1	1	1
6	3	3	3	2	2
7	3	3	3	3	3
8	8	7	7	6	4
9	9	8	7	6	5
10	15	12	11	10	9
11	21	16	16	14	13
12	23	20	20	18	14
13	29	25	25	20	16
14	30	27	27	22	17
15	36	31	30	24	18
16	39	32	31	26	20
17	41	37	37	28	22
18	46	39	38	29	28
19	50	43	42	35	35
20	60	50	50	42	40
21	62	57	56	45	45
22	69	60	60	48	47
23	74	68	67	55	55
24	82	71	70	61	60
25	88	76	75	71	70
26	93	82	82	76	76
27	95	92	90	81	81
28	97	92	91	85	85
29	98	97	95	94	94
30	99	98	98	98	97
31	99	99	99	99	98
32	99	99	99	99	99
33	99	99	99	99	99

Single word reading

raw score	11 year centile	12 year centile	13 year centile	14 year centile	15 year centile
0	1	1	1	1	1
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	2	1	1	1	1
14	2	1	1	1	1
15	2	1	1	1	1
16	2	1	1	1	1
17	2	1	1	1	1
18	2	1	1	1	1
19	2	1	1	1	1
20	2	1	1	1	1
21	2	1	1	1	1
22	3	2	1	1	1
23	4	2	1	1	1
24	5	3	2	1	1
25	6	4	3	1	1
26	7	5	4	2	2
27	9	7	5	4	4
28	14	12	8	7	5
29	25	24	17	16	15
30	99	99	99	99	99

Sentence reading

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
99	0.3000	0.3000	0.3000	0.3000	0.3000
98	0.3611	0.3452	0.3452	0.3451	0.3450
97	0.3645	0.3525	0.3500	0.3452	0.3451
96	0.3681	0.3603	0.3549	0.3453	0.3452
95	0.3712	0.3604	0.3579	0.3454	0.3453
94	0.3756	0.3605	0.3602	0.3455	0.3454
93	0.3899	0.3610	0.3604	0.3456	0.3455
92	0.4006	0.3618	0.3611	0.3457	0.3456

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
91	0.4105	0.3687	0.3685	0.3577	0.3527
90	0.4203	0.3688	0.3686	0.3606	0.3588
89	0.4241	0.3789	0.3747	0.3622	0.3594
88	0.4282	0.3795	0.3772	0.3638	0.3613
87	0.4443	0.3839	0.3779	0.3707	0.3616
86	0.4636	0.3872	0.3787	0.3772	0.3619
85	0.4989	0.3877	0.3815	0.3773	0.3627
84	0.5371	0.3882	0.3849	0.3774	0.3635
83	0.5497	0.3920	0.3915	0.3775	0.3655
82	0.5624	0.3949	0.3947	0.3776	0.3683
81	0.5753	0.4024	0.3999	0.3777	0.3728
80	0.5890	0.4112	0.4106	0.3778	0.3754
79	0.5935	0.4214	0.4205	0.3804	0.3782
78	0.5963	0.4320	0.4298	0.3843	0.3817
77	0.5982	0.4440	0.4399	0.3893	0.3846
76	0.6000	0.4460	0.4400	0.3949	0.3900
75	0.6011	0.4516	0.4425	0.4006	0.3956
74	0.6079	0.4596	0.4482	0.4007	0.4002
73	0.6222	0.4625	0.4522	0.4035	0.4030
72	0.6447	0.4651	0.4544	0.4081	0.4063
71	0.6544	0.4762	0.4684	0.4151	0.4131
70	0.6642	0.4977	0.4926	0.4221	0.4176
69	0.6789	0.5065	0.5057	0.4253	0.4237
68	0.6984	0.5106	0.5089	0.4396	0.4302
67	0.7071	0.5189	0.5111	0.4423	0.4376
66	0.7160	0.5279	0.5240	0.4450	0.4439
65	0.7200	0.5317	0.5371	0.4752	0.4513
64	0.7273	0.5610	0.5598	0.5135	0.4589
63	0.7309	0.5723	0.5645	0.5356	0.4627
62	0.7383	0.5836	0.5789	0.5617	0.4687
61	0.7394	0.5901	0.5893	0.5720	0.4865
60	0.7407	0.5947	0.5926	0.5824	0.5043
59	0.7482	0.6002	0.5975	0.5899	0.5155
58	0.7512	0.6249	0.6224	0.5992	0.5359
57	0.7553	0.6303	0.6288	0.6054	0.5475
56	0.7602	0.6413	0.6350	0.6128	0.5590
55	0.7617	0.6495	0.6386	0.6159	0.5782
54	0.7632	0.6505	0.6424	0.6190	0.5971
53	0.7687	0.6599	0.6459	0.6279	0.5997
52	0.7762	0.6603	0.6495	0.6315	0.6028
51	0.7793	0.6687	0.6657	0.6387	0.6035
50	0.7824	0.6722	0.6709	0.6421	0.6042
49	0.7872	0.6841	0.6810	0.6442	0.6087
48	0.7920	0.6957	0.6921	0.6501	0.6128
47	0.7944	0.7086	0.7052	0.6605	0.6235
46	0.7968	0.7188	0.7122	0.6710	0.6355
45	0.8010	0.7251	0.7232	0.6745	0.6450
44	0.8059	0.7314	0.7241	0.6778	0.6547

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
43	0.8125	0.7420	0.7377	0.6782	0.6622
42	0.8191	0.7527	0.7401	0.6785	0.6688
41	0.8215	0.7600	0.7504	0.6854	0.6721
40	0.8251	0.7778	0.7608	0.6917	0.6768
39	0.8294	0.7882	0.7748	0.6944	0.6779
38	0.8335	0.7904	0.7879	0.6973	0.6813
37	0.8446	0.8059	0.7995	0.7021	0.6866
36	0.8558	0.8189	0.8109	0.7149	0.6896
35	0.8659	0.8250	0.8201	0.7258	0.7156
34	0.8760	0.8291	0.8238	0.7474	0.7447
33	0.8856	0.8384	0.8374	0.7650	0.7605
32	0.8952	0.8458	0.8450	0.7844	0.7836
31	0.9062	0.8569	0.8523	0.7876	0.7870
30	0.9172	0.8681	0.8593	0.7937	0.7906
29	0.9202	0.8706	0.8634	0.7981	0.7922
28	0.9242	0.8794	0.8762	0.8062	0.7941
27	0.9329	0.8885	0.8807	0.8177	0.7955
26	0.9418	0.8932	0.8858	0.8289	0.7975
25	0.9421	0.8979	0.8969	0.8306	0.8027
24	0.9422	0.9076	0.8987	0.8470	0.8097
23	0.9439	0.9107	0.9066	0.8651	0.8208
22	0.9458	0.9141	0.9105	0.8862	0.8394
21	0.9511	0.9196	0.9178	0.8991	0.8566
20	0.9568	0.9201	0.9199	0.9108	0.8723
19	0.9581	0.9339	0.9317	0.9128	0.8748
18	0.9593	0.9381	0.9358	0.9147	0.8759
17	0.9620	0.9429	0.9379	0.9233	0.8862
16	0.9647	0.9471	0.9401	0.9320	0.8965
15	0.9655	0.9543	0.9433	0.9373	0.9032
14	0.9664	0.9596	0.9465	0.9427	0.9114
13	0.9698	0.9628	0.9501	0.9486	0.9179
12	0.9675	0.9645	0.9564	0.9513	0.9223
11	0.9676	0.9670	0.9613	0.9538	0.9273
10	0.9677	0.9675	0.9671	0.9563	0.9327
9	0.9686	0.9680	0.9673	0.9595	0.9435
8	0.9696	0.9690	0.9675	0.9643	0.9514
7	0.9698	0.9695	0.9688	0.9658	0.9567
6	0.9700	0.9699	0.9698	0.9675	0.9611
5	0.9705	0.9700	0.9699	0.9681	0.9636
4	0.9711	0.9710	0.9709	0.9705	0.9659
3	0.9712	0.9711	0.9710	0.9708	0.9660
2	0.9713	0.9712	0.9711	0.9710	0.9661
1	0.9999	0.9999	0.9999	0.9999	0.9999

Spelling

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
99	0.1000	0.1000	0.1000	0.1000	0.1000
98	0.2305	0.1751	0.1734	0.1433	0.1362
97	0.2452	0.1752	0.1742	0.1438	0.1375
96	0.2600	0.1972	0.1930	0.1451	0.1389
95	0.3022	0.2054	0.2010	0.1634	0.1408
94	0.3445	0.2148	0.2092	0.1925	0.1426
93	0.3551	0.2454	0.2156	0.1945	0.1454
92	0.3701	0.2831	0.2353	0.1966	0.1480
91	0.3772	0.2953	0.2552	0.2271	0.1555
90	0.3839	0.3144	0.2751	0.2479	0.1630
89	0.3942	0.3361	0.2851	0.2546	0.1745
88	0.4114	0.3483	0.3018	0.2701	0.1860
87	0.4356	0.3533	0.3257	0.2776	0.1941
86	0.4434	0.3583	0.3507	0.2853	0.2021
85	0.4566	0.3628	0.3581	0.3036	0.2156
84	0.4756	0.3673	0.3653	0.3209	0.2294
83	0.4824	0.3697	0.3694	0.3357	0.2553
82	0.4887	0.3721	0.3718	0.3512	0.2941
81	0.4925	0.3745	0.3735	0.3599	0.3050
80	0.4961	0.3754	0.3753	0.3689	0.3230
79	0.5042	0.3809	0.3795	0.3704	0.3275
78	0.5134	0.3869	0.3836	0.3717	0.3324
77	0.5139	0.3913	0.3845	0.3718	0.3369
76	0.5261	0.3962	0.3869	0.3719	0.3414
75	0.5353	0.3975	0.3876	0.3726	0.3494
74	0.5497	0.4028	0.3961	0.3754	0.3573
73	0.5554	0.4157	0.4040	0.3817	0.3649
72	0.5735	0.4336	0.4120	0.3877	0.3722
71	0.5765	0.4387	0.4340	0.3936	0.3755
70	0.5795	0.4538	0.4390	0.3976	0.3788
69	0.5859	0.4742	0.4709	0.4020	0.3815
68	0.6023	0.4846	0.4756	0.4038	0.3842
67	0.6042	0.4927	0.4794	0.4265	0.3952
66	0.6062	0.5009	0.4837	0.4484	0.4062
65	0.6153	0.5035	0.4893	0.4556	0.4253
64	0.6388	0.5172	0.4948	0.4627	0.4593
63	0.6456	0.5324	0.5055	0.4654	0.4635
62	0.6674	0.5466	0.5171	0.4681	0.4659
61	0.6780	0.5561	0.5265	0.4718	0.4687
60	0.6885	0.5756	0.5459	0.4752	0.4718
59	0.6977	0.5964	0.5754	0.4854	0.4757
58	0.7171	0.6173	0.6032	0.5018	0.4797
57	0.7281	0.6356	0.6254	0.5229	0.4843
56	0.7387	0.6540	0.6448	0.5444	0.4890
55	0.7679	0.6791	0.6667	0.5545	0.4925
54	0.7971	0.6985	0.6885	0.5645	0.4959

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
53	0.8074	0.7128	0.7055	0.5701	0.4979
52	0.8177	0.7271	0.7234	0.5779	0.5001
51	0.8255	0.7420	0.7328	0.5784	0.5027
50	0.8323	0.7565	0.7422	0.5787	0.5053
49	0.8384	0.7645	0.7515	0.5855	0.5115
48	0.8437	0.7725	0.7609	0.6064	0.5318
47	0.8456	0.7751	0.7645	0.6157	0.5612
46	0.8489	0.7805	0.7684	0.6249	0.5907
45	0.8539	0.7879	0.7740	0.6314	0.5935
44	0.8592	0.7909	0.7768	0.6373	0.5963
43	0.8654	0.7964	0.7872	0.6453	0.6019
42	0.8896	0.8069	0.7977	0.6613	0.6094
41	0.8959	0.8078	0.8002	0.6718	0.6130
40	0.9145	0.8186	0.8028	0.6821	0.6171
39	0.9216	0.8275	0.8072	0.6892	0.6208
38	0.9278	0.8364	0.8113	0.6959	0.6242
37	0.9284	0.8397	0.8158	0.7041	0.6351
36	0.9287	0.8432	0.8210	0.7122	0.6530
35	0.9307	0.8455	0.8314	0.7349	0.6652
34	0.9332	0.8484	0.8419	0.7638	0.6874
33	0.9381	0.8556	0.8538	0.7765	0.7154
32	0.9432	0.8678	0.8658	0.7894	0.7584
31	0.9433	0.8756	0.8725	0.7956	0.7866
30	0.9434	0.8894	0.8791	0.8187	0.8124
29	0.9436	0.9044	0.8974	0.8223	0.8210
28	0.9436	0.9215	0.9169	0.8351	0.8339
27	0.9447	0.9253	0.9209	0.8379	0.8356
26	0.9459	0.9290	0.9253	0.8405	0.8367
25	0.9474	0.9351	0.9253	0.8504	0.8473
24	0.9541	0.9373	0.9280	0.8618	0.8531
23	0.9594	0.9442	0.9312	0.8746	0.8654
22	0.9648	0.9488	0.9343	0.8907	0.8855
21	0.9681	0.9511	0.9391	0.8972	0.8956
20	0.9732	0.9526	0.9437	0.9038	0.9008
19	0.9743	0.9545	0.9454	0.9156	0.9117
18	0.9754	0.9575	0.9483	0.9262	0.9226
17	0.9757	0.9631	0.9562	0.9309	0.9298
16	0.9760	0.9691	0.9642	0.9358	0.9329
15	0.9783	0.9732	0.9653	0.9391	0.9321
14	0.9807	0.9768	0.9664	0.9425	0.9413
13	0.9814	0.9789	0.9705	0.9462	0.9407
12	0.9820	0.9805	0.9750	0.9540	0.9534
11	0.9827	0.9815	0.9755	0.9698	0.9623
10	0.9833	0.9822	0.9761	0.9760	0.9758
9	0.9898	0.9876	0.9771	0.9767	0.9760
8	0.9923	0.9893	0.9781	0.9779	0.9771
7	0.9928	0.9845	0.9799	0.9784	0.9777
6	0.9933	0.9901	0.9817	0.9812	0.9806

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
5	0.9936	0.9913	0.9826	0.9817	0.9811
4	0.9940	0.9925	0.9844	0.9833	0.9828
3	0.9941	0.9933	0.9897	0.9861	0.9850
2	0.9942	0.9940	0.9929	0.9921	0.9918
1	0.9999	0.9999	0.9999	0.9999	0.9999

Reasoning

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
99	0.2000	0.2000	0.2000	0.2000	0.2000
98	0.3332	0.2506	0.2428	0.2274	0.2059
97	0.3427	0.2724	0.2656	0.2479	0.2267
96	0.3519	0.2892	0.2884	0.2688	0.2476
95	0.3555	0.2971	0.2965	0.2956	0.2674
94	0.3649	0.3118	0.3096	0.3085	0.2871
93	0.3751	0.3227	0.3155	0.3143	0.2953
92	0.3906	0.3436	0.3384	0.3287	0.3026
91	0.3920	0.3655	0.3555	0.3457	0.3111
90	0.3933	0.3924	0.3735	0.3637	0.3193
89	0.4244	0.3984	0.3737	0.3701	0.3246
88	0.4550	0.4034	0.3740	0.3731	0.3296
87	0.4932	0.4242	0.3850	0.3781	0.3453
86	0.5309	0.4450	0.3961	0.3880	0.3521
85	0.5335	0.4497	0.4251	0.3897	0.3655
84	0.5360	0.4548	0.4521	0.3917	0.3777
83	0.5419	0.4658	0.4606	0.3938	0.3846
82	0.5475	0.4838	0.4796	0.3959	0.3896
81	0.5576	0.4971	0.4955	0.4064	0.3992
80	0.5678	0.5203	0.5124	0.4169	0.4088
79	0.5769	0.5246	0.5216	0.4235	0.4156
78	0.5860	0.5389	0.5307	0.4301	0.4232
77	0.5897	0.5455	0.5343	0.4324	0.4309
76	0.5914	0.5541	0.5480	0.4368	0.4353
75	0.5922	0.5587	0.5577	0.4481	0.4463
74	0.6183	0.5654	0.5610	0.4588	0.4542
73	0.6255	0.5704	0.5643	0.4690	0.4575
72	0.6440	0.5754	0.5706	0.4792	0.4605
71	0.6487	0.5811	0.5724	0.4855	0.4682
70	0.6554	0.5872	0.5849	0.5064	0.4759
69	0.6574	0.5982	0.5937	0.5152	0.4856
68	0.6595	0.6033	0.6022	0.5340	0.5074
67	0.6629	0.6085	0.6051	0.5520	0.5153
66	0.6658	0.6106	0.6081	0.5700	0.5347
65	0.6683	0.6128	0.6107	0.5812	0.5358
64	0.6733	0.6189	0.6174	0.5924	0.5369

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
63	0.6741	0.6245	0.6222	0.5982	0.5579
62	0.6749	0.6330	0.6320	0.6056	0.5789
61	0.6821	0.6372	0.6359	0.6135	0.5867
60	0.6906	0.6475	0.6448	0.6215	0.5946
59	0.6927	0.6500	0.6485	0.6245	0.5977
58	0.6948	0.6585	0.6529	0.6275	0.6010
57	0.6979	0.6632	0.6610	0.6325	0.6035
56	0.7036	0.6745	0.6709	0.6375	0.6069
55	0.7085	0.6763	0.6741	0.6419	0.6112
54	0.7113	0.6781	0.6765	0.6457	0.6162
53	0.7121	0.6793	0.6782	0.6492	0.6191
52	0.7237	0.6808	0.6802	0.6534	0.6228
51	0.7258	0.6829	0.6819	0.6573	0.6305
50	0.7279	0.6857	0.6831	0.6612	0.6383
49	0.7336	0.6882	0.6854	0.6644	0.6473
48	0.7413	0.6911	0.6878	0.6676	0.6563
47	0.7446	0.6979	0.6939	0.6702	0.6626
46	0.7479	0.7009	0.6980	0.6729	0.6689
45	0.7542	0.7041	0.7018	0.6782	0.6714
44	0.7607	0.7053	0.7042	0.6821	0.6739
43	0.7660	0.7141	0.7092	0.6881	0.6748
42	0.7718	0.7187	0.7152	0.6960	0.6757
41	0.7737	0.7243	0.7182	0.7020	0.6776
40	0.7760	0.7260	0.7211	0.7082	0.6795
39	0.7791	0.7308	0.7256	0.7156	0.6827
38	0.7817	0.7352	0.7325	0.7227	0.6851
37	0.7827	0.7389	0.7363	0.7281	0.6865
36	0.7837	0.7434	0.7402	0.7352	0.6880
35	0.7869	0.7456	0.7410	0.7396	0.6894
34	0.7921	0.7477	0.7432	0.7421	0.6915
33	0.7963	0.7521	0.7497	0.7469	0.7001
32	0.8005	0.7555	0.7535	0.7479	0.7086
31	0.8055	0.7645	0.7633	0.7495	0.7177
30	0.8105	0.7676	0.7656	0.7509	0.7277
29	0.8168	0.7746	0.7674	0.7575	0.7332
28	0.8227	0.7800	0.7752	0.7641	0.7380
27	0.8287	0.7850	0.7792	0.7699	0.7385
26	0.8347	0.7900	0.7846	0.7761	0.7391
25	0.8397	0.7967	0.7907	0.7882	0.7436
24	0.8397	0.7996	0.7952	0.7943	0.7504
23	0.8422	0.8053	0.8022	0.8014	0.7659
22	0.8452	0.8120	0.8075	0.8067	0.7823
21	0.8624	0.8181	0.8143	0.8137	0.8011
20	0.8806	0.8241	0.8162	0.8157	0.8100
19	0.8934	0.8361	0.8284	0.8246	0.8229
18	0.9078	0.8482	0.8396	0.8375	0.8256
17	0.9185	0.8555	0.8477	0.8428	0.8351
16	0.9293	0.8729	0.8658	0.8580	0.8548

centile	11 year raw score	12 year raw score	13 year raw score	14 year raw score	15 year raw score
15	0.9351	0.8854	0.8728	0.8652	0.8616
14	0.9402	0.8991	0.8808	0.8733	0.8674
13	0.9403	0.9190	0.9091	0.9052	0.8971
12	0.9404	0.9289	0.9188	0.9170	0.9168
11	0.9409	0.9346	0.9345	0.9328	0.9189
10	0.9415	0.9391	0.9364	0.9307	0.9217
9	0.9468	0.9445	0.9408	0.9402	0.9296
8	0.9507	0.9482	0.9413	0.9409	0.9377
7	0.9509	0.9504	0.9470	0.9460	0.9454
6	0.9512	0.9507	0.9506	0.9501	0.9493
5	0.9523	0.9519	0.9514	0.9504	0.9501
4	0.9545	0.9532	0.9529	0.9526	0.9515
3	0.9558	0.9548	0.9533	0.9530	0.9526
2	0.9586	0.9585	0.9580	0.9547	0.9537
1	0.9999	0.9999	0.9999	0.9999	0.9999

Appendix 12: Pupil Information Sheet

Pupil Information Sheet

Full Name: ______________________________

School: ______________________________

Date of Birth: ______________ Age: ______________

Sex:
- ❐ Male
- ❐ Female

Ethnicity:
- ❐ White
- ❐ Black Caribbean
- ❐ Black African
- ❐ Black Other
- ❐ Indian
- ❐ Pakistani
- ❐ Bangladeshi
- ❐ Chinese
- ❐ Other

Handedness:
- ❐ Left
- ❐ Right
- ❐ Ambidextrous or no clear preference

Language:
- ❐ English first language
- ❐ English second language

SEN Register:
- ❐ Not on SEN Register
- ❐ Stage 1
- ❐ Stage 2
- ❐ Stage 3
- ❐ Stage 4
- ❐ Stage 5

SEN Type: ☐ Severe Learning Difficulties

☐ Moderate Learning Difficulties

☐ Specific Learning Difficulties (e.g. dyslexia/dyspraxia)

☐ Speech and Communication Difficulties

☐ Physical Difficulties

☐ Hearing Impaired

☐ Visually Impaired

☐ Emotional and Behavioural Difficulties

☐ Autistic Spectrum Disorder

☐ Attention Deficit Hyperactivity Disorder

If the pupil has been identified as having dyslexia, please give the following information, where possible:

Age at which dyslexia was identified: ______________

Severity of dyslexia: ☐ Mild

☐ Moderate

☐ Severe

How much support does he/she receive? ________ (hours / week)

How long has he/she been receiving support? ____ yrs ____ mths

If the pupil has not been identified as having dyslexia, do you suspect that he / she has dyslexia? ☐ Yes

☐ No

Pupils in the Baseline Project or Secondary Validity Project

Which tests were preferred by the pupil?

☐ Computer

☐ Pen and Paper

☐ No Preference

Appendix 13: Discrepancy calculations for dyslexic students

ID: 1				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	14	-1.08	-1.84*	-1.52*
Mobile	61	0.28	-0.49	-0.16
Non-words	46	-0.10	-0.87*	-0.54
Segments	28	-0.58	-1.35*	-1.02*
Reading	34	-0.41	-1.18*	-0.85*
Spelling	37	-0.33	-1.10*	-0.77*
SW Reading				
Reasoning	78	0.77		
Reasoning (rtm)	67	0.44		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	1	-2.32	-1.51*	-1.85*
Digit span	26	-0.64	+0.17	-0.17
PhAB non-words	22	-0.77	+0.04	-0.30
PhAB spoonerisms	20	-0.84	-0.03	-0.37
NFER SCT	3	-1.88	-1.07*	-1.41*
BSTS 3	16	-0.99	-0.18	-0.52
BAS reading				
MAT	21	-0.81		
MAT (rtm)	32	-0.47		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	14*			
Spoonerisms	20			
Non-words	22			
Picture naming	28			
Digit naming	28			
Fluency (alliteration)	68			
Fluency (rhyme)	16			
[Fluency (semantic)]	8			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 2				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	3	-1.88	-2.03*	-2.01*
Mobile	15	-1.04	-1.19*	-1.17*
Non-words	3	-1.88	-2.03*	-2.01*
Segments	25	-0.67	-0.82*	-0.80*
Reading	32	-0.47	-0.62	-0.60
Spelling	21	-0.81	-0.96*	-0.94*
SW Reading				
Reasoning	56	0.15		
Reasoning (rtm)	55	0.13		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	26	-0.64	-1.08*	-0.84*
Digit span	5	-1.64	-2.08*	-1.84*
PhAB non-words	1	-0.81	-1.25*	-1.01*
PhAB spoonerisms	34	-0.41	-0.85*	-0.61
NFER SCT	1	-2.32	-2.76*	-2.52*
BSTS 3	1	-2.32	-2.76*	-2.52*
BAS reading				
MAT	67	0.44		
MAT (rtm)	58	0.20		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	9*			
Spoonerisms	34			
Non-words	21			
Picture naming	2*			
Digit naming	2*			
Fluency (alliteration)	8*			
Fluency (rhyme)	10*			
[Fluency (semantic)]	4			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 3				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	74	0.64	-0.70*	-0.17
Mobile	34	-0.41	-1.75*	-1.22*
Non-words	53	0.08	-1.26*	-0.73*
Segments	39	-0.28	-1.62*	-1.09*
Reading	47	-0.08	-1.42*	-0.73*
Spelling	70	0.52	-0.82*	-0.29
SW Reading				
Reasoning	91	1.34		
Reasoning (rtm)	79	0.81		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	63	0.33	+0.48	+0.38
Digit span	25	-0.67	-0.52	-0.62
PhAB non-words	24	-0.71	-0.56	-0.66
PhAB spoonerisms	20	-0.84	-0.69*	-0.79*
NFER SCT	23	-0.74	-0.59	-0.69*
BSTS 3	28	-0.58	-0.43	-0.53
BAS reading				
MAT	44	-0.15		
MAT (rtm)	48	-0.05		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	9*			
Rhyme	58			
Spoonerisms	20			
Non-words	24			
Picture naming	90			
Digit naming	63			
Fluency (alliteration)	24			
Fluency (rhyme)	8*			
[Fluency (semantic)]	26			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 4				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	51	0.03	-1.45*	-0.85*
Mobile	26	-0.64	-2.12*	-1.52*
Non-words	20	-0.84	-2.32*	-1.72*
Segments	20	-0.84	-2.32*	-1.72*
Reading	34	-0.41	-1.89*	-1.29*
Spelling	18	-0.92	-2.40*	-1.80*
SW Reading				
Reasoning	93	1.48		
Reasoning (rtm)	81	0.88		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	63	0.33	+1.81	+1.21
Digit span	9	-1.34	+0.14	-0.46
PhAB non-words	11	-1.23	+0.25	-0.35
PhAB spoonerisms	1	-2.32	-0.84*	-1.44*
NFER SCT	1	-2.32	-0.84*	-1.44*
BSTS 3	1	-2.32	-0.84*	-1.44*
BAS reading				
MAT	7	-1.48		
MAT (rtm)	19	-0.88		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	8*			
Spoonerisms	1*			
Non-words	11*			
Picture naming	72			
Digit naming	12*			
Fluency (alliteration)	80			
Fluency (rhyme)	52			
[Fluency (semantic)]	95			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 5				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	83	0.95	+0.54	+0.75
Mobile	28	-0.58	-0.99*	-0.78*
Non-words	20	-0.84	-1.25*	-1.04*
Segments	18	-0.92	-1.33*	-1.12*
Reading	58	0.20	-0.21	-0.01
Spelling	19	-0.88	-1.29*	-1.08*
SW Reading				
Reasoning	66	0.41		
Reasoning (rtm)	58	0.20		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	25	-0.67	-0.54	-0.67*
Digit span	1	-2.32	-2.19*	-2.32*
PhAB non-words	30	-0.52	-0.39	-0.52
PhAB spoonerisms	40	-0.25	-0.12	-0.25
NFER SCT	39	-0.28	-0.15	-0.28
BSTS 3	4	-1.75	-1.62*	-1.75*
BAS reading				
MAT	45	-0.13		
MAT (rtm)	50	0		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	9*			
Rhyme	1*			
Spoonerisms	40			
Non-words	30			
Picture naming	8*			
Digit naming	20			
Fluency (alliteration)	32			
Fluency (rhyme)	3*			
[Fluency (semantic)]	28			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 6				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	38	-0.31	-0.98*	-0.72*
Mobile	26	-0.64	-1.31*	-1.05*
Non-words	74	0.64	-0.03	+0.23
Segments	55	0.13	-0.54	-0.28
Reading	59	0.23	-0.44	-0.18
Spelling	44	-0.14	-0.81*	-0.55
SW Reading				
Reasoning	75	0.67		
Reasoning (rtm)	66	0.41		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Not dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	37	-0.33	-0.43	-0.43
Digit span	37	-0.33	-0.43	-0.43
PhAB non-words	28	-0.58	-0.68*	-0.68*
PhAB spoonerisms	3	-1.88	-1.98*	-1.98*
NFER SCT	14	-1.08	-1.18*	-1.18*
BSTS 3	13	-1.13	-1.23*	-1.23*
BAS reading				
MAT	54	0.10		
MAT (rtm)	54	0.10		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	1*			
Spoonerisms	3*			
Non-words	28			
Picture naming	22			
Digit naming	7*			
Fluency (alliteration)	16			
Fluency (rhyme)	11*			
[Fluency (semantic)]	1			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 7				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	14	-1.08	-1.72*	-1.49*
Mobile	67	0.44	-0.20	+0.03
Non-words	20	-0.84	-1.48*	-1.25*
Segments	71	0.55	-0.09	+0.14
Reading	27	-0.61	-1.25*	-1.02*
Spelling	18	-0.92	-1.56*	-1.33*
SW Reading				
Reasoning	74	0.64		
Reasoning (rtm)	66	0.41		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	25	-0.67	+0.97	+0.25
Digit span	25	-0.67	+0.97	+0.25
PhAB non-words	40	-0.25	+1.39	+0.67
PhAB spoonerisms	24	-0.71	+0.93	+0.21
NFER SCT	5	-1.64	0	-0.72*
BSTS 3	1	-2.32	-0.68*	-1.40*
BAS reading				
MAT	5	-1.64		
MAT (rtm)	18	-0.92		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Not dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	7*			
Spoonerisms	24			
Non-words	40			
Picture naming	3*			
Digit naming	5*			
Fluency (alliteration)	48			
Fluency (rhyme)	50			
[Fluency (semantic)]	16			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 8				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	84	0.99	+1.98	+1.57
Mobile	8	-1.40	-0.41	-0.82*
Non-words	13	-1.13	-0.14	-0.55
Segments	18	-0.92	+0.07	-0.34
Reading	15	-1.04	-0.05	-0.46
Spelling	17	-0.95	+0.04	-0.37
SW Reading	2	-2.05	-1.06*	-1.47*
Reasoning	16	-0.99		
Reasoning (rtm)	28	-0.58		
Decision using LASS: Not dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	2	-2.05	+0.27	-0.71*
Digit span	16	-0.99	+1.33	+0.35
PhAB non-words	1	-2.32	0	-0.98*
PhAB spoonerisms	3	-1.88	+0.44	-0.54
NFER SCT	1	-2.32	0	-0.98*
BSTS 3	1	-2.32	0	-0.98*
BAS reading	81	0.88	+3.20	+2.22
MAT	1	-2.32		
MAT (rtm)	9	-1.34		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	9*			
Rhyme	1*			
Spoonerisms	3*			
Non-words	1*			
Picture naming	1*			
Digit naming	1*			
Fluency (alliteration)	16			
Fluency (rhyme)	4*			
[Fluency (semantic)]	8			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 9				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	37	-0.33	-0.23	-0.23
Mobile	22	-0.77	-0.67*	-0.67*
Non-words	50	0	+0.10	+0.10
Segments	50	0	+0.10	+0.10
Reading	17	-0.95	-0.85*	-0.85*
Spelling	30	-0.52	-0.42	-0.42
SW Reading				
Reasoning	46	-0.10		
Reasoning (rtm)	46	-0.10		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	16	-0.99	-0.71*	-0.86*
Digit span	16	-0.99	-0.71*	-0.86*
PhAB non-words	20	-0.84	-0.56	-0.71*
PhAB spoonerisms	60	0.25	+0.03	+0.38
NFER SCT	16	-0.99	-0.71*	-0.86*
BSTS 3	12	-1.17	-0.89*	-1.04*
BAS reading				
MAT	39	-0.28		
MAT (rtm)	45	-0.13		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	24			
Spoonerisms	60			
Non-words	20			
Picture naming	40			
Digit naming	32			
Fluency (alliteration)	63			
Fluency (rhyme)	45			
[Fluency (semantic)]	68			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 10				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	63	0.33	0	+0.13
Mobile	34	-0.41	-0.74*	-0.61
Non-words	45	-0.13	-0.46	-0.33
Segments	43	-0.18	-0.51	-0.38
Reading	33	-0.44	-0.77*	-0.64
Spelling	16	-0.99	-1.32*	-1.19*
SW Reading				
Reasoning	63	0.33		
Reasoning (rtm)	58	0.20		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Not dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	63	0.33	-0.06	+0.13
Digit span	9	-1.34	-1.73*	-1.54*
PhAB non-words	26	-0.64	-1.03*	-0.84*
PhAB spoonerisms	22	-0.77	-1.16*	-0.97*
NFER SCT	4	-1.75	-2.14*	-1.95*
BSTS 3	1	-2.32	-2.71*	-2.52*
BAS reading				
MAT	65	0.39		
MAT (rtm)	58	0.20		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	56			
Rhyme	24			
Spoonerisms	22			
Non-words	26			
Picture naming	55			
Digit naming	40			
Fluency (alliteration)	37			
Fluency (rhyme)	91			
[Fluency (semantic)]	52			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 11				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	94	1.56	+1.38	+1.43
Mobile	20	-0.84	-1.02*	-0.97*
Non-words	5	-1.64	-1.82*	-1.77*
Segments	30	-0.52	-0.70*	-0.65
Reading	18	-0.92	-1.10*	-1.05*
Spelling	25	-0.67	-0.85*	-0.80*
SW Reading	4	-1.75	-1.93*	
Reasoning	57	0.18		
Reasoning (rtm)	55	0.13		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	25	-0.67	-2.72*	-1.84*
Digit span	25	-0.67	-2.72*	-1.84*
PhAB non-words	22	-0.77	-2.82*	-1.94*
PhAB spoonerisms	22	-0.77	-2.82*	-1.94*
NFER SCT	1	-2.32	-4.37*	-3.49*
BSTS 3	1	-2.32	-4.37*	-3.49*
BAS reading	1	-2.32	-4.37*	-3.49*
MAT	98	2.05		
MAT (rtm)	88	1.17		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	9*			
Rhyme	2*			
Spoonerisms	22			
Non-words	22			
Picture naming	26			
Digit naming	2*			
Fluency (alliteration)	60			
Fluency (rhyme)	3*			
[Fluency (semantic)]	48			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 12				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	73	0.61	+1.02	+0.81
Mobile	8	-1.40	-0.99*	-1.20*
Non-words	49	-0.03	+0.38	+0.17
Segments	55	0.13	+0.54	+0.33
Reading	29	-0.55	+0.14	-0.35
Spelling	15	-1.04	-0.63	-0.84*
SW Reading				
Reasoning	34	-0.41		
Reasoning (rtm)	42	-0.20		
Decision using LASS: Not dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	5	-1.64	-0.51	-0.97*
Digit span	5	-1.64	-0.51	-0.97*
PhAB non-words	3	-1.88	-0.75*	-1.21*
PhAB spoonerisms	12	-1.17	-0.04	-0.50
NFER SCT	5	-1.64	-0.51	-0.97*
BSTS 3	3	-1.88	-0.75*	-1.21*
BAS reading				
MAT	13	-1.13		
MAT (rtm)	25	-0.67		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	14*			
Spoonerisms	12*			
Non-words	3*			
Picture naming	52			
Digit naming	6*			
Fluency (alliteration)	58			
Fluency (rhyme)	52			
[Fluency (semantic)]	1			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 13				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	92	1.40	+1.12	+1.27
Mobile	44	-0.15	-0.43	-0.28
Non-words	46	-0.10	-0.38	-0.23
Segments	76	0.71	+0.43	+0.58
Reading	67	0.44	+0.16	+0.31
Spelling	55	0.13	-0.15	0
SW Reading				
Reasoning	61	0.28		
Reasoning (rtm)	55	0.13		
Decision using LASS: Not dyslexic				
Decision using LASS (allowing for regression to the mean): Not dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	37	-0.33	+1.72	+0.84
Digit span	1	-2.32	-0.27	-1.15*
PhAB non-words	5	-1.64	+0.41	-0.47
PhAB spoonerisms	3	-1.88	+0.17	-0.71*
NFER SCT	23	-0.74	+1.31	+0.43
BSTS 3	18	-0.92	+1.13	+0.25
BAS reading				
MAT	2	-2.05		
MAT (rtm)	12	-1.17		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Not dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	2*			
Spoonerisms	3*			
Non-words	5*			
Picture naming	20			
Digit naming	3*			
Fluency (alliteration)	6*			
Fluency (rhyme)	22			
[Fluency (semantic)]	8			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 14				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	6	-1.56	-0.79*	-1.15*
Mobile	1	-2.32	-1.55*	-1.91*
Non-words	63	0.33	+1.10	+0.74
Segments	27	-0.61	+0.16	-0.20
Reading	38	-0.31	+0.46	+0.10
Spelling	21	-0.81	-0.04	-0.40
SW Reading				
Reasoning	22	-0.77		
Reasoning (rtm)	34	-0.41		
Decision using LASS: Not dyslexic				
Decision using LASS (allowing for regression to the mean): Not dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	1	-2.32	0	-0.98*
Digit span	2	-2.05	+0.27	-0.71*
PhAB non-words	12	-1.17	+1.15	+0.17
PhAB spoonerisms	16	-0.99	+1.33	+0.35
NFER SCT	1	-2.32	0	-0.98*
BSTS 3	1	-2.32	0	-0.98*
BAS reading				
MAT	1	-2.32		
MAT (rtm)	9	-1.34		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	1*			
Spoonerisms	16			
Non-words	12*			
Picture naming	48			
Digit naming	11*			
Fluency (alliteration)	6*			
Fluency (rhyme)	7*			
[Fluency (semantic)]	12			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 15				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	30	-0.52	-1.07*	-0.85*
Mobile	69	0.50	-0.05	+0.17
Non-words	48	-0.05	-0.60	-0.38
Segments	41	-0.23	-0.78*	-0.56
Reading	4	-1.75	-2.30*	-2.08*
Spelling	26	-0.64	-1.19*	-0.97*
SW Reading	5	-1.64	-2.19*	-1.97*
Reasoning	71	0.55		
Reasoning (rtm)	63	0.33		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	50	0	-0.58	-0.33
Digit span	37	-0.33	-0.91*	0
PhAB non-words	32	-0.47	-1.05*	-0.80*
PhAB spoonerisms	18	-0.92	-1.50*	-1.25*
NFER SCT	1	-2.32	-2.90*	-2.65*
BSTS 3	4	-1.75	-2.33*	-2.08*
BAS reading	30	-0.52	-1.10*	-1.43*
MAT	72	0.58		
MAT (rtm)	63	0.33		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	13*			
Spoonerisms	18			
Non-words	32			
Picture naming	24			
Digit naming	28			
Fluency (alliteration)	14*			
Fluency (rhyme)	22			
[Fluency (semantic)]	16			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 16				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	8	-1.40	-0.69*	-0.99*
Mobile	15	-1.04	-0.33	-0.63
Non-words	3	-1.88	-1.17*	-1.47*
Segments	32	-0.47	+0.24	-0.06
Reading	23	-0.74	-0.03	-0.33
Spelling	23	-0.74	-0.03	-0.33
SW Reading				
Reasoning	24	-0.71		
Reasoning (rtm)	34	-0.41		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	1	-2.32	-0.44	-1.24*
Digit span	1	-2.32	-0.44	-1.24*
PhAB non-words	26	-0.64	+1.24	+0.44
PhAB spoonerisms	28	-0.58	+1.30	+0.50
NFER SCT	4	-1.75	+0.13	-0.67*
BSTS 3	1	-2.32	-0.44	-1.24*
BAS reading				
MAT	3	-1.88		
MAT (rtm)	14	-1.08		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	9*			
Rhyme	13*			
Spoonerisms	28			
Non-words	26			
Picture naming	5*			
Digit naming	20			
Fluency (alliteration)	2*			
Fluency (rhyme)	14*			
[Fluency (semantic)]	1			
Decision using PhAB: Dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 17				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	23	-0.74	-1.51*	-1.15*
Mobile	34	-0.41	-1.18*	-0.82*
Non-words	27	-0.61	-1.38*	-1.02*
Segments	43	-0.18	-0.95*	-0.59
Reading	51	0.03	-0.74*	-0.38
Spelling	32	-0.47	-1.24*	-0.88*
SW Reading				
Reasoning	78	0.77		
Reasoning (rtm)	66	0.41		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	37	-0.33	-0.08	-0.20
Digit span	16	-0.99	-0.74*	-0.86*
PhAB non-words	25	-0.67	-0.42	-0.54
PhAB spoonerisms	37	-0.33	-0.08	-0.20
NFER SCT	68	0.47	+0.72	+0.60
BSTS 3	13	-1.13	-0.88*	-1.00*
BAS reading				
MAT	40	-0.25		
MAT (rtm)	45	-0.13		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	37			
Spoonerisms	37			
Non-words	25			
Picture naming	73			
Digit naming	58			
Fluency (alliteration)	50			
Fluency (rhyme)	66			
[Fluency (semantic)]	53			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 18				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	51	0.01	-0.70*	-0.40
Mobile	90	1.28	+0.57	+0.87
Non-words	83	0.95	+0.24	+0.54
Segments	55	0.13	-0.58	-0.28
Reading	69	0.50	-0.21	+0.09
Spelling	46	-0.10	-0.81*	-0.51
SW Reading				
Reasoning	76	0.71		
Reasoning (rtm)	66	0.41		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Not dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	75	0.67	+0.98	+0.80
Digit span	63	0.33	+0.64	+0.46
PhAB non-words	92	1.40	+1.71	+1.53
PhAB spoonerisms	47	-0.08	+0.23	+0.05
NFER SCT	47	-0.08	+0.23	+0.05
BSTS 3	12	-1.17	-0.86*	-1.04*
BAS reading				
MAT	38	-0.31		
MAT (rtm)	45	-0.13		
Decision using validation tests: Not dyslexic				
Decision using validation tests (allowing for regression to the mean): Not dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	37			
Spoonerisms	47			
Non-words	92			
Picture naming	92			
Digit naming	63			
Fluency (alliteration)	84			
Fluency (rhyme)	34			
[Fluency (semantic)]	81			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB

ID: 19				
LASS Test	*Centile*	*z score*	*z (reasoning) - z (test)*	*z -z (rtm)*
Cave	55	0.13	-0.39	-0.20
Mobile	28	-0.58	-1.10*	-0.91*
Non-words	52	0.05	-0.47	-0.28
Segments	81	0.88	+0.36	+0.55
Reading	64	0.36	-0.16	+0.03
Spelling	33	-0.44	-0.96*	-0.77*
SW Reading				
Reasoning	70	0.52		
Reasoning (rtm)	63	0.33		
Decision using LASS: Dyslexic				
Decision using LASS (allowing for regression to the mean): Dyslexic				
Validation Test	*Centile*	*z score*	*z (MAT) - z (test)*	*z -z (rtm)*
Spatial span	84	0.99	+0.84	+0.91
Digit span	16	-0.99	-1.14*	-1.07*
PhAB non-words	32	-0.47	-0.62	-0.55
PhAB spoonerisms	63	0.33	+0.18	+0.25
NFER SCT	32	-0.47	-0.62	-0.55
BSTS 3	16	-0.99	-1.14*	-1.07*
BAS reading				
MAT	56	0.15		
MAT (rtm)	53	0.08		
Decision using validation tests: Dyslexic				
Decision using validation tests (allowing for regression to the mean): Dyslexic				
PhAB Test	Centile			
Alliteration	50			
Rhyme	90			
Spoonerisms	63			
Non-words	32			
Picture naming	78			
Digit naming	16			
Fluency (alliteration)	63			
Fluency (rhyme)	66			
[Fluency (semantic)]	42			
Decision using PhAB: Not dyslexic				

rtm: allowing for regression to the mean

*significant discrepancy or low score on PhAB