6 | The Politics of Reading Assessment

If we are concerned that all children learn to read to the best of their abilities, like any high-risk enterprise, trial-and-error methods of operation are undesirable. In the field of early reading instruction, such methods are also unnecessary. Scientific research has demonstrated precisely what beginning readers need to know and which methods produce the best results. If we are to ensure that reading methods are as closely aligned as possible with this research knowledge, the careful measurement and systematic evaluation of children's reading performance is of key importance. Information gained from such measurement permits the fine-tuning of reading methods, so that maximum success in learning to read is guaranteed for all children.

However, what we find in practice is that reading is assessed in ways which are unreliable and unnecessarily expensive; and that the monitoring of teachers' performance in enabling children to read is entirely subjective and, again, exceedingly expensive. Moreover, it is clear that the the nature of teacher training and the 'research' on which this training is based is suspect. Much of the inadequacy can be traced to the politicisation of reading assessment, the subject of this chapter.

Assessing Children's Reading

How should pupils' reading achievement be measured? It is important that test measures yield the sort of information that will help teachers perfect their teaching methods in order to produce the highest reading standards possible. With this purpose in mind, what kind of test measures are available and what are their specific advantages and disadvantages?

Essentially there are two criteria involved with testing children's early reading achievement: first, the test should actually measure what it sets out to, that is, how well the child is able to read (*validity*); and second, the test should give

consistent results, from one use of the test to another and from one child to another of the same reading ability (*reliability*). A reading test, for example, that requires a child to read a story that has been memorised will not be testing how well the child can read, but how well the child has memorised, and hence will lack validity as a test of reading. Similarly, lack of reliability can arise if a test uses a very coarse scale of measurement, dividing readers into, for example, three broad levels, which may result in children of similar reading ability being awarded different level scores, or those very different in ability being awarded identical scores.

For the purposes of our discussion here, we need to contrast the procedures of the National Curriculum Standard Attainment Tests (SATs) with those of norm-referenced or standardised tests of reading. The SATs use criterion-referencing, where what is measured is how well a child meets certain criteria: for example, the child is able to 'read and understand three typical classroom signs or captions' (English Key Stage 1, Level 1), or the child is able to 'retell the content of a passage read including at least two of the main points and makes a sensible prediction about what might happen next' (English Key Stage 1, Level 2, National Curriculum reading tests) (SCAA, 1994e). Children are assessed individually to determine whether or not they meet the criteria demanded of the level description. The teacher then makes a subjective judgement as to whether or not the criteria have been met.

Norm-referenced or standardised reading tests, on the other hand, furnish tables of norms, based on prior sampling of a population; such tables permit teachers to convert a child's score into a reading quotient which takes chronological age into account; this quotient gives some indication of the child's standing relative to that of the population to which he belongs. Standard procedures for conducting such tests are clearly stipulated. Timing of the test, the test materials used, and objective marking procedures are the same for all pupils. These tests usually yield scores on a relatively fine scale of measurement, they require limited administration time since they are usually administered to the class as a whole; and in scoring, each item is either right or wrong. The *Young's Group Reading Test* is an example of this type; it measures a child's ability to read and understand single words, and short sentences of increasing difficulty.

Value of Different Kinds of Tests

Each type of test measure has advantages and disadvantages. Criterion-referenced tests such as the SATs can be designed to match curriculum targets closely, with the result that they may help teachers to define specific teaching goals. They also permit every student to achieve at least some of the specified targets. On the other hand, they may be very time-consuming and expensive to administer and mark. They do not furnish any reliable information about relative standing, and criteria must be worded so specifically that there is no doubt about what constitutes attainment of each criterion. Judgements about whether or not criteria have been met, or whether the test actually measures reading ability, for example, are totally subjective in nature, and therefore, validity and reliability are reduced.

Standardised reading tests can save time both in administration and marking, involve, therefore, less cost, and yield useful information about relative standing between pupils, schools, school districts, and about comparative standing over time. Standardised tests are regarded by statisticians as 'the products of a high degree of professional competence and skill in test-writing and, as such, are usually quite reliable and generally valid' (Kerlinger, 1986, p. 451). These tests are subjected to rigorous item analysis in order to ensure that the test actually measures what it sets out to. For this reason, standardised tests are more likely to provide a valid measure of a child's reading ability; and because marking of the test is objective, the reliability of the scores obtained is high. Standardised scores, represented as they are on a relatively fine scale, may be useful in allowing teachers to evaluate different teaching methods and in encouraging teachers to produce even the smallest improvements.

The one primary advantage standardised tests have over the criterion-referenced SATs is the more reliable comparisons of results that can be made over time, making them more suitable as tools for helping to improve reading standards. Since to a large extent SATs outcomes are based on subjective judgement, the danger is that, with the use of such tests, judgements could gradually become more liberal, and reading standards could decline, perhaps dramatically, without being detected. On the other hand, standardised tests can be used to make more reliable comparisons of results over months or

years and they thus permit more careful monitoring of trends in reading attainment. These tests can also be used to draw tentative conclusions: conclusions, for example, about the relative effectiveness of different teaching methods between schools, in different years, or under different circumstances. In short, the knowledge gained, under the more controlled conditions of standardised tests, can be much more valuable in helping to shape effective reading practice over the long term.

Ineffective Tuning Methods?

What is happening in practice with regard to the assessment of readers? Turner's investigations into reading standards in 1990 helped to raise concern in England that reading standards among 7-year-olds were too low and in serious need of improvement (Turner, 1990). Since that time, the government has introduced a national curriculum along with compulsory tests for 7-, 11- and 14-year-olds. Will the National Curriculum tests in reading for 7-year-olds help to raise expectations as to what is possible and help to improve reading standards among this age group? How reliable and valid is this form of assessment? And what other forms of assessment of reading are being used in classrooms?

Some assessment of children's reading abilities is contained in teachers' own reports. These, however, are extremely subjective and descriptive in nature. In a recent Ofsted report (1995), school inspectors complained that the type of reporting contained in children's Records of Achievement (required when children transfer from or leave a school) is often unclear, jargon-filled and overly positive; inspectors were concerned that more than half the reports examined failed to diagnose problems and that the nature of these reports gave the false impression that standards were higher than they actually were. Although some standardised reading assessment does take place in schools, a 1991 report on reading standards revealed that only 59 out of 116 LEAs carried out standardised testing of primary school reading at some point in the previous ten years (Cato & Whetton, 1991).

In general, whether standardised testing is conducted or not, school inspectors report that very few schools have 'effective' systems for monitoring the standards of work achieved by pupils (Ofsted, 1994). In this one report, they

concluded that in only 4% of schools did headteachers make an attempt to monitor the standards of work achieved by pupils, and among teachers inspected, only 50% made provision for systematic monitoring and evaluation of pupils' performance.

Shortcomings of the National Curriculum Tests

As the only reading results available for 7-year-olds on a national scale at the present time are those from the National Curriculum Key Stage 1 tests in Reading, these clearly deserve serious attention here. While some form of testing may possibly be better than no testing at all, there are in fact serious shortcomings with these particular tests. The remarkable degree to which progressive attitudes towards reading extend is seen in the government's choice of a measuring device to assess reading standards among 7-year-olds. For we see that the National Curriculum assessments use a whole-word, meaning-emphasis procedure to measure reading ability: this is none other than the 'miscue' inventory, invented by the popular whole language advocate, Kenneth Goodman.

The fact that the (now) Department of Education & Employment has decided to test in this manner is significant. The method used to measure how well children read is symptomatic of how they are being taught, or not being taught, to read: teachers attest that this manner of assessment is suited to the type of instruction being given and to the philosophies adhered to. When teachers were asked if the National Curriculum forms of assessment provided grades which were a valid reflection of children's reading abilities, 74% of teachers agreed that they did (NUT, 1992).

Why has the Department of Education chosen to measure reading ability with techniques approved by progressive, whole-word advocates rather than with a conventional standardised reading test? In explaining how this happened, Seaton (1994a) has pointed out how 'the Government unwittingly appointed educationists to the National Curriculum Council' (to design the curriculum), 'and the School Examinations and Assessment Council' (to devise the tests) 'who were mainly progressive in outlook'. Many of those chosen for these tasks are members of NATE (National Association for the Teaching of English) and/or LATE (London

Association for the Teaching of English), organisations that are vociferous in their opposition to traditional, research-based forms of teaching and testing. There are four main shortcomings of the National Curriculum SATs tests for seven-year olds:

(a) Assessment is Entirely Subjective in Nature

There is opportunity for a wide margin of error since the reading material used, the timing of test administration, the child's age when tested, and the teacher evaluation of performance can all vary widely; none of these aspects has been standardised. Teachers can choose from a list of ordinary story books (which vary widely in difficulty) whatever text they want a particular child to read; teachers may then 'familiarise' the child with the chosen text to whatever degree they like; they can conduct tests at any time they choose over several weeks; they may vary with each child tested the amount of time and assistance allowed during testing; they do not have to consider the age of the child when scoring; and they are required to make subjective judgements about each child's reading fluency, accuracy, and understanding.

(b) Reliability of the Tests is Low

The non-standardised, imprecise nature of these tests make it impossible to determine in a reliable way what 'national standards' are; reliable comparison between children, between schools, or between different school years cannot be made.

The tests invite variation. Subjective evaluation means that teachers may vary from year to year, from to school to school, or from child to child in the manner in which they award a score. The reading passages designated for use in these tests (derived from a long list of story books) vary widely on a number of factors, so that a child's performance may depend entirely on which story passage is selected. In 1994, for example, there were fourteen story books with designated reading passages from which teachers could choose. Since each passage is different, the skills being tested by each passage are likely to vary. One child may score highly because a particular word that the child has memorised well is liberally repeated throughout the selected test passage. Another child may score highly because the test passage, selected subjectively by the teacher, is familiar to the child or

almost completely committed to memory. Yet another child may achieve an inflated score because the passage selected is shorter than others, allowing more time to discuss the content with the teacher and guess from the pictures.

Such children may not have the ability to decode unfamiliar words, and may not understand that reading is more than recognising memorised words on the page or guessing from the pictures. This inability is acceptable in the eyes of whole-word advocates: 'children should know what they are going to read before they read it' (Waterland 1985, p. 14). The vast opportunities for teacher variation in judging what good reading ability actually is mean that one can depend very little on the accuracy of the results obtained.

Furthermore, there has been year-to-year variation in test content and procedures. In 1996, for example, the list of twelve story books to be used for Key Stage 1 tests differed to a considerable degree from the nineteen books listed in 1995, and from the fourteen books listed in 1994. In 1995, teachers were to determine an accuracy level based on a child's number of oral reading errors in a particular book passage, whereas in 1996, accuracy scores were no longer included in the assessment procedures. In 1994, teachers were to select the book for testing, while in 1996, children were to select from among three or four books chosen by the teacher.

In every year, teachers were to determine whether a child had reached Level 1, 2 or 3 in his or her reading ability, Level 2 being the standard expected of a 7-year-old. There is, however, a huge leap in performance represented between Levels 1 and 2. Those assessed as Level 1 in ability need only recognise one word and a few letters, while those who are categorised as Level 2 must be able to read a simple piece of text with 'reasonable accuracy' (SCAA, 1994d).

More specifically, requirements for various levels are as follows:

Level 1 shows interest in a book, verbally or non-verbally, talks about the content of a book, shows understanding that print carries meaning, and recognises at least one word and at least three letters.

Level 2 reads aloud from one of a number of designated books with no more than eight words told (words that make sense are permitted).

In some cases, depending on the particular book, a child need have only six to eight words marked correct in order to attain Level 2 (based on 1994 material). Some of the words permitted as correct may not be the actual words on the page, but words permitted because meaning is preserved. Words omitted that do not disturb the sense of the passage are not counted as errors either (SCAA, 1994e, p. 8). When accuracy rate was part of the test procedures, the accuracy rate for different story passages required to pass Level 2 could range from six to twenty words correct. But this method of compensation for variance in passage difficulty does not account for factors such as the number of irregularly spelled words, the number of multisyllabic words, the degree of advantage provided by picture cues, or the amount of repetition in words or phrases that may occur in a passage. An in-depth analysis of these passages shows that they are not comparable in difficulty, and an examination of the present system for dealing with the differences in passage difficulty is shown to be inadequate (Morris, 1993).

Thus, among the 76% of children who attained Level 1 on the national curriculum tests in 1992, for example, it is likely that a very wide range of reading abilities was represented. Addressing this issue, an investigation comparing pupils' standardised reading scores with their level of attainment on national curriculum tests is starkly illustrative (Pumfrey et al., 1991). Children who had been assessed as having attained Level 2 (the average expected for their age) on national curriculum tests, were found to have reading ages, determined from standardised testing, ranging from 5·7 to 12·9 years. That is, within the group of pupils all categorised as Level 2, there was an *incredible 7-year range* in the actual reading abilities represented. Similarly, those categorised as Level 1, were found to have reading ages ranging from 5·0 to 9·6 years.

The Secretary for State for Education and Employment, Gillian Shephard, has stated that 'educational reforms' (such as the national curriculum testing and performance tables) 'are helping our children to achieve higher standards' (Shephard, 1995, p. 1). How accurate is this assertion? Are the reading standards among 7-year-olds improving, are teachers becoming more adept at administering the tests, or are the subjective judgements of teachers in deciding whether a child qualifies for Level 1 or 2 perhaps becoming more relaxed?

A recent study sheds a great deal of light on these questions and shows the danger in accepting the results from the National Curriculum assessments without a considerable degree of caution. A comparison of the reading performance of children tested on the National Curriculum assessment tasks in the years 1991 and 1992 with the reading levels attained by the same children on standardised reading tests revealed a discrepancy. Confirming the suspicion that the results from the curriculum tests are inflated, children's performance on the standardised reading test was much poorer than their National Curriculum assessment indicated (Davies, Brember, & Pumfrey, 1995).

Based on the proportions of children found to have a reading age of less than 6 in the samples that were randomly chosen for this study, and comparing these with the children's attainment on the National Curriculum assessments, the evidence would suggest that national reading standards are much worse than they appear. Although National Curriculum assessments indicate 28% (1991) and 24% (1992) of children failing to reach Level 2, evidence extrapolated from this study suggests that the proportion of children failing to learn how to read after two years in school is likely to be closer to 31% in 1991, and 34% in 1992. Most alarming, based on standardised reading comprehension scores obtained in this study, an average drop in reading standard of 2·6 months of reading age occurred over this one-year period. This represents a current annual rate of decline more than three times greater than that reported by past NFER investigations.

Thus, instead of finding that reading standards were improving between the years 1991 and 1992, as the results from the National Curriculum assessments would suggest, these researchers found the opposite trend. Analysis of the raw test scores from a standardised reading test revealed that the means (representing an equivalent standard to that of Level 2 on the national curriculum tests) were lower in 1992 than in 1991, a difference that was highly significant (at the 1% level).

As Professor Peter Pumfrey observes, 'perhaps no one in government wants or needs to know what reading standards are, and whether they are or are not changing? The public *illusion* that this information is important to government could be *all* that is politically required' (Pumfrey 1991, p. 57). If this is so, it is suggested that the national tests are the ideal

instruments because, as has been noted by a number of researchers, they do not provide valid or reliable information about national standards (Davies, Brember, & Pumfrey, 1995; Morris, 1993; Pumfrey, 1991; Pumfrey, Elliot, & Tyler, 1991).

(c) Validity of the Tests is Suspect

According to the *Parents Guide to Testing* (DES, 1991b, p. 2.), 'the point of the new tests is to give you and the teachers an exact picture of what your child has learned' so that it will be known 'how your child measures up against *national* standards.' While these may be the stated aims, the truth is that the National Curriculum tests fail to achieve either of these goals. Not only are the test results obtained unreliable for the purpose of making comparisons, they are also of limited use, either for assessing reading performance or as a means to raising reading standards.

The tests for reading are not capable of giving an 'exact picture' of what reading skills have been learned; they may furnish a 'picture' of how the child is *pretending to be a reader*, but this is not reading and the picture furnished is not 'exact'. Because the measuring scale for determining a particular reading level covers such a large range of performance, only a very vague picture of what is regarded as 'reading ability' is obtained.

How 'reading ability' is defined by these tests is not in line with current research-based models of reading; the most important skill which must be acquired during early reading development is not even measured – the ability to decode print alphabetically. Research consistently underlines the fact that fluent reading requires accurate and fluent decoding. Yet instead of using a standardised reading test which measures context-free word-recognition skills, the national curriculum tests focus on assessing the child's ability to make sense of a passage by guessing, by using context or picture cues, or by recognising whole words by their shape. It is more an assessment of the range of a child's learned repertoire of whole-word guessing strategies than an assessment of his or her ability to read.

Indeed, these tests may not be measuring reading ability at all. They may, instead, be measuring the ability to memorise. There is evidence that parents may be increasingly active in helping their children memorise the story books used to test 7-year-olds. Booksellers cannot keep up with the demand.

Schools, having to compete with parents' demand for the books, are having difficulty obtaining the books they require for testing. One book club has been 'actively selling the books from the lists to parents, running a 'buy-two-and-get-one-free' promotion' (Williams, 1996, p. 15).

(d) Financial Costs are High

The national curriculum tests are expensive. Apart from the development and marking costs, the administration costs (in terms of teacher time) are also very high.

The estimated development costs for the years 1991, 1992, and 1993 were approximately £4 million per year. However, in 1991 the eventual costs were £6 million. The cost of hiring outside markers, necessary to secure the co-operation of teachers' unions, is another £30 million per year. Taking into consideration the amount of teacher time required (about forty hours), the National Curriculum assessments are estimated to cost £60 million; the cost per year, then, is roughly (4 + 30 + 60 =) £94 million. If these tests were replaced by standardised, whole-class, paper and pencil tests (taking about two hours in total for administration and marking), the estimated cost is £3 million (Turner, 1991a). Thus, the comparative costs are £94 million (for the present criterion-referenced testing) versus £3 million (for standardised reading testing).

Opposition to Standardised Testing

Teachers' unions have objected strenuously to national testing on both ideological and practical grounds, and consequently, only limited testing was carried out prior to 1994. And, although 90% of schools used the Key Stage 1 tests for testing the reading of 7-year-olds in 1994, only 52% of schools reported the results. Objections to the tests have ranged from the heavy demand made on teachers' time to administer and mark the tests, to the disrespect implied for teachers' own evaluations of student progress. These objections are puzzling, if not illogical; for the teachers' unions themselves demanded criterion-referenced testing, a form of testing which they felt would more closely match their concept of good practice. It should, therefore, come as no surprise that tests which have to be individually administered and which have no right or wrong answers are time-consuming to administer and mark; nor should it be surprising that the information gained is

superfluous since teachers' current mode of evaluating pupils' abilities tends to be identical in nature.

However, in spite of the shortcomings of subjective measurement, many teachers are even more opposed to objective forms of testing. A speaker at the 1991 conference of the United Kingdom Reading Association (UKRA) gave three reasons why standardised tests are 'inappropriate': 'Teachers who use a whole language approach to reading instruction in their classrooms have discovered that traditional standardized testing is not an appropriate assessment mode ... since it lacks validity, suitability, and availability' (Leland, p. 238). How justified are these assertions?

(a) Validity

Whether standardised tests are 'valid' or not is dependent on one's concept of what reading is. Traditional standardised reading tests include measures that assess both decoding and reading comprehension ability; they require both the reading of words in lists and the comprehension of short passages read silently. These tests reflect a definition of 'reading' which includes both the ability to decode print and to understand it. However, if 'reading' is now seen purely as the ability to make guesses at words through the use of context, syntax, or pictures, then indeed, such tests would not be valid. The fact that mainstream teachers view standardised reading tests as lacking in validity is a testament to just how widespread this new concept of reading has become. Reading is no longer seen as the dual process of deciphering print *and* comprehending it, as a process where print is decoded in order to achieve the main aim of understanding it. The Nelson-NFER Reading Ability tests are selected for use in NFER surveys because they reflect 'modern conceptions of reading' where the *only* reading-related ability that is seen worthy of measurement is how well children 'make sense of coherent and complete texts' (Gorman & Fernandes, 1992, pp. 3-4); whether children are able to decode print accurately or not is no longer seen as important.

Pertinent to this issue, in 1991, England and Wales accepted an invitation to take part in an international survey of the reading ability of 9-year olds. Ultimately, however, they withdrew from the study, which went ahead in twenty-seven countries, because government officials and researchers at the NFER were 'dissatisfied' with the tests that were to be used.

They protested that the questions measured mainly literal comprehension, were almost entirely objective in nature, and therefore represented an 'outmoded and inadequate model of the reading process' (Brooks, Pugh, Schagen, 1996, p.3).

(b) Suitability

Standardised reading tests are accused of being 'unsuitable' for teachers using a whole-language reading approach. Standardised reading tests measure a child's ability to decode and understand words, short sentences, or short passages. This is accomplished through a variety of tasks: children are required to read words in isolation without the aid of context, and/or to read and understand hitherto unseen short sentences or passages without the aid of accompanying pictures. Under a whole-language régime, the child is not taught how to perform either of these tasks. Thus, in this sense standardised tests do not 'suit' the teaching approach.

On the other hand, standardised tests could be viewed as especially suitable to the whole-language teacher as they may provide information about a pupil that might not otherwise be revealed. One of the most frequent teacher complaints about the National Curriculum tests is that they do not provide any further information about a pupil beyond the teacher's own assessment. Traditional standardised tests, therefore, are particularly 'suitable' in this context. As whole-word methods do not focus directly, or in a systematic fashion on the learning of letter-sound correspondences, for example, standardised reading tests could provide a useful warning of children's lack of progress in this area. Furthermore, standardised reading tests may be particularly 'suitable' in a whole-word classroom, in helping to draw a teacher's attention to some of the shortcomings of such teaching methods.

(c) Availability

As for the 'lack of availability' claim, this assertion is patently untrue. Inexpensive standardised tests are readily available, and have been used by most schools to some degree in the past. It is good quality, criterion-referenced tests that are in short supply. The National Curriculum tests are a case in point, where the results obtained are questionable and the costs in time and money are huge.

Politics and the Reading Debate

Validity, suitability and availability aside, perhaps there is another, overriding reason why such teachers are opposed to testing in any form? It has been frequently noted that the debate on reading methods has become highly politicised (McKenna, Stahl, & Reinking, 1994). Current orthodoxy, and in particular, the whole-language philosophy, is concerned with passing on certain values, engendering certain attitudes and teaching ways of behaving; what it is least concerned with is passing on specific subject knowledge. Reading has been redefined by whole-word advocates to such an extent that reading is no longer viewed in a way that makes testing of reading ability possible. Instead, it has become more 'suitable' to assess various random behaviours, none of which are supported by research as useful strategies for the beginning reader: what kind of guessing strategies children are able to use, whether meaningful word substitutions are made, and whether or not a child has memorised some words by shape.

Whole-word advocates see external tests as a threat to the search for an egalitarian society, a 'more just world' (Goodman, 1992a). External imposition of tests and standards are seen as 'disempowering' children who must be allowed to direct their own learning, to interact with texts individually in their own personal manner and at their own rate. Competition is seen as encouraging élitism; tests are seen as an interfering affront to teacher professionalism. And yet by encouraging reading methods which result in high rates of reading failure especially among boys, are not whole-word supporters causing individuals to be ranked in a way that actually *highlights* individual differences? In their anxiety not to make any child feel inferior, whole-word advocates actually guarantee that certain children will be consigned to a lower status by adhering to methods of instruction which are inadequate, methods which fail to teach *all* children to read. As Gough, a supporter of research-based instruction, states, 'I would "devalorize" no one; I would teach everyone to read' (Gough, 1995, p. 86).

'Progressive' Ideology

Child-centred, discovery, or so-called 'progressive' methods have been practised widely in the UK over the last twenty years, although the idea that children should be engaged in

gaining experiences while at school rather than knowledge or facts may have begun as early as 1931 with the Hadow Report. Such methods have been particularly popular during the 1980s, a period when reading standards among 7-year-olds declined (Cato & Whetton, 1991; Turner, 1990). From 1987 to 1991, reading standards revealed evidence of a continued decline (Gorman & Fernandes, 1992).

As Professor Robin Alexander and his colleagues point out, 'over the last few decades the progress of primary pupils has been hampered by the influence of highly questionable dogmas' (Alexander et al. 1994, p. 1). Clarke expressed his concern, as the Secretary of State for Education, that 'present primary practice is not well adapted to effective teaching' (Clarke 1991 p. 9). Long before this, it was the Plowden Report of 1967 that endorsed and helped to enshrine many of the child-centred, progressive tenets: it focused on 'active learning', concentrated on processes rather than the learning of specific subjects or knowledge, and favoured children discovering things for themselves, rather than being directly taught. The advent of this document led to 'an all-embracing, and dogmatic orthodoxy about how children should be taught' (Clarke, 1991, pp. 2-3).

It is perhaps ironic, given the well-known views of Her Majesty's Chief Inspector, Chris Woodhead, that Her Majesty's Inspectorate have, intentionally or otherwise, helped to reinforce the progressive stance, through their establishment of the concept of 'good practice'. Rather than judge teaching practice by the academic results produced, inspectors have judged classroom procedures on how well they fit the mould of 'good practice'.

What is 'good practice' exactly? An HMI report defines classes characterised by 'good practice' as providing opportunities for 'independent work', 'individual initiative', accommodation of the 'needs of individual pupils', and an 'eclectic' approach to the teaching of reading (HMI 1989-90 pp. 13-14).

Throughout the 1970s and 1980s, however, observations made by HMI and others inadvertently reveal some of the undesirable consequences of this 'good practice'. Inspectors note unacceptable noise levels, fragmented reading activities, extreme difficulty managing to cater to the individual needs of all pupils, shortage of time to listen to all children read,

assessments that are too generalised (HMI, 1990), decreasing enthusiasm for reading among the less able, lack of systematic attention to poorer readers (HMI, 1990-91), lack of progression, lack of guidance in choosing books, underchallenged readers, and inadequate monitoring of reading progress (DFE, 1989). Researchers note a substantial variance in the use of phonics (Cato et al., 1992), and clear evidence that topic work leads to fragmentary, and superficial teaching and learning (Alexander et al., 1994).

It would be hardly surprising if reading standards are declining, since 'good primary practice' appears to support precisely those practices identified by scientific research as the most detrimental to reading progress. Findings from experimental research have clear and consistent implications for designing instruction that will most enhance reading achievement. In direct contrast to these findings, however, 'good primary practice' encourages the following: incidental learning rather than direct instruction; learning dictated by the child rather than by logical sequence or order of progression; an emphasis on enjoyable reading activities in collaboration with an adult rather than on teaching children how to read so that they are able to derive the real satisfaction that only comes with knowing how to read for themselves; group projects in favour of more productive, whole-class teaching; trying to decode words by context or meaning, instead of decoding words by their sounds; and reading by osmosis, or slow progress whole-word discovery methods (which are usually described as a 'mixture of methods') instead of an early emphasis on systematic, phonological, code-emphasis instruction.

Evidence from researchers based at Bristol University suggests that since the introduction of the new National Curriculum, the majority of teachers still continue to favour progressive, child-centred methods; these are described as methods whereby teacher and pupil are equal partners, tasks are negotiated with pupils, and the teacher offers assistance as the need arises (Osborn & Broadfoot, 1991; Webster, Beveridge, & Reed, 1996).

Assessing Teachers

Just as it is possible to measure children's reading achievement through validity- and- reliability-enhancing

objective means, the quality of reading instruction teachers deliver can also be measured this way. If quality of teaching is judged not by subjective evaluation but by the reading results produced, the danger of the wholesale adoption of unproven but fashionable methods can be avoided. Such a policy would allow teachers to direct their attention to producing the highest standards of reading by whatever methods they found to be the most effective. Otherwise, the danger is that attention is directed to satisfying subjective criteria instead, such as how well the teacher is fulfilling the requirements of the National Curriculum, how well the teacher is shaping his or her teaching practice to satisfy the current definition of 'good practice', or how well practice matches what is currently expected by school inspectors, school heads, or governors.

If the most important goal is to ensure that all children learn to read early and well, then the criteria by which teacher performance is judged should be related first and foremost to this issue; teachers should be judged on their ability to perform this task. Children's standardised reading scores are an objective indicator of teacher performance in this regard. Judging the quality of teacher effectiveness by the reading results produced, as well as increasing the validity and reliability of teacher assessment, would have the added advantage of permitting comparisons to be made between different teaching techniques. Under this system, the evaluation of teacher performance would be both more objective and more clearly defined. Effective teachers could be identified and rewarded.

In measuring overall reading achievement over time, it is important to take initial differences into account. As the research demonstrates, being able to hear and identify speech sounds, along with having a knowledge of letter-sound correspondences are the two most important indicators of future reading achievement. An early screening test of these skills could be given as an initial routine to establish children's ability prior to instruction; studies show that such a procedure can accurately identify children at risk, alerting the teacher, right at the outset, to those children who will require more intensive teaching (Hoien et al., 1995; Hurford, Schauf, Bunce, Blaich, & Moore, 1994; Majsterek & Ellenwood, 1995: Torgesen, Wagner, & Rashotte, 1994). Intervening at this stage with intensive phonological and letter-sound instruction

has been shown to produce impressive differences in reading achievement, compared to those children at risk who may or may not receive help at a much later stage (Blachman et al., 1994; Felton, 1993). To adopt this procedure as a routine would not only circumvent the need for determining a value-added score for school league tables, but would also considerably reduce the need for reading remediation that may be difficult to arrange, is often expensive and usually limited in its effectiveness.

How does practice in schools measure up to these principles? A teacher appraisal scheme for newly-qualified teachers that was introduced by the government in 1991 fails to monitor teacher progress adequately. Currently such teacher appraisal may consist of one meeting with the training school department head every two years. The National Association of Head Teachers (NAHT) has recently set up a scheme to improve monitoring of teacher effectiveness, through workshops that teach specific target setting and close monitoring of progress. At present, however, teachers do not have a record of their progress, making it difficult for teachers to know what they ought to aim for, or how they might improve.

What many school heads and teachers must aim for is satisfying school inspectors. Inspections reveal that up to 30% of lessons are 'inadequate', 20% of new teachers' performance is 'unsatisfactory', and 10% of teaching is so poor that teachers should not be teaching. What do these descriptions mean? On what basis are lessons and teachers judged? These terms represent inspectors' subjective judgements and lack the precision of, for example, the reading achievement scores produced.

A school's 7+ Average Reading Quotient (ARQ) is the most important indicator of how effectively teachers are carrying out their primary task. The ARQ, a standardised measure, is similar to an Intelligence Quotient in that the average score is 100, a score of 94 indicating that a child is approximately one year retarded in reading, and a score of 106 indicating that he or she is one year advanced.

However, instead of adopting the school's 7+ ARQ as a useful index of teacher performance, confusion reigns. Confusion is created because targets are not specific (what criteria are used to determine if performance is

'satisfactory'?), and the particular biases of the inspector may not be known. In fact, the present system is such that far from being judged on their ability to produce good reading results, teachers are judged on their ability to fit the mould of 'good practice'.

Recently the government (primarily due to the efforts of its Chief Inspector, Chris Woodhead) has attempted to change this concept of 'good practice' where child-centred approaches, and the teacher-as-facilitator view are much applauded, where direct, whole-class teaching are much frowned upon, where results obtained are not important, and where mention of the word 'phonics' is, in the words of one headteacher, 'a capital offence ... certainly it was made clear you wouldn't get promoted using words like that' (Kent, 1996). However, in spite of the government's best efforts to shatter the 'conspiracy of silence' surrounding educational methods (Woodhead, 1996b), to expose the fact that teachers are clinging to reading methods that are 'self-evidently not working' (Ofsted, 1996), and to encourage early systematic phonics instruction, its attempts to bring about change are hampered. Although some teachers may be resistant to change, those who judge teachers' performance must also be considered. Among the inspectorate who once approved of and strongly encouraged this old concept of 'good practice' and whose views one cannot be certain about, are many who now are in positions of power, leading Ofsted teams, in charge of literacy centres, working for government teacher training bodies, or advising important educational officials.

Each school inspection costs £30,000 to £40,000; the Ofsted budget for 1996-7 is £118 million (DfEE, 1996e, p. 165). This huge expenditure on Ofsted inspections perpetuates a system that is, thus far at least, committed to descriptive measurement, and subjective evaluation. If objective, reliable measurement of teacher performance is the desired outcome, then subjective methods of assessment, open to bias, are better avoided. A staggering amount of money could be saved if the standardised testing of reading were to replace national curriculum testing, but even more money could be saved if the results from such tests were used as indicators of the quality of teaching, thus eliminating the need for subjective inspections from government officials. Schools' reading results could be monitored instead; poor reading attainment in a

school would then be a situation which would alert government officials to the school's need for inspection, advice, and/or assistance.

Teacher Trainers

The roots of the politicisation of reading are likely to lie with the teacher trainers and educational researchers. To ensure that teacher trainees are equipped with the ability to deliver high- quality reading instruction in the schools, it should be vital that teacher trainers give students a comprehensive theoretical and practical grounding in reading methods, with a focus on what works and what does not. Reading is a field that is heavily researched, so that knowing how children learn to read and how best to teach them are areas of inquiry that now have a very solid research base. Teacher trainers need to ensure not only that their students are well acquainted with this research but also that they are taught the practical details of the particular sequence of instruction that one needs to follow in order to teach all children to read as efficiently as possible. Teachers in training should also be given the opportunity to observe and work in schools with the most effective teachers of reading, those teachers who have a proven record of producing outstanding reading achievement among their pupils.

How do the teacher trainers measure up? After up to four years of teacher training, nearly half of newly trained primary-school teachers feel inadequately prepared to teach reading, with almost two-thirds taught little or nothing about phonics (Brooks et al., 1992). Many frequently express their general dissatisfaction with the quality of training received (Blatchford et al., 1994). They report insecurity at the prospect of facing a child who cannot read. Some students have accused college teachers of putting forward their own individual perspectives, prejudices and progressive theories, at the expense of providing students with the practical, step-by-step details of how best to teach a child to read (Hadfield, 1992).

Indeed, many universities and colleges currently stress the social and political issues related to education, rather than providing students with an objective analysis of current reading research and teaching them the concrete procedures that comprise a well-structured, effectively sequenced

programme of reading instruction. In an Institute of Education, University of London course guide, for example, it states that 'literacy should thus be regarded as socio-cultural practice and not a neutral technical skill'; 'learning to read and write thus means being socialised into particular views of the meaning of literacy events' (University of London, 1992, p. 29).

Such views are not isolated; evidence suggests that they represent the majority view adopted by teacher training institutions today. Recently, leaflets were handed out at a King's College, London, conference in which Brian Cox comments on the National Curriculum, a document that he sees as 'muddled on vital issues' owing to continual 'political interference'; he goes on to state, 'The new curriculum will do much harm ... sensible teachers will adapt the proposals according to their knowledge of good practice' (Cox, 1994). Students at this conference, who were offered the chance to discuss teaching methods, refused on the grounds that they were 'too stressed and too frightened to accept'; although aware of the bias in the education they were receiving, they were reluctant to speak against it for fear of jeopardising their prospects of qualifying.

Experimental research is strongly at variance with the child-centred, progressive views upheld by most of today's universities, but if attacked for ignoring research findings, these colleges are enormously defensive. The progressive bias present in the teaching of reading in three teacher-training centres, Exeter University, London University and the University College of Ripon and York St John, was exposed through an analysis of their publications, reading lists, pamphlets and brochures (Seaton, 1993). The reaction of academics at London and Exeter universities was to threaten legal action (Salmon, 1993).

It is suggested that there is 'deep insecurity and confusion among teacher trainers that induces ... close-minded antagonism to any reforms that challenge ... existing assumptions' (Hargreaves, 1994). As one recent example of the general stance taken, a student Diploma in Education booklet, *Curriculum Studies and the Primary School* (from the London Institute of Education), asks students what they think about the strongly progressive Plowden Report and various attempts on the part of the present government to sabotage the Plowden ideal. This booklet, transparently manipulative,

then proceeds to enquire of students, 'Are you offended by the suggestion that teachers lack the intelligence to become the Plowden ideal?' (Institute of Education, 1992, p. 14).

Furthermore, there is no doubt that a large industry has grown up in connection with education, not least in the field of publishing. There are powerful vested interests at stake. A brief description of the birth of the International Reading Association, one of the most influential bodies today in the field of reading education, serves as just one case in point. In 1956, William Scott Gray and his American colleagues formed the International Reading Association (IRA). This event came one year after Rudolph Flesch's best-selling book, *Why Johnny Can't Read,* was published. Flesch's book identified and publicised widely for the first time why so many children were failing to learn how to read: they were being taught by the look-and-say reading programmes of Gray and others. The Scott Foresman 'Dick and Jane' readers, first published in 1930, were the dominant texts used throughout American primary schools. Their use represented a multi-million dollar industry; teaching children to read with the new look-and-say basal readers, with their controlled vocabularies, was far more expensive than teaching children phonics, three to eight times more expensive according to some estimates (Armstrong, 1989).

However, parents were beginning to be unhappy with the results; in schools, 'remedial reading' and 'reading disability' were suddenly new phenomena. With the advent of Flesch's book, parents became more vociferous in their disapproval of look-and-say methods. It was necessary for those such as Gray, who had established an extremely lucrative industry by this time, to fight back; amalgamating two previously formed reading organisations (formed in response to growing reading problems), the International Reading Association was created. It was to become 'the impregnable citadel' of the look-and-say method (Blumenfeld, 1990, p. 121). Gray was elected as its first president. Subsequent presidents have been strong proponents of look-and-say methods, more recently termed *whole language*. In 1956, in its first year, the IRA had 7,000 members; in 1995, it had more than 92,000 members in ninety-nine countries and recent conferences have been attended by more than 16,000 people (IRA, 1995).

An organisation of such size wields a powerful influence in shaping the attitudes of educational professionals. A total of

1,609 reading professors on an IRA list of American and Canadian university and college professors were asked to list which reading authorities of all time, in their opinion, had written the most significant, worthy, 'classic' studies in reading; the top three individuals listed were: Frank Smith, Kenneth Goodman, and Edmund Huey (Froese, 1982). Needless to say, these three represent some of the most dedicated advocates of whole-word, look-and-say methods. Kenneth Goodman, who succeeded Gray as a senior author of the Scott Foresman basal reading series (Flesch, 1983), became president of the IRA in 1981.

In England, the United Kingdom Reading Association (UKRA) contributed their support and encouragement of look-and-say methods in 1962 by becoming a branch affiliation of the IRA. By the early 1970s, teacher trainers began to develop closer links with publishers. Up until this time, publishers in England had been cautious and had resisted the progressive, child-centred movement. Organisations such as NATE, the National Association for the Teaching of English, played an influential rôle, in collaboration with universities, colleges, and publishers, in establishing the progressive movement over a number of decades (NATE, 1986). Currently, whether because vested interests are at stake or not, academics continue in their steadfast refusal to acknowledge the leading empirical research in the field, research which is conducted, and consistently confirmed, not only in the United States and England, but in many other countries of the world.

In their concern to preserve the status quo it is now quite likely that many teacher trainers have neither the knowledge nor the skills needed to ensure that their students are taught how to teach reading effectively. Prominent professionals express the popular view that phonics is too complicated, abstract or 'overrated' (Clay, 1991). Some others express their pride at being totally ignorant about the subject (Hynds, 1994).

As for giving trainee teachers the opportunity to observe effective teachers in action and to practise the skills of such teachers, financial concerns may play a rôle here too. Since September 1994, teacher trainees are required to spend most of their time in schools. In the past, schools were happy to have students because they came for a limited period and were supervised by college tutors. Much of this supervisory work now falls to already overburdened teachers, with the

result that many are reluctant to take students. Colleges pay schools that are willing to take students at £600 to £1600 per place. As a result, instead of placing students in classrooms where high standards of reading are achieved, the criterion determining placement are how much colleges are willing or able to pay, and which schools are prepared to accept the offers.

The concept of offering trainee teachers a more practically based training is a positive step, but only so long as quality is assured. At present, the quality of placements students receive is based purely on chance. There is a high probability that the shortcomings of the teacher training to which students are subjected in college are simply compounded by school-based training that may be similarly biased, inadequate, or of poor quality.

Evaluating Reading Research

Any research on which teachers are to base their teaching practice should be dependable. A scientific approach to research has one characteristic that no other method of gaining knowledge has: self-correction. Checks are designed and used to control and verify scientific procedures and conclusions, to satisfy the one ultimate aim of obtaining dependable knowledge. Even where an hypothesis is apparently supported by experiment, a scientist will not accept statements as true, even though first evidence appears promising. Hypotheses are tested and retested and open to public scrutiny. The checks are designed to prevent the possibility of the scientist's personal beliefs, opinions, attitudes, and biases from influencing results. They are designed to ensure objectivity. The level of objectivity brought about by using such checks helps to ensure that the findings of an experiment can be replicated and that any predictions made from these are as dependable as possible.

In the field of education, there are two kinds of reading research conducted: experimental and non-experimental. Experimental research is also known as *controlled, scientific*, or *quantitative* research; non-experimental research may be termed *descriptive, naturalistic, ethnographic,* or *qualitative* research. These terms hint at the differences between them.

Experimental research is characterised by the checks described above; investigations are controlled and objective

test measures (including standardised tests) are used. Non-experimental research, on the other hand, is characterised by a lack of control, and the use of descriptive, subjective measures.

In experimental research it is not assumed that because there is a correlation between two variables, that A causes B; checks are made to rule out other influences that could possibly have contributed to the correlation. If, for example, one wanted to investigate if the amount of letter-sound instruction a child received was correlated with later reading achievement, it would be necessary to consider what other factors might have influenced the measured reading achievement. Some examples of the sort of checks or controls imposed to help rule out the effects of outside factors or extraneous variables are: the random assignment of subjects to experimental groups, the selection of large sample sizes or the use of statistical procedures that remove or partial out the effects of extraneous factors.

In non-experimental research, however, it often seems to be assumed, on the basis of descriptive data analysis (and in the absence of controls or checks), that A is correlated with B or even that A causes B. However, such assertions cannot be made with the same degree of confidence as under experimental conditions, and reliable conclusions cannot be drawn. For example, the observation might be made that children appear to read words better in the context of a story than they do in lists, and it is therefore concluded that context assists decoding. However, if during such observations, the children are always required to read the words in lists first, and words in context, second, this unvarying presentation order could be a factor influencing results. Children may find the words appearing in context easier to read simply because of the practice afforded them during the prior viewing of these same words in lists. By failing to take this factor into account, counterbalancing practice effects by varying the presentation order, the experimenter could easily draw the wrong conclusion.

This is exactly what happened in the case of Kenneth Goodman's non-experimental research on context effects (Goodman, 1965), which constitutes the only evidence ever offered by whole-word advocates in support of the rôle of context in reading (Vellutino, 1991). Goodman not only failed

to take practice effects into account, he also failed to account for differences between good and poor readers. Others have tried to verify Goodman's findings, but under experimental research conditions where such factors have been controlled, the effects reported by him have not been replicated (Nicholson, 1991). In fact, the contradictory evidence is definitive and highly reliable (Gough et al., 1981; Perfetti, 1985; Stanovich, 1980); unfortunately, compared to Goodman's flawed study, it appears to be little known. The detrimental effects of Goodman's erroneous conclusion are seen in today's classrooms where children are encouraged to follow the fruitless practice of guessing at words from context, a practice which not only has been proven unproductive but which also contradicts the scientific research demonstrating what skilful readers do.

This example helps to highlight one of the inherent weaknesses of non-experimental research: the risk of drawing erroneous conclusions. The danger of improper interpretation is high in non-experimental research where data is simply collected, and then interpreted. There is no attempt to control for the difference factors between groups being compared, no attempt to control or equalise experimental treatment conditions of groups under comparison, no attempt to ensure that sample sizes are large, and no random assignment of subjects to groups, which would help reduce the bias likely to occur when none of the foregoing procedures are used. Under these uncontrolled conditions, plausible explanations may be compelling, but often quite wrong. A further difficulty is that once plausible explanations are found and believed, they are often difficult to test, and at the same time, new interpretations can invariably be found to fit the facts (Kerlinger, 1986).

As far as determining teaching practice is concerned, if results from experimental research were not available, we would have to make do with the less objective findings from non-experimental research; but they are. Most of the references in this book are to sophisticated, large-sample studies, conducted under controlled conditions, where the data has been subjected to proper statistical analyses and the results obtained can be replicated.

Again, we can ask: what sort of research do teacher training institutions seem to favour?

As mentioned earlier, the top thirty most recommended teacher training texts (Brooks et al., 1992) do not include one text that provides an objective review of experimental research findings, and the majority of these texts do not even mention this type of research. It is absolutely extraordinary, for example, that a text such as *Beginning to Read: Thinking and Learning about Print* by Marilyn Jager Adams (1990), which provides one of the most contemporary, comprehensive and objective reviews of research findings in the field of reading education, does not appear anywhere on the list of the thirty most recommended teacher training texts.

The texts that *are* on this list repeatedly cite the 'research' of such authors as Frank Smith, Kenneth Goodman, Margaret Meek, and Liz Waterland. None of these people has conducted experimental research; a careful examination of their writing reveals that their research consists of 'assumptions', personal 'beliefs', 'theories', or 'observations'. In an endless cycle, these four authors cite the 'research' of others, or even each other, which upon investigation is found to consist once again of anecdote or speculation. Students reading their books are exposed, perhaps very often without realising it, to an extraordinary bias against experimental studies. As a result, students in teacher training programmes today, tutored as they are in popular theories, opinions, or personal beliefs, are most unfortunately deprived of a wealth of valuable and practical information.

The following statements show each author's bias against empirical research (Words antithetical to experimental research have been italicised):

Margaret Meek

> 'Any significant research I have done rests on my having treated *anecdotes* as evidence' (1979, p.8).
>
> 'Where the partnership of home and school takes into account the *views* of parents, teachers and children and records these as words, *not numbers*, we can already see what modern literacy looks like.' (1982, p. xiii).

Kenneth Goodman

> 'For me research is never neutral. It is always for or against something or somebody. I could never do *amoral*

and atheoretical research. Nor could I do contrived studies on bits and pieces ...' (1992a, p. 192).

Goodman dismisses people who favour experimental research as 'academic elitists (sic) who view many learners as incapable'. The experimental researchers themselves are condemned as being 'amoral' people who depend on '*one-legged models of inquiry*' (1992a, p. 198).

'I am weary, oh so weary, of attacks on whole language ... their disagreement does not make my view wrong ... whole language is ...*beliefs expressed in texts*' (1994, pp. 340, 345).

Frank Smith

'The main instructional implication of *the analysis of this book* is that children learn to read by reading' (1978, p. 3).

'*My* present *assertion* is that any written language is read as Chinese is read, directly for meaning' (1988, p.153).

There have been other indications of a desire to replace experimental research with subjective observations, anecdotes, or 'naturalistic research'. The changes seen in the content of a number of reading research journals and the content of government curriculum or policy documents reveal an acceptance of concepts which are based on theories and assumptions shown by empirical research to be false.

Journals such as *Reading* (published by the United Kingdom Reading Association), *The Reading Teacher*, and *The Reading Research Quarterly* (both published by the International Reading Association, whose past presidents include Marie Clay, and Kenneth Goodman) show a definite bias towards non-experimental research.

The April 1994 issue of *Reading* was devoted entirely to 'teacher research', teachers' personal observations, or 'reflections' (Manning & Harste, 1992, p. 2). During the period from 1986 to 1991, *The Reading Teacher*, hitherto providing a relatively open forum on different points of view as to the best method of teaching reading, published 115 articles in favour of 'whole-language' teaching (an approach based on popular beliefs and anecdote), and 24 articles that described teaching methods opposed by whole-language advocates (methods based on experimental research) (Groff, 1992). The April 1990 issue of *The Reading Teacher* was entirely devoted to 'whole-

language'. The winter 1994 issue of *The Reading Research Quarterly*, no longer able to ignore the popularity of whole-language methods of teaching, and descriptive research, devoted almost the entire journal to a debate over the 'rhetoric of whole language' (Moorman, Blanton, & McLaughlin, 1994, p. 309).

For many professionals, publication means promotion. The fact that such articles are being published does not merely show the current popularity with anecdotal, 'classroom research' (and also, the current tendency to ignore scientific research), but it also helps to demonstrate how vitally dependent many professionals are on journal bodies for furthering their careers. The kind of articles that are published certainly play a part in determining the fashionable attitudes which evolve and become powerfully entrenched.

Government documents also influence attitudes. The Hadow Report (1931), the Plowden Report (1967), the Bullock Report (1975), the Cox Report (1989) and the final draft orders for the National Curriculum (1994) are government policy directives, all of which endorse progressive, child-centred methods to some degree. As each of these documents demonstrates support for theories that have been shown to be false when subjected to strict empirical investigation, they all represent at least indirectly a denigration of experimental research.

As an example, the new National Curriculum final draft orders for English Key Stage 1, Reading, contain the wholesale endorsement of 'real books'. The first statement that appears is: 'Pupils should be given extensive experience of children's literature' (p. 6). Instead of stating that the main goal is to teach children how to read, the primary aim is to have children 'experience' literature. In support of using guessing and context to decode words, the orders state that pupils are to be 'taught to recognise the value of surrounding text in identifying unknown words' (p. 7). In support of whole-word, meaning-emphasis approaches such as shared reading, apprenticeship reading or paired reading, where the child is to 'read' texts over and over with an adult until he or she has memorised the story and is able to join in, the orders state that pupils should be 'focusing on meaning' (p. 7) and given opportunities to 'reread favourite stories and poems, learning some by heart' (p. 8).

Although the curriculum orders do contain a number of concepts that are supported by empirical research, all of the dictated procedures outlined above are not. On the contrary, there is a substantial amount of research which shows that directing a child to behave like a reader or 'experience' literature, to decode words by using guessing and context, and to treat reading as if it were a memorisation task, are all procedures which have been shown to be ineffective in, and even detrimental to, teaching a child how to read.

In addition, the government's national curriculum tests (SATs) are measures which are derived from subjective research. The government's decision not to use objective tests and to use this sort of measure instead, is seen as 'another vote of confidence in teacher professionalism' (Rowan, 1994); in other words, another concession to the progressive establishment, further support of their rejection of scientific research findings.

Thus both the curriculum orders and the national tests represent, to some degree, government support of fashionable theory; indirectly, these also represent the rejection of experimental research findings and scientific investigation itself as a valuable method of enquiry.

Summary

Although research indicates that the most valid and reliable assessment is through the use of objective, standardised measures, in practice children's reading ability, teachers' ability to teach reading effectively, and teacher-trainers' ability to produce effective teachers of reading are all evaluated through the use of subjective, error-prone assessment measures. Reading standards will not improve so long as there is no year-to-year system of assessment in place that is capable of giving an accurate and reliable picture of what reading standards really are, there is no monitoring of classroom teachers' reading methods linking these with pupil reading attainment produced, and there is no routine, objective assessment of the quality of teacher-training courses.

In addition, the type of research on which teachers and others in the field of education most often judge acceptable enough to base practice consists of descriptive, non-experimental research. As long as practice is to be based on fashionable anecdotal or 'naturalistic research', the reading standards among 7-year-olds will continue to be under threat.