Published in final edited form as:

Br Educ Res J. 2007 February 1; 33(1): 47–59.

Improvement in national test reading scores at Key Stage 1; grade inflation or better achievement?

Sara Meadows*, David Herrick, Anthony Feiler, and the ALSPAC Study Team
University of Bristol, UK

Abstract

The aim of the UK National Literacy Strategy is to raise standards in literacy. Strong evidence for its success has, however, been lacking: most of the available data comes from performance on tests administered in schools or from Office for Standards in Education reports and is vulnerable to suggestions of bias. An opportunistic analysis of data from a population cohort study extending over three school years compares school-based scores at school entry and at age 7–8 with independently administered scores on similar tests. The results show a small but statistically significant rise between 1998 and 1999 and between 1998 and 2000 in scores on both Key Stage 1 Reading Standard Assessment Tasks taken in schools and the reading component of the WORD test taken independently. This is clear evidence for a real rise in reading attainment over this period, which may be attributable to the children's experience of the National Literacy Strategy.

Introduction

Few would disagree that it is of paramount importance that children achieve good levels of literacy in modern society. At an individual level, literacy has a profound effect on overall quality of life, self-identity, and the capacity to function in an ever more complex world; and at a macro-level, the success of society depends on a well-educated, highly literate and adaptable workforce (Riley, 2001). Improving literacy learning has been a dominant theme ever since the development of statutory education over the last one hundred years or so. However, with the recent advent of the digital age and the need for skilful use of information communications technology, the importance of children becoming proficient with literacy skills is becoming ever more pressing.

As part of their ambitious school improvement programme, the newly elected Labour Government implemented the National Literacy Strategy (NLS) in English primary schools in 1998. The aim of the NLS was to raise literacy standards in primary schools, and improving literacy was seen as the key to raising educational levels generally across the curriculum. In the 'Introduction' to the NLS 'Framework for Teaching' (Department for Education and Employment [DfEE], 1998) the reader is informed that 'Literacy is at the heart of the drive to raise standards in schools' (p. 2). Improving literacy was to be achieved by implementing a major programme of reform of literacy teaching. The NLS introduced an approach to literacy that, according to Office for Standards in Education (Ofsted) reports, was not consistently or uniformly in use in schools; namely, a move away from individualised methods to whole-class and group teaching; making more links between reading and writing; introducing more systematic use of phonics; and a strong emphasis on more direct or explicit teaching of literacy skills (Beard, 1990, 2000).

Although aspects of the NLS have been welcomed, including the focus on whole-school strategies (Frater, 2000), the literature critiquing the NLS is beginning to grow. Mroz et al. (2000) raise concerns that teachers have been pressurised into using more directive forms of teaching that reduce opportunities for pupils to question or explore ideas in depth. They argue that some of the styles of teaching prescribed in the NLS are at odds with social constructivist theories of learning that emphasise the importance of pupils playing an active part in their learning. This, it is argued, results in interactions between teachers and pupils that are cognitively undemanding, where pupils are often 'mere listeners or respondents' (Mroz et al., 2000, p. 387). Evidence that the NLS, with its emphasis on pace and speed, may militate against cognitively rich interactions between teachers and pupils comes from observational research conducted by English et al. (2002). These authors found that since the introduction of the NLS some teachers appeared to ask fewer challenging questions and had fewer sustained interactions with pupils.

Another criticism relates to the lack of emphasis given to meaning in the theoretical model of reading outlined in NLS documents, which, it is argued, is fundamental to the process of reading being purposeful and rewarding for children. Indeed, Riley (2001) points to the cursory manner in which reading is theorised in the NLS documentation, giving teachers only a superficial insight into the complex field of literacy acquisition. Wyse (2003) argues that the pedagogic strategies of the NLS Framework for Teaching do not rest on adequate empirical evidence, and that the causes of any improvement are therefore ambiguous, a point also made in the evaluations by Earl and colleagues from the Ontario Institute for Studies in Education (Earl et al., 2000, 2001, 2003). The debate about whether the pedagogies of the NLS should be expected to bring about improvements in literacy continues (e.g. Hiebart, 2000; Beard, 2003), and there is further debate about whether there is good evidence of real improvements.

The success of the Government's pressure to improve the standards achieved by school pupils is often assessed by looking for rises in the performance of children as measured on Standardised Assessment Tasks (SATs) at the end of each Key Stage of education. The evidence is mixed. A government report on the NLS (Ofsted, 2002) provides an overview of the first four years of its implementation, and includes a summary of the standards attained by pupils. In 1998, 64% of 11-year-olds achieved level 4 in English, and by 2000 this rose to 75%. Ofsted reports on literacy standards (Ofsted, 2003; 2004) indicates that although government targets for primary English have still not been met, there is some evidence of a slight improvement in the results for pupils in Years 3, 4 and 5 in the 2004 national tests (after a period when results had been static).

Despite primary literacy targets not being met, Ofsted has maintained that the NLS has had a significant impact on pupils' literacy levels. This assertion has been questioned. Not all commentators are willing to take this claim as evidence of real improvement in pupils' achievement. Those dismissing the reality of improvement on SATs use a number of arguments against them (Smith et al., 1999; Davies, 2000; Gold, 2002). One basic concern is about the independence of the teaching and assessment processes. SATs are administered in school by class teachers, the people who have been responsible for teaching the pupils being assessed, and thus are vulnerable to charges of bias in the direction of judging pupils' performance more favourably than it objectively deserves. Furthermore, schools are judged on their pupils' success rate on SATs and claims have been made that schools have submitted falsified records in order to boost figures for league tables. Additionally, there have been suggestions that SATs measure a rather limited amount of learning as related to the NLS. Improvement might have been brought about through the introduction of 'high stakes' tests, resulting in teachers teaching exactly what the test requires, in which case the

pupils' improved performance might be limited to the tests rather than reflecting real improvements in the quality of pupils' learning (Black et al., 2002).

It is possible that there could have been a real improvement on SATs tests and in children's reading attainment. However, those who want to see reliable evidence of a real improvement in pupils' performance, especially those who want education to extend beyond the formally assessed curriculum, need evidence independent of teachers and extending beyond the syllabus. Data from an ongoing cohort study provide a unique and robust answer to this. The data set allows us to compare children's achievement at the beginning of the primary school, their achievement on Key Stage 1 (KS1) reading in successive school year cohorts, and their achievement on the reading element of a standardised test administered outside the school system. The reality of a rise in SATs scores in successive years can be checked in two ways: first, by examining the same pupils' performance at school entry and their family background, to exclude the possibility that the SATs improvement is due to each successive year arriving in school more ready for learning; and secondly, by looking at whether a rise in SATs scores is accompanied by a rise in the scores of the same pupils on similar tests administered independently of the schools. There are three possibilities:

1. If children are initially at the same educational level and there is no sign of improvement on SATs scores between 1998 and 1999, and between 1999 and 2000, this suggests that there has been no rise in standards of reading.
2. For children who are at the same educational standard at the start, an improvement only in SATs (and not in independently administered tests) could suggest a narrow increase in skills that only shows up on the SATs. This might show marker bias, teaching for the test or easier SATs.
3. If pupils in successive years are at the same educational level at the start of their schooling but are better on SATs *and* on other independent assessments by the end of KS1, then this is stronger evidence that they are really becoming better at literacy skills.

The present study tests these hypotheses by looking for changes in reading scores on KS1 SATs over three school years, checking whether any changes are due to changes in children's characteristics at school entry, and examining whether changes are also found on tests independent of schools.

Sample and method

The present study uses the Avon Longitudinal Study of Parents and Children (ALSPAC) data set (Golding et al., 2001) to address these questions. ALSPAC is a prospective longitudinal cohort study that includes families of all demographic categories, from inner-city, suburban, seaside, rural and small town areas. To be eligible for the study, mothers had to be resident in Avon during their pregnancy and have an expected date of delivery between 1 April 1991 and 31 December 1992. The core ALSPAC sample consists of 14,541 pregnancies. At one year, 13,988 infants were alive; this is the baseline figure for these analyses. The focus of the study has been the effects of the physical, social and psychological environment on the health and development of the children and their families. Relevant data for the present study include:

1. children's scores on reading assessed by the teacher in the School Entry Assessment (EA) during the first month in the reception class, when the child was aged 4–5;
2. SATs scores in reading at the end of KS1, when the child was aged about 7;
3. scores on the WORD test of reading, assessed in a research clinic when the child was aged about 7½.

The actual birth dates of the majority of the ALSPAC children extended over the 21-month period of expected dates of delivery, and their entry to school fell in three successive school years, with the oldest children, born April–August 1991, entering school during 1995 as the youngest part of that year's intake, the whole school intake of 1996 comprising the middle of the sample (born September 1991–August 1992), and the youngest ALSPAC children, born September–December 1992, forming the oldest part of the 1997 school intake. The sample is not ideal for comparisons of school cohorts because only the central school year covers the entire year of births. However, comparisons between children with the same ages and the same length of school experience are possible for two pairs of successive school years: 1995 and 1996 entry for children born between April and August, and 1996 and 1997 entry for children born between September and December. In most of our analyses, we control for the age of the child at the time of testing. The analyses are confined to children attending schools in the state sector.

At the time of the birth of the ALSPAC sample they were all resident in one local authority, Avon. Local government reorganisation in 1996 divided the Avon area into four local authorities (Bath and North-East Somerset [BANES]; South Gloucestershire; North Somerset; and the County and City of Bristol). The four new local education authorities continued to cooperate in various ways, including the use of the same entry assessment used by Avon.

Measures

Entry assessment (EA) and Key Stage 1

Scores on the school-based tests were made available by the four local education authorities (LEAs) covering the study region: Bristol; South Gloucestershire; North Somerset; BANES.

Although it was not compulsory, most of the local schools took part in an entry assessment exercise, with the same criteria used by all four local LEAs for all three entry cohorts. The Avon Reception Entry Assessment was developed by a working group of heads, teachers, early years advisers and educational psychologists to provide baseline information on entry to school. It was standardised on local samples (10,000 pupils) and as well as being used by the LEAs for value-added work it has been used by a number of independent research studies (e.g. Blatchford et al., 2003). Children were rated by their teacher on a number of scales provided by the LEA (South Gloucestershire Professional and Curriculum Support Service, 1996), including early reading, during the first half of their first term in the reception class. For each topic the child was given an integer score between 2 and 7 indicating the level that they had reached (note that these levels are not related to National Curriculum levels). Illustrative items used for rating children's early reading levels are outlined in Appendix 1. For this project binary variables were created indicating whether each child had reached level 5 or not (see Table 1).

At the end of KS1, during the summer term of the children's third year in school (that is, in their ninth term of schooling) and at an age of approximately 6 years 9 months to 7 years 8 months, all children in state schools took nationally set tests (SATs) including reading (Qualifications and Curriculum Authority, 1998). The assessments were administered and marked by the children's own teachers in their schools and rate each child according to which National Curriculum level they have reached (1, 2 or 3). Children who failed to reach level 1 were marked as W and levels higher than 3 could optionally be assessed using KS2 materials. Level 2, the target level for KS1, was split into three grades A, B and C. Reading was assessed by two instruments: a reading task and a reading comprehension test. The reading task awards levels 1 and 2 only. Those children achieving level 2 in the reading task were then entered for the reading comprehension test. The reading comprehension test first

awards level 2. Children achieving grade A at level 2 in both the reading task and reading comprehension test were then assessed with a level 3 reading comprehension test. A composite reading measure was derived by taking the results of the reading task, and adding level 3/4+ from the reading comprehension test. For this project binary variables were created indicating whether each child had reached the target of level 2 or not (see Table 2).

The reading measures at EA and KS1 were available for 9475 and 11,218 children (67.7% and 80.2% of the ALSPAC children respectively), with 9032 (64.6%) children having both measures. Performance at KS1 was not biased by availability of an EA measure, with 16.1% of both those with and without EA failing to achieve the expected standard ($p=.965$). A further 14 children had taken either EA or KS1 in the wrong school year for their date of birth (i.e. were outliers in the relevant age at testing distribution). Excluding these children ensures that all children were taking their EA in their first month in school and their KS1 SATS during their ninth term in school. The sample for the first part of the reading analyses was therefore 9018.

WORD

At the target age of 7 years 6 months, the ALSPAC children attended a research clinic, called Focus@7, during which they took the WORD test (Rust et al., 1993). The actual age at attendance varied from 6 years 10 months to 9 years 5 months, with 95% attending between 7 years 4 months and 8 years 7 months. The score used in this study is from the Basic Reading part of WORD, which required the child to read out loud single words of increasing difficulty. The basic reading subtest of the WORD uses pictures and printed words to assess decoding and word reading. The child is first shown a number of pictures with words underneath and asked to point to the word with the same beginning or ending sound as the picture. The child is then asked to point to words that name the pictures shown. Finally, the child is asked to read aloud a series of words in order of increasing difficulty. The task is stopped after six consecutive errors (Rust et al., 1993). The score used here is an integer score on a scale of 0 to 52. It is approximately normally distributed with a mean score of 28.22 and a standard deviation of 9.35. The scale was standardised on national samples of children (Rust et al., 1993) and these data have been used in other investigations of children's reading skills (e.g. Gathercole et al., 2005).

Focus@7 was attended by 7841 (56.1%) of the study children, with 7623 having a valid score on the reading test. Of these 7623, 5105 also have EA and KS1 reading data and were in the correct school year when they took them. This is the sample for the second part of the reading analyses. This is a biased sub-sample due to selective drop out affecting availability of data collected at Focus@7. Only 10.1% failed to achieve the target level for reading at KS1, as opposed to 16.1% in the whole population ($p<.001$). Table 3 shows the mean WORD scores for children scoring at different levels on the KS1 assessment.

Analysis

The first stage of the analysis was to look for cohort effects in the EA and KS1 results. The raw effects are illustrated in Table 4 in which the 1996 entry cohort is split into three groups by month of birth to allow comparison with the partial data available from the 1995 and 1997 entry cohorts.

Comparison of rows 2 to 4 in Table 4 (i.e. comparing the performance of different age groups within a single entry cohort) shows that there were strong season of birth effects at both EA and KS1, as is usual on all tests taken at a fixed date in the year when performance on the tests is age-related. Comparison of rows 1 and 4 and rows 2 and 5 in Table 4 (i.e. comparing the performance of equivalent age groups in different school entry cohorts)

indicates that there is no clear pattern for EA in early reading; summer born children do worse in 1996 than in 1995, while autumn born children do better in 1997 than in 1996. However, there were cohort effects for reading at KS1, with a noticeable improvement between successive cohorts. There was a strong association between scores on the EA and scores at KS1 (chi square $[df=1]=956.9$, $p<.000$).

The differences between cohorts were formally examined by fitting logistic regression models.

After adjustment for age at testing and sex, the p -value for including school cohort in the model predicting EA reading performance was $<.001$ (see Table 5). There was strong evidence for a cohort effect in early reading at EA. Compared with the 1997 entry cohort, children in the 1996 cohort were more likely to be low scorers (Odds Ratio (OR)=1.193, 95% Confidence Interval (CI)=[1.005, 1.417]) whilst children in the 1995 cohort were marginally less likely to be low scorers (OR=0.955, 95% CI=[0.765, 1.191]). Thus, although the cohorts differ, there is no sign of a systematic improvement at EA between 1995 and 1997.

We applied the same methodology to the KS1 reading results and also adjusted for children's performance at the EA, to adjust for possible baseline differences in different years. After adjustment for performance at EA, age at testing and sex, the p -value for including school entry cohort in the model predicting KS1 reading performance was $<.001$ (see column relating to the full sample in Table 6). There was strong evidence for a cohort effect in reading at KS1. Compared with the 1997 entry cohort, children in the 1996 cohort were more likely to fail to attain the expected level (OR=1.500, 95% CI=[1.218, 1.847]) with the same being true, but to a greater extent, of children in the 1995 entry cohort (OR=1.875, 95% CI=[1.437, 2.446]). This shows a systematic improvement in KS1 reading results between the 1995 and 1997 entry cohorts, and between the 1996 and 1997 entry cohorts.

In order to determine whether the systematic improvement in KS1 reading performance is reflected in the independent assessment of reading performance it is necessary to restrict the sample of 9018 children used in the first part of the analysis to the group of 5015 children for whom the WORD score is also available. Since, as noted above, this is a biased restriction, it is necessary to first check that the systematic improvement in KS1 reading performance noted above is reflected in the restricted sample (results also displayed in Table 6). The p -value for including school cohort in the model predicting KS1 reading performance was .038. There is some evidence for a cohort effect on reading at KS1 in the restricted data set. As before, the fitted odds ratios indicate a systematic improvement: for 1996 vs. 1997 OR=1.428, 95% CI=[1.023, 1.994] and for 1995 vs. 1997 OR=1.719, 95% CI=[1.130, 2.615]. Although the p value is inevitably larger in the much smaller restricted sample, the pattern in the odds ratios is similar in the full and restricted samples.

The next stage was to perform a multiple regression predicting the WORD reading score to see if a similar rise in those results could be identified. Since attendance at the research clinic was targeted at a fixed age, children had had varying amounts of schooling at the time of the WORD assessment. It was decided to use the time between starting school and attending the clinic to adjust for this amount of schooling effect. After correcting for age at testing, amount of schooling, sex and performance at EA, the p -value for including the KS1 cohort in the multiple regression model predicting WORD score was .012 (see Table 7). There is some evidence for a school cohort effect in performance on the WORD test. Compared with the children in the 1997 school entry cohort, the children in the 1996 cohort had a mean WORD score deficit of 0.891 (95% CI=[0.174, 1.609]) and the children in the 1995 cohort had a mean WORD score deficit of 1.532 (95% CI=[0.514, 2.550]). Thus, the

systematic rise in SATs scores is paralleled by a systematic rise in reading scores on an independent test. As the standard deviation of the WORD scores was 9.35, this suggests an improvement of about one-twelfth of a standard deviation over one year, and a sixth over two years.

Discussion

Small, but systematic and statistically significant, improvements in performance in KS1 reading performance year on year were observed in the 1995 to 1997 school entry cohorts. These improvements were mirrored in independent tests of reading ability, where there were also small but systematic and significant improvements. These differences are not due to initial differences in the children or by age at testing or by sex, as these factors have all been adjusted for. They are not caused by factors in the tests and assessments, because although differences in KS1 tests are confounded by changes in test content between years, the independent standardised tests given were identical for all three year groups. They are unlikely to be random fluctuations in test score, since they were both statistically significant and in a consistent direction for all tests over all years. They are found in a large population sample spanning four local authorities and broadly representative of the national population. While they may reflect teachers preparing children better for the KS1 tests, the children appear to have shown a benefit from this preparation as they tackle the reading tasks in the ALSPAC clinics. Our data do not, of course, show whether they would have done better on other reading tasks.

ALSPAC was not set up in order to carry out studies of the effects of the introduction of the National Literacy Strategy, and consequently there are limitations in the present study. We think there are three main limitations. Firstly, we do not have complete cohorts for all school years, though by adjusting for the age of the child at each occasion of testing we have reduced this difficulty. Secondly, we do not have observations of the children's exposure to, and their work under, the National Literacy Strategy. We know that the National Literacy Project was piloted in the local schools in 1996–98 and the NLS fully implemented in them from 1998. We think it is reasonable to infer that this was a period when teachers were developing increasing expertise with the NLS, and that the younger ALSPAC children had more exposure to it (3 years) than the older (1 year), but we are not able to document this more fully. Thirdly, we do not have tests of unimpeachable psychometric exactness assessing the children's performance, but as Tymms (Tymms, 2004; Tymms *et al.*, 2004) points out, few if any studies have such measures. Our measures are public, standardised, and well used by both education authorities and researchers. Unreliability of tests with young children is an issue, but using a regression analysis which adjusts for earlier attainment to assess year-on-year change in later attainment substantially controls for unreliability in earlier measures. It may be an advantage that the data we have used were not collected on purpose for a study of the implementation of the NLS: rather, they were collected for the normal purposes of schools or of an independent epidemiological research study. This means that bias due to being led by the expected research outcomes is much less likely. The large size and comprehensive data set of ALSPAC are great strengths.

Bearing these limitations and strengths in mind, we think that there is a small, but real, rise in scores over the three years of the ALSPAC cohort. The crucial issue is how to interpret this. It is conceivable that the rise in reading scores has nothing to do with the implementation of the National Literacy Strategy in schools. We do not think it was due to a change in the cohorts of children independent of their experience in school: despite a comprehensive knowledge of the relevant research literature (Meadows, 1996, 2005), we are not aware of any child-based or family-based variables that could both differ between the three year groups of the ALSPAC sample and also cause a difference in scores at KS1

without having caused a difference in scores at EA. The only possibility of this sort that we can see is that the rise is an instance of the 'Flynn effect' (Flynn, 1987, 1998, 1999, 2003; Meadows, 2006), i.e. the general rise in IQ scores (especially on tests involving general intelligence or 'g') found in many samples over the last half century. Our findings may be a sign of the Flynn effect in samples of children very much younger than in any other research, over a very much shorter time period, and on a test which is not heavily dependent on 'g'. However, what has led to the Flynn effect is not known, and its likeliest cause may be improvements in education reflecting more effective teaching. We therefore believe that the rises in scores at KS1 which we have found are most probably caused by improvements in school teaching of reading over the period. While such improvements in school teaching of reading might have occurred independent of the implementation of the National Literacy Strategy, the implementation and bedding-down of the NLS occurred during the period in which the ALSPAC children's scores were rising, and we are not aware of other changes in these schools over this period which were focused on improving children's reading. We think that in default of a better explanation it is plausible to link the change in teaching and the improvement in scores. We therefore believe that these results suggest that in this sample there is a real, generalised and statistically significant improvement in early reading attainment associated with the National Literacy Strategy.

Appendix 1. Early reading scale from the Avon Reception Entry Assessment

The items listed below are illustrative of the skills assessed in the Avon Reception Entry Assessment early reading scale (Woodspring Educational Resource Centre, 1996):

Stage 3: Listens to a story lasting 5–10 minutes, can look at books alone, can talk about pictures in a book.

Stage 4: Knows the direction in which pages are turned, i.e. from front to back of a book, knows that you read words not pictures, can retell part of a story using pictures in the book.

Stage 5: Arranges a sequence of up to four pictures and tells the story which makes sense to the child, when shown one word can find the word on the page, can select two objects that begin with the same sound from a collection of objects (3–4 sounds).

Stage 6: Recognises at least 15 letters by sound, has a sight vocabulary of at least 15 words from different contexts, can draw inferences from stories, e.g. how a person might be feeling, can suggest more than one possible ending for a story.

References

- Beard, R. Developing Reading 3–13. 2nd edn. Hodder & Stoughton; London: 1990.
- Beard R. The National Literacy Strategy: review of research and other related evidence. Oxford Review of Education. 2000; 26:421–436.
- Beard R. Not the whole story of the National Literacy Strategy: a response to Dominic Wyse. British Educational Research Journal. 2003; 29(6):917–928.
- Black, P.; Broadfoot, P.; Daugherty, R., et al. Testing, motivation and learning. Cambridge: Cambridge University Press; 2002.
- Blatchford P, Edmonds S, Martin C. Class size, pupil attentiveness and peer relations. British Journal of Educational Psychology. 2003; 73:15–36. [PubMed: 12639275]
- Davies N. Fiddling the figures to get the right results. Education Guardian. Jul 11.2000
- Department for Education and Employment (DfEE). The National Literacy Strategy: framework for teaching. London: HMSO; 1998.

- Earl, L.; Fullan, M.; Leithwood, K., et al. Watching and learning: OISE/UT evaluation of the implementation of the National Literacy and Numeracy Strategies. Nottingham: DfEE Publications; 2000.
- Earl, L.; Fullan, M.; Leithwood, K., et al. Watching and learning 2: OISE/UT evaluation of the implementation of the National Literacy and Numeracy Strategies. Nottingham: DfEE Publications; 2001.
- Earl, L.; Watson, N.; Levin, B., et al. Watching and learning 3: OISE/UT evaluation of the implementation of the National Literacy and Numeracy Strategies. Nottingham: DfES Publications; 2003.
- English E, Hargreaves L, Hislam J. Pedagogical dilemmas in the National Literacy Strategy: primary teachers' perceptions, reflections and classroom behaviour. *Cambridge Journal of Education*. 2002; 32(1):9–26.
- Flynn JR. Massive IQ gains in 14 nations: what IQ tests really measure. *Psychological Bulletin*. 1987; 101:171–191.
- Flynn, JR. IQ gains over time: toward finding the causes. In: Neisser, U., editor. *The rising curve: long-term gains in IQ and related measures*. Washington, DC: American Psychological Association; 1998.
- Flynn JR. Searching for justice: the discovery of IQ gains over time. *American Psychologist*. 1999; 54(1):5–20.
- Flynn JR. Movies about intelligence: the limitations of *g*. *Current Directions in Psychological Science*. 2003; 12(3):95–99.
- Frater G. Observed practice: English in the National Literacy Strategy. *Reading*. 2000; 34(3):107–112.
- Gathercole S, Tiffany C, Briscoe J, Thorn A, the ALSPAC Study Team. Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry*. 2005; 46(6):598–611. [PubMed: 15877766]
- Gold K. Great results fudge true picture. *Times Education Supplement*. May 10.2002
- Golding J, Pembrey M, Jones R. ALSPAC—the Avon Longitudinal Study of Parents and Children—I. Study methodology. *Paediatric and Perinatal Epidemiology*. 2001; 15(1):74–87. [PubMed: 11237119]
- Hiebart E. The National Literacy Strategy from an international perspective. *Journal of Research in Reading*. 2000; 23(3):308–313.
- Meadows, S. *Parenting behaviour and children's cognitive development*. Hove: The Psychology Press; 1996.
- Meadows, S. *The child as thinker: the development and acquisition of cognition in childhood*. London: Psychology Press; 2006.
- Mroz M, Smith F, Hardman F. The discourse of the Literacy Hour. *Cambridge Journal of Education*. 2000; 30:379–390.
- Office for Standards in Education (Ofsted). *The National Literacy Strategy: the first four years 1998–2002*. London: Ofsted; 2002.
- Office for Standards in Education (Ofsted). *The National Literacy and Numeracy Strategies and the primary curriculum*. London: Ofsted; 2003.
- Office for Standards in Education (Ofsted). *The National Literacy and Numeracy Strategies and the primary curriculum*. London: Ofsted; 2004.
- Qualifications and Curriculum Authority (QCA). *Key Stage 1 1999—assessment and reporting arrangements*. London: QCA; 1998.
- Riley J. The National Literacy Strategy: success with literacy for all? *The Curriculum Journal*. 2001; 12(1):29–58.
- Rust, J.; Golombok, S.; Trickey, G. *WORD Wechsler objective reading dimensions manual*. Sidcup: The Psychological Corporation; 1993.
- Smith, F.; Hardman, F.; Mroz, M. Evaluating the effectiveness of the National Literacy Strategy: identifying indicators of success. paper presented at the European Conference on Educational Research; Lahti, Finland. September 1999; 1999.

- South Gloucestershire Professional and Curriculum Support Service. Entry assessment in South Gloucestershire. Weston-super-mare: Woodspring Educational Resource Centre; 1996.
- Tymms P. Are standards rising in English primary schools? *British Educational Research Journal*. 2004; 30(4):477–494.
- Tymms P, Merrell C, Jones P. Using baseline assessment data to make international comparisons. *British Educational Research Journal*. 2004; 30(5):673–689.
- Wyse D. The National Literacy Strategy: a critical review of empirical evidence. *British Educational Research Journal*. 2003; 29(6):904–916.

Table 1
Entry assessment reading results

Level	Reading			
	Full variable		Binary split	
	<i>n</i>	%	<i>n</i>	%
2	72	0.8	2091	23.2
3	201	2.2		
4	1818	20.2		
5	4470	49.6	6927	76.8
6	2056	22.8		
7	401	4.4		
Total	9018	100.0	9018	100.0

Table 2**KS1 reading results**

Level	Reading			
	Full variable		Binary split	
	<i>n</i>	%	<i>n</i>	%
W	216	2.4	1451	16.1
1	1235	13.7		
2C	1455	16.1	7567	83.9
2B	1863	20.7		
2A	1496	16.6		
3/4+	2753	30.5		
Total	9018	100.0	9018	100.0

Table 3
WORD and KS1 scores

KS1 level	<i>n</i>	WORD Mean score	Standard deviation
W	68	8.4	5.9
1	628	14.6	5.8
2C	948	20.4	5.9
2B	1257	25.2	6.3
2A	1141	29.8	6.0
3+	2565	34.8	6.2

Table 4
EA & KS1 reading results split by grouped month of birth

Date of birth	<i>n</i>	School entry cohort	% achieving level 5+ at EA	% achieving level 2+ at KS1
April–August 1991	1856	1995	71.9	78.4
September–December 1991	1735	1996	83.5	86.3
January–March 1992	1346	1996	77.0	84.7
April–August 1992	2284	1996	68.2	80.4
September–December 1992	1797	1997	86.1	91.0
All children	9018		76.8	83.9

Table 5
Cohort effects for EA reading. Displayed are the *p* values for inclusion in logistic regression models predicting failure to achieve level 5 after correction for age at testing and sex and also adjusted ORs and 95% CIs

Entry cohort	Reading (<i>n</i> =9018) <i>p</i> <0.001
1995	0.955 (0.765, 1.191)
1996	1.193 (1.005, 1.417)
1997	1 Ref

Table 6

Cohort effects for KS1 reading. Displayed are p values for inclusion of cohort in logistic regression models predicting KS1 reading, after adjustment for EA performance, age at testing and sex, with adjusted ORs and 95% CIs. Results are displayed for both the full samples and the restricted samples for which independent test results are available

Entry cohort	Reading	
	Full ($n=9018$) $p<0.001$	Restricted ($n=5015$) $p=0.038$
1995	1.875 (1.437, 2.446)	1.719 (1.130, 2.615)
1996	1.500 (1.218, 1.847)	1.428 (1.023, 1.994)
1997	1 Ref	1 Ref

Table 7
Cohort effects for WORD score. Displayed are the p values for inclusion in regression models after correction for EA performance, age at testing, amount of schooling and sex and also adjusted parameter estimates and 95% CIs

Entry cohort	WORD ($n=5105$) $p=0.012$
1995	-1.532 (-2.550, -0.514)
1996	-0.891 (-1.609, -0.174)
1997	0