National Research and Development Centre
for adult literacy and numeracy

The levels of attainment in literacy and numeracy of 13- to 19-year-olds in England, 1948–2009

Research report

Sammy Rashid and Greg Brooks, University of Sheffield

The University Of Sheffield.

Leading education and social research
Institute of Education
University of London

Published by the National Research and Development Centre
for Adult Literacy and Numeracy

For information on alternative formats, or to give feedback on the content and accessibility of this publication, please contact:

Publications
NRDC
Institute of Education
20 Bedford Way
London WC1H 0AL
Telephone: +44 (0)20 7612 6476
Fax: +44 (0)20 7612 6671
Email: publications@nrdc.org.uk

Contents

Date of research and writing of this report

This report is based on research funded by the Department for Education and Skills and undertaken between April 2005 and March 2007. The text was updated in March 2010 to include reference to additional research made available since the initial research was completed.

Project team and authorship

Professor Greg Brooks, University of Sheffield, project director
Dr Sammy Rashid, University of Sheffield, research fellow

Sammy Rashid assembled all the data and all the level descriptions, carried out all necessary statistical analyses, and wrote the first draft of this report (80% of the research, 40% of authorship).

Greg Brooks directed the project and edited this report (20% of the research, 60% of authorship).

Acknowledgments

The project was funded by the *Skills for Life* Strategy Unit within the (then) Department for Education and Skills, now subsumed into the Department for Business, Innovation and Skills, via the National Research and Development Centre for adult literacy and numeracy, which is led by the Institute of Education, University of London.

Most of the information used was publicly available. However, the Yellis data (see Sections 2.1.4, 2.3.3, 3.1.3 and 3.3.3) were supplied by Dr Robert Coe (for 1993–2006) and Neil Defty (for 2007–09) of CEM, the Centre for Evaluation and Monitoring (formerly the Curriculum, Evaluation and Management Centre) at Durham University (www.cemcentre.org); this was authorised by Professor Peter Tymms, Director of CEM. We are most grateful to them, to Peter Burke of the University of Nottingham, who responded to a query and thus provided the comment reproduced towards the end of Chapter 3, to Prof. Anna Vignoles and Claire Crawford of the Institute of Education, University of London, for providing the correlations from their Widening Participation Project mentioned towards the end of Chapter 3, and to Dr Siobhán Carey who provided the IALS correlations also mentioned towards the end of Chapter 3.

Executive summary

Headline findings

The reading, writing and numeracy attainments of this age group at the top end are among the best in the world. Except in number skills at age 13 over 20 years ago and in writing between 1980 and 1993 there is no evidence of a decline in any area. Overall, the data suggest that average attainments in all three areas have improved over roughly the last decade. Most young people do have functional to good literacy and numeracy, but more needs to be done. In particular, about 17% of young people aged 16–19 have poorer literacy, and about 22% have poorer numeracy, than is needed for full participation in today's society.

Method

Information on attainment levels of 13- to 16-year-olds was gathered from national monitoring surveys for the period 1948–88, international surveys for the period 1960–2007, and test and examination results for 1995–2009. Data on 16- to 19-year-olds were extracted from adult literacy and numeracy studies for the period 1981–2003. More information was available for reading than for numeracy, and least for writing. Where possible, the data were analysed to provide time series, and detailed criteria for various levels of literacy and numeracy from more recent surveys were assembled and analysed.

Conclusions on trends in attainment in reading

- There was an improvement in average reading scores from 1948 to 1960.
- Average levels remained remarkably constant from 1960 to 1988.
- There is a gap in the evidence base from 1988 to about 1997.
- Some data suggest a gentle rise between 1997 and 2004, then a further plateau.
- But the proportion of young adults with poor reading (below Level 1) seems to have remained stubbornly at about 17%. A lifecourse trend (improvement into early middle age, then usually a plateau, then decline) will have lifted some people into adequate literacy levels by their early to mid 20s, but many still have poor literacy at all later ages.

Conclusions on trends in attainment in writing

- There is no evidence before 1979.
- There was no significant change between 1979 and 1988.
- GCSE results suggest a rise from 1989 to 1990, then a plateau from 1990 to 1998, then a gently rising trend up to about 2005 and a rather faster increase up to 2009. The increase over the entire period was substantial.
- KS3 test results show a substantial rise between 1995 and 2007.

- But studies of examination scripts suggest little change between 1980 and 2004 (with a dip in between).
- There may be no contradiction: achievement at particular levels may not have changed (as suggested by studies of marking in KS3 and GCSE), even if overall levels have.

Conclusions on trends in attainment in numeracy

- There is very little evidence before 1978, and none at age 13 after 1995.

At age 13

- There appears to have been a slight decline between 1964 and 1988, especially in arithmetic.
- There appears to have been no significant change between 1988 and 1995.

At age 14

- International surveys show no significant change between 1995 and 2003, but
- National data show a significant improvement between 1997 and 2007/09 (Yellis and KS3 maths), even though a part of the rise in KS3 maths between 1996 and 2000 may have been due to grade drift.

No satisfactory way of resolving most of this contradiction suggests itself, but

- There was a significant improvement between 2003 and 2007 (TIMSS).

At age 15/16

- There was a small improvement between 1978 and 1982.
- There was no significant change between 1982 and 1987.
- There was a substantial increase in the GCSE ‘pass rate’ between 1989 and 2005.

At age 16+

- Though there are four surveys spanning the period 1981–2003 they cannot be used to identify trends.
- However, they all show that substantial proportions of young people (16–19 or 16–24/25) have poor numeracy (below Entry level 3), of the order of 22%.
- A lifecourse trend (improvement into early middle age, then usually a plateau, then decline) will have lifted some people into adequate numeracy levels by their early to mid 20s, but many still have poor numeracy at all later ages.

Overall

- On average, number skills in England are poorer than in many other countries, especially industrialised ones, though other aspects of numeracy are better.

Criteria and standards

- There is no evidence to suggest that the criteria against which literacy and numeracy attainment is judged have been lowered. On the contrary, there is some evidence that the criteria have been maintained or even made more demanding.
- The criteria used to judge poor literacy focus on reading, and essentially define it as the ability to handle only simple texts and straightforward questions on them where no distracting information is present or nearby. Making inferences and understanding forms of indirect meaning (e.g. allusion, irony) are likely to be difficult or impossible.
- The criteria used to judge poor numeracy essentially define it as very basic competence in maths, mainly limited to arithmetical computations and some ability to comprehend and use other forms of mathematical information.
- While these are useful skills, they are clearly not enough to deal confidently with many of the literacy and numeracy challenges of contemporary life.
- That said, most young people in England do have functional skills, and those with the highest skills are up with the best in the world.
- Moreover, since criteria for judging both literacy and numeracy appear to have become more demanding in recent years, *steadily or gently rising and even quite flat graphs can be seen as a success story.*

Implications for policy

- Efforts to improve the initial teaching of literacy and numeracy to young children must continue.
- Children who fall behind in the early stages must be identified and given targeted catch-up programmes immediately.
- Family literacy and numeracy programmes can make a contribution to preventing early failure.
- Effective programmes should be maintained and not funded only in the short term.
- The search for effective ways of raising levels of functional literacy and numeracy should continue.
- Given that a light sample monitoring system seems to be being established at adult level there is a case for re-establishing one at school level.

Caveat

- All the findings are based on the assumption that experts in the field know what other people should be able to do. Little research has been done to establish what people

actually need to be able to do, though the next international survey of adult literacy and numeracy in 2011 will begin to remedy this.

1. Introduction

1.1 Context

How well can 13- to 19-year-olds in England read, write and use mathematics, now and in the past? These are the questions this project set out to investigate. It arose from the considerable interest within the (then) Department for Education and Skills in 2004–05 in the literacy and numeracy levels of this age-range, especially in the light of the results from the PISA (Programme for International Student Assessment) survey of 2000. A brief paper on literacy levels at age 15/16, 1948–2003 was delivered to the Department in December 2004 (Brooks, 2004); this project began in the autumn of 2005, and ended in March 2007. In the interim before publication, this report was updated in December 2007 and again in November–December 2009 in order to add data that had appeared since the project formally ended.

The topic is relevant to:

- the fact that half of the 860,000+ achievements under the *Skills for Life* Strategy to the end of academic year 2003/04 were provided by 16- to 18-year-olds
- the achievement of the government's *Skills for Life* targets for 2004, 2007 and 2010, in particular the achievement of the 2010 target two years early, by the summer of 2008
- the Leitch review of skills and the ambitious targets it set for adult literacy and numeracy levels to be achieved by 2020
- the government's subsequent setting of an ambitious interim target for 2011
- the high rate of drop-out from education at age 17
- progress towards the Government's targets for GCSE (age 16)
- the gradual introduction of diplomas following the Tomlinson proposals for reform of 14–19 education
- the Foster review of further education
- what employers can reasonably expect of young people entering the labour market
- the relationship between public examination pass rates and other evidence such as surveys and standardised tests
- international comparisons, and
- the questions of whether and, if so, in which direction, levels of attainment of 13- to 19-year-olds have moved over time.

1.2 Method

Data from existing national and international surveys and other sources were re-analysed to provide findings on the literacy and numeracy levels of 13- to 19-year-olds both now and over time. Some data on older age groups were also analysed for comparison. Concurrently, the criteria against which the levels were judged were analysed in order to give some sense of what 'functional' and 'less than functional' literacy and numeracy mean.

1.3 Structure of this report

The sources which were used (and some which were not) are described in Chapter 2, and the findings on the literacy and numeracy levels of 13- to 19-year-olds (and in some cases of older age groups) in Chapter 3. The criteria against which the levels were judged are presented and analysed in Chapter 4.

2. Sources of data

The questions the project addressed inherently required the use of quantitative data, which could have come from various sources:

- Self-report
- Studies of progress in adult literacy and numeracy
- Qualifications achieved in adult literacy and numeracy
- National monitoring surveys
- Lifetime cohort studies
- International surveys
- Repeated use of standardised tests
- Trends in national test results
- Trends in examination results
- Analyses of examination scripts
- Adult basic skills surveys

Examples of all of these were used, except the first three:

- For present purposes, self-report was considered inherently less reliable than performance on a test.
- There have been four national studies of progress in adult literacy, the last two of which also covered adult numeracy:
    - A study by the National Foundation for Educational Research in 1976–79 (Gorman, 1981; Gorman and Moss, 1979)
    - The Basic Skills Agency's *Progress in Adult Literacy* study in 1998–99 (Brooks *et al*., 2001)
    - A study by the National Institute of Economic and Social Research in 2002–07 (Metcalf *et al*., 2009)
    - The National Research and Development Centre for adult literacy and numeracy's Learner Study in 2003–07 (Brooks and Pilling, forthcoming; Rhys Warner *et al*., 2008; Vorhaus *et al*., 2009).

However, these studies by definition sampled only adults with poorer literacy and numeracy; since this means that the samples were not representative of the full population or of the full range of attainment levels the findings were not used.

- Similarly, the national Learning and Skills Council (LSC) and its predecessor, the Further Education Funding Council, have for many years collected data on qualifications in literacy and numeracy achieved by adult learners (Individual Student Records, ISRs, up to 2002/03; Individual Learner Records, ILRs, from 2003/04), and in recent years there have been national tests in basic skills and key skills. Again, however, the samples achieving these qualifications are not representative of either the population or the range of levels, and therefore the data were not used.

The sources that were used are now listed and briefly described; some that were not used in one area or another (in addition to those already ruled out for all areas) are also listed, and the reasons for not using them are given. Details are given first for reading, then for writing, and finally for numeracy. In most cases it is clear whether literacy results should be included under reading or writing since the assessment focused on one or the other. The key stage 3 (KS3) national tests in English at age 14 did have separate reading and writing papers, but results for them were not published separately. It was decided to include the results from these tests under writing because they are most akin to GCSE English results. The forms of evidence used in each area are listed in Table 2.1.

Table 2.1: Forms of evidence used for reading, writing and numeracy

	Used for:		
Form of evidence	**Reading**	**Writing**	**Numeracy**
National monitoring surveys	✓	✓	✓
Lifetime cohort studies	✓		
International surveys	✓		✓
Repeated use of standardised tests	✓		✓
Trends in national test results		✓	✓
Trends in examination results		✓	✓
Analyses of examination scripts		✓	
Adult basic skills surveys	✓		✓

None of the sources available for England is as comprehensive as the main source for France would be. Every year the French Ministry of Defence organises a 'jour de préparation' which all 17-year-olds must attend, young women as well as young men, and at which their literacy is tested. (This is a survival from the days of compulsory military service for young men.) In 2004 the results showed that 11% of French 17-year-olds struggled to understand a simple text, and 4.5% had extreme difficulty.

For caveats on the use of national test and examination results for monitoring trends over time see Green and Oates (2007) and other authors cited by them. They also advocate the re-establishment of a light sample monitoring system at school level (the last such survey in England took place in 1988) to provide evidence on trends that is not subject to text or examination pressures or attempts to achieve government targets.

For parallel data on attainments over time at primary level see the two analyses prepared for the Cambridge review of primary education in England directed by Robin Alexander: Tymms and Merrell (2007) for the national evidence, and Whetton, Ruddock and Twist (2007) for the international evidence. The latter provides a summary of arguments for and against the reliability of international comparative data, and arguments for the robustness of international data are elaborated in Whetton, Twist and Sainsbury (2007) – these arguments are also, of course, relevant to the age range considered in this report and should be taken into account when judging the results summarised here, particularly the numeracy results.

2.1 Reading

No information was available from trends in national test results (but see the note above about KS3 test results), trends in examination results or analyses of examination scripts.

2.1.1 National monitoring surveys

Between 1948 and 1988 there were three overlapping series of surveys which monitored the average and range of attainment in reading of pupils aged 15/16 (Year 11) in England:

- A series carried out in 1948/1952/1956/1961(twice)/1971 using the Watts-Vernon test
- A series carried out in 1955/1960/1971/1979 using the National Survey 6 test
- A series carried out annually from 1979 to 1983 and with a final survey in 1988 by the Assessment of Performance Unit Language (i.e. English) Monitoring Project.

(For most of the period these surveys also covered Wales and, in 1979–88, Northern Ireland. When they were available, separate figures for England were used for this report; when they were not, the combined figures for England and Wales were used. Given that over 90 per cent of the population of the two countries lives in England, removing the data for Wales would have made rather little difference.)

Watts-Vernon (WV) and National Survey 6 (NS6) surveys

For a summary of the results of all these surveys except the last see Start and Wells (1972). Pupils aged 15/16 (Year 11) were targeted in all these surveys, and in many later ones, as being, generally, in their last year of secondary education. The survey findings could therefore be seen as estimates of the effectiveness of 11 years of schooling. The surveys of 1948–71 were conducted either by HM Inspectorate or by the National Foundation for Educational Research (NFER).

The WV test was developed from 1938 by Dr Watts. He piloted it with London schoolchildren, but no records appear to have survived of any testing carried out. The Ministry of Education (1950) calibrated WV using 'accurate' pre-war norms of other tests, and by testing pupils in six secondary schools in London and the Home Counties chosen to be 'fairly representative'. The sample was 432 Year 10 pupils (14- to 15-year-olds). By modern standards this 'standardisation' was rudimentary, and caution therefore needs to be exercised in interpreting findings from this test. It had 35 items. Each was a multiple-choice sentence-completion item with five options. The first item was:

> You can buy stamps at a post (station, house, shop, man, office).

The NS6 test, devised in 1955, appears to have been standardised more rigorously (by NFER). It had 60 items. Each was again a multiple-choice sentence-completion item with five options. The first item was:

> Mary has one brother called (Ann / Susan / John / Hilda / Jane).

In the 1971 survey both tests were taken by all pupils in the sample, thus enabling a statistical link between the two series.

When first developed WV and NS6 were considered state-of-the-art, but by the time they were last used (in 1971 and 1979 respectively) had come to seem both out of date (for example, an item in NS6 used the word 'mannequin') and to under-represent the complex nature of reading and its uses, not to mention other aspects of language. However, for the period 1948–71 the results from these tests are the only national monitoring survey information available for any age group in the range 14 to 19 (and for any aspect of the curriculum) in England.

Assessment of Performance Unit (APU) surveys

For the results of these surveys see Gorman *et al*. (1988, 1991) and other reports cited there. From 1979 to 1988, the APU Language Monitoring Project based at NFER carried out six surveys of the reading (and writing – for this see below) attainment of pupils aged 15/16 (Year 11) in England, Wales and Northern Ireland. The reading tests were all specially devised, and attempted to present authentic literacy tasks. They typically contained one or more passages in a reading booklet, with questions in an answer booklet, and inaugurated a

new era in reading tests. Some items were multiple-choice but most were 'supply' type, that is, pupils had to write their own short answers. Roughly half the booklets contained fiction and half factual material. Several tests were used in more than one year, in particular in 1979, 1983 and 1988, thus enabling statistical equating across years. Similarly, NS6, on its final outing, was taken by a sub-sample of pupils in the 1979 survey, thus enabling a statistical link between the APU surveys and both previous series.

2.1.2 Lifetime cohort studies

There are three possibly relevant studies of this type ongoing in Britain:

- The Medical Research Council National Survey of Health and Development (NSHD), based on all those born in Britain in a week in March 1946. This cohort took the WV reading test at age 15 in 1961 (Douglas *et* al., 1968);
- The National Child Development Study (NCDS), based on all those born in Britain in a week in March 1958. This cohort took a reading test at age 16 in 1974 (Fogelman, 1976, 1983);
- The British Cohort Study 1970 (BCS70), based on all those born in Britain in a week in April 1970; this cohort took a reading test at age 16 in 1986 in what was called the 'Youthscan' study (www.cls.ioe.ac.uk/studies.asp?section=00010002000200070001).

Unfortunately, the results of NCDS and BCS70 could not be used. NCDS used a variant of the WV test which cannot be statistically equated with it, and the BCS70 'Youthscan' results were unreliable and incomplete due to a teachers strike. However, the NSHD result was used, because the test was the Watts-Vernon – the result has therefore been included with those from the series of national monitoring surveys which used this test (and the heading 'lifetime cohort studies' does not re-appear in Chapter 3).

2.1.3 International surveys

There have been six such surveys that could be relevant here:

- a survey in 1960 of the reading attainment of pupils aged 13/14 in 12 countries, carried out by the International Association for the Evaluation of Educational Achievement (IEA) (Foshay *et al*., 1962)
- a survey in 1971 of the reading attainment of pupils aged 13/14 and 15/16 in 15 countries, also carried out by IEA (Thorndike, 1973)
- four surveys in 2000, 2003, 2006 and 2009 of the reading attainment of pupils aged 15 called the Programme for International Student Assessment (PISA), sponsored by the Organisation for Economic Cooperation and Development, and carried out by an international consortium (Gill *et al*., 2002; OECD, 2005, 2007; Bradshaw *et al*., 2007).

The UK's PISA 2003 results were declared unreliable by OECD because of an inadequate sample, and were therefore not used, and the 2009 results were not available in time to be included. The results of the other four surveys just listed were used.

The IEA surveys, 1960 and 1971

Details of the 1960 survey are difficult to obtain; however, it is known that all the test items were of four-option multiple-choice type, and that none of the items were repeated in the 1971 survey.

In the 1971 survey all the test items were again of four-option multiple-choice type. For 13/14-year-olds there were two tests containing eight passages and 52 items; for 15/16-year-olds, also two tests containing eight passages, but 54 items.
England and Wales took part jointly in both IEA surveys.

PISA 2000, 2006
England, Northern Ireland and Scotland took part in 2000 and all four countries of the UK in 2006, and the surveys were designed to enable calculation of separate results for England. PISA 2000 was carried out in 32 countries and the main focus was on reading literacy, with sub-samples of pupils also taking tests of mathematical literacy and scientific literacy. PISA 2006 was carried out in 57 countries and the main focus was on scientific literacy, with sub-samples of pupils also taking tests of mathematical literacy and reading literacy. The reading tests were all of the 'new era' type pioneered by NFER in the APU surveys. The definition of reading set out for the 2000 survey and also used later was:

> the ability to understand, use, and reflect on written texts to participate effectively in life (PISA does not seek to measure the extent to which students are fluent readers or their ability in spelling or word recognition). Reading literacy was assessed using a series of texts, and a number of tasks set for students on each text. Just over half of the tasks were in the form of multiple-choice questions; the rest required students to construct their own answers…
>
> In each domain, a student's score is expressed as a number of points on a scale, and shows the highest difficulty of task that the student is likely to be able to complete. The scales are constructed so that the average score for students from all OECD countries participating in PISA 2000 is 500 and its standard deviation is 100 – that is, about two-thirds of students internationally score between 400 and 600. Each country contributes equally to this average irrespective of its size…
>
> In [the 2000 survey in] England, co-operation was obtained from 4,120 young people born in 1984, in a representative sample of 155 schools throughout the country: 59% of those initially selected and 82% after including replacement schools for those that refused to take part. Thirty-five students who were born in 1984 were randomly selected to take part from each school. In March 2000 each student took a written assessment lasting two hours, which was administered in his or her own school using standardised methodology and in test conditions… The response rate among students who were selected to take part and were eligible for the survey was 81%. (Gill *et al*., 2002)

In PISA 2006, because the main focus was science, only a sub-sample of pupils took the reading tests. However, in each country their results were weighted to estimate the performance of the full sample. In England, 169 schools and 4935 pupils participated in PISA 2006.

2.1.4 Repeated use of standardised tests

One possible source was not used. In 2004 the children aged 6 to 16 of half the BCS70 cohort members were given a word recognition test, and the average scores of the different ages of children involved were compared with those of the original standardisation sample (Bynner and Parsons, 2006: 87–89). Just over 400 children aged 13–16 were involved (number estimated from Bynner and Parsons, 2006: 80, Figure 6.2) – these were children born in 1987–91, when their mothers were aged 17–21 – and the children's average scores were slightly lower than those of the standardisation sample (Bynner and Parsons, 2006: 89, Figure 6.10). However, this cannot be taken as a reliable over-time result because the samples at each age were rather small and because, as Bynner and Parsons (2006: 89) pointed out, they 'reflect the fact that older children in our sample are a product of young

motherhood, and young mothers are more likely to have less education and associated qualifications – hence the reduction in the ... children's test scores.'

Yellis

Therefore the only, but very useful, source of standardised test data over time was the Yellis vocabulary test devised by the CEM Centre at Durham University (www.cemcentre.org). A written vocabulary test can be considered as a measure of one part of reading ability. The Yellis test has been taken by nationally representative samples of year 10 pupils (age 14) annually since 1997; data for 1993–2009 were made available to this project (for details see the Acknowledgments). All items are four-option multiple-choice, and the test is computer-administered and adaptive.

2.1.5 Adult literacy surveys

Between 1972 and 2003 there were 10 national surveys and one international survey which assessed the reading attainment of nationally representative samples of adults in England, but only three included young people under 20:

- the International Adult Literacy Survey (IALS), 1996 (Carey *et al*., 1997)
- the Basic Skills Agency's survey of need, 1996–97 (Basic Skills Agency, 1997a)
- the *Skills for Life* survey of need (SfL), 2002–03 (Williams *et al*., 2003),

and only two of these reported separate findings for an age-group relevant to this project, the exception being the Basic Skills Agency's survey.

IALS, 1996

This was the first of only two international adult literacy surveys which have so far been mounted, and the only one in which the UK took part. (A policy decision was made in 2000 not to participate in the Adult Literacy and Lifeskills (ALL) survey of 2002–03; England instead carried out its own survey – see below. However, all four countries of the UK are scheduled to take part in the next such survey, the Programme of International Assessment of Adult Competencies (PIAAC), in 2011.)

Like PISA, IALS was sponsored by OECD and carried out by an international consortium. It took place in 23 countries in three sweeps between 1994 and 1998. The UK took part in the second sweep in 1996.

The definition of literacy used in IALS was:

> using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.

Literacy was examined across three domains:

- Prose literacy: the knowledge and skills required to understand and use information from texts such as passages of fiction and newspaper articles;
- Document literacy: the knowledge and skills required to locate and use information contained in various formats such as timetables, graphs, charts, and forms; and
- Quantitative literacy: the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials such as advertisements, or working out the interest required to achieve a desired return on an investment.

For this report, the prose and document domains were taken to represent literacy, and results for them are reported in Section 3.1.4 under reading. 'Quantitative literacy' was taken as a proxy for numeracy – results for this are given in Section 3.3.2 under numeracy. The tests were all specially devised, and most were based on pieces of authentic text (that is, authentic in the source language; all were subject to translation). All responses were of supply type.

In Britain, the survey sampled a total of 3811 adults aged between 16 and 65. Of these, 372 respondents were from England and in the age group 16 to 25 – no further differentiation within that group was available. Weighting was used to allow for sample design and adjustments were made for any non-response bias.

The SfL survey, 2002–03
This survey treated literacy as one domain rather than two. The survey covered only England, but the age-range 16–64. The data were broken down by age-groups, of which one was 16–19. The test items were all specially devised, by CDELL, the Centre for Developing and Evaluating Lifelong Learning at the University of Nottingham. A pilot version of the instrument was heavily criticised by Brooks *et al*. (2005) for the inauthentic nature of some of the items. All responses were multiple-choice. This was one of the few surveys covered in this report to be computer-administered. It was to an extent adaptive: after a short and very simple initial screening section, test-takers were routed to three further blocks of items intended to be appropriate to their reading level as estimated from their percentage score on the previous block. A new *Skills for Life* survey scheduled for 2010 is intended to provide the first over-time evidence on reading performance at adult level. This and the decision to take part in PIAAC (see above) seem to represent the establishment at adult level of a light sample monitoring system of the sort that was abolished at school level in 1988.

2.2 Writing

No information was available from lifetime cohort studies, international surveys or repeated use of standardised tests.

There has been one international study specifically of written composition at school level, of pupils aged 13/14 in 1983. England and Wales took part (Gubb *et al*., 1987), but the survey was deliberately designed and reported to prevent the calculation of average national scores and therefore the creation of a 'league table' (Gorman *et al*., 1988). Even if that had been possible, there have been no further attempts to carry out such a survey, so that it would not be possible to comment on trends over time.

2.2.1 National monitoring surveys

The only surveys which covered writing which provided usable results for this report were the APU surveys of 1979–83 and 1988 (see again Gorman *et al*., 1988, 1991). As with reading, the writing tests were all specially devised, and attempted to present authentic literacy tasks. Several tests were used in more than one year, in particular in 1979, 1983 and 1988, thus enabling statistical equating across years. The principal results from the APU surveys were measures of attainment in written composition, but a small amount of information was also available on spelling, based on re-analyses of scripts from the 1980 and 1983 surveys.

Also, several sources of data which were not available for reading were useful for writing, as described under the next three headings.

Use was also made of a detailed study of the comparability of marking standards in 1996 and 2001 (Massey *et al*., 2002).

2.2.2 National test results

Overall national results for the KS3 (age 14) English tests were obtained for 1995 to 2008 (www.standards.dfes.gov.uk/performance/ap/?version=1, www.standards.dfes.gov.uk/performance/archive/ and www.dcsf.gov.uk/rsgateway/DB/SFR/s000847/index.shtml), that is, for the whole period when these tests were in operation. The results for 2008 were not used here because the marking problems which occurred in that year (and which led to these tests being discontinued after 2008) made them insufficiently comparable to earlier results. The form of these tests varied slightly over this period, but they always contained a reading test, focusing largely on a Shakespeare play, and a paper testing writing; there was sometimes also a spelling test. Only aggregated results were published. The test was therefore an assessment of response to literature as well as of literacy as such; nevertheless it has been counted here as a literacy test.

Use was also made of a detailed study of the comparability of marking standards in 1996 and 2001 (Massey *et al*., 2002).

2.2.3 National examination results

The relevant data under this heading are results for GCSE English (Language), which for present purposes has been counted, like the KS3 English tests, as a literacy test, even though an assessment of oracy (speaking and listening) skills has contributed to candidates' overall GCSE results since 1994 (in 1989–93 oracy was separately certificated). Although three different examination boards now offer this exam in England (and earlier in the period covered by the data gathered there were one or two more), the national data are aggregated across boards. Also, although people of other ages may take this examination, the published national data relate only to pupils aged 15/16 and were therefore usable for this project.

Each cohort of pupils in these years numbered between 520,000 and 660,000. It should be noted that two slightly different definitions of pupils eligible to be counted in the '15-year-old' age-group were used during this period: for 1989 to 2004, those aged 15 on 31 August at the start of the academic year; for 2005 to 2009, those at the end of KS4. This does not appear to have caused any significant change in the overall results.

Information was obtained for the period 1989–2009, that is, for the entire history of GCSE from its inception to the present (sources: www.dcsf.gov.uk/rsgateway/DB/SFR/ for various years).

Use was also made of a study of examination standards in GCSE English between 1989 and 1995 (SCAA, 1996).

2.2.4 Analyses of examination scripts

Public attention to GCSE results tends to focus on the 'pass rate' (the number of candidates achieving at least grade C, the (intended) equivalent of the pre-1989 O-Level pass). One way to sidestep this and provide a more nuanced account would be to carry out detailed analyses of differences over time in the quality of the writing in examinations in English. The purpose would be to determine whether later candidates, compared to earlier ones, showed better or

worse or similar control of (for example) spelling, punctuation, paragraphing, overall text structure, and the content of what they wrote.

There have been three such studies:

- The first was carried out by the University of Cambridge Local Examinations Syndicate, now part of the OCR Board, in 1995, and compared aspects of the writing of candidates taking examinations in English at 16+ in 1980, 1993 and 1994 (Massey and Elliott, 1996)
- QCA (2004) studied standards of performance in GCSE English in the period 1999–2002
- the Cambridge study was later extended to 2004 (Massey *et al*., 2005).

All three were obtained and used.

2.3 Numeracy

For numeracy, information was available under all the headings listed under reading and/or writing except analyses of examination scripts (there were none) and lifetime cohort studies.

For brief details of the three ongoing lifetime cohort studies see under reading, Section 2.1.2. There were no numeracy results from them that could be used for this project. The NCDS (1958) and BCS70 (1970) studies did not administer a maths test at any relevant age. The NSHD (1946) cohort did take a maths test at age 15, but the test had been specially devised and could not be calibrated with any others analysed here. Also, the distribution of the raw scores was severely skewed and indicated that the test was too difficult.

2.3.1 National monitoring surveys

There were no maths surveys comparable to the reading surveys of the period up to 1971. However, the APU Mathematics Monitoring Project, also based at NFER, carried out surveys at age 15/16 (Year 11) in England, Wales and Northern Ireland annually from 1978–82, with a final survey in 1987 (APU, 1988; Foxman *et al*., 1990a, b).

The tests were all specially devised. The modes of assessment were written and practical, and the tests covered number, measures, algebra, geometry, and probability and statistics. Again, items were repeated systematically in order to enable calculation of trends over time.

2.3.2 International surveys

For numeracy, this is the major source of information, because there have been 10 relevant studies at school level (and two at adult level with data for a relevant age-group – see Section 2.3.6):

- First International Mathematics Survey (FIMS), 1964 – age 13
- Second International Mathematics Survey (SIMS), 1981 – age 13
- International Assessment of Educational Progress 1 (IAEP1), 1988 – age 13
- International Assessment of Educational Progress 2 (IAEP2), 1991 – age 13
- The Third International Mathematics and Science Study (TIMSS), 1995 – ages 13 and 14 (separately)

- The Third International Mathematics and Science Study-Repeat (TIMSS(R)), 1999 – age 14
- Trends in International Mathematics and Science Study (TIMSS 2003), 2003 – age 14
- Trends in International Mathematics and Science Study (TIMSS 2007), 2007 – age 14
- Programme for International Student Assessment (PISA), 2000 – age 15
- Programme for International Student Assessment (PISA), 2006 – age 15

Results from all of these were gathered and used; brief details of each survey follow.

FIMS, 1964
England and Wales (together) along with eleven other countries (including Scotland) participated in the First International Mathematics Survey, which surveyed 13-year-olds (attending school in the (school) year where the majority of students had attained the age of 13 by the middle of the school year). Those in the final year of (non-compulsory) secondary education were also surveyed; in Britain this meant what is now called Year 13, but since the sample was by definition not representative of all young people of that age the results have not been used in this report. The survey was conducted by IEA. Categories of mathematics tested were new maths (e.g. groups, matrices), basic arithmetic, advanced arithmetic, algebra and geometry (Husén, 1967).

SIMS, 1981
Twenty education systems, including England and Wales (together) participated in the Second International Mathematics Survey, seventeen years after FIMS. The main sample consisted of 13-year-olds. (Year 13 students were again tested, but their results are not reported here.) The testing covered arithmetic, algebra, geometry, descriptive statistics and measurement (Robitaille and Garden, 1989).

IAEP1, 1988
IAEP1 assessed the mathematics achievement of 13-year-old students (born 1 January to 31 December 1974) in six countries. England, Scotland and Wales all contributed to the UK sample. Sixty-three questions were asked in the test (Lapointe *et al*., 1989).

IAEP2, 1991
IAEP2 assessed the mathematics skills of samples of 13-year-old students from 20 countries. 75 questions were set (Lapointe *et al*., 1992).

TIMSS 1995
This study was conducted at five grade levels (including the seventh and eighth grades, UK Years 8 and 9) in more than 40 countries. Students were tested, and extensive information about the teaching and learning of mathematics was collected from students, teachers, and school principals. Altogether, TIMSS tested and gathered contextual data for more than half a million students and administered questionnaires to thousands of teachers and school principals (http://isc.bc.edu/timss1995.html).

TIMSS(R), 1999
This study was conducted by the International Study Center at Boston College, Massachusetts and included 38 countries. It measured the mathematics achievement of eighth-grade students (ages 13 and 14 years) and collected extensive information from students, teachers, and school principals about mathematics curricula, instruction, home contexts, and school characteristics and policies. Of the 38 participating countries, 26 had also participated in the 1995 TIMSS assessment, which enabled those countries to measure trends in their children's mathematics achievement (http://isc.bc.edu/timss1999.html).

TIMSS 2003
This study collected educational achievement data at the eighth grade (Year 9) to provide information about trends in performance over time, together with extensive background information to address concerns about the quantity, quality, and content of instruction. Approximately 50 countries from all over the world participated (http://isc.bc.edu/timss2003.html).

TIMSS 2007
This is the most recent study in this ongoing series (the next is planned for 2011). Again, data were collected on educational achievement at the eighth grade (Year 9), together with extensive background information. A total of 48 countries from all over the world participated (http://nces.ed.gov/timss/table07_1.asp#f4).

PISA 2000, 2006
For details of the samples see above under reading, Section 2.1.3. The basic definition of mathematical literacy for the 2000 survey and later was:

> the ability to formulate and solve mathematical problems in situations encountered in life

and a more elaborate definition was also given:

> Mathematical literacy is defined as the capacity to identify, understand and engage in mathematics and to make well-founded judgements about the role that mathematics plays in an individual's current and future private life, occupational life, social life with peers and relatives, and life as a constructive, concerned and reflective citizen. As with reading, mathematical literacy indicates the ability to put mathematical knowledge and skills to use, rather than just mastering them within a school curriculum. Mathematical literacy was assessed through a combination of question types. As with reading literacy, there were a number of texts in which a situation or problem was explained, and then more than one question or task based on the text was presented to students. Various combinations of diagrams and written information were used. Some questions were multiple-choice, but open-ended items were used for assessing higher-order mathematical processes. PISA measures students' capacity to analyse, reason, and communicate ideas effectively by posing, formulating, and solving mathematical problems in a variety of contexts. (Gill *et al*., 2002; see also OECD, 2001)

Both in 2000 and in 2006 mathematics was a minor domain and only sub-samples of pupils took the maths tests. However, their results were weighted to estimate the performance of the full sample. (The mathematics results of PISA 2003, in which maths was the major domain, could not be used for the reason stated under reading, Section 2.1.3.)

2.3.3 Repeated use of standardised tests

Again (cf. under reading, Section 2.1.4), one possible source was not used. In 2004 the children of BCS70 cohort members were also given a number skills test, and the average scores of the different ages of children involved were again compared with those of the original standardisation sample (Bynner and Parsons, 2006: 91–94). At ages 13 and 14, the children's average scores were about the same as those of the standardisation sample, but at ages 15 and 16 they were slightly lower (Bynner and Parsons, 2006: 94, Figure 6.15). However, these cannot be taken as reliable over-time results either, for the same reasons as given under reading.

Yellis
Again, therefore, the only, but very useful, source of large-scale data was the Yellis maths test devised by the CEM Centre in Durham (http://www.cemcentre.org). This test has also

been taken by nationally representative samples of year 10 pupils (age 14) annually since 1997; data for 1993–2009 were made available to this project (for details see the Acknowledgments). All items are four-option multiple-choice, and the test is computer-administered and adaptive.

2.3.4 National test results

Overall national results for the KS3 (age 14) maths tests were obtained for 1995 to 2008 (www.standards.dfes.gov.uk/performance/ap/?version=1, www.standards.dfes.gov.uk/performance/archive/ and www.dcsf.gov.uk/rsgateway/DB/SFR/s000847/index.shtml), that is, for the whole period when these tests were in operation. The results for 2008 were not used here because the marking problems which occurred in that year (and which led to these tests being discontinued after 2008) made them insufficiently comparable to earlier results.

Use was also made of a detailed study of the comparability of marking standards in 1996 and 2000 (Massey *et al*., 2002).

2.3.5 National examination results

The relevant data under this heading are results for GCSE mathematics, which for present purposes has been counted as a numeracy test. As with GCSE English, although people of other ages may take this examination, the published national data relate only to pupils aged 15/16 (see Section 2.2.3 for a slight change in 2005 concerning which pupils were counted as 'aged 15') and are aggregated across exam boards, and were therefore usable for this project. Information was obtained for the period 1989–2009, that is, for the entire history of GCSE from its inception to the present (sources: www.dcsf.gov.uk/rsgateway/DB/SFR/ for various years).

2.3.6 Adult numeracy surveys

Between 1981 and 2003 there were seven national and two international surveys which assessed the numeracy attainment of adults in England, of which four included young people under 20 and gave separate results for a relevant age-group:

- the ACACE (Advisory Committee on Adult Continuing Education, later the National Institute of Adult Continuing Education, NIACE) survey, 1981
- IALS, 1996
- the International Numeracy Survey, 1996
- the *Skills for Life* (*SfL*) survey of need, 2002–03.

ACACE, 1981
This survey was carried out by Social Surveys (Gallup Poll) Ltd for ACACE. It attempted to measure the numeracy attainment of a nationally representative sample of adults selected according to ITV regions (ACACE, 1982; Sewell, 1981; Social Surveys (Gallup Poll) Ltd, 1981). However, the instrument used was clearly inadequate – just 11 very simple items.

IALS, 1996
For main details see under reading, Section 2.1.5. The 'quantitative literacy' data from IALS were treated as a proxy for numeracy for this report.

International Numeracy Survey, 1996

This study involved testing people aged 16–60 with 12 numeracy tasks in 7 industrialised countries: Australia, Denmark, France, Japan, Netherlands, Sweden, and the UK. Just under 6,000 people were involved in the study. They were representative of the populations of the participating countries involved. The UK section was carried out by Opinion Research Business on behalf of the Basic Skills Agency (Basic Skills Agency, 1997b).

In the UK the number of people who refused to tackle the tasks at all was higher than in any of the other countries. This may have been, however, for a variety of reasons unconnected with their mathematical skill. The instrument used has been widely criticised for its banality.

The SfL survey, 2002–03

For main details again see under reading, Section 2.1.5. A pilot version of the instrument was mildly criticised by Brooks *et al*. (2005). Like the literacy test, the numeracy test was to an extent adaptive: after a short and very simple initial screening section, test-takers were routed to three further blocks of items intended to be appropriate to their numeracy level as estimated from their percentage score on the previous block. The areas of maths covered were number, estimation, measures, data-handling, and percentages.

3. Findings

As in Section 2, information is presented first for reading, then writing, then numeracy. However, this section also has a final short subsection on correlations in attainment between literacy and numeracy. There is a good deal of information for reading and numeracy, much less for writing.

3.1 Reading

3.1.1 National monitoring surveys, 1948–88

This is the longest and most comprehensive section of the data analysed in this report. The data for the WV, NS6 and APU series are shown in Tables 3.1–3 respectively. The WV and APU results are for England only; those for NS6 sometimes include Wales. The WV and NS6 results are average raw scores; those for the APU series are average standardised scores.

Table 3.1: Results for Watts-Vernon reading test, age 15, 1948–71, England only

Year	Mean score	Change since previous survey	School types *	Sample size	Sample details
1948	20.79	n/a	Maintained and direct grant grammar	3314	Judgment sample
1952	21.25	+ 0.46	Maintained and direct grant grammar	(not known)	Stratified random
1956	21.71	+ 0.46	Maintained and direct grant grammar	1741	Stratified random
1961a	24.10**	+ 2.39	Secondary moderns and comprehensives only	18156	Stratified random
1961b	23.53	(from 1956) +1.82	In principle, all, including independent schools	c.4400	Sub-sample of NSHD life-time cohort
1971	23.46	(from 1961a) –0.64 (from 1961b) –0.07	Maintained and direct grant grammar	1844	Stratified random

* For this period, maintained schools were grammar, secondary technical, secondary modern and comprehensive schools, whether operated by local education authorities (LEAs) or as 'voluntary controlled' or 'voluntary aided' schools; 'voluntary' schools were mostly church schools. 'Direct grant' grammar schools took a minimum of 25%, and in some cases up to 100%, LEA pupils (those who had passed the '11+'), and received funding for those pupils direct from the Ministry of Education/ Department for Education and Science, and not from LEAs.
** Estimate for total maintained school population based on the assumption that other schools made the same advance as secondary modern schools between 1956 and 1961.
Notes: (1) The sample size for 1952 could not be retrieved from the literature, and that for 1961b is stated only approximately. (2) The rise from 1956 to 1961a is known from the literature to have been statistically significant, and the fall from 1961a to 1971 to have been statistically non-significant. The statistical significances of other changes over time were not stated in the literature, and could not be calculated because the necessary information (especially the standard deviations) was missing.

Table 3.2: Results for NS6 reading test, age 15, 1955–79

Year	Mean score	Change since previous survey	School types *	Sample size	Sample details
1955	42.18**	n/a	Maintained	(not known)	Stratified random
1960	44.57	+2.39	Maintained	(not known)	Stratified random
1971	44.65*	+0.08	Maintained	2194	Random, but very low response rate because of a postal strike
1979	44.03**	–0.62**	Maintained	973	Based on sub-sample in first APU survey

* For the definition of maintained schools see footnote to Table 3.1. In 1971 direct grant grammar schools were also included, to allow direct comparison with the simultaneous WV survey. However, the scores shown are for maintained schools only; the scores including direct grant grammars were almost exactly the same (mean = 44.96).
** Mean and change scores for England AND Wales. Given that over 90% of the population of the two countries lives in England, the figures are probably valid for England alone.
Notes: (1) The sample sizes for 1955 and 1960 could not be retrieved from the literature. (2) The rise from 1955 to 1960 is known from the literature to have been statistically significant, and the changes from 1960 to 1971 and from 1971 to 1979 to have been statistically non-significant.

At this distance in time perhaps the most interesting fact about the 1971 results is the reaction to the small drop in the mean score on the WV test relative to the 1961a survey. Despite the difference being statistically non-significant (and the non-significant rise on NS6 and the obviously non-significant fall from the 1961b survey) a considerable moral panic over 'falling standards' ensued, fuelled to an extent by the presentation of the results in Start and Wells (1972) in a graph with a grossly truncated and therefore misleading y-axis scale. However, the positive outcome was the setting up of the Bullock Committee, whose report in turn led to the setting up of the APU.

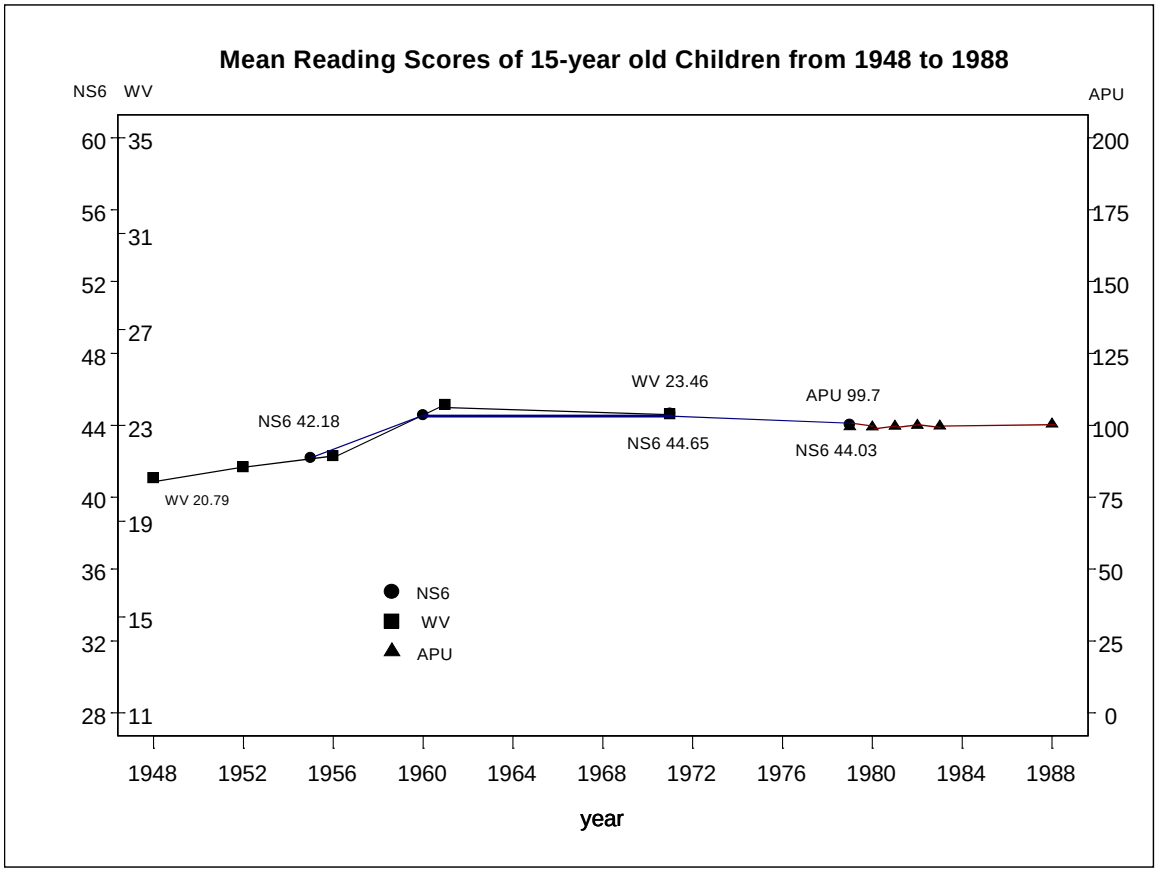
Table 3.3: Results for APU reading tests, age 15, 1979–88, England only, pupils with English as 1st language only

Year	Mean score
1979	99.7
1980	99.6
1981	99.8
1982	100.1
1983	99.9
1988	100.6

None of the differences were statistically significant.

It proved possible to display the data from all three series of surveys on one graph (Figure 3.1). This is because most of the pupils in the 1971 survey took both the WV and the NS6 test, and a sub-sample of those in the 1979 survey took both NS6 and a subset of the APU tests. It was assumed that the average scores for WV and NS6 in 1971 could be treated as equivalent and made to coincide, despite their different score scales; and a parallel assumption was made about the 1979 results for the NS6 and APU tests.

Figure 3.1: Average reading scores of 15-year-old children in England, 1948–88



Scores below the baseline values of 11 and 28 for WV and NS6 respectively (on the left y-axis) could not be calibrated with the APU scale (on the right). Very few pupils scored below those marks, and in any case a score of less than 11 on the WV was described (in earlier and less PC days) as indicating 'illiteracy'.

Figure 3.1 shows a general increase in attainment from 1948 until 1960/61 and then effectively no change for the rest of the period examined. It is widely thought that there was a decline in standards due to the Second World War and that this took some years to rectify afterwards; see, for example, Start and Wells (1972) and Hurry (1999). The virtually flat line from 1960 to 1988 is striking.

3.1.2 International surveys, 1960–2006

The IEA surveys, 1960 and 1971

Given that national surveys were also going on at the time, 1960–61 and 1971 were busy years for reading surveys in schools.

Because different tests were used on the two occasions, the international survey results do not permit calculation of a trend over time. However, some findings of interest emerge from the ranking data, and from the distribution of scores. Though there were problems of sampling in some countries, the 1960 study placed the average scores of the age 13/14 (Year 9) pupils in the 12 participating countries in the following order: Yugoslavia, Scotland, Finland, England and Wales, United States, Switzerland, West Germany, Sweden, France, Israel, Belgium, Poland (Foshay, 1962). The 1971 study produced the following rank orders for the two relevant ages (Thorndike, 1973):

Age 13/14 (Year 9)	Age 15/16 (Year 11)

New Zealand	New Zealand
Italy	Scotland
United States	England & Wales
Belgium (French-speaking)	Netherlands
Finland	Finland
Scotland	Belgium (French-speaking)
Sweden	Sweden
Hungary	Israel
England & Wales	Belgium (Flemish-speaking)
Netherlands	Italy
Belgium (Flemish-speaking)	Hungary
Israel	United States
Chile	Chile
Iran	Iran
India	India

On its own, the position of England and Wales in the 1971 age 13/14 list could have been seen as slippage from the 1960 position, but the higher placing at age 15/16 suggests that it was a cohort effect.

Inspection of the distribution of scores reveals another point of interest. Even in the 1960 study it was noted that England and Wales had 'by far the largest dispersion of test scores' (Pidgeon, in Foshay, 1962, p.59), with Scotland close behind. In the 1971 study, the standard deviation for England and Wales was among the largest at age 13/14 (though not at age 15/16). Pidgeon had a theory on this in 1962:

> The general aim of the grade class teacher may tend to result in a relatively smaller dispersion. Perhaps exerting a greater influence, however, is the belief a teacher may have that innate ability is of paramount importance in determining the level of attainment to be expected from a child. Streaming by ability, which is viewed as an administrative device resulting from the acceptance of this belief, will merely tend to enhance its effects. When all these factors act in the same direction the effect will clearly be greatest and this is what happens in England. Here, it is claimed, the aims and, more especially, the beliefs of most teachers and educational administrators lead them to expect wide differences in performance, and this is what is therefore achieved. Where, on the other hand, the grade placement system operates and especially where, within such a system, teachers do not attempt to measure innate ability and therefore do not expect their pupils' attainments to be matched to it, then the dispersion of achievement will be much less. (Pidgeon, in Foshay, 1962, pp.61–62)

In other words, low expectations of some children contribute to their low achievement; and in this respect very little seems to have changed in British education. If this expectation effect is true, it would seem to imply a 'devil take the hindmost' attitude, possibly rooted in Britain's class structure, and a need to counteract this by concentrating special help on those most in need.

PISA 2000, 2006

The results are shown in Figures 3.2 and 3.3.

Figure 3.2: Reading attainment of 15-year-olds in PISA 2000

Source: Gill *et al*. (2002)

Figure 3.3: Reading attainment of 15-year-olds in PISA 2006

	Mean score		significance
	Mean	S.E.	
Korea	556	3.8	▲
Finland*	547	2.1	▲
Hong Kong-China	536	2.4	▲
Canada	527	2.4	▲
New Zealand	521	3.0	▲
Republic of Ireland*	517	3.5	▲
Australia	513	2.1	▲
Liechtenstein	510	3.9	*NS*
Poland*	508	2.8	*NS*
Sweden*	507	3.4	*NS*
Netherlands*	507	2.9	*NS*
Belgium*	501	3.0	*NS*
*Estonia**	501	2.9	*NS*
Switzerland	499	3.1	*NS*
Japan	498	3.6	*NS*
Chinese Taipei	496	3.4	*NS*
England	496	2.7	
United Kingdom*	495	2.3	
Germany*	495	4.4	*NS*
Denmark*	494	3.2	*NS*
*Slovenia**	494	1.0	*NS*
Macao-China	492	1.1	*NS*
OECD average[1]	**492**	**0.6**	*NS*
Austria*	490	4.1	*NS*
France*	488	4.1	*NS*
Iceland	484	1.9	▼
Norway	484	3.2	*NS*
Czech Republic*	483	4.2	*NS*
Hungary*	482	3.3	*NS*
*Latvia**	479	3.7	▼
Luxembourg*	479	1.3	▼
Croatia	477	2.8	▼
Portugal*	472	3.6	▼
*Lithuania**	470	3.0	▼
Italy*	469	2.4	▼
Slovak Republic*	466	3.1	▼
Spain*	461	2.2	▼
Greece*	460	4.0	▼
Turkey	447	4.2	▼
Chile	442	5.0	▼
Russian Federation	440	4.3	▼
Israel	439	4.6	▼
Mexico	410	3.1	▼
*Bulgaria**	402	6.9	▼
*Romania**	396	4.7	▼

key
▲ significantly higher
NS no significant difference
▼ significantly lower

OECD countries (not italicised)
Countries not in OECD (italicised)
*EU countries

13 countries with scores below 430 omitted
Multiple comparison P value = 0.045%
[1] Simple comparison P-value = 5%

Source: Bradshaw *et al*. (2007), p.81

The authors of the 2000 report for England commented:

> The proficiency in reading … literacy of 15-year-olds in England compares well with that of young people of the same age in other countries. Students in England scored an average of 523 points on the reading literacy scale, significantly higher than students in OECD countries as a whole, where the mean score was set at 500. English students were at a similar level of achievement as those in, for example, Australia, Japan and Sweden. Countries with lower average achievement than England included France and Germany, and in only two countries, Finland and Canada, did 15-year-olds do significantly better than in England. (Gill *et al*., 2002)

No statistical comparison between the 2000 and 2006 mean scores for the UK was given in either the 2006 international report or the 2006 national report for England, apparently because the OECD (2007, p.325, note 4) retrospectively declared the 2000 result unreliable. It is therefore not possible to state whether the fall of 27 points in England's mean score between the two surveys was or was not statistically significant.

However, it may be possible to address the question of a trend over time in a rough-and-ready fashion by considering rank orders. At age 15/16 England and Wales were 3rd out of 15 countries participating in 1971, the UK (and therefore probably England because the 'United Kingdom' sample was largely from England) was 7th out of 27 in 2000 – see Figure 3.2, and England was 17th out of 57 in 2006 – see Figure 3.3. The 2006 international report (OECD, 2007, Figure 6.8b, p.298) estimated that the UK's rank was between 11th and 16th of the 30 OECD countries (all OECD countries took part), and between 14th and 22nd among all 57 participating countries. Given the increased number of countries participating in 2000 and 2006, these placings are probably roughly equivalent, and therefore consistent with the level graph from 1970 onwards in the national data.

3.1.3 Repeated use of standardised tests, 1997–2009

The results for the Yellis vocabulary test taken by pupils in Year 10 (age 14) 1993–2009 are shown in Table 3.4. Details of the test itself are at www.cemcentre.org/RenderPage.asp?LinkID=11518000.

Each cohort of pupils in these years numbered close to 600,000. From 1997 onwards, therefore, the test was taken by over 20% of each cohort. If these samples are taken as nationally representative (and the earlier ones are disregarded), the average score rose gently from 49% to 53% between 1997 and 2004, and has increased by less than 1 percentage point since then. And the increase over the whole period from 1997, while no doubt worthwhile, was not dramatic: about 5 percentage points.

Table 3.4: Average Yellis vocabulary scores, 1993–2009

Year	average (%)	sd	N
2009	53.86	17.00	104,012
2008	53.57	16.86	120,650
2007	53.62	16.74	143,878
2006	53.14	16.63	171,478
2005	53.10	16.86	188,175
2004	53.04	16.95	200,869
2003	52.11	17.24	217,461
2002	51.22	17.54	212,671
2001	51.66	17.40	202,622
2000	50.26	16.71	193,845
1999	50.06	16.60	184,511
1998	48.56	15.25	153,622
1997	48.62	15.48	127,261
1996	50.38	15.09	86,103
1995	52.64	16.79	48,347
1994	43.49	19.07	5,190
1993	54.84	18.54	1,846

NB – The Yellis test changes only by one or two questions (of 120) each year, based on statistical analysis.

3.1.4 Adult literacy surveys, 1996–2003

International Adult Literacy Survey (IALS), 1996

Results for the prose and document literacy domains are shown in Table 3.5. Note that the numbering of IALS and UK National Qualifications Framework levels is out of step by one.

Table 3.5: IALS prose and document literacy results by age group, England only

	Age-group	16–25	26–35	36–45	46–55	56–65	Total
	Sample size	372	1172		928		2472
IALS level		%	%	%	%	%	%
	UK NQF level	Prose literacy					
3/4/5	Level 2 or above	52	54	54	49	24	48
2	Level 1	30	29	29	29	37	30
1	Entry level	17	18	17	22	39	21
		Document literacy					
3/4/5	Level 2 or above	56	56	58	48	27	51
2	Level 1	26	25	24	28	33	26
1	Entry level	18	19	18	24	40	23

Source: Carey *et al.* (1997)

The authors of the report for Britain (and these remarks apply fully to England) commented:

> The prose literacy and document literacy levels of people aged 16–45 were more or less uniform. The incidence of low basic skills in these two domains increased at ages 46–55 and – much more sharply – at ages 56–65. The gap was most marked at Level 1 [UK Entry level]: about 40% of the 56–65 group were at this level, compared with 17%–19% of 16- to 45-year-olds. (Carey *et al*., 1997)

It should also be noted that these results were the major source of the Moser Committee's estimate (GB. D*f*EE, 1999) that about 20% of adults in England (roughly 7 million people) had less than functional literacy, where the minimum level of functional literacy is defined as Level 1, and 'less than functional literacy' therefore as Entry level. However, in the youngest age-group slightly fewer than 20% were in Entry level.

The Skills for Life (SfL) survey of need, 2002/03

This survey treated literacy as a single domain. Results by age-group are given in Table 3.6. The survey also gathered data on participants' qualifications; though not presented here, these data are commented on in the quotation below the Table.

Table 3.6: *Skills for Life* survey results for literacy, by age group

Age-group	16–19	20–24	25–34	35–44	45–54	55–64	Total
Sample size	**444**	**613**	**1774**	**2044**	**1509**	**1488**	**7874**
	%	%	%	%	%	%	%
Level 2 or above	43	43	47	46	45	38	44
Level 1	41	45	40	39	36	40	40
Entry level overall	**17**	**13**	**14**	**15**	**19**	**22**	**16**
Entry 3	12	8	9	10	12	15	11
Entry 2	3	4	3	3	4	4	2
Entry 1 or below	2	1	2	2	3	3	3

Source: Williams *et al.* (2003)

The authors of the report commented:

> There were large differences in educational achievement between the various age-groups. Younger respondents were much more likely to hold qualifications than older respondents. Around 10 per cent of respondents under the age of 35 held no qualifications at all but this proportion climbs steadily with each subsequent age group: 16 per cent of 35- to 44-year-olds, 27 per cent of 45- to 54-year-olds and 41 per cent of 55- to 65-year-olds held no qualifications. However, the relatively flat age data for literacy test performance suggests that the difference between age groups in underlying ability was minimal. The sharp increase in qualification acquisition since the war – particularly in the 1960s and 1970s – has not led to a sharp increase in literacy skills. (Williams *et al*., 2003)

Williams *et al*. could equally have put the last point the other way round: the 'sharp increase in qualification acquisition since the war' has been achieved despite a generally steady level of underlying reading skills (for a related point about writing see Section 3.2.5). That could be interpreted as meaning that test demands have fallen – but Section 4 shows that this is not the case; more positively, it could be taken to show that young people have increasingly learnt how to give of their best under test/examination conditions, which in turn implies that their teachers have improved their preparation of pupils to take the tests/exams.

Also, what Williams *et al*. did not remark on is that these data also show an increase in low skills in the oldest age-group, though not as sharp as in the IALS data, and a decrease in the percentage with low skills from 16- to 19-year-olds to the next age-group. Both are compatible with a 'lifecourse' trend in attainment, namely that literacy skill increases in early middle age (perhaps as a result of honing the skills needed in employment), then plateaus for some decades, then declines again. A similar pattern was apparent across the age-range 22 to 74 in the results of the *Older and Younger* survey conducted in 1993–94 (Basic Skills Agency, 1995).

An alternative explanation would be that these differences are cohort effects, meaning (roughly speaking) that these age-groups left the school system with skills already at the level they still exhibit at later ages, or (to put this another way) that some cohorts leave the school system with better or worse average levels than others. However, the fact that the

same 'inverted U-shape' graph appears in surveys conducted years apart argues against this interpretation.

Another finding which supports the lifecourse theory is this. In 1972, a sub-sample of the NSHD cohort who had taken the Watts-Vernon reading test in 1961 when they were aged 15 (see Section 3.1.1) took it again, at the age of 26. The results showed 'a substantial general increase in reading scores' over the 11-year gap, and an 'illiteracy rate … as low as one per cent'; the corresponding 'illiteracy rate' for these same people at age 16 had been 3.5 per cent (Rodgers, 1986).

3.1.5 Conclusions on trends in reading attainment

- There was an improvement in average reading scores from 1948 to 1960.
- Average levels remained remarkably constant from 1960 to 1988.
- There is a gap in the evidence base from 1988 to about 1997.
- The Yellis data suggest a gentle rise between 1997 and 2004, then a further plateau.
- There was a fall in the average score between the two PISA surveys of 2000 and 2006, but it is not known whether this was statistically significant.
- But the proportion of young adults with poor reading (Entry level) seems to have remained stubbornly at about 17%. The lifecourse trend (improvement into early middle age, then usually a plateau, then decline) will have lifted some people into adequate literacy levels by their early to mid 20s, but many still have poor literacy at all later ages.

3.2 Writing

3.2.1 National monitoring surveys, 1979–88

There were none before 1979, and results for the APU writing surveys were not (and could not have been) expressed, like those for reading, as average standardised scores. However, results for repeated tasks are given in the various reports, especially those for 1979, 1983 and 1988. These show quite clearly that there had been no overall change in the average level of attainment in writing of Year 11 pupils in England in that period.

Brooks *et al*. (1993) analysed attainment in spelling in samples of repeated writing tasks undertaken by Year 11 pupils in 1980 and 1983. The level of attainment in spelling was the same in both years.

3.2.2 National test results, 1995–2007

The results for KS3 English 1995–2007 are shown in Table 3.7 and Figure 3.4. Both versions of the data show that the curve moved steadily to the right, that is, towards higher overall levels. The headline figure for this age-group is the percentage achieving Level 5 or above, which rose from 55% to 74% in this period, a substantial increase of 19 percentage points.

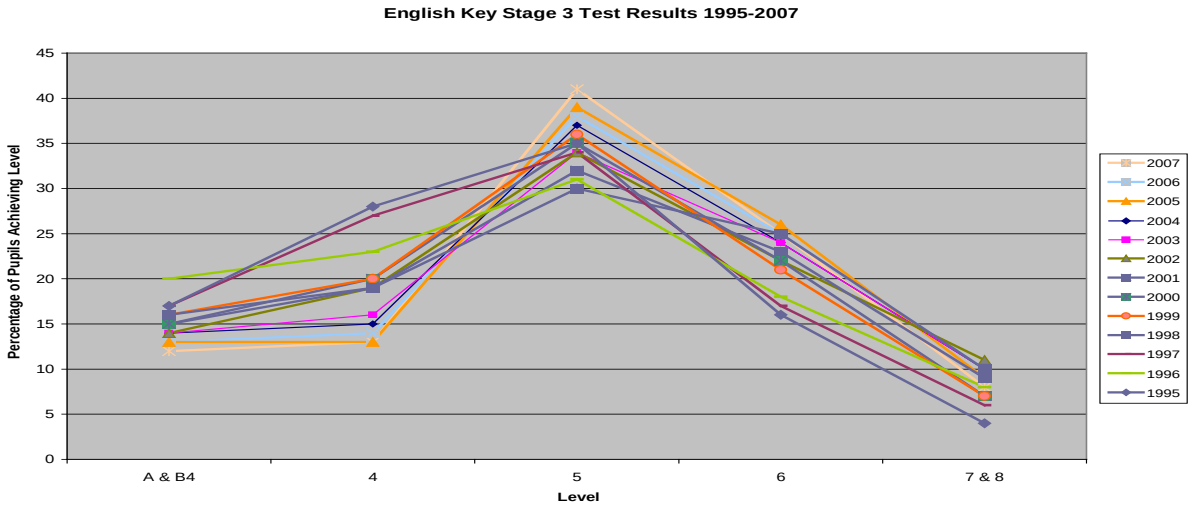
Using a very detailed and sophisticated method to compare pupils' results on KS3 English papers for 1996 and 2001, Massey *et al*. (2002) found it unlikely that marking standards had changed between those two years. This suggests that the rise of seven percentage points in

those achieving Level 5 or above in that period was genuine, and not caused by drift in markers' expectations or award of grades.

Table 3.7: KS3 English results, 1995–2007

	National Curriculum levels (%)					
	Below 4	4	5	6	7 & 8	**5 or above**
2007	12	13	41	25	8	**74**
2006	13	14	38	25	10	**73**
2005	13	13	39	26	9	**74**
2004	14	15	37	24	10	**71**
2003	14	16	34	24	10	**68**
2002	14	19	34	22	11	**67**
2001	15	19	32	23	9	**64**
2000	15	20	35	22	7	**64**
1999	16	20	36	21	7	**64**
1998	16	19	30	25	10	**65**
1997	17	27	34	17	6	**57**
1996	20	23	31	18	8	**57**
1995	17	28	35	16	4	**55**

Figure 3.4: KS3 English results, 1995–2007



Source: Based on preceding table

3.2.3 National examination results, 1989–2009

The results for GCSE English (Language) for the period 1989–2009 are shown in Table 3.8. For notes on the numbers and age of the pupils involved, see Section 2.2.3. The proportion of the cohort taking GCSE English (Language) rose over the period from 88% to 95%, thus making the results very representative of the full population of Year 11 pupils. The percentage achieving grade C or above shows an immediate increase from 1989 to 1990, then a plateau up to 1998, then a gently rising trend up to about 2005 and a rather faster increase in recent years. The increase over the entire period was substantial, at 16 percentage points.

Table 3.8: GCSE English results, 15-year-olds, England only, 1989–2009

Year	% grades A*–C
2009	66
2008	65
2007	63
2006	62
2005	61
2004	60
2003	60
2002	60
2001	59
2000	59
1999	58
1998	56
1997	55
1996	56
1995	56
1994	57
1993	56
1992	54
1991	55
1990	54
1989	50

Sources: www.dcsf.gov.uk/rsgateway/DB/SFR/for various years

A study of examination standards in GCSE English between 1989 and 1995 (SCAA, 1996) was carried out jointly by Ofsted and SCAA (the (then) Schools Curriculum and Assessment Authority, later merged into the Qualifications and Curriculum Authority, QCA, now the Qualifications and Curriculum Development Agency, QCDA). The study concluded that there was little evidence of any significant change in examination standards (criteria) in that period. This suggests that the rise of six percentage points in those achieving grade C or above in that period was genuine and not caused by grade drift.

3.2.4 Analyses of examination scripts, 1980–2004

Since the first of the three relevant studies (Massey and Elliott, 1996) was later extended it is dealt with below.

QCA (2004) found no change in standards of attainment in GCSE English in the period 1999–2002. This largely confirms the findings from the overall exam grades, above.

Massey and Elliott (1996) took samples of age 16 English examination scripts from 1980 (GCE O-Level), 1993 and 1994 (GCSE), and Massey *et al*. (2005) added GCSE scripts from 2004. For each exam they analysed a similar question (across years) from the scripts of 30 boys and 30 girls. The question was a composition of some kind (a story or opinion). The fourth sentence was taken from each script, and the sentences were compared using five criteria:

1. vocabulary
2. spelling
3. punctuation

4. sentence structure
5. use of 'non-standard' structure.

Massey and Elliott (1996) argued that, using the above criteria, quantitative analysis was possible, whereas a qualitative analysis (e.g. of imagination, content and style) was impossible without further evidence.
All samples were taken from exam scripts rather than coursework because for the latter more time and help would have been available (and in 1980 there was no coursework). No CSE scripts from 1980 were analysed, so comparison with all grades of GCSE was also impossible.

Massey and Elliott (1996) concluded that (1) they could not validly compare the content, structure and stylistic qualities of the scripts over time because the form of the paper and the task demands had changed too much; (2) they could, however, validly compare the technical aspects; (3) on these, candidates in 1980 seemed rather more accomplished than those in 1993 and 1994; but (4) the researchers 'therefore lack[ed] sufficient empirical evidence to conclude safely that, overall, writing in 1980 was better, grade for grade, or that grading standards … ha[d] changed.'

The results of the extended study showed that, while 1993 and 1994 were relatively poor years, the 2004 scripts had returned to the 1980 level (though not in spelling, where 1980 pupils were much better), and in some cases (e.g. punctuation) exceeded 1980. However, the use of non-standard English had increased through the years, and Massey *et al*. (2005) suggested that there was a case for an attempt to reverse this trend.

The later study still avoided analysis of the compositional aspect of the scripts. Ongoing work at Cambridge Assessment has produced promising approaches which can tackle this in addition to 'surface' features (Green *et al*., 2008), but these have yet to be applied to a representative sample of scripts.

3.2.5 Conclusions on trends in writing attainment

Various findings seem clear:

- There is no evidence before 1979.
- There was no change between 1979 and 1988 (APU results).
- GCSE exam results suggest a rise from 1989 to 1990, then a plateau from 1990 to 1998, then a gently rising trend up to about 2005 and a rather faster increase up to 2009. The increase over the entire period was substantial.
- KS3 test results show a substantial rise between 1995 and 2007.
- But studies of exam scripts suggest little change between 1980 and 2004.

There may be no contradiction: achievement at particular levels may not have changed (as the Massey *et al*. and SCAA studies suggest for KS3 and GCSE respectively), even if overall levels have.

3.3 Numeracy

3.3.1 National monitoring surveys, 1978–87

Results from successive from APU Mathematics Monitoring Project surveys cannot be compared as a whole because they are not available as raw data and overall weighted means are not given in any documentation. However, the following general statements are made in the two key reports:

> The results from comparing identical sets of items from the 1978 to the 1982 survey indicate a general improvement in the level of performance. The order of magnitude is about 2%. (APU, 1988, p.763)
>
> In the interval between the surveys of 1982 and 1987, there was a very general pattern of performance changes at … age [15/16], with improvements in the APU categories of Geometry, Probability and Statistics, and Measures, and a decline in Number and in Algebra (there were some variations in the detail of the general pattern). (Foxman, 1998)

This would seem to indicate a small overall improvement between 1978 and 1982, but no overall change between 1982 and 1987.

3.3.2 International surveys, 1964–2007

FIMS, 1964 (age 13)
England and Wales were ranked 6th overall with a mean score just above the average of the 12 countries. However, the standard deviation was the highest, suggesting a wider range of performance for England and Wales than for the other countries (Husén, 1967).

SIMS, 1981 (age 13)
SIMS did not publish country rankings in the way that FIMS and subsequent mathematics studies have reported their data. Part of the reason for this was that there were concerns about curriculum coverage among the participating countries, and also an attempt to move away from international rankings of achievement (cf. the study of writing, two years later – see the beginning of Section 2.2). However, the relative ranking of England and Wales dropped in all categories when compared with FIMS (Foxman, 1998). Foxman also states that 30 questions were common to both studies and that, amongst the 10 countries participating in both surveys, there was a general decline, except in algebra, where only England and Wales declined and all nine other countries participating on both occasions improved.

Badly documented IEA raw data were obtained for all countries, but it proved impossible to reconcile these with published figures. Cresswell and Gubb (1987) reported that there were 37 items common to FIMS and SIMS and that for 29 items there were significantly lower success rates in 1981 than in 1964. (Curiously, Robitaille and Garden, 1989 reported that 35 items were repeated from the previous survey.)

However, the tests were taken 3 months earlier in the year in 1981 than in 1964, therefore a term's worth of teaching was lacking in 1981. Also the curriculum had changed significantly in 17 years. Cresswell and Gubb (1987) stated that the effects of these two factors were difficult to quantify.

IAEP1, 1988 (age 13)
The UK was ranked 9th out of the 12 countries (Lapointe *et al*., 1989).

IAEP2, 1991 (age 13)

Table 3.9 shows a summary of the results. Many countries attempted to include all eligible children, but England and Wales had a very low participation rate and were deemed ineligible for comparison purposes. However, the mean score for England and Wales was similar to that of Scotland, and would have put England and Wales in the bottom half of the rankings.

Table 3.9: Country rankings in IAEP2, 1991, by mean score

Country	Mean (s.d.)
Korea	73 (0.6)
Taiwan	73 (0.7)
Switzerland	71 (1.3)
Soviet Union	70 (1.0)
Hungary	68 (0.8)
France	64 (0.8)
Emilia-Romagna, Italy	64 (0.9)
Israel	63 (0.8)
Canada	62 (0.6)
Scotland	61 (0.9)
Ireland	61 (0.9)
Slovenia	57 (0.8)
Spain	55 (0.8)
US	55 (1.0)
Jordan	40 (1.0)
Exclusions/Low participation	
China	80 (1.0)
England & Wales	61 (2.2)
Portugal	48 (0.8)
São Paulo, Brazil	37 (0.8)
Bortaleza, Brazil	32 (0.6)
Maputo and Beira, Mozambique	28 (0.3)

Source: Lapointe *et al.* (1992)

Six countries took part in both IAEP surveys. ‘England’s position relative to the five other countries … remain[ed] the same: first in statistics/data-handling and sixth in arithmetic/number and operations’ (Foxman *et al*., 1993: 147).

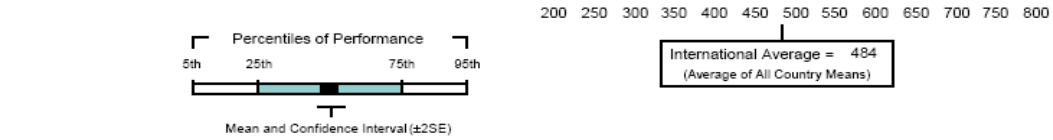
TIMSS 1995 (ages 13 and 14)

The league tables for ages 13 and 14 are shown in Tables 3.10 and 3.11 respectively.

Table 3.10: Results of TIMSS 1995, age 13

Distributions of Mathematics Achievement - Lower Grade (Seventh Grade*)

Country	Mean	Years of Formal Schooling	Average Age	Mathematics Achievement Scale Score
Singapore	601 (6.3)	7	13.3	
Korea	577 (2.5)	7	13.2	
Japan	571 (1.9)	7	13.4	
Hong Kong	564 (7.8)	7	13.2	
† Belgium (Fl)	558 (3.5)	7	13.0	
Czech Republic	523 (4.9)	7	13.4	
Slovak Republic	508 (3.4)	7	13.3	
† Belgium (Fr)	507 (3.5)	7	13.2	
1 Switzerland	506 (2.3)	6 or 7	13.1	
Hungary	502 (3.7)	7	13.4	
Russian Federation	501 (4.0)	6 or 7	13.0	
Ireland	500 (4.1)	7	13.4	
Canada	494 (2.2)	7	13.1	
France	492 (3.1)	7	13.3	
Sweden	477 (2.5)	6	12.9	
†2 England	476 (3.7)	8	13.1	
† United States	476 (5.5)	7	13.2	
New Zealand	472 (3.8)	7.5 - 8.5	13.0	
† Scotland	463 (3.7)	8	12.7	
1 Latvia (LSS)	462 (2.8)	7	13.3	
Norway	461 (2.8)	6	12.9	
Iceland	459 (2.6)	7	12.6	
Spain	448 (2.2)	7	13.2	
Cyprus	446 (1.9)	7	12.8	
1 Lithuania	428 (3.2)	7	13.4	
Portugal	423 (2.2)	7	13.4	
Iran, Islamic Rep.	401 (2.0)	7	13.6	
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):				
Australia	498 (3.8)	7 or 8	13.2	
Austria	509 (3.0)	7	13.3	
Bulgaria	514 (7.5)	7	13.1	
Netherlands	516 (4.1)	7	13.2	
Countries Not Meeting Age/Grade Specifications (High Percentage of Older Students; See Appendix A for Details):				
Colombia	369 (2.7)	7	14.5	
†1 Germany	484 (4.1)	7	13.8	
Romania	454 (3.4)	7	13.7	
Slovenia	498 (3.0)	7	13.8	
Countries With Unapproved Sampling Procedures At The Classroom Level (See Appendix A for Details):				
Denmark	465 (2.1)	6	12.9	
Greece	440 (2.8)	7	12.6	
† South Africa	348 (3.8)	7	13.9	
Thailand	495 (4.8)	7	13.5	



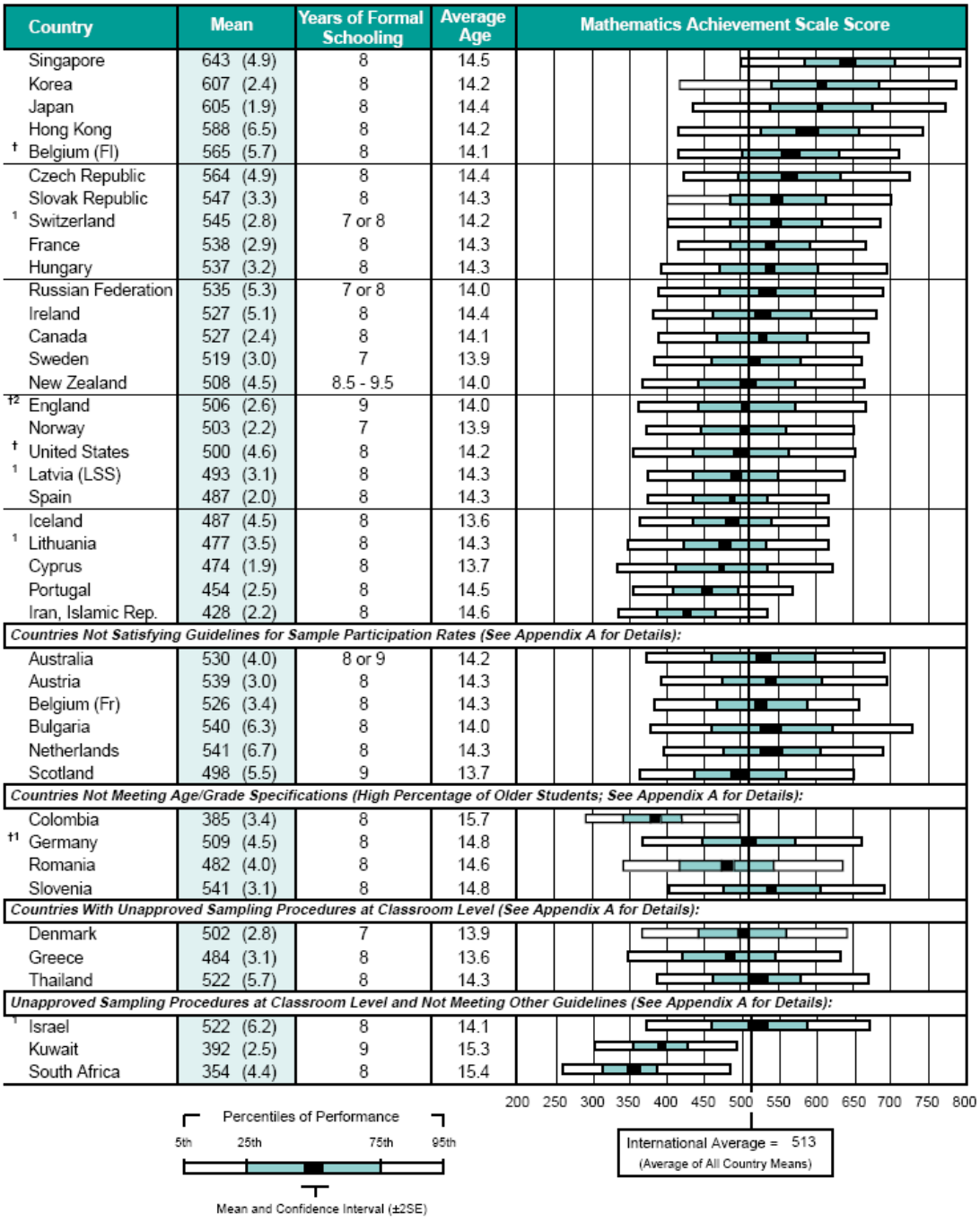
*Seventh grade in most countries; see Table 2 for information about the grades tested in each country.
†Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).
1National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
2National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).
( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table 3.11: Results of TIMSS 1995, age 14

Distributions of Mathematics Achievement - Upper Grade (Eighth Grade*)

Country	Mean	Years of Formal Schooling	Average Age
Singapore	643 (4.9)	8	14.5
Korea	607 (2.4)	8	14.2
Japan	605 (1.9)	8	14.4
Hong Kong	588 (6.5)	8	14.2
† Belgium (Fl)	565 (5.7)	8	14.1
Czech Republic	564 (4.9)	8	14.4
Slovak Republic	547 (3.3)	8	14.3
1 Switzerland	545 (2.8)	7 or 8	14.2
France	538 (2.9)	8	14.3
Hungary	537 (3.2)	8	14.3
Russian Federation	535 (5.3)	7 or 8	14.0
Ireland	527 (5.1)	8	14.4
Canada	527 (2.4)	8	14.1
Sweden	519 (3.0)	7	13.9
New Zealand	508 (4.5)	8.5 - 9.5	14.0
†2 England	506 (2.6)	9	14.0
Norway	503 (2.2)	7	13.9
† United States	500 (4.6)	8	14.2
1 Latvia (LSS)	493 (3.1)	8	14.3
Spain	487 (2.0)	8	14.3
Iceland	487 (4.5)	8	13.6
1 Lithuania	477 (3.5)	8	14.3
Cyprus	474 (1.9)	8	13.7
Portugal	454 (2.5)	8	14.5
Iran, Islamic Rep.	428 (2.2)	8	14.6
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):			
Australia	530 (4.0)	8 or 9	14.2
Austria	539 (3.0)	8	14.3
Belgium (Fr)	526 (3.4)	8	14.3
Bulgaria	540 (6.3)	8	14.0
Netherlands	541 (6.7)	8	14.3
Scotland	498 (5.5)	9	13.7
Countries Not Meeting Age/Grade Specifications (High Percentage of Older Students; See Appendix A for Details):			
Colombia	385 (3.4)	8	15.7
†1 Germany	509 (4.5)	8	14.8
Romania	482 (4.0)	8	14.6
Slovenia	541 (3.1)	8	14.8
Countries With Unapproved Sampling Procedures at Classroom Level (See Appendix A for Details):			
Denmark	502 (2.8)	7	13.9
Greece	484 (3.1)	8	13.6
Thailand	522 (5.7)	8	14.3
Unapproved Sampling Procedures at Classroom Level and Not Meeting Other Guidelines (See Appendix A for Details):			
1 Israel	522 (6.2)	8	14.1
Kuwait	392 (2.5)	9	15.3
South Africa	354 (4.4)	8	15.4



*Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
†Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).
1National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
2National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).
( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Thus 13-year-olds in England ranked 16th out of 22 countries in the main list in TIMSS 1995 (Table 3.11). This seems consistent with 9th out of 12 in IAEP1 and about equal to the 10th and 11th out of 14 countries in the main list in IAEP2.

An attempt was made to link TIMSS 1995 (presumably at age 13) back to SIMS. However,

> There were only 22 link items between SIMS and TIMSS… The relative position of England in mathematics appeared to have deteriorated since previous comparative studies were carried out. (Keys *et al*., 1996b)

> There were only 23 link items between SIMS and TIMSS. England's overall mean success rate on these items declined by about 1.5 per cent … between the two studies, 1981 to 1995. (Foxman, 1998)

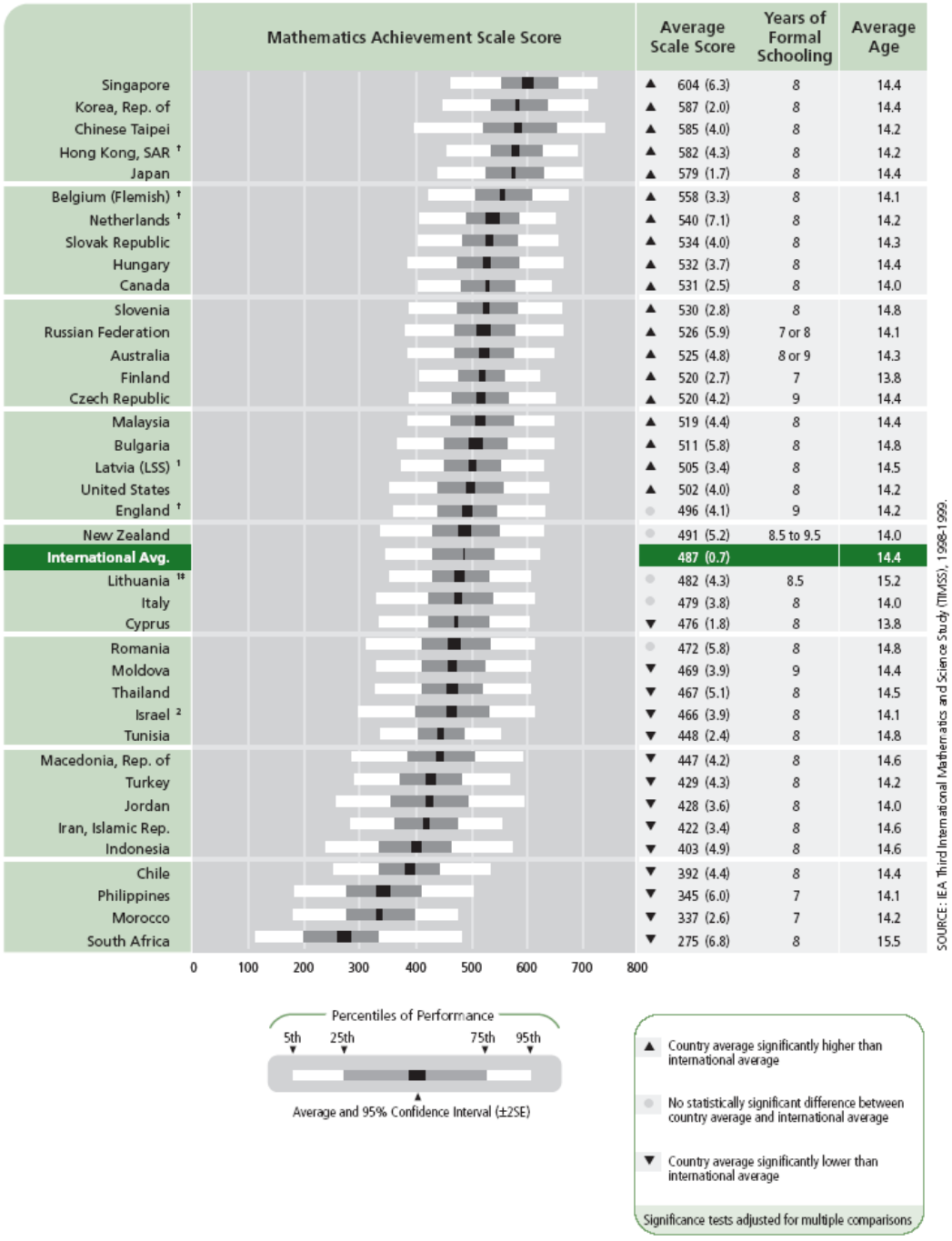
TIMSS 1995 was the first survey to assess the mathematical skills of 14-year-olds; England's placing in Table 3.12 was 16th out of 25 in the main list.

TIMSS(R), 1999 (age 14)

The league table is shown in Table 3.12. The authors of a report for England defined a comparison group of 11 other advanced industrialised countries, and commented on the performance of the English sample with reference to both this group and all countries taking part:

> England's score of 498 was significantly higher than the international average, 467, but significantly lower than the average for the comparison group, 529. Nine countries out of the 46 participating, including six from the comparison group – Singapore (605), Hong Kong, Japan, Belgium (Flemish), the Netherlands and Hungary (529) – scored significantly higher than England. Twelve other countries, including Australia, United States, Scotland and New Zealand from the comparison group performed at a similar level to England (between 493 and 508). All other countries scored at a significantly lower level than England. These included Italy from the comparison group countries. In summary, the performance of secondary school pupils in England in mathematics is below that of a number of developed countries, and similar to that in a range of others. (Ruddock *et al*., 2004)

Table 3.12: Results of TIMSS(R), 1999, age 14



† Met guidelines for sample participation rates only after replacement schools were included (see Exhibit A.8).

1 National Desired Population does not cover all of International Desired Population (see Exhibit A.5). Because coverage falls below 65%, Latvia is annotated LSS for Latvian-Speaking Schools only.

2 National Defined Population covers less than 90 percent of National Desired Population (see Exhibit A.5).

‡ Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

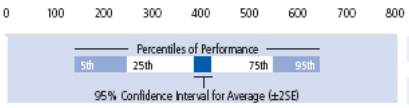
Source: Mullis *et al*. (2000)

TIMSS 2003 (age 14)

The league table is shown in Table 3.13. The sample size for England fell below the international standard, and England's result is therefore shown near the foot of the Table.

Table 3.13: Results of TIMSS 2003, age 14

Countries	Years of Schooling*	Average Age	Average Scale Score	Human Development Index**
Singapore	8	14.3	605 (3.6) ▲	0.884
⧓ Korea, Rep. of	8	14.6	589 (2.2) ▲	0.879
† Hong Kong, SAR	8	14.4	586 (3.3) ▲	0.889
Chinese Taipei	8	14.2	585 (4.6) ▲	–
Japan	8	14.4	570 (2.1) ▲	0.932
Belgium (Flemish)	8	14.1	537 (2.8) ▲	0.937
† Netherlands	8	14.3	536 (3.8) ▲	0.938
Estonia	8	15.2	531 (3.0) ▲	0.833
Hungary	8	14.5	529 (3.2) ▲	0.837
Malaysia	8	14.3	508 (4.1) ▲	0.790
Latvia	8	15.0	508 (3.2) ▲	0.811
Russian Federation	7 or 8	14.2	508 (3.7) ▲	0.779
Slovak Republic	8	14.3	508 (3.3) ▲	0.836
Australia	8 or 9	13.9	505 (4.6) ▲	0.939
‡ United States	8	14.2	504 (3.3) ▲	0.937
1 Lithuania	8	14.9	502 (2.5) ▲	0.824
Sweden	8	14.9	499 (2.6) ▲	0.941
† Scotland	9	13.7	498 (3.7) ▲	0.930
2 Israel	8	14.0	496 (3.4) ▲	0.905
New Zealand	8.5 - 9.5	14.1	494 (5.3) ▲	0.917
Slovenia	7 or 8	13.8	493 (2.2) ▲	0.881
Italy	8	13.9	484 (3.2) ▲	0.916
Armenia	8	14.9	478 (3.0) ▲	0.729
1 Serbia	8	14.9	477 (2.6) ▲	–
Bulgaria	8	14.9	476 (4.3) ▲	0.795
Romania	8	15.0	475 (4.8)	0.773
International Avg.	**8**	**14.5**	**467 (0.5)**	**–**
Norway	7	13.8	461 (2.5) ▼	0.944
Moldova, Rep. of	8	14.9	460 (4.0)	0.700
Cyprus	8	13.8	459 (1.7) ▼	0.891
2 Macedonia, Rep. of	8	14.6	435 (3.5) ▼	0.784
Lebanon	8	14.6	433 (3.1) ▼	0.752
Jordan	8	13.9	424 (4.1) ▼	0.743
Iran, Islamic Rep. of	8	14.4	411 (2.4) ▼	0.719
1 Indonesia	8	14.5	411 (4.8) ▼	0.682
Tunisia	8	14.8	410 (2.2) ▼	0.740
Egypt	8	14.4	406 (3.5) ▼	0.648
Bahrain	8	14.1	401 (1.7) ▼	0.839
Palestinian Nat'l Auth.	8	14.1	390 (3.1) ▼	0.731
Chile	8	14.2	387 (3.3) ▼	0.831
1 ‡ Morocco	8	15.2	387 (2.5) ▼	0.606
Philippines	8	14.8	378 (5.2) ▼	0.751
Botswana	8	15.1	366 (2.6) ▼	0.614
Saudi Arabia	8	14.1	332 (4.6) ▼	0.769
Ghana	8	15.5	276 (4.7) ▼	0.567
South Africa	8	15.1	264 (5.5) ▼	0.684
‡ England	9	14.3	498 (4.7) ▲	0.930
Benchmarking Participants				
Basque Country, Spain	8	14.1	487 (2.7) ▲	–
Indiana State, US	8	14.5	508 (5.2) ▲	–
Ontario Province, Can.	8	13.8	521 (3.1) ▲	–
Quebec Province, Can.	8	14.2	543 (3.0) ▲	–



SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 200:

* Represents years of schooling counting from the first year of ISCED Level 1.

** Taken from United Nations Development Programme's *Human Development Report 2003*, p. 237-240.

† Met guidelines for sample participation rates only after replacement schools were included (see Exhibit A.9).

‡ Nearly satisfied guidelines for sample participation rates only after replacement schools were included (see Exhibit A.9).

‖ Did not satisfy guidelines for sample participation rates (see Exhibit A.9).

1 National Desired Population does not cover all of International Desired Population (see Exhibit A.6).

2 National Defined Population covers less than 90% of National Desired Population (see Exhibit A.6).

⧓ Korea tested the same cohort of students as other countries, but later in 2003, at the beginning of the next school year.

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A dash (–) indicates comparable data are not available.

The authors of a report for England commenting on the first three rounds of TIMSS had this to say:

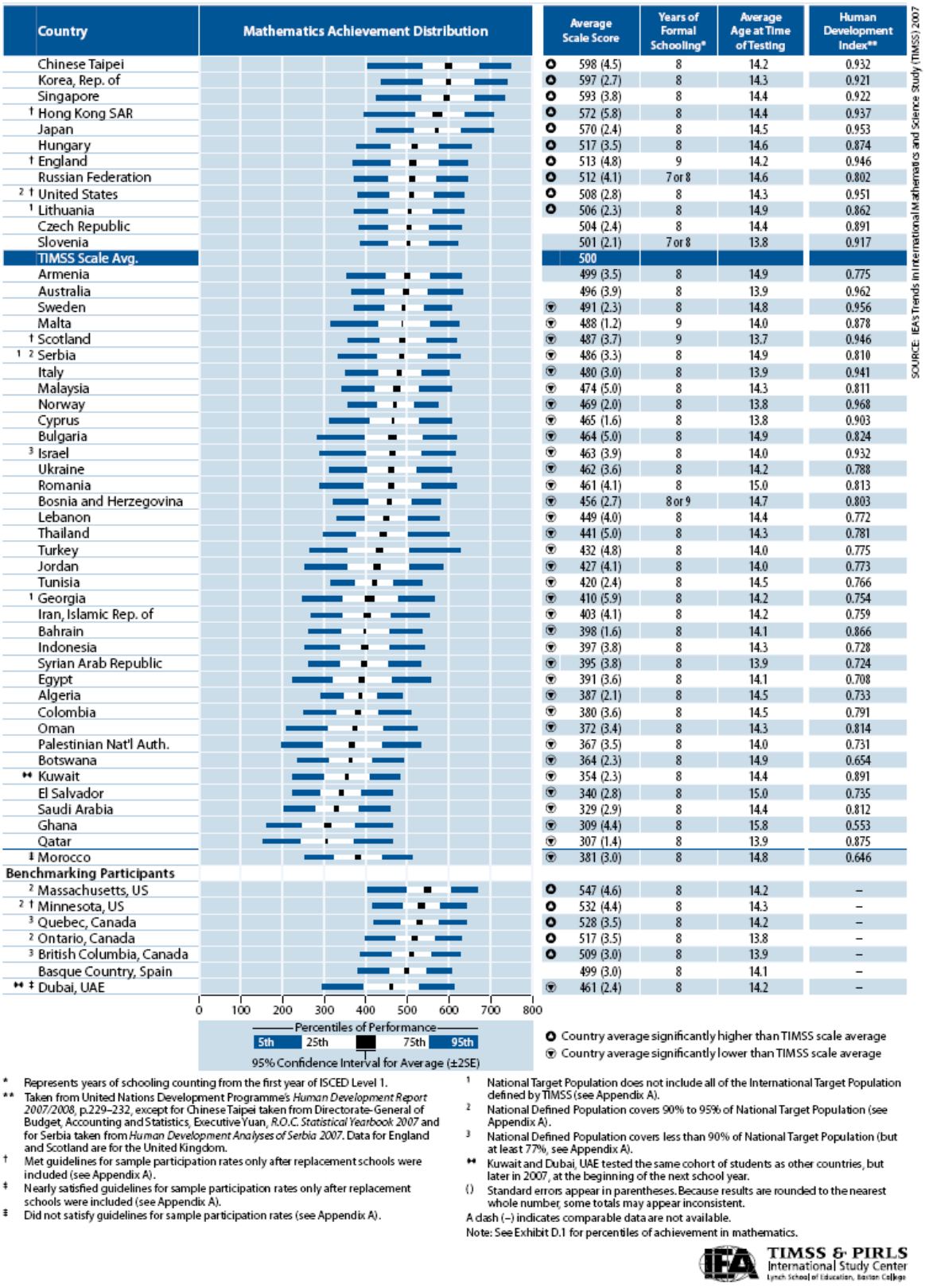
> England's mathematics scores [at age 14] were 498 in 1995, 496 in 1999 and 498 in [2003], a very consistent pattern. No change in performance was the most common pattern in the comparison group countries, England being one of eight of these 12 countries to show no change from either 1995 or 1999 to 2003. Only two showed an increase, both from 1995 to 2003.
> In the 12 comparison group countries the average scale score fell by 2 scale points from 1999 to 2003.
> No change in performance was also the most common pattern overall in grade 8 mathematics; 17 of the 35 countries showed no change from either 1995 or 1999 to 2003, while 11 showed a decline in performance.
> (Ruddock *et al*., 2004)

TIMSS 2007 (age 14)

The league table is shown in Table 3.14. The sample size for England met the international standard, and England's result is therefore shown near the top of the Table.

The average score for England in 2007 was 513. This was significantly above the scale average of 500 and the average scores of all countries from Slovenia downwards, and only significantly lower than the average scores of the five Asian countries at the top of the Table. It was also significantly higher than the average scores for England in all three previous age 14 TIMSS surveys.

Table 3.14: Results of TIMSS 2007, age 14



Country	Mathematics Achievement Distribution	Average Scale Score	Years of Formal Schooling*	Average Age at Time of Testing	Human Development Index**
Chinese Taipei		598 (4.5)	8	14.2	0.932
Korea, Rep. of		597 (2.7)	8	14.3	0.921
Singapore		593 (3.8)	8	14.4	0.922
† Hong Kong SAR		572 (5.8)	8	14.4	0.937
Japan		570 (2.4)	8	14.5	0.953
Hungary		517 (3.5)	8	14.6	0.874
† England		513 (4.8)	9	14.2	0.946
Russian Federation		512 (4.1)	7 or 8	14.6	0.802
2 † United States		508 (2.8)	8	14.3	0.951
1 Lithuania		506 (2.3)	8	14.9	0.862
Czech Republic		504 (2.4)	8	14.4	0.891
Slovenia		501 (2.1)	7 or 8	13.8	0.917
TIMSS Scale Avg.		**500**			
Armenia		499 (3.5)	8	14.9	0.775
Australia		496 (3.9)	8	13.9	0.962
Sweden		491 (2.3)	8	14.8	0.956
Malta		488 (1.2)	9	14.0	0.878
† Scotland		487 (3.7)	9	13.7	0.946
1 2 Serbia		486 (3.3)	8	14.9	0.810
Italy		480 (3.0)	8	13.9	0.941
Malaysia		474 (5.0)	8	14.3	0.811
Norway		469 (2.0)	8	13.8	0.968
Cyprus		465 (1.6)	8	13.8	0.903
Bulgaria		464 (5.0)	8	14.9	0.824
3 Israel		463 (3.9)	8	14.0	0.932
Ukraine		462 (3.6)	8	14.2	0.788
Romania		461 (4.1)	8	15.0	0.813
Bosnia and Herzegovina		456 (2.7)	8 or 9	14.7	0.803
Lebanon		449 (4.0)	8	14.4	0.772
Thailand		441 (5.0)	8	14.3	0.781
Turkey		432 (4.8)	8	14.0	0.775
Jordan		427 (4.1)	8	14.0	0.773
Tunisia		420 (2.4)	8	14.5	0.766
1 Georgia		410 (5.9)	8	14.2	0.754
Iran, Islamic Rep. of		403 (4.1)	8	14.2	0.759
Bahrain		398 (1.6)	8	14.1	0.866
Indonesia		397 (3.8)	8	14.3	0.728
Syrian Arab Republic		395 (3.8)	8	13.9	0.724
Egypt		391 (3.6)	8	14.1	0.708
Algeria		387 (2.1)	8	14.5	0.733
Colombia		380 (3.6)	8	14.5	0.791
Oman		372 (3.4)	8	14.3	0.814
Palestinian Nat'l Auth.		367 (3.5)	8	14.0	0.731
Botswana		364 (2.3)	8	14.9	0.654
↤ Kuwait		354 (2.3)	8	14.4	0.891
El Salvador		340 (2.8)	8	15.0	0.735
Saudi Arabia		329 (2.9)	8	14.4	0.812
Ghana		309 (4.4)	8	15.8	0.553
Qatar		307 (1.4)	8	13.9	0.875
‡ Morocco		381 (3.0)	8	14.8	0.646
Benchmarking Participants					
2 Massachusetts, US		547 (4.6)	8	14.2	–
2 † Minnesota, US		532 (4.4)	8	14.3	–
3 Quebec, Canada		528 (3.5)	8	14.2	–
2 Ontario, Canada		517 (3.5)	8	13.8	–
3 British Columbia, Canada		509 (3.0)	8	13.9	–
Basque Country, Spain		499 (3.0)	8	14.1	–
↤ ‡ Dubai, UAE		461 (2.4)	8	14.2	–

* Represents years of schooling counting from the first year of ISCED Level 1.

** Taken from United Nations Development Programme's *Human Development Report 2007/2008*, p.229–232, except for Chinese Taipei taken from Directorate-General of Budget, Accounting and Statistics, Executive Yuan, *R.O.C. Statistical Yearbook 2007* and for Serbia taken from *Human Development Analyses of Serbia 2007*. Data for England and Scotland are for the United Kingdom.

† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A).

‡ Nearly satisfied guidelines for sample participation rates only after replacement schools were included (see Appendix A).

‡ Did not satisfy guidelines for sample participation rates (see Appendix A).

1 National Target Population does not include all of the International Target Population defined by TIMSS (see Appendix A).

2 National Defined Population covers 90% to 95% of National Target Population (see Appendix A).

3 National Defined Population covers less than 90% of National Target Population (but at least 77%, see Appendix A).

↤ Kuwait and Dubai, UAE tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A dash (–) indicates comparable data are not available.

Note: See Exhibit D.1 for percentiles of achievement in mathematics.

Source: Mullis *et al*. (2008)

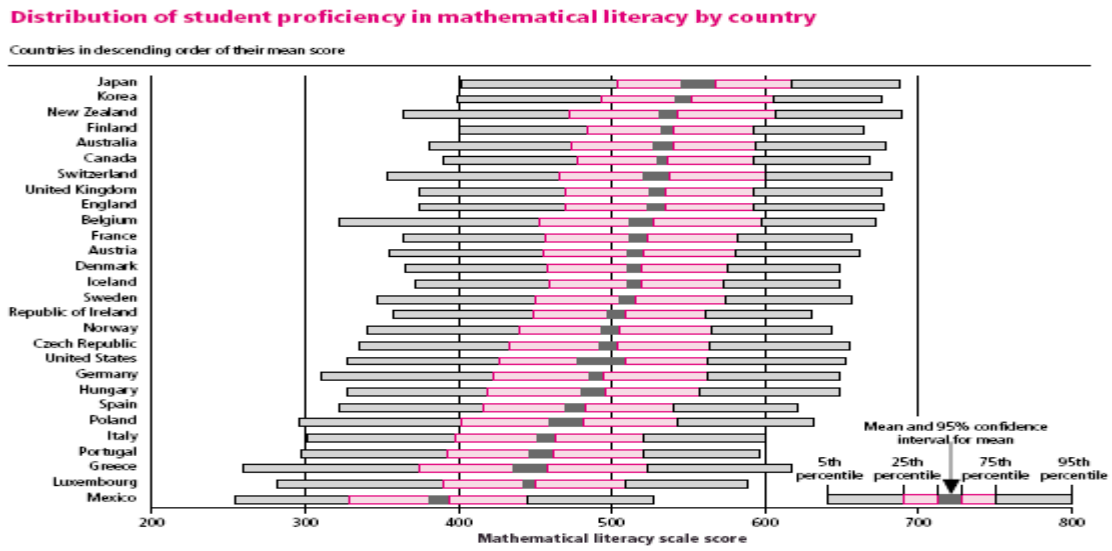
PISA 2000, 2006 (age 15)

Figures 3.5 and 3.6 show the results for maths.

> [In 2000] English students … did significantly better than the OECD average in … mathematical … literacy, averaging 529 … points … similar … to … students in Australia, Canada, Finland and New Zealand. Only Japan and Korea did significantly better ... English students did significantly better … than those in, for example, Denmark, Ireland, the US and Germany. (Gill *et al*., 2002; see also OECD, 2001)

In 2006, England's mean score was 495, just but not significantly below the international average of 498. As for reading, there are no statements in either the international or national report on whether the fall of 34 points since 2000 was or was not statistically significant.

Figure 3.5: PISA 2000 results for mathematics



Source: Gill *et al*. (2002)

Figure 3.6: PISA 2006 results for mathematics

	Mean score		significance
	Mean	S.E.	
Chinese Taipei	549	4.1	▲
Finland*	548	2.3	▲
Hong Kong-China	547	2.7	▲
Korea	547	3.8	▲
Netherlands*	531	2.6	▲
Switzerland	530	3.2	▲
Canada	527	2.0	▲
Macao-China	525	1.3	▲
Liechtenstein	525	4.2	▲
Japan	523	3.3	▲
New Zealand	522	2.4	▲
Belgium*	520	3.0	▲
Australia	520	2.2	▲
*Estonia**	515	2.7	▲
Denmark*	513	2.6	▲
Czech Republic*	510	3.6	▲
Iceland	506	1.8	▲
Austria*	505	3.7	*NS*
*Slovenia**	504	1.0	▲
Germany*	504	3.9	*NS*
Sweden*	502	2.4	*NS*
Republic of Ireland*	501	2.8	*NS*
OECD average[1]	**498**	**0.5**	*NS*
France*	496	3.2	*NS*
United Kingdom*	495	2.1	
Poland*	495	2.4	*NS*
England	495	2.5	
Slovak Republic*	492	2.8	*NS*
Hungary*	491	2.9	*NS*
Luxembourg*	490	1.1	*NS*
Norway	490	2.6	*NS*
*Lithuania**	486	2.9	*NS*
*Latvia**	486	3.0	*NS*
Spain*	480	2.3	▼
Azerbaijan	476	2.3	▼
Russian Federation	476	3.9	▼
United States	474	4.0	▼
Croatia	467	2.4	▼
Portugal*	466	3.1	▼
Italy*	462	2.3	▼
Greece*	459	3.0	▼
Israel	442	4.3	▼
Serbia	435	3.5	▼
Turkey	424	4.9	▼
*Romania**	415	4.2	▼
*Bulgaria**	413	6.1	▼
Mexico	406	2.9	▼

key
▲ significantly higher
NS no significant difference
▼ significantly lower

OECD countries (not italicised)
Countries not in OECD (italicised)
*EU countries

12 countries with scores below 430 omitted
Multiple comparison P-value = 0.045%
[1] Simple comparison P-value = 5%

Source: Bradshaw *et al*. (2007), p.75

3.3.3 Repeated use of standardised tests, 1997–2009

The results for the Yellis maths test taken by pupils in Year 10 (age 14) 1993–2009 are shown in Table 3.15. Details of the test itself are at www.cemcentre.org/RenderPage.asp?LinkID=11518000.

Table 3.15: Average Yellis maths scores, 1993–2009

Year	average (%)	sd	N
2009	47.48	20.32	104,012
2008	46.34	19.70	120,650
2007	46.14	19.19	143,878
2006	45.40	18.58	171,478
2005	45.02	18.59	188,175
2004	44.27	18.28	200,869
2003	43.46	18.11	217,461
2002	42.71	18.12	212,671
2001	41.82	17.96	202,622
2000	41.86	17.85	193,845
1999	41.10	17.62	184,511
1998	39.76	17.43	153,622
1997	39.06	17.49	127,261
1996	39.22	17.42	86,103
1995	39.77	17.24	48,347
1994	32.48	17.86	5,190
1993	43.95	20.24	1,846

NB – The Yellis test changes only by one or two questions (of 120) each year, based on statistical analysis.

Each cohort of pupils in these years numbered close to 600,000. From 1997 onwards, therefore, the test was taken by over 20% of each cohort. If these samples are taken as nationally representative (and the earlier ones are disregarded), the average score rose gently over the whole period, and the total increase was about 8.5 percentage points, which seems significant.

3.3.4 National test results, 1995–2007

The results for KS3 maths 1995–2007 are shown in Table 3.16 and Figure 3.7.

Both versions of the data show that the curve moved very steadily to the right, that is, towards higher overall levels. The headline figure for this age-group is the percentage achieving Level 5 or above, which rose from 57% to 76% in this period, a substantial increase of 19 percentage points.

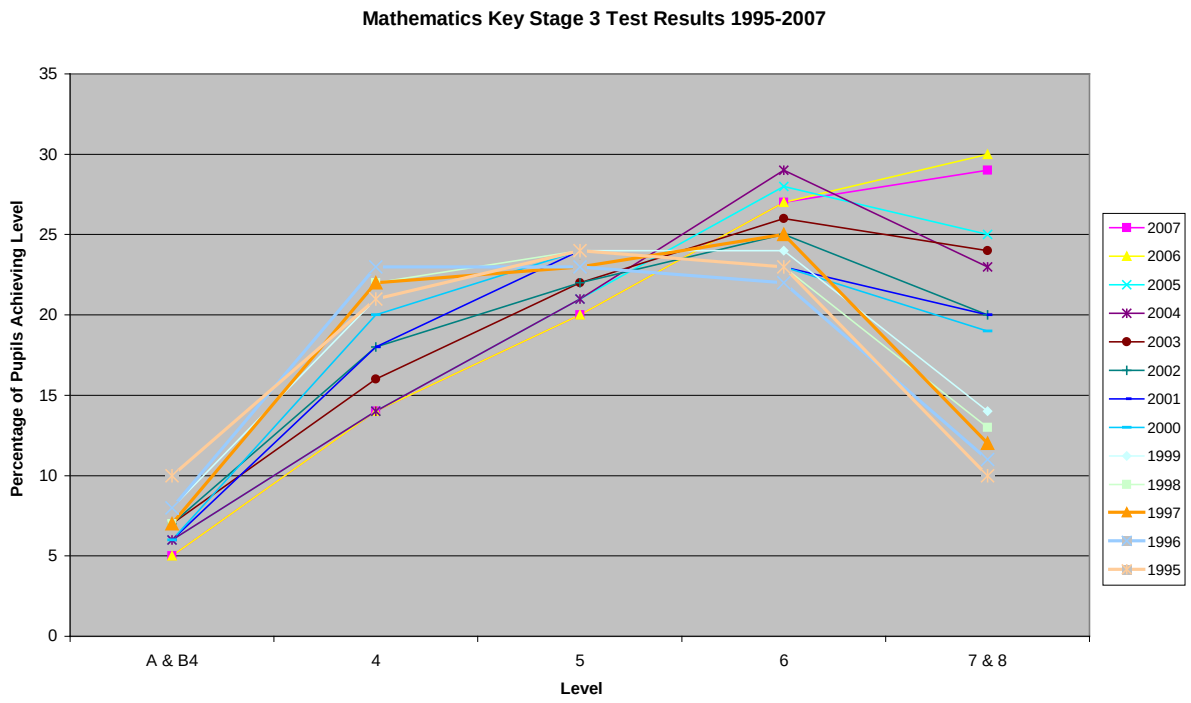
Massey *et al*. (2002) used a very detailed and sophisticated method to compare pupils' results on KS3 maths papers for 1996 and 2000, a period during which the proportion achieving Level 5 or above rose by 10 percentage points. The introduction of a mental arithmetic test from 1998 made the comparison more difficult than for English, but it would seem that achieving Level 5 or above in maths may have become somewhat easier. The implication (not stated by the researchers) may be that *part* of the 10-point rise may have

been caused by drift in markers' expectations or award of grades, even though the rest was genuine. No estimate of the balance between drift and genuine rise seems feasible.

Table 3.16: KS3 maths results, 1995–2007

	National Curriculum levels (%)					
	Below 4	4	5	6	7 & 8	**5 or above**
2007	5	14	20	27	29	**76**
2006	5	14	20	27	30	**77**
2005	6	14	21	28	25	**74**
2004	6	14	21	29	23	**73**
2003	7	16	22	26	24	**72**
2002	7	18	22	25	20	**67**
2001	6	18	24	23	20	**67**
2000	6	20	24	23	19	**66**
1999	8	21	24	24	14	**62**
1998	7	22	24	23	13	**60**
1997	7	22	23	25	12	**60**
1996	8	23	23	22	11	**56**
1995	10	21	24	23	10	**57**

Figure 3.7: KS3 maths results, 1995–2007



Source: Based on preceding Table

3.3.5 National examination results, 1989–2009

The results for GCSE maths for the period 1989–2009 are shown in Table 3.17. For notes on the numbers and age of the pupils involved, see Section 2.2.3. The proportion of the cohort taking GCSE maths rose over the period from 84% to 94%, thus making the results very representative of the full population of Year 11 pupils.

The percentages achieving grade C or above show a steadily rising trend over the entire period, and the increase over the whole period was a substantial rise of 21 percentage points.

Table 3.17: GCSE maths results, 15-year-olds, England only, 1989–2009

Year	% grades A*–C
2009	61
2008	59
2007	57
2006	56
2005	55
2004	53
2003	51
2002	52
2001	51
2000	50
1999	49
1998	47
1997	47
1996	46
1995	45
1994	46
1993	46
1992	44
1991	44
1990	41
1989	40

Sources: www.dcsf.gov.uk/rsgateway/DB/SFR/ for various years

3.3.7 Adult numeracy surveys, 1981–2003

ACACE, 1981

The instrument used was very poor. For what it is worth, 21 per cent of the sample scored less than 6 out of 11. The total sample was 2,890. Of these, 529 were in the age range 16–24, of whom 93% answered Q1 correctly and only 59% answered Q6 correctly:

Q1). How much would it cost you altogether to buy a cup of coffee at 17p and a sandwich at 24p?

Q6). The respondent was given a railway timetable and this text: ‘I live in Leicester and have arranged to meet a friend in London at 4 o’clock in the afternoon. Assuming the trains run on time, which is the latest train that I can get from Leicester to arrive in time for the meeting?’ The timetable below is an edited version of the one given to respondents.

Mondays to Fridays	
Leicester	London
dep.	arr.
01.36	03.52
.	.
13.44	15.42
.	.
21.24	23.51

IALS, 1996

Results for the 'quantitative literacy' domain are shown in Table 3.18. Note that the numbering of IALS and UK National Qualifications Framework levels is out of step by one.

Table 3.18: IALS, 1996, 'quantitative literacy' results by age group, England only

	Age-group	**16–25**	**26–35**	**36–45**	**46–55**	**56–65**	**Total**
IALS level	**Sample size**	372	1172		928		2472
	UK NQF level	%	%	%	%	%	%
3/4/5	Level 2 or above	48	52	57	50	31	49
2	Level 1	29	28	24	26	34	27
1	Entry level	22	20	19	24	35	23

Source: Carey *et al.* (1997)

An apparent 'lifecourse' trend is apparent in these data as in the IALS reading data: skills levels improve from the youngest age-group to the next two, then plateau, then decline, particularly after age 55.

International Numeracy Survey, 1996

The full set of tasks used (Basic Skills Agency, 1997b) is listed below, and the results for the UK are given in Table 3.19.

Subtract 1.78 from 5.

Take away 2.43 from 5.

Add together 5.5, 7.25 and 3.75.

The total of 4.25, 6 and 7.74.

Multiply 6 x 21.

Multiply 16 x 21.

Area of a room 11m x 18m.

Number of apples each person gets if a box of 72 is shared by six people.

Work out 15% of 700.

Number of children in a crowd of 7900 if the proportion is 10%.

What is 5/6 of 300?

Number of books not in the sale if a third are in the sale and the total number of books is

420.

While 43% of those included in the survey in Japan got all 12 tasks right, only 20% did so in the UK. At the other end of the scale, while more than a fifth of those in the UK (22%) could only answer up to 5 of the questions posed, only 4% of those in the Netherlands were as poor as this with numbers. While overall the UK sample could only get an average of 7.9 answers correct out of 12, all the other countries managed to get 9 or more. In the UK, 16- to 24-year-olds did worse than people from other age groups, and 45- to 54-year-olds did best – but the age-group sub-samples were so small that these differences were probably unreliable.

Table 3.19: International Numeracy Survey, 1996, results by age group, UK

Age-group	**16–24**	**25–34**	**35–44**	**45–54**	**55–60**	**Total**
Sample size	**148**	**164**	**161**	**133**	**52**	**660**
Score	%	%	%	%	%	%
12 correct	16	16	19	29	29	20
11 correct	11	18	18	14	8	15
10 correct	7	17	7	16	12	12
9 correct	8	13	11	9	6	10
8 correct	9	7	7	6	6	7
7 correct	7	7	7	4	4	6
6 correct	8	3	2	2	4	4
5 correct	7	2	3	5	6	5
4 correct	4	4	4	4	2	4
3 correct	5	3	4	2	2	4
2 correct	7	1	3	1	8	3
1 correct	2	2	3	5	4	3
0 correct	5	3	5	1	4	3
Average number correct (out of 12)	7.1	8.3	7.7	8.8	7.6	7.9

Columns may not sum to 100% because of questions not answered
Source: Basic Skills Agency (1997b)

The Skills for Life (SfL) survey of need, 2002/03

Results by age-group are given in Table 3.20.

Table 3.20: *Skills for Life* survey results for numeracy, by age group

Age-group	**16–19**	**20–24**	**25–34**	**35–44**	**45–54**	**55–64**	**Total**
Sample size	**444**	**613**	**1774**	**2044**	**1509**	**1488**	**7874**
	%	%	%	%	%	%	%
Level 2 or above	23	24	29	27	26	20	25
Level 1	27	27	28	29	27	27	28
Entry level overall	**50**	**49**	**43**	**44**	**47**	**53**	**47**
Entry 3	29	30	24	24	24	26	25
Entry 2	15	14	14	15	16	19	16
Entry 1 or below	6	4	4	5	6	8	5

Source: Williams *et al*. (2003)

The authors of the report commented:

> It should be noted that the performance of 20- to 24-year-olds was very similar to that of 16- to 19-year-olds, suggesting either a cohort effect among 16- to 24-year-olds as a whole, or that numeracy skills are further developed by the labour market. Many 16- to 24-year-olds either had not yet entered the labour market or had not settled into their 'career' occupations. As an example of the latter, only one in four 16- to 24-year-olds with Level 2 or above numeracy were in managerial/professional occupations, compared to six

in ten 25- to 34-year-olds with Level 2 or above numeracy. (Williams *et al*., 2003)

There is the usual falling-off after age 55.

It should be noted that the estimate for Entry level numeracy at age 16–19 in this survey (50%) was much higher than in any other numeracy survey mentioned in this report. In fact, the *Skills for Life* survey estimate for numeracy at Entry level 2 and below, 21%, is much closer to other estimates for the whole of Entry level. As discussed in Section 4, the grade criteria seem to have been set higher than in other surveys.

There is one other result which was (initially) closer to the *Skills for Life* survey result. The Basic Skills Agency's 1996–97 survey of adult literacy and numeracy needs throughout England showed 33% of adults aged 16–60 nationally having numeracy below Level 1 of the Agency's Numeracy Standards, based on the threshold for Level 1 being defined as 14 out of the 18 items in the test having been answered correctly (Basic Skills Agency, 1997a). In 2001 the Agency re-issued the results with the threshold for Level 1 lowered to 13 answers correct out of 18: the proportion of adults having literacy below (the new) Level 1 was 24%, now almost exactly equal to the IALS figure of 23% (Basic Skills Agency, 2001). What this illustrates is the fragility of some estimates of poor numeracy (and see Section 4.2 for a contrary adjustment in reading).

3.3.8 Conclusions on trends in numeracy attainment

The findings are complex:

- There is very little evidence before 1978 (one isolated international survey, FIMS 1964), and none at age 13 after 1995.

At age 13

- There may have been a fall between 1964 and 1981 if the repeated items in FIMS and SIMS are to be taken as reliable despite the vast changes in curricula.
- There appears to have been no change between 1988 and 1995 (IAEP1 – IAEP2 – TIMSS 1995).
- But there was a small decline between 1981 and 1995 (SIMS – TIMSS 1995), therefore presumably before 1988 if the previous result is reliable.

At age 14

- There was no change between 1995 and 2003 (TIMSS 1995 – TIMSS(R) – TIMSS 2003), or
- There was a significant improvement between 1997 and 2009 (Yellis), or between 1995 and 2007 (KS3 maths), even though part of the rise in KS3 maths between 1996 and 2000 may not have been genuine.

No satisfactory way of resolving most of this contradiction suggests itself, but

- There was a significant improvement between 2003 and 2007 (TIMSS).

At age 15/16

- There was a small improvement between 1978 and 1982 (APU).
- There was no change between 1982 and 1987 (APU).

- There was a substantial increase in the 'pass rate' between 1989 and 2009 (GCSE maths).
- There was a fall in the average score between the two PISA surveys of 2000 and 2006, but it is not known whether this was statistically significant.

At age 16+

- Though the four surveys span the period 1981–2003 they cannot be used to identify trends.
- However, they all show that substantial proportions of young people (16–19 or 16–24/25) have poor numeracy, of the order of 22% (provided the *Skills for Life* survey estimate for Entry level 2 or below is used, rather than its estimate for the whole of Entry level).
- The lifecourse trend (improvement into early middle age, then usually a plateau, then decline) will have lifted some people into adequate numeracy levels by their early to mid 20s, but many still have poor numeracy at all later ages.

Overall

- Internationally, number skills in England are poorer than in many other countries, especially industrialised ones, though other aspects of numeracy are better.

3.4 Correlations between literacy and numeracy achievement

Where there are literacy and numeracy results from the same set of students, an attempt can be made to see if there is any correlation between them and, if so, also to establish whether the correlation is curvilinear, i.e. you need a certain basic level of literacy before you can begin to score on numeracy (perhaps especially if the numeracy items are 'problems' embedded in text, as in IALS 'quantitative literacy' and PISA).

Literacy and numeracy data were gathered from the same samples of people in the 1974 age 16 sweep of NCDS, IALS (taking 'quantitative literacy' as a proxy for numeracy), PISA 2000, PISA 2006, Yellis, and the *Skills for Life* needs survey, plus all state school pupils in England born between 1/9/85 and 31/8/87. These pupils formed the two school-year cohorts which were studied in a Economic and Social Research Council Teaching and Learning Research Programme project at the Institute of Education, University of London (IoE) called 'Widening participation in higher education: A quantitative analysis' and directed by Prof. Anna Vignoles. She and her colleagues were able to obtain and merge all available educational data on those cohorts (numbering about a million pupils in all), including national test results at age 14 and GCSE results at 16.

However,

- For NCDS 1974 the only published data (Fogelman, 1976, 1983) are aggregated, and for Yellis only aggregated data were supplied – individual-level data would have been needed to carry out the correlation;
- No information could be found on correlations in PISA 2006;
- Peter Burke (personal communication, 2007) of the University of Nottingham, who led the team which designed the questions for the *SfL* survey, commented that:

  > It was not part of the design brief for the survey given to us by D*f*ES to be able to correlate the literacy and numeracy responses in any way that would facilitate your

> research. This should however be part of the brief for any future survey of this nature and it may be worth making this point as part of your report on your current project.

Despite all this, some findings were obtained. Using their 'Widening participation' dataset, Anna Vignoles and Claire Crawford at IoE were able to supply the following correlations:

> Between KS3 English and KS3 maths: 0.69
>
> Between GCSE English and KS3 maths: 0.80

For IALS, the correlations between the three domains across the three countries of Britain and all age-groups (Carey *et al*., 1997, p.144) were:

	Prose	Document	Quantitative
Prose	1		
Document	0.946	1	
Quantitative	0.912	0.960	1

Thus all the domains were largely tapping the same underlying trait.

And for PISA 2000, Gill *et al*. (2002) stated:

> Students who performed well in the assessment of mathematical literacy in England also tended to perform well on the reading literacy scales. The overall correlation between the scores attained in mathematical literacy and on the combined reading literacy scale was high, 0.87.

However, as Goldstein (2004) pointed out about PISA 2000:

> This causally oriented approach to interpretation continues when the report quotes high (0.8–0.9) correlations between reading, maths and science test scores and uses these data to suggest that 'reading is a prerequisite for successful performance in any school subject.' This may have some truth, but the existence of high simple correlations does not demonstrate this. It would be more relevant to look, for example, at how progress in maths correlates with progress in reading and we do know from other studies that such correlations are much lower. Thus, in Inner London (Goldstein *et al*., 1993) the simple correlation between English and maths at 16 years is 0.62 but only 0.40 in terms of progress between 11 and 16 years.

Much work remains to be done on this topic, and Peter Burke's point about the need to provide for such correlational analyses within the design of future surveys is well made.

4. Interpretation

Copious data were presented in Chapter 3, but what do they mean? In other words, what do the various numbers and the reports they are drawn from tell us about what 13- to 19-year-olds can do? This chapter contains some attempted answers to these questions.

However, the first gap in the evidence to be pointed out is a serious one: there is effectively no evidence on what attainment at different levels means for the compositional (authorial) aspects of writing, as opposed to the secretarial (surface) features. The APU surveys at age 15/16 gave detailed descriptions of performance on specific tasks but few or no generalisations, the QCA study was only concerned with overall standards, and the Cambridge studies of examination scripts have so far analysed only surface features.

Secondly, all the earlier reports on reading and numeracy (up to 1993), and some of the more recent ones, provided rather little information that is useful in this context. However, some more recent reports do provide detailed descriptions. The main reason for the growth of information in some later surveys seems to be the exponential increase in computer power for both number-crunching and word-processing. A contributory reason seems to be the increasing need within international surveys to make everything as clear as possible in English to national organisers across a widening range of countries, languages and cultures.

For this reason, the main focus of this section is those more recent studies which have provided descriptions of performance in reading and numeracy at various levels. Information from other studies will be summarised very briefly, not only because there is much less of it but also because recent information is clearly more relevant to the current situation.

4.1 'Single-criterion' studies

This section heading is meant to capture the fact that the authors of most of the earlier studies of literacy and numeracy provided only brief or even no descriptions of what 13- to 19-year-olds could do. The reports concentrated on discussing a single criterion, or at most a very few criteria, for acceptable performance, and/or the average and range of standardised scores, and/or 'pass rates', and/or the construction of a 'league table'. In the early WV reading surveys (including the NSHD age 15 survey of 1961), for example, 'illiteracy' was defined as a score of 11 or fewer correct answers on the 35-item test and/or a reading age of 7 or below, and 'semi-literacy' was defined as a reading age of between 7 and 9.

Studies which focused on one or a few criteria were the WV and NS6 reading surveys. The APU reading surveys mainly reported in terms of average standardised scores and standard deviations; individual reports did give richer descriptions of performance on some tasks, but provided few generalisations across tasks and studiously avoided value judgments.

There are several sources of useful over-time trends in attainment at various levels which however provide no interpretation at all: the Yellis vocabulary and maths tests, the KS3 national tests, and GCSE results – the last of these is translated by the media every year into a 'pass rate'.

Almost all the earlier international studies provided league tables and not much else; in this group are the IEA reading surveys of 1960 and 1971 and the maths surveys called FIMS, IAEP1 and IAEP2, though some of these maths surveys (and SIMS) do give some

historically useful information on performance in separate areas of maths. The international survey of written composition of 1983, as mentioned at the beginning of Chapter 2, and the SIMS maths survey of 1981 were deliberately designed to prevent the construction of even a league table.

All the sources of data so far mentioned in this section provided school-level information. At adult level, surveys which provided only attainment data and no interpretation were the ACACE numeracy survey (1981), the International Numeracy Survey (1996) and the Basic Skills Agency's survey of literacy and numeracy needs (1996–97).

Other sources did provide fuller descriptions, and these are now presented, first for reading, then for numeracy, with an intervening glance at the dearth of evidence on writing.

Note on scales and levels
As pointed out in Section 3.1.4, the numbering of IALS levels (which also apply in PISA) and UK National Qualifications Framework levels is out of step by one. Where necessary in what follows, both scales are shown.

4.2 Reading

There are four relevant surveys: PISA 2000 and PISA 2006 for 15-year-olds, and IALS and the *Skills for Life* survey for young adults.

PISA 2000, 2006
Virtually the same descriptions of performance in reading at different levels were given in the reports of PISA 2000, 2003 and 2006; therefore, although the results from PISA 2003 are not used in this report, in this section the descriptions from the 2003 survey are reproduced, in Table 4.1 (source: OECD, 2004, Figure 6.1). Included are some percentages from PISA 2000 and PISA 2006.

IALS, 1996
Desjardins *et al.* (2005) gave the definitions shown in Table 4.2 of performance at each of the five levels in prose and document literacy.

The Skills for Life Survey, 2002–03
Unlike the two previous surveys, the *Skills for Life* survey did not differentiate performance above (UK) Level 2, but did differentiate performance at three sub-levels within UK Entry level (IALS Level 1). The brief characterisations of each level are shown in Table 4.3 (source: Williams *et al.*, 2003).

Table 4.1: Descriptions of the five levels of proficiency in reading literacy in PISA 2003

General	Continuous texts	Non-continuous texts
IALS/international Level 5 (UK Level 4)		
Locate and possibly sequence or combine multiple pieces of deeply embedded information, some of which may be outside the main body of the text. Infer which information in the text is relevant to the task. Deal with highly plausible and/or extensive competing information. Either construe the meaning of nuanced language or demonstrate a full and detailed understanding of a text. Critically evaluate or hypothesise, drawing on specialised knowledge. Deal with concepts that are contrary to expectations and draw on a deep understanding of long or complex texts.	Analyse texts whose discourse structure is not obvious or clearly marked, in order to discern the relationship of specific parts of the text to its implicit theme or intention.	Identify patterns among many pieces of information presented in a display which may be long and detailed, sometimes by referring to information external to the display. The reader may need to realise independently that a full understanding of the section of text requires reference to a separate part of the same document, such as a footnote.
IALS/international Level 4 (UK Level 3)		
Locate and possibly sequence or combine multiple pieces of embedded information, each of which may need to meet multiple criteria, in a text with familiar context or form. Infer which information in the text is relevant to the task. Use a high level of text-based inference to understand and apply categories in an unfamiliar context, and to construe the meaning of a section of text by taking into account the text as a whole. Deal with ambiguities, ideas that are contrary to expectation and ideas that are negatively worded. Use formal or public knowledge to hypothesise about or critically evaluate a text. Show accurate understanding of long or complex texts.	Follow linguistic or thematic links over several paragraphs, often in the absence of clear discourse markers, in order to locate, interpret or evaluate embedded information or to infer psychological or metaphysical meaning.	Scan a long, detailed text in order to find relevant information, often with little or no assistance from organisers such as labels or special formatting, to locate several pieces of information to be compared or combined.

Table 4.1: Descriptions of the five levels of proficiency in reading literacy in PISA 2003, cont.

General	Continuous texts	Non-continuous texts
IALS/international Level 3 (UK Level 2) Percentage of age-group in England at this level or above in PISA 2000, 2006: 67%, 59%		
Locate, and in some cases recognise the relationship between, pieces of information, each of which may need to meet multiple criteria. Deal with prominent competing information. Integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. Compare, contrast or categorise taking many criteria into account. Deal with competing information. Make connections or comparisons, give explanations, or evaluate a feature of text. Demonstrate a detailed understanding of the text in relation to familiar, everyday knowledge, or draw on less common knowledge.	Use conventions of text organisation, where present, and follow implicit or explicit logical links such as cause and effect relationships across sentences or paragraphs in order to locate, interpret or evaluate information.	Consider one display in the light of a second, separate document or display, possibly in a different format, or combine several pieces of spatial, verbal and numeric information in a graph or map to draw conclusions about the information represented.
IALS/international Level 2 (UK Level 1) Percentage of age-group in England at this level in PISA 2000, 2006: 20%, 23%		
Locate one or more pieces of information, each of which may be required to meet multiple criteria. Deal with competing information. Identify the main idea in a text, understand relationships, form or apply simple categories, or construe meaning within a limited part of the text when the information is not prominent and low-level inferences are required. Make a comparison or connections between the text and outside knowledge, or explain a feature of the text by drawing on personal experience and attitudes.	Follow logical and linguistic connections within a paragraph in order to locate or interpret information; or synthesise information across texts or parts of a text in order to infer the author's purpose.	Demonstrate a grasp of the underlying structure of a visual display such as a simple tree diagram or table, or combine two pieces of information from a graph or table.
IALS/international Level 1 (UK Entry level) Percentage of age-group in England at this level or below in PISA 2000, 2006: 13%, 19%		
Locate one or more independent pieces of explicitly stated information, typically meeting a single criterion, with little or no competing information in the text. Recognise the main theme or author's purpose in a text about a familiar topic, when the required information in the text is prominent. Make a simple connection between information in the text and common, everyday knowledge.	Use redundancy, paragraph headings or common print conventions to form an impression of the main idea of the text, or to locate information stated explicitly within a short section of text.	Focus on discrete pieces of information, usually within a single display such as a simple map, a line graph or a bar graph that presents only a small amount of information in a straightforward way, and in which most of the verbal text is limited to a small number of words or phrases.

Table 4.2: Descriptions of the five levels of proficiency in prose and document literacy in IALS 1996

Level	Prose literacy	Document literacy
IALS/ international Level 5 (UK Level 4)	Some tasks in this level require the respondent to search for information in dense text which contains a number of plausible distractors. Others ask respondents to make high-level inferences or use specialized background knowledge. Some tasks ask respondents to contrast complex information.	Tasks in this level require the respondent to search through complex displays that contain multiple distractors, to make high-level text-based inferences, and to use specialized knowledge
IALS/ international Level 4 (UK Level 3)	These tasks require respondents to perform multiple-feature matches and to integrate or synthesize information from complex or lengthy passages. More complex inferences are needed to perform successfully. Conditional information is frequently present in tasks at this level and must be taken into consideration by the respondent.	Tasks in this level, like those at the lower levels, ask respondents to perform multiple-feature matches, cycle through documents, and integrate information; however, they require a greater degree of inferencing. Many of these tasks require respondents to provide numerous responses but do not designate how many responses are needed. Conditional information is also present in the document tasks at this level and must be taken into account by the respondent.
	Percentage of 16–25 age-group in England at UK Level 2 or above: 52%	**Percentage of 16–25 age-group in England at UK Level 2 or above: 56%**
IALS/ international Level 3 (UK Level 2)	Tasks in this level tend to require respondents to make literal or synonymous matches between the text and information given in the task, or to make matches that require low-level inferences. Other tasks ask respondents to integrate information from dense or lengthy text that contains no organizational aids such as headings. Respondents may also be asked to generate a response based on information that can be easily identified in the text. Distracting information is present, but is not located near the correct information.	Some tasks in this level require the respondent to integrate multiple pieces of information from one or more documents. Others ask respondents to cycle through rather complex tables or graphs which contain information that is irrelevant or inappropriate to the task.

Table 4.2: Descriptions of the five levels of proficiency in prose and document literacy in IALS 1996, cont.

Level	Prose	Document
	Percentage of 16–25 age-group in England at UK Level 1: 30%	**Percentage of 16–25 age-group in England at UK Level 1: 26%**
IALS/ international Level 2 (UK Level 1)	Some tasks in this level require respondents to locate a single piece of information in the text; however, several distractors or plausible but incorrect pieces of information may be present, or low-level inferences may be required. Other tasks require the respondent to integrate two or more pieces of information or to compare and contrast easily identifiable information based on a criterion provided in the question or directive.	Tasks in this level are more varied than those in Level 1 (UK Entry level). Some require the respondents to match a single piece of information; however, several distractors may be present, or the match may require low-level inferences. Tasks in this level may also ask the respondent to cycle through information in a document or to integrate information from various parts of a document.
	Percentage of 16–25 age-group in England at UK Entry level: 17%	**Percentage of 16–25 age-group in England at UK Entry level: 18%**
IALS/ international Level 1 (UK Entry level)	Most of the tasks in this level require the respondent to read relatively short text to locate a single piece of information which is identical to or synonymous with the information given in the question or directive. If plausible but incorrect information is present in the text, it tends not to be located near the correct information.	Tasks in this level tend to require the respondent either to locate a piece of information based on a literal match or to enter information from personal knowledge onto a document. Little, if any, distracting information is present.

Table 4.3: Descriptions of levels of proficiency in reading in the *Skills for Life* survey, 2002–03

An adult classified at this (UK) level…	… has these skills	Percentage of 16–19 age-group in England at this level
Level 2 or above	• Understands a range of texts of varying complexity accurately and independently • Can obtain information of varying length and detail from different sources	43%
Level 1	• Understands short straightforward texts of varying length on a variety of topics accurately and independently • Can obtain information from different sources	41%
Entry level 3	• Understands short straightforward texts on familiar topics accurately and independently • Can obtain information from everyday sources	12%
Entry level 2	• Understands short straightforward texts on familiar topics • Can obtain information from short documents, familiar sources and signs and symbols	3%
Entry level 1	• Understands short texts with repeated language patterns on familiar topics • Can obtain information from common signs and symbols	2%

Commentary

The authors of the report on the British section of IALS (Carey *et al*., 1997) noted that the IALS measurement framework:

> is only one of many such classifications … Not only do different typologies identify different core or basic skills, but they also define attainment in those skills in different ways … there is no consensus on what skills comprise core/ basic or key skills and no easy way to map one classification system onto another.

Nevertheless, some attempt must be made, especially since the Leitch report (Leitch, 2006) defined functional literacy as 'Level 1 or better' and functional numeracy as 'Entry level 3 or better'. The most obvious first point is that all the early surveys, and some later ones, relied on very simple criteria for functional literacy which seem relatively simple to achieve. Also, as soon as more detailed descriptions are available they seem to define not just more detailed but more demanding capabilities at various levels.

If there has been such a rise in demands, then a steadily or gently rising graph, or even a largely flat one, in average levels in literacy could be seen as a success story: despite the raising of demands, underlying abilities were adequate to meet them. And there have been some substantial rises in recent years, at least in writing (KS3 English, GCSE English).

In one case it is known that a 'grade boundary' was deliberately adjusted upwards: relative to the Entry level/Level 1 boundary for reading in the previous Basic Skills Agency Communication Standards, the UK boundary with the same nomenclature in the year 2000 National Standards is distinctly higher, having been moved up to correspond with the IALS Level 1/Level 2 boundary (see Brooks *et al*., 2001, pp.121–2). In 1996–97 the BSA carried out a survey of literacy and numeracy needs throughout England; the results were not reported by age-group and are therefore not included in Chapter 3. However, nationally the results showed 15% of adults aged 16–60 having literacy below Level 1 of the Agency's Communication Standards, based on the threshold for Level 1 being defined as 19 out of the 22 items in the test having been answered correctly (Basic Skills Agency, 1997a). Once the

new National Standards had raised the Entry level/Level 1 boundary, in 2001 the Agency re-issued the results of the 1996–97 survey with the threshold for Level 1 raised to 20 answers correct out of 22: now the proportion of adults having literacy below (the new) Level 1 was 24%, almost exactly equal to the IALS figure of 23% (Basic Skills Agency, 2001). What this illustrates is the fragility of some estimates of poor literacy (and see Section 3.3.7 for the contrary adjustment to the figures for poor numeracy).

However, there does appear to be a basic correspondence between the three sets of criteria listed above at the levels where they are most easily compared, UK Levels 1 and 2. The detailed descriptions in IALS seem like expansions of those in *Skills for Life*, and those in PISA seem like even more elaborate versions of those in IALS. Level 1 can be seen as basically competent comprehension of not very complicated information, and Level 2 as somewhat more competent comprehension of rather more complex information.

The comparisons also reveal that Entry level does betoken weakness in reading comprehension. People at this level can handle only simple texts and straightforward questions on them where no distracting information is adjacent or nearby. Making inferences and understanding forms of indirect meaning (e.g. allusion, irony) are likely to be difficult or impossible. This is less than the functional literacy needed to partake fully in employment, family life and citizenship and to enjoy reading for its own sake. About 17% of young people in England are at this level. While this is lower than at some older ages in England, it is higher than in many other industrialised countries.

4.3 Numeracy

There are eight relevant surveys: TIMSS 1995, TIMSS(R) 1999, TIMSS 2003 and TIMSS 2007 for 14-year-olds, PISA 2000 and 2006 for 15-year-olds, and IALS and the *Skills for Life* survey for young adults.

The TIMSS surveys, 1995, 1999, 2003, 2007

In each of these surveys performance was described at four levels, namely the 90th, 75th, 50th and 25th percentiles, that is, for a typical pupil who is just in the top 10%, just in the top quarter, at the average, or just above the bottom quarter, **internationally**. The relevant descriptions are all shown in Table 4.4, and Table 4.5 gives the percentages of students in England who performed at or above each benchmark in each survey.

Table 4.4: Level descriptions for TIMSS surveys

Level	TIMSS 1995	TIMSS(R) 1999	TIMSS 2003	TIMSS 2007
Top 10% / Advanced International Benchmark	Organize information in problem-solving situations; solve time-distance-rate problems involving conversion of measures within a system; apply relationships – fractions and decimals, ratios, properties of geometric figures, and algebraic rules – to solve problems; solve word problems involving the percentage of increase.	Students can organize information, make generalizations, and explain solution strategies in non-routine problem-solving situations. They can organize information and make generalizations to solve problems; apply knowledge of numeric, geometric and algebraic relationships to solve problems (e.g. among fractions, decimals, and percents; geometric properties; and algebraic rules); and find the equivalent forms of algebraic expressions.	Students can organize information, make generalizations, solve non-routine problems and draw and justify conclusions from data. They can compute percent change and apply their knowledge of numeric and algebraic concepts and relationships to solve problems. Students can solve simultaneous linear equations and model simple situations algebraically. They can apply their knowledge of measurement and geometry in complex problem situations. They can interpret data from a variety of tables and graphs, including interpolation and extrapolation.	*Students can organize information, make generalizations, and solve non-routine problems.* They can solve a variety of ratio, proportion and percent problems. They can apply their knowledge of numeric and algebraic concepts and relationships. Students can express generalizations algebraically and model situations. They can apply their knowledge of geometry in complex problem situations. Students can derive and use data from several sources to solve multi-step problems.
Upper Quarter / High International Benchmark	Order, relate, multiply, and divide fractions and decimals; relate area and perimeter; understand simple probability; use knowledge of geometric properties to solve problems; identify algebraic expressions and solve equations with two variables.	Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They can order, relate and compute with fractions and decimals to solve word problems; solve multi-step word problems involving proportions with whole numbers; solve probability problems; use knowledge of geometric properties to solve problems; identify and evaluate algebraic expressions and solve equations with one variable.	Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They can order, relate, and compute with fractions and decimals to solve word problems, operate with negative integers, and solve multi-step word problems involving proportions with whole numbers. Students can solve simple algebraic problems including evaluating expressions, solving simultaneous linear equations, and using a formula to determine the value of a variable. Students can find areas and volumes of simple geometric	*Students can apply their understanding and knowledge in a wide variety of relatively complex situations.* They can relate and compute with fractions, decimals and percents, operate with negative integers, and solve word problems involving proportions. Students can work with algebraic expressions and linear equations. Students use knowledge of geometric properties to solve problems, including area, volume and angles. They can interpret data in a variety of graphs and tables and solve simple problems involving probability.

Level	TIMSS 1995	TIMSS(R) 1999	TIMSS 2003	TIMSS 2007
			shapes and use knowledge of geometric properties to solve problems. They can solve probability problems and interpret data in a variety of graphs and tables.	
Median / Intermediate International Benchmark	Use understanding of rounding in problem situations; perform basic operations with familiar fractions; understand place value of decimal numbers; understand measurement in several settings; locate data in charts and graphs to solve word problems; know and use simple properties of geometric figures to solve problems; identify algebraic expressions and solve equations with one variable.	Students can apply basic mathematical knowledge in straightforward situations. They can add or subtract to solve one-step word problems involving whole numbers and decimals; identify representations of common fractions and relative sizes of fractions; solve for missing terms in proportions; recognize basic notions of percents and probability; use basic properties of geometric figures; read and interpret graphs, tables, and scales; and understand simple algebraic relationships.	Students can apply basic mathematical knowledge in straightforward situations. They can add, subtract, or multiply to solve one-step word problems involving whole numbers and decimals. They can identify representations of common fractions and relative sizes of fractions. They understand simple algebraic relationships and solve linear equations with one variable. They demonstrate understanding of properties of triangles and basic geometric concepts including symmetry and rotation. They recognize basic notions of probability. They can read and interpret graphs, tables, maps, and scales.	*Students can apply basic mathematical knowledge in straightforward situations.* They can add and multiply to solve one-step word problems involving whole numbers and decimals. They can work with familiar fractions. They understand simple algebraic relationships. They demonstrate understanding of properties of triangles and basic geometric concepts. They can read and interpret graphs and tables. They recognize basic notions of likelihood.
Lower Quarter / Low International Benchmark	Understand different representations of fractions – verbal and decimal; add and subtract decimals with the same number of decimal places; read, locate, and compare data in charts and graphs; calculate average of whole numbers.	Students can do basic computations with whole numbers. The few items that anchor at this level provide some evidence that students can add, subtract, and round with whole numbers. When there are the same numbers of decimal places, they can subtract with multiple regrouping. Students can round whole numbers to the nearest hundred. They recognize some	Students have some basic mathematical knowledge.	*Students have some knowledge of whole numbers and decimals, operations, and basic graphs.*

Level	TIMSS 1995	TIMSS(R) 1999	TIMSS 2003	TIMSS 2007
		basic notation and terminology.		

Sources: TIMSS 1995 – Kelly *et al*. (2000); TIMSS(R) – Mullis *et al*. (2000); TIMSS 2003 – Ruddock *et al*. (2004); TIMSS 2007 – Mullis *et al*. (2008)

Table 4.5: Percentage of Year 9 students in England at or above each international benchmark in mathematics in the TIMSS surveys

At or above:	Inter-nationally	TIMSS 1995	TIMSS(R) 1999	TIMSS 2003	TIMSS 2007
Top 10% Benchmark	10%	7%	7%	5%	8%
Upper Quarter Benchmark	25%	20%	24%	26%	35%
Median Benchmark	50%	48%	58%	61%	69%
Lower Quarter Benchmark	75%	77%	89%	90%	90%

Sources: (1995) http://timss.bc.edu/timss1995i/TIMSSPDF/PSA_M_all.pdf
(1999) http://timss.bc.edu/timss1999i/pdf/T99i_Math_All.pdf
(2003) http://timss.bc.edu/PDF/t03_download/T03INTLMATRPT.pdf
(2007) http://timssandpirls.bc.edu/TIMSS2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf

In isolation, Table 4.5 might suggest that there was a substantial rising trend, especially in the middle and lower ranges of the distribution. However, this is true only of the changes from 2003 to 2007 since, as shown in Section 3.3.2, England's average score barely changed between the first three surveys. Table 4.5 does, however, show that in all four surveys more students in England were above the lower quarter benchmark than was the case internationally.

PISA 2000

Descriptions were provided for three levels for PISA 2000. These are given in Table 4.6.

Table 4.6: Level descriptions for mathematical literacy in PISA 2000

Towards the top end, around 750 points, students typically take a creative and active role in their approach to mathematical problems. They interpret and formulate problems in terms of mathematics, can handle more complex information, and can negotiate a number of processing steps. Students at this level identify and apply relevant tools and knowledge (frequently in an unfamiliar problem context), use insight to identify a suitable way of finding a solution, and display other higher-order cognitive processes such as generalisation, reasoning and argumentation to explain and communicate results.
At around 570 points on the scale, students are typically able to interpret, link and integrate different representations of a problem or different pieces of information; and/or to use and manipulate a given model, often involving algebra or other symbolic representations; and/or verify or check given propositions or models. Students typically work with given strategies, models or propositions (e.g. by recognising and extrapolating from a pattern), and they select and supply relevant mathematical knowledge in order to solve a problem that may involve a small number of processing steps.
At the lower end of the scale, around 380 points, students are usually able to complete only a single processing step consisting of reproducing basic mathematical facts or processes, or applying simple computational skills. Students typically recognise information from diagrammatic or text material that is familiar and straightforward and in which a mathematical formulation is provided or readily apparent. Any interpretation or reasoning typically involves recognition of a single familiar element of a problem. The solution calls for application of a routine procedure in a single processing step.

Source: OECD (2001)

In PISA 2000, 750 points represented the highest-performing students of all; 570 points was about 100 points above the international average – about 30% of UK students were at or above this point (25% internationally); and only about 6% of UK students were at or below 380 points (about 15% internationally).

PISA 2006

For PISA 2003 and 2006 the framework was expanded to six levels; their descriptions (from a source on PISA 2006) are given in Table 4.7, and the percentages of pupils in England and across all OECD countries in 2006 are shown in Table 4.8. As noted earlier for reading, the numbering of levels in the international and UK scales is out of step by one.

Table 4.7: Level descriptions for mathematical literacy in PISA 2006

Summary descriptions of the six proficiency levels in mathematics

Level	Lower score limit	What students can typically do
6	669.3	At Level 6 students can conceptualise, generalise, and utilise information based on their investigations and modelling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations.
5	607.0	At Level 5 students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.
4	544.7	At Level 4 students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic ones, linking them directly to aspects of real-world situations. Students at this level can utilise well-developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.
3	482.4	At Level 3 students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning.
2	420.1	At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.
1	357.8	At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Source: OECD (2007), p.312

Table 4.8: Percentage of pupils at each level in mathematics in PISA 2006, in England and internationally

International level	UK NQF level	England	All OECD
6	5	3%	3%
5	4	9%	10%
4	3	18%	19%
3	2	26%	24%
2	1	25%	22%
1 or below	Entry or below	20%	21%

Thus, consistently with the fact that the mean score for England was very close to the international mean, the distribution of pupils across levels in England was almost identical to that for the OECD as a whole.

IALS, 1996

Desjardins *et al.* (2005) gave the definitions shown in Table 4.9 of performance at each of the five levels in 'quantitative literacy'.

Table 4.9: Descriptions of performance at five levels of 'quantitative literacy', IALS 1996

IALS/ international Level 5 (UK Level 4)	Tasks in this level require respondents to understand complex representations and abstract and formal mathematical and statistical ideas, possibly embedded in complex texts. Respondents may have to integrate multiple types of mathematical information, draw inferences, or generate mathematical justification for answers.	
IALS/ international Level 4 (UK Level 3)	Tasks at this level require respondents to understand a broad range of mathematical information of a more abstract nature represented in diverse ways, including in texts of increasing complexity or in unfamiliar contexts. These tasks involve undertaking multiple steps to find solutions to problems and require more complex reasoning and interpretation skills, including comprehending and working with proportions and formulas or offering explanations for answers.	
IALS/ international Level 3 (UK Level 2)	Tasks in this level require the respondent to demonstrate understanding of mathematical information represented in a range of different forms, such as in numbers, symbols, maps, graphs, texts and drawings. Skills required involve number and spatial sense, knowledge of mathematical patterns and relationships, and the ability to interpret proportions, data and statistics embedded in relatively simple texts where there may be distractors. Tasks commonly involve undertaking a number of processes to solve problems.	**Percentage of 16–25 age-group at or above this level in England: 48%**
IALS/ international Level 2 (UK Level 1)	Tasks in this level are fairly simple and relate to identifying and understanding basic mathematical concepts embedded in a range of familiar contexts where the mathematical content is quite explicit and visual with few distractors. Tasks tend to include one-step or two-step processes and estimations involving whole numbers, benchmark percents and fractions, interpreting simple graphical or spatial representations, and performing simple measurements.	**Percentage of 16–25 age-group at this level in England: 29%**
IALS/ international Level 1 (UK Entry level)	Tasks in this level require the respondent to show an understanding of basic numerical ideas by completing simple tasks in concrete, familiar contexts where the mathematical content is explicit with little text. Tasks consist of simple, one-step operations such as counting, sorting dates, performing simple arithmetic operations, or understanding common and simple percents such as 50%.	**Percentage of 16–25 age-group at this level in England: 22%**

The Skills for Life Survey, 2002–03

Unlike all the previous surveys, the *Skills for Life* survey did not differentiate performance above (UK) Level 2, but did differentiate performance at three sub-levels within UK Entry level (IALS Level 1). The brief characterisations of each level are shown in Table 4.10 (source: Williams *et al.*, 2003).

Table 4.10: Descriptions of levels of proficiency in numeracy in the *Skills for Life* survey, 2002–03

An adult classified at this (UK) level…	… has these skills	Percentage of 16–19 age-group in England at this level
Level 2 or above	Understands mathematical information used for different purposes and can independently select and compare relevant information from a variety of graphical, numerical and written material	23%
Level 1	Understands straightforward mathematical information used for different purposes and can independently select relevant information from given graphical, numerical and written material	27%
Entry level 3	Understands information given by numbers, symbols, diagrams and charts used for different purposes and in different ways in graphical, numerical and written material	29%
Entry level 2	Understands information given by numbers, symbols, simple diagrams and charts in graphical, numerical and written material	15%
Entry level 1	Understands information given by numbers and symbols in simple graphical, numerical and written material	6%

Commentary

As for reading, the most obvious first point is that all the early surveys, and some later ones, relied on very simple criteria for functional numeracy which seem relatively simple to achieve. Also, as soon as more detailed descriptions are available they seem to define not just more detailed but more demanding capabilities at various levels. This seems to be especially true of the *Skills for Life* survey, where the description of Entry level 2 seems parallel to the description of IALS level 1 – this would appear to justify the decision in Chapter 3, and in recent government documents, to treat the *Skills for Life* survey figure for Entry level 2 and below as its estimate of poor numeracy. **If demands were raised, then, as for literacy, the steadily or gently rising or flat graphs can be seen as a success story, and this applies even more strongly to the substantial rises in recent years (KS3 maths, GCSE maths).**

Given this, there does appear to be a basic correspondence between the criteria for the lowest levels of mathematical attainment listed above, namely the TIMSS Lower Quarter Benchmark, the PISA 2000 lower end of the scale, PISA 2006 level 1, IALS Level 1, and *Skills for Life* Entry levels 2 and 1. What all these describe is very basic competence in maths, mainly limited to arithmetical computations and some ability to comprehend and use other forms of mathematical information. While this is valuable, it is clearly not enough to deal confidently with many of the mathematical challenges of contemporary life. From the various surveys at age 16+ it can be estimated that about 22% of young people in England are at this level. While this is lower than at some older ages in England, it is higher than in many other industrialised countries.

Before leaving the topic of criteria and standards, the point must be made that most young people in England do have functional skills in both literacy and numeracy, and that those with the highest skills are up with the best in the world.

4.4 Implications for policy

- Efforts to improve the initial teaching of literacy and numeracy to young children must continue.
- Children who fall behind in the early stages must be identified and given targeted catch-up programmes immediately.
- Family literacy and numeracy programmes can make a contribution to preventing early failure.
- Effective programmes should be maintained and not funded only in the short term.
- The search for effective ways of raising levels of functional literacy and numeracy should continue.
- Given that a light sample monitoring system seems to be being established at adult level there is a case for re-establishing one at school level.

4.5 Caveat

Who decides what adequate literacy and numeracy are? The findings in this report are based on the criteria described (or implied), and the descriptions are based on what 'experts' in the field think other people should be able to do. Manifestly, neither the findings nor the criteria are based on evidence of what people actually need to be able to do. Ideally, judgments on adequate literacy and numeracy should be based on large-scale surveys of what people actually need to be able to do with reading, writing and maths as employees, family members and citizens and for their own purposes, and on the skills that they are observed using to meet those needs.

The criteria used in IALS were based, historically, on just such a survey that is known to have been carried out in the USA in the 1980s. No further details of this could be obtained, but it would in any case be thoroughly out of date, particularly with respect to the exponential growth of computer-based skills. A similar survey was conducted in Canada in 1989 by Statistics Canada; this was called the Literacy Skills Used in Daily Activities survey – see www.statcan.ca/bsolc/english/bsolc?catno=89M0008X – again this would not be up to date, and neither survey is necessarily directly relevant to England.

If the relevant field in England is defined as the 2x4 matrix crossing literacy and numeracy with people as employees, family members, citizens and private individuals, only one of the eight cells has been adequately surveyed for England. A detailed study of the demands on numeracy in the workplace was carried out several years ago – see Hoyles *et al*. (2002) – and trends in this area over the 20-year period 1986–2006 were analysed by Felstead *et al*. (2007). The other research remains to be done.

In this connection it is encouraging to note that the PIAAC survey in 2011 will include an element in which participants self-report the demands which everyday literacy and numeracy make on them, including in the workplace. This would ideally provide a reality check on the assumptions made by the experts devising the domain specifications and tasks for the study.

Meanwhile, all ascriptions of poor literacy and numeracy, whether to 13- to 19-year-olds or to adults, should be made with due humility – those who have the power to decide what other people should be able to do have imposed their views on those who do not.

References

Assessment of Performance Unit (APU) (1988). *Mathematical Development: a review of monitoring in mathematics, 1978–1982*, Parts 1 and 2. Slough: NFER.

Basic Skills Agency (1995). *Older and Younger: the basic skills of different age groups.* London: Basic Skills Agency.

Basic Skills Agency (1997a). *Adults' Basic Skills: Benchmark Information on the Scale of Need in Different Areas.* [CD-ROM]. London: Basic Skills Agency.

Basic Skills Agency (1997b). *International Numeracy Survey: a Comparison of the Basic Numeracy Skills of Adults 16–60 in Seven Countries*. London: Basic Skills Agency.

Basic Skills Agency (2001). *Adults' Basic Skills: Benchmark Information on the Scale of Need in Different Areas*, revised version. [CD-ROM]. London: Basic Skills Agency.

Bradshaw, J., Sturman, L., Vappula, H., Ager, R. and Wheater, R. (2007). *Achievement of 15-year-olds in England: PISA 2006 national report*. Slough: NFER.

Brooks, G., Davies, R., Duckett, L., Hutchison, D., Kendall, S. and Wilkin, A. (2001). *Progress in Adult Literacy: Do learners learn?* London: The Basic Skills Agency.

Brooks, G., Gorman, T.P. and Kendall, L. (1993). *Spelling It Out: the spelling abilities of 11- and 15-year-olds*. Slough: NFER.

Brooks, G., Heath, K. and Pollard, A. (2005). *Assessing adult literacy and numeracy: a review of research instruments (research report).* London: National Research and Development Centre for adult literacy and numeracy.

Brooks, G. and Pilling, M. (forthcoming). *The impact of Skills for Life on adult literacy, language and numeracy learners. Final report on analysis of new quantitative data*. London: NRDC.

Bynner, J. and Parsons, S. (2006). *New Light on Literacy and Numeracy. Full report*. London: NRDC. http://www.nrdc.org.uk/publications_details.asp?ID=78

Carey, S., Low, S. and Hansbro, J. (1997). *Adult Literacy in Great Britain: a survey of adults aged 16–65 in Great Britain carried out by Social Survey Division of ONS as part of the International Adult Literacy Survey (IALS).* London: The Stationery Office.

Cresswell, M. and Gubb, J. (1987). *The Second International Mathematics Study in England and Wales (International Studies in Pupil Performance Series).* Windsor: NFER-Nelson.

Desjardins, R., Murray, S., Clermont, Y. and Werquin, P. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Statistics Canada and OECD, Ottawa and Paris.
http://www.statcan.ca/english/freepub/89-603-XIE/2005001/pdf/89-603-XWE-part1.pdf

Douglas, J.W.B., Ross, J.M. and Simpson, H.R. (1968). *All Our Future.* London: Peter Davies Ltd.

Felstead, A., Gallie, D., Green, F. and Zhou, Y. 2007. *Skills at work, 1986–2006*. Oxford and Cardiff: Economic and Social Research Council Centre on Skills, Knowledge and Organisational Performance.

Fogelman, K. (ed.) (1976). *Britain's Sixteen-Year-Olds*. London: National Children's Bureau.

Fogelman, K. (ed.) (1983). *Growing Up in Great Britain.* London and Basingstoke: Macmillan.

Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A. and Walker, D.A. (1962). *Educational Achievements of Thirteen-year-olds in Twelve Countries*. Hamburg: UNESCO Institute for Education.

Foxman, D. (1998). *TOPIC*, Issue no.19 (Spring), Item 2, *Monitoring trends in Numeracy, 1953–1995*. Slough: NFER.

Foxman, D., Hutchison, D. and Bloomfield, B. (1990a). *The APU Experience 1977–1990.* London: SEAC.

Foxman, D., Ruddock, G. and McCallum, I. (1990b). *APU Mathematics Monitoring, 1984–88 (Phase 2): a summary of findings, conclusions and implications* (Assessment Matters No. 3). London: SEAC.

Gill, B., Dunn, M. and Goddard , E. (2002). *Student achievement in England: Results in reading, mathematical and scientific literacy among 15-year-olds from OECD PISA 2000 study*. London: The Stationery Office.
http://www.statistics.gov.uk/downloads/theme_education/PISA_2000.pdf

Goldstein, H. (2004). 'International comparisons of student attainment: some issues arising from the PISA study.' *Assessment in Education: principles, policy and practice*, **11**, 3, 319–30.
http://www.mlwin.com/hgpersonal/PISA%20reading%20literacy.pdf

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., *et al*. (1993). 'A multilevel analysis of school examination results.' *Oxford Review of Education*, **19**, 425–33.

Gorman, T.P. (1981). 'A survey of attainment and progress of learners in adult literacy schemes', *Educational Research*, **23**, 3, 190–8.

Gorman, T. and Moss, N. (1979). *Survey of Attainment and Progress in Adult Literacy Schemes.* Slough: NFER.

Gorman, T.P., Purves, A.C. and Degenhart, R.E. (eds) (1988). *The IEA Study of Written Composition I: the International Writing Tasks and Scoring Scales.* Oxford: Pergamon.

Gorman, T.P., White, J., Brooks, G. and English, F. (1991). *Language for Learning: a summary report on the 1988 APU surveys of language performance* (Assessment Matters No.4). London: School Examinations and Assessment Council.

Gorman, T.P., White, J., Brooks, G., Maclure, M. and Kispal, A. (1988). *Language Performance in Schools: a review of APU language monitoring, 1979–83.* London: HMSO.

Great Britain. Department for Education and Employment. Working group on post-school basic skills (1999). *A Fresh Start: Improving Literacy and Numeracy* (Moser Report). London: D*f*EE.

Green, S., Elliott, G. and Johnson, N. (2008). ' "Aspects of Writing": Beyond an atomistic approach to evaluate qualities of features of writing.' Paper presented at International Association for Educational Assessment Annual Conference, Cambridge, September. Available at http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/149773_Microsoft_Word_-_Green.pdf

Green, S. and Oates, T. (2007). 'How to promote educational quality through national assessment systems.' Paper presented at International Association for Educational Assessment Annual Conference, Baku, Azerbaijan, September. Available at http://www.cambridgeassessment.org.uk/ca/digitalAssets/113988_IAEApaper07FINAL_SGr_2_.pdf

Gubb, J., Gorman, T.P. and Price, E. (1987). *The Study of Written Composition in England and Wales.* Windsor: NFER-Nelson.

Harmon, M., Smith, T.A., Martin, M.O, Kelly, D.L., Beaton, A.E, Mullis, I.V.S, Gonzalez, E.J. and Orpwood, G. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMSS)*. International Association for the Evaluation of Educational Achievement. http://timss.bc.edu/timss1995i/TIMSSPDF/PAreport.pdf

Hoyles, C., Wolf, A., Molyneux-Hodgson, S. and Kent, P. 2002. *Mathematical skills in the workplace: final report to the Science, Technology and Mathematics Council*. London: Institute of Education, University of London/STM Council.

Hurry, J. (1999). 'Children's reading levels.' *Journal of Child Psychology and Psychiatry*, **40**, 2, 143–50.

Husén, T. (ed.) (1967*). International Study of Achievement in Mathematics: A Comparison of Twelve Countries. (International Project for the Evaluation of Education Achievement (IEA) Phase I.)* New York: John Wiley & Sons.

Kelly, D.L., Mullis, I.V.S. and Martin, M.O. (2000). *Profiles of Student Achievement in Mathematics at the TIMSS International Benchmarks: U.S. Performance and Standards in an International Context.* Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Keys, W., Harris, S. and Fernandes, C. (1996a). *Third International Mathematics and Science Study, National Reports. Appendices: Additional Information Relating to the Study in England.* Slough: NFER.

Keys, W., Harris, S. and Fernandes. C. (1996b). *Third International Mathematics and Science Study, First National Report. Part 1: Achievement in Mathematics and Science at Age 13 in England.* Slough: NFER.

Lapointe, A.E., Mead, N.A and Askew, J.M. (1992). *Learning Mathematics, IAEP Report No. 22-CAEP-01*. Princeton NJ: Educational Testing Service.

Lapointe, A.E., Mead, N.A. and Phillips, G.W. (1989). *A World of Differences: An International Assessment of Mathematics and Science. Report of the First International Assessment of Educational Progress.* Report No. 19-CAEP-01. Princeton NJ: Educational Testing Service.

Leitch, S. (2006). *Prosperity for All in the Global Economy – world class skills. (Leitch Review of Skills Final Report)* Norwich: The Stationery Office. http://www.hm-treasury.gov.uk/independent_reviews/leitch_review/review_leitch_index.cfm

Massey, A.J., Elliott, G.L. and Johnson, N.K. (2005). *Variations in Aspects of Writing in 16+ English Examinations between 1980 and 2004: Vocabulary, Spelling, Punctuation, Sentence Structure, Non-Standard English*. Cambridge: University of Cambridge Local Examinations Syndicate. http://www.cambridgeassessment.org.uk/research/confproceedingsetc/publication.2005-10-31.1289126542/file/

Massey, A., Green, S., Dexter, T. and Hamnett, L. (2002). *Comparability of national tests over time: key stage test comparisons between 1996 and 2001*. London: QCA. http://www.ofqual.gov.uk/files/6301_compare_nat_tests_96-01.pdf

Metcalf, H., Meadows, P., Rolfe, H. and Dhudwar, A. (2009). *Evaluation of the Impact of Skills for Life Learning: longitudinal survey of adult learners on College-based literacy and numeracy courses.* Final report. London: National Institute of Economic and Social Research.

Ministry of Education (1950). *Reading ability: some suggestions for helping the backward.* Ministry of Education pamphlet no.18. London: HMSO.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J. and Smith, T. A. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* International Association for the Evaluation of Educational Achievement. http://timss.bc.edu/timss1999i/math_achievement_report.html

Mullis, I.V.S., Martin, M.O., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Boston MA: TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/TIMSS2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf

OECD (2001) *Knowledge and Skills for Life. First results from PISA 2000*. Paris: OECD. http://www.pisa.oecd.org/dataoecd/44/53/33691596.pdf

OECD (2004). *Learning for tomorrow's world. First results from PISA 2003.* Paris: OECD. http://www.pisa.oecd.org/dataoecd/1/60/34002216.pdf

OECD (2005). *PISA 2003 Technical Report.* Paris: OECD. http://www.pisa.oecd.org/dataoecd/49/60/35188570.pdf

OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World.* Paris: OECD. http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

Rhys Warner, J., Vorhaus, J., Appleby, Y., Bathmaker, A-M., Brooks, G., Cole, P., Pilling, M. and Pearce, L. (2008). *The Learner Study: the impact of the Skills for Life strategy on adult literacy, language and numeracy learners. Summary report*. London: NRDC. http://www.nrdc.org.uk/publications_details.asp?ID=158

Robitaille, D.F. and Garden. R.A. (eds) (1989). *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics.* Oxford: Pergamon Press.

Rodgers, B. (1986). 'Change in the reading attainment of adults: a longitudinal study.' *British Journal of Developmental Psychology*, **4**, 1, 1–17.

Ruddock, G., Sturman, S., Schagen, I., Styles, B., Gnaldi, M. and Vappula, H. (2004). *Where England stands in the Trends in International Mathematics and Science Study (TIMSS): National report for England.* Slough: NFER. http://www.nfer.ac.uk/publications/other-publications/downloadable-reports/pdf_docs/timssfullreport.PDF

Sewell, B. (1981). *Use of Mathematics by Adults in Daily Life*. Leicester: ACACE.

Start, K. and Wells, B. (1972). *The Trend of Reading Standards.* Slough: NFER.

Statistics Canada (1989). *Literacy Skills Used in Daily Activities survey*. http://www.statcan.ca/bsolc/english/bsolc?catno=89M0008X

Thorndike, R.L. (1973). *Reading Comprehension Education in Fifteen Countries.* Stockholm: Almqvist and Wiksell.

Tymms, P. and Merrell, C. (2007). *Standards and Quality in English Primary Schools over Time: the national evidence* (Primary Review research survey 4/1). Cambridge: University of Cambridge Faculty of Education.

Vorhaus, J., Howard, U., Brooks, G., Bathmaker, A-M., and Appleby, Y. (2009). 'The impact of the "Skills for Life" infrastructure on learners: a summary of methods and findings.' In Reder, S. and Bynner, J. (eds) *Tracking Adult Literacy and Numeracy Skills: findings from longitudinal research.* New York and Abingdon, Oxon.: Routledge, 200–21.

Whetton, C., Ruddock, G. and Twist, L. (2007). *Standards in English Primary Education: the international evidence* (Primary Review research survey 4/2). Cambridge: University of Cambridge Faculty of Education.

Whetton, C., Twist, L. and Sainsbury, M. (2007). 'Measuring standards in primary English: the validity of PIRLS: a response to Mary Hilton.' *British Educational Research Journal*, **33**, 6.

Williams, J., with Clemens, S, Oleinikova, K. and Tarvin, K. (2003). *The Skills for Life survey: a national needs and impact survey of literacy, numeracy and ICT skills*. Norwich: The Stationery Office. http://www.dfes.gov.uk/research/data/uploadfiles/RR490.pdf