THE UNIVERSITY
of LIVERPOOL

ENGLAND'S EDUCATION

what can be learned by comparing countries?

Alan Smithers
Centre for Education and Employment Research
May 2004

Contents

Executive Summary

The Centre for Education and Employment Research at the University of Liverpool has been commissioned by the Sutton Trust to carry out an independent assessment of the methods and interpretations of international comparisons of educational performance. In particular, it was asked to consider the OECD's Programme for International Student Assessment (PISA) whose results in 2000 have been presented as indicating that education in England has been improving. Our evaluation has been shaped by the following questions:

- Is PISA technically sound?
- How do the results compare with other studies?
- Can reliable policy conclusions be drawn?
- Have the results been interpreted fairly?
- What is the value of international studies?
- What can reasonably be inferred about England's educational performance?

Methods and Findings

The history of international rankings is traced through to the two current main players, the International Association for the Evaluation of Educational Achievement (IEA) with its Trends in International Mathematics and Science Study (TIMSS) and the OECD with its PISA surveys.

The maths and science results of TIMSS and PISA from the same cohort of pupils tested one year apart are compared and found to be very different. The discrepancies are attributable mainly to differences in the aims, tests, types of questions, target populations, and response rates.

Although Finland does consistently well and other countries regularly perform poorly in international comparisons of literacy, the intercorrelations between the various studies are low reflecting different interpretations of literacy and different methodologies.

The maths, science and reading scores in PISA were found to vary similarly between and within countries pointing to considerable common ground between the tests.

PISA's indices of status and wealth do not correlate, and the description of England as a low equity country in which the poor perform worse than in many other countries is not supported by the evidence. The assumption that it is better to have a low spread of scores is also challenged. Higher dispersion, as in England, can arise through the top performers doing particularly well.

The differing performance of native and non-native pupils can have a big impact on overall scores. For example, while in maths England appears to have closed the gap on Switzerland in PISA 2000, there is still a 20 points difference in that country's favour when only native students are considered.

Major differences have been found between regions within countries, and particularly intriguing is the consistent but contrasting performance of Flemish and French Belgium across a range of studies.

Results, little discussed in PISA's reports, show that private schools, both independent and government funded, tended to do better than wholly maintained schools in most countries.

Countries differ considerably in the proportions of their populations shown as having successfully completed upper secondary education, though the criterion for doing so differs from attendance to achieving qualifications at a specified level. England which has one of the toughest criteria has a low rank, particularly for the youngest age group.

A country's performance in PISA was not found to correlate with its GDP per capita or its spend on education.

Evaluation

Technically Sound?

It is suggested the PISA programme has not done enough to demonstrate that its literacy tests are measures of 'knowledge and skills for life', and there has to be the suspicion that the maths and science tests are more tests of reading centred on elementary mathematical and scientific concepts.

The relative standing of countries in the different studies will have been affected by the countries taking part, the types of questions asked, whether the target population was age-based or grade-based, and poor response rates in some countries including England.

It was possible to identify a number of sources of unintended bias, including ignoring the degree of curriculum match, and the removal of difficult questions so capping high performing systems.

More meaning is sometimes imputed to the results than they hold, such as treating as real ranks the apparent ranks of mean scores which do not differ significantly.

Comparisons Between Studies

Differences between international studies do not, in the main, definitively demonstrate that the results of one are more credible than those of another, but they do make it difficult to conclude that there have been changes over time.

Even if differences can be substantiated, it should not be assumed that they reflect the school system or education policy since they could arise in other ways, for example, from economic conditions or immigration policies. Differences between countries could also be due to such non-school factors as inherited abilities, parental support or cultural values.

Relevance to Policy

The logic of PISA in seeking to derive policy conclusions for education systems from avowedly non-curriculum-based tests is questionable. Ignoring the curriculum does not eliminate it as a factor. The absence of curriculum match analysis severely reduces PISA's explanatory power.

Both TIMSS and PISA have been disappointing in explaining differences between countries. They have included measures of a number of school characteristics such as the ethos, the availability of teachers and the physical infrastructure, but hardly any consistent relationships have been found.

Interpretation

There has been a temptation on the part of both the PISA analysts and politicians to over-interpret the findings. The cloud of data generated becomes a canvas on to which the committed can project what they want to see.

The danger in PISA is that ambitious politicians in their desire to move up league tables will distort their countries' education systems in the direction of PISA's contestable view of education as it has been operationalized.

Value of International Studies

International studies have enabled broad-brush pictures to be drawn, but they also raise questions which they themselves may not be capable of answering. To tackle these there should be detailed studies within and between countries to move from description to practical understanding.

England's Educational Performance

Although results from international studies must be treated with caution, it would be wrong to dismiss them out of hand. Taking the findings at face value, it looks as though maths education should be the chief concern for England. Apart from the short test in PISA 2000 about which there have to be doubts, the country consistently comes well down the maths listings.

England has performed relatively well in international science tests since the 1990s, and there are hopeful signs from recent literacy studies. The independent schools, in fact, achieve the best scores in the world, but with the biggest gap from the maintained sector of any country.

However, England has not shown up well in studies of adult literacy and it emerges as having a relatively low rate of successful completion of upper secondary education. It is arguable, therefore, that any gains in school education run the risk of dissipation in the post-school years.

Introduction

1. Large-scale comparisons of performance of children in different countries on the same educational tests have been around for forty years, but outside the research community they have not attracted much attention. The two which have been published most recently, however, the Programme for International Student Assessment (PISA), a study of 15-year-olds, and Progress in International Reading Literacy (PIRLS), a study of 10-year-olds, have been very much in the news. In England this may not be unconnected with the country suddenly appearing to do very well, whereas during the nineties the messages were mixed. While previous science scores have been encouraging, maths performance has been consistently below average, and there have been some dire warnings about adult literacy.

2. Ministers and DfES officials are naturally delighted by the apparent upturn. David Normington (2002a), Permanent Secretary at the DfES, wrote in *The Education Journal,* in June 2002, that: "For those doubters who constantly seek to run down (our education performance), we now have the OECD/PISA study – the biggest ever international study of comparative performance of 15-year-olds in 32 countries – which shows UK fourth in science, seventh in literacy and eighth in maths. Only Finland and Canada are consistently ahead of the UK – and major countries like Germany, Italy and Spain are well behind". Charles Clarke, the Secretary of State for Education and Skills, in a speech to the annual conference of the National Association of Schoolmasters and Union of Women Teachers, on 24 April 2003 said, "The result of the commitment of your profession is that standards are rising. Recent international reports show that this country is third out of 35 developed nations for literacy standards at age 9, and seventh and eighth respectively for literacy and maths at age 15".

3. The league tables have been a recurring theme in the speeches and interviews of David Miliband, the Minister for School Standards. In his address (Miliband, 2003a) to the annual conference of the Association of Teachers and Lecturers in Blackpool, on 15 April 2003, he said, "International studies show our ten-year-olds achieving more than every other country in the industrialised world except Sweden and the Netherlands. At 15, students achieve in the top quartile in international comparisons." In an interview with *The Guardian* newspaper (Miliband, 2003b), published on 5 May 2003, he is reported as citing a study of the performance of 10-year-olds published in 1996 (just before his government came to office). "We were average: I can't remember, sixteenth or eighteenth. It was just redone and it's been published last week: we are the third best achievers for primary schools achievement at age 10 out of 35 countries". The uncertainty about where we were is understandable. The previous international study of reading had been in 1991 (Elley, 1994), and England had not taken part. The National Foundation for Educational Research (Brooks, Pugh and Schagen, 1996) did undertake a study in 1996 to try to see where we would have been if we had participated and concluded we would have fallen within a group of 13 countries with average scores.

4. At times ministers in their enthusiasm have come close to inferring improvements from what have been cross-sectional surveys, not longitudinal studies. But beyond this obvious criticism there are some important questions to be asked. Just what do

the numbers mean? Can the tests be taken as accurate measures of educational attainment in the way that, say, a thermometer is capable of precisely recording temperature, or are the numbers arrived at more to do with the nature of the tests or who has been tested? Then, even supposing the numbers do carry a lot of meaning, can the differences recorded between countries be attributed directly to their educational systems? It is at least possible that the results could reflect other differences such as culture, parental support or the gene pool. Although there is a reluctance to contemplate some of these possibilities this should not lead us to leap to the convenient conclusion (in the sense that it would be possible to make changes) that the differences in educational performance were caused by the educational system.

5. The Sutton Trust has commissioned the Centre for Education and Employment Research at the University of Liverpool to consider PISA in relation to other international studies to see what can reasonably be concluded from them. This paper reports our findings. We begin with a brief look at the historical background, before considering in detail two studies, PISA 2000 and TIMSS 1999 (the Third International Maths and Science Study Repeat). These tested children from the same age cohort, those born in 1984, one year apart in overlapping rafts of countries, but came up with very different pictures.

Background

6. International comparative studies of educational attainment, in their present form, have their origins in the 1950s when researchers from a dozen countries came together, sponsored by the United Nations Educational, Scientific and Cultural Organisation (UNESCO), to consider the feasibility of devising common tests for children in different countries. The embryonic International Association for the Evaluation of Educational Achievement (IEA), as the group became known, satisfied themselves that it was possible and, in 1964, embarked on a survey of mathematical attainment in 12 countries. As other major cross-national surveys have followed this became known as the First International Maths Study (FIMS). The whole programme, following the third sweeps, has been re-designated recently as Trends in International Mathematics and Science Study (TIMSS).

7. Table 1 sets out a chronology of the major international comparative studies with the sponsor and the number of countries and educational systems participating (some countries have more than one educational system as in the UK[*] itself). The table shows that many of the subsequent international surveys have also been conducted by the IEA (now based at the International Study Centre, Boston College, USA). The Educational Testing Service (ETS), a private non-profit education testing service, with its main offices in Princeton and Ewing, New Jersey, USA, entered the field in the late 1980s.

[*] The UK sometimes participates as such and sometimes as individual countries. This report focuses on England. Where there are results for England these are used. But when only UK data are available they are taken as applying to England since it comprises 83.6 per cent of the UK population.

8. More recently a major new programme has come on the scene. The Programme for International Student Assessment (PISA), a collaboration by the member countries of the Organisation for Economic Co-operation and Development (OECD), has been established "to measure how well young adults, age 15 and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's knowledge societies" (OECD, 2001). A Board of Participating Countries has been set up to determine policy priorities within which the PISA Consortium, led by the Australian Council for Educational Research (ACER), is responsible for the design and implementation of the surveys.

Table 1: Selected International[1] Comparative Studies in Education

Year(s) Conducted	Sponsor	Description	Countries/ Systems
1964	IEA	First International Maths Study (FIMS)	12
1970-71	IEA	First International Science Study (FISS) (part of six subjects study including reading, comprehension, French, English, civics)	19
1980-82	IEA	Second International Maths Study (SIMS)	20
1983-84	IEA	Second International Science Study (SISS)	24
1985	IEA	Written Composition	14
1988	ETS	First International Assessment of Education Progress (Maths and Science)	6
1989,1992	IEA	Computers in Education	22
1989-91	IEA	Pre-Primary Education Phase I	11
1990-91	IEA	Reading Literacy	32
1991	ETS	Second International Assessment of Education Progress (Maths and Science)	20
1994-98	Statistics Canada	International Adult Literacy Survey (IALS)	20
1994-95	IEA	Third International Maths and Science Study (TIMSS)	45
1998-99	IEA	Third International Maths and Science Study – Repeat (TIMSS-R)	38
1999	IEA	Civic Education	28
2000	OECD	Programme for International Student Assessment (PISA)	33
2001	IEA	Progress in International Reading Literacy (PIRLS)	35

Sources: IEA completed studies, www.iea.nl/Studies/; Reynolds, D. and Farrell, S. (1996) *Worlds Apart? A Review of International Surveys of Educational Achievement Involving England.* London: HMSO; OECD (2001) *Knowledge and Skills for Life, First Results from PISA 2000.* Paris: OECD; Twist, L., Sainsbury, M., Woodthorpe, A. and Whetton, C. (2003) *Reading All Over The World.* National Report for England on PIRLS, Progress in International Reading Literacy Study (PIRLS). Slough: NFER.

9. PISA has deliberately set out to establish itself as different. It made no attempt to carry forward TIMSS or the 1991 IEA Literacy Study. Like TIMSS it has included measures of school characteristics and attitudes to learning, but it ignored the only two school-based factors, 'opportunity to learn' and 'time on task', that the IEA studies had found to be linked to test scores. Micklewright (2003) at a seminar held

at the Royal Statistical Society on 3 December 2003 commented that, "PISA through its reluctance to acknowledge TIMSS seemed to have departed from normal good scientific practice of building on what had gone before." Both PISA and TIMSS were separately in the field in 2003, testing the same subjects in overlapping rafts of countries, which speaks of rivalry rather than collaboration. They are testing different years and it is unlikely that the same schools will be involved in both, but the duplication does add to the assessment burden on schools evident in their increasing reluctance to participate.

Comparison of TIMSS 1999 and PISA 2000

10. We begin our comparative analysis with the IEA's TIMSS study in 1999 and the OECD's PISA study in 2000 since they tested children from the same age cohort – those born in 1984. Maths and science were, in fact, minor parts of PISA 2000, and it was always the plan to assess them in more depth in later studies. But PISA in its reports has accorded the 2000 findings in these subjects almost as much prominence as those in reading, the main part of the study. This is surprising since, as we shall see, only some of the sample was tested in maths and science, the tests were a maximum of 30 minutes (compared to a maximum of two hours in reading), and the range of topics was very restricted.

Table 2: Maths Rankings

Country	TIMSS 1999[1]		PISA 2000[2]	
	Score	Rank	Score	Rank
Korea	587	1	547	2
Japan	579	2	557	1
Belgium (Flemish)	558	3	543	3
Hungary	532	4	488	11
Canada	531	5	533	6=
Russia	526	6	478	12
Australia	525	7	533	6=
Finland	520	8=	536	5
Czech Republic	520	8=	498	9
Latvia	505	10	463	13
USA	502	11	493	10
England	496	12	529	8
New Zealand	491	13	537	4
Italy	479	14	457	14

1. *TMISS 1999 International Student Achievement in Maths* http://www.iea.nl, Exhibit 1.1.

2. OECD (2001) *Knowledge and Skills for Life, First Results from PISA 2000.* Paris: OECD, Table 3.1, except for Belgium (Flemish) from Table B2.1 and England from *First Release International Student Assessment 2000,* London: DfES, 2001, Table 2.

11. Tables 2 and 3 show the respective rankings for maths and science in the fourteen countries which participated in both studies. The rankings for TIMSS 1999 and PISA 2000 do correlate significantly at the 5 per cent level of significance - in

maths at +0.59 and in science at +0.60. In other words, one test explains about a third of the variation in the other. Table 2 shows that in maths some countries did well in both TIMSS 1999 and PISA 2000, notably Korea, Japan and Flemish-speaking Belgium. Italy was at the foot of both rankings. But there were also some major discrepancies. New Zealand and England score well in PISA but not in TIMSS, but the reverse is true for Hungary and Russia.

12. The pattern is similar in science. Table 3 shows that again Korea and Japan did well in both, and Italy badly. As in maths, New Zealand was a big gainer, with England also apparently improving from a higher baseline. But the falls were even more dramatic. Hungary dropped from first to eleventh out of 14. The Czech Republic also fell appreciably. Since these gains or falls involved the same cohort of children tested within a year, it is inconceivable that they could be the result of changes in education policy. A first hypothesis has to be that it is the methods used which are producing the different scores.

Table 3: Science Rankings

Country	TIMSS 1999[1]		PISA 2000[2]	
	Score	Rank	Score	Rank
Hungary	552	1	496	11
Japan	550	2	550	2
Korea	549	3	552	1
Australia	540	4	528	6=
Czech Republic	539	5	511	9
England	538	6	533	4
Finland	535	7=	538	3
Belgium (Flemish)	535	7=	519	8
Canada	533	9	529	5
Latvia	503	10	460	13=
USA	515	11	499	10
New Zealand	510	12	528	6=
Russia	529	13	460	13=
Italy	493	14	478	12

1. *TMISS 1999 International Student Achievement in Science* http://www.iea.nl, Exhibit 1.1.

2. OECD (2001) *Knowledge and Skills for Life, First Results from PISA 2000.* Paris: OECD, Table 3.3, except for Belgium (Flemish) from Table B2.1 and England from *First Release International Student Assessment 2000,* London: DfES, 2001, Table 3.

13. It becomes interesting, therefore, to compare the results of TIMSS 1999 and PISA 2000 with those that have been obtained in previous studies. By and large the TIMSS scores are the more consistent with what went before. In Table 4 we summarize as percentile ranks England's relative standing from the earliest study in 1964 (the detailed rankings are given in Tables A1 and A2 in the Appendix). The position achieved is, of course, dependent on which other countries were taking part, and the full listing in the appendix makes it clear that the participants varied considerably from study to study. Nevertheless, it is possible to discern a pattern.

Up to and including TIMSS 1999 England tended to come in the bottom half for maths, but leapt up the table in PISA 2000. In the case of science, while England had fared better than in maths, there was also a jump in the relative position in PISA 2000.

Table 4: England's Relative Standing in Maths and Science in Different Studies[1]

Study[2]	Age	Maths			Science		
		N[3]	Rank		N[3]	Rank	
			Actual	%[4]		Actual	%[4]
PISA 2000	15	33	8	24	33	4	12
TIMSS 1999	14	28	20	71	26	9	35
TIMSS 1995	14	24	16	67	23	6	26
IAEP2 1991	13	17	11=	65	17	9=	53
SIMS 1980-82/SISS 1983-84	13/14	16	10	63	17	16	94
FIMS 1964/FISS 1970-71	13/14	12	6	50	13	9	69

1. Table summarizes data of A1 and A2 in the Appendix, which also give the sources.
2. For full titles see Table 1
3. Number of countries included in comparison.
4. Ranks adjusted to base 100 for comparison.

14. What appears to be the case for England is borne out by other countries, as we can see from Tables A1 and A2 (in the Appendix). Other Anglophone countries, including Australia, Canada, Scotland, New Zealand also showed up much better in PISA. Conversely, the Eastern European nations fared badly compared with their previous results. The contrast is particularly stark in the case of Hungary, but is also evident for Russia and the Czech Republic.

15. Not all countries showed this shift. Japan and Korea were consistently in the top group, often joined by other Asian nations when they took part. This raises the possibility that inherited ability, character and culture are more important than the education system *per se*. This thought is further prompted by the success of the Chinese in English education. Data provided by the DfES to the House of Commons Education and Skills Committee (2003, Annex G, Ev 120) show that, in maintained schools, 73 per cent of pupils from a Chinese background achieved 5+ GCSEs at grades A*-C compared with an average of 51 per cent across all ethnic groups, including home students. But there are also data which point to the potential impact of educational systems. The TIMSS results for Hong Kong showed it to be among the top performers in maths, but to do less well in science. This has been attributed to Hong Kong's 'English-style' education system where it was possible to drop science after age 14 (Law, 2002).

16. Some other trends in Tables A1 and A2 are also notable. Curiously Israel - which was among the highest performers in the early studies - has progressively dropped down the tables. This could be associated with the inclusion of children in the occupied territories, though there have also been changes to the maths curriculum. Italy has been stuck at the bottom except for the IAEP2 study when only one district, Emilio-Romagna in the north-east, was sampled. In PISA 2000, the north east of Italy was not far behind high-performing Finland, so the representativeness

of the sampling within a country has an important bearing on the results obtained. Canada's uneven placing across the studies has much to do with which provinces participated.

Differences in Approach and Methods

17. Given the lack of continuity between the TIMSS and PISA studies, it is possible that the explanation for the contrasting results lies more in the different approaches than actual changes in educational policies or practices. The surveys differ in the countries participating, and in aims, tests, types of questions and target populations.

Aims

18. The aims of TIMSS and PISA are explicitly different. Whereas TIMSS focused on the extent to which students have mastered maths and science as they appear in school curricula, PISA aimed to capture "the ability to use knowledge and skills to meet real-life challenges." (OECD, 2001). In evidence to the House of Commons Education and Skills Committee, Barry McGaw (2002a), the Director for Education of the OECD, characterised the difference as TIMSS being interested to discover, "what science have you been taught and how much have you learned?", while for PISA it was "what can you do with the science you have been taught?"

19. The OECD is an economic body, so it is not surprising that its prime interest in education should be its contribution to higher productivity. But, less obviously, it has operationalized applied skills as 'literacy'. The notion has been broadened from its everyday meaning of being able to read and write to include 'mathematical literacy' and 'scientific literacy'. 'Mathematical literacy' has been defined as "the capacity to identify, to understand, and to engage in maths and make well-founded judgements about the role maths plays, as needed for 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". 'Scientific literacy' has been defined as "the capacity to use scientific knowledge, to identify questions and to draw evidenced-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity." (OECD, 2002a, p 12). In practice, the constructs look very much like the ability to apply mathematical and scientific knowledge in literary contexts.

20. In contrast, TIMSS 1999 was, as were earlier IEA studies, designed to test understanding in relation to an agreed curriculum framework. The curricula of countries differ - sometimes markedly - but the IEA was able to achieve agreement across the national co-ordinators on what it would be appropriate to include. PISA has argued that a curriculum framework based on common denominators is unduly restrictive and "not likely to be of high value or interest to educational policy-makers or practitioners" (Adams, 2003). However, deliberately ignoring school curricula does not save the PISA tests from being more closely related to some than others. As we shall be seeing, how well the questions fit a particular curriculum is a likely explanation of some of the differences in performance between countries, but PISA is not geared to explore this.

Tests

21. TIMSS and PISA also differed in the length and content of their tests. In TIMSS 1999 each student completed one 90 minute test booklet. In order to achieve broad coverage a total of 308 items (162 mathematics and 146 science) were systematically distributed across 8 test booklets which were assigned randomly to students. Mathematics covered five content areas: 'fractions and number sense'; 'measurement'; 'algebra'; 'geometry'; and 'data representation, analysis and probability'. Science consisted of six content areas: 'earth science' (15 per cent); 'life science' (27 per cent); 'physics' (27 per cent); 'chemistry' (14 per cent); 'environmental and resource issues' (9 per cent); and 'scientific inquiry and the nature of science' (8 per cent).

22. Mathematical and scientific literacy in PISA were tested to a maximum of 30 minutes each. Mathematical literacy consisted of a total of 32 items (one of which was not used in the analysis), 18 on the theme of 'growth and change' and 14 on 'space and shape'. PISA's scientific literacy test consisted of 35 items on 'science in life and health' (37 per cent), 'science in Earth and environment' (37 per cent) and 'science in technology' (26 per cent). The science questions seem very much tilted towards 'nature' and everyday science in keeping with PISA's interest in general knowledge and reading.

23. The mathematical and scientific literacy items were each organised into four 15-minute clusters which were aggregated in various combinations with nine 15-minute reading clusters to produce nine linked two-hour test booklets, one of which contained no maths or science, six either maths or science, and two both (Adams and Wu, 2002). In addition, pupils answered a background questionnaire, taking about 30 minutes, to provide information about themselves and their homes. The schools also completed a 20-minute questionnaire. In England, 2,292 pupils were assessed in maths and 2,284 in science compared with 4,120 in reading. The testing of any one individual is very limited, but the hope is that by a process of statistical scaling (Item Response Theory) it is possible to derive a score for populations based on all the items. Item Response Theory was also used in TIMSS 1999 to scale up from its matrix samples to the population of items.

Questions

24. Not only does length and content balance of the PISA and TIMSS tests differ, but there is a different style to the questions. In keeping with its emphasis on literacy, PISA's questions tend to be lengthy and wordy, whereas the TIMSS questions tend to be more direct and abstract, and to give due weight to operations and calculations. Boxes 1 and 2 show example questions that the two groups have released.

Maths

25. The TIMSS maths questions were designed to test five kinds of performance ('knowing', 'using routine procedures', 'investigating and problem solving', 'mathematical reasoning' and 'communicating') and involve three kinds of response: 'multiple choice'; 'short answer'; and 'extended response'. The first question in Box 1 is an example of 'using routine procedures' and it is in multiple

response format. The second question is designed to test 'investigating and problem solving'.

26. PISA emphasized context and five were identified: 'community'; 'educational'; 'occupational'; 'personal'; and 'scientific'. Of the questions illustrated in Box 1, the first under PISA is an example of a 'space and shape' question in a 'person context', and the second of a 'change and relationships' question in a 'scientific context'.

Box 1: Example Maths Questions

TIMSS 1999

Divide 15.45 by 0.003. Respondents are asked to tick one of five possible answers ranging from 0.515 to 5150.

A diagram is provided of a rectangular garden with a path of constant width around three sides of it. The dimensions are given of two sides of the garden and the path plus garden. Respondents are asked to calculate the area of the path and tick one of four answers ranging from 16 m^2 to 144m^2.

PISA 2000

A map of Antarctica is provided together with a scale. Respondents are asked to estimate the area of Antarctica, showing their working and explaining how they made the estimate. Two marks are given for drawing a square or rectangle and getting the correct result. One mark is awarded for using the correct method but not getting the right result.

A graph shows how the speed of a racing car varies along a flat 3 kilometre track during its second lap. Among other things respondents are asked to do is to choose from diagrams of five tracks the one that would have resulted in the speed graph.

Sources: International Study Centre, Boston College (2002) *TMISS 1999 Maths Items*: *Released Set for Eighth Grade.* http://isc.bc.edul; OECD (2002a) *Sample Tasks from the PISA 2000 Assessment*: *Reading, Mathematical and Scientific Literacy.* Paris: OECD.

27. In PISA each question contributed several items so a test booklet had very limited coverage. The first PISA maths question in Box 1, for example, was one of just three questions making up the booklet. The second question contributed four items to the eight of its booklet. The extent of maths testing was so limited that one wonders if it can bear the weight that has been put on it. Furthermore, items were designed so as not to need particular calculation skills and the wrong answer could be awarded the marks if an appropriate method had been adopted. With arithmetic being a particular hazard for pupils in England this could have worked to the country's advantage. Braams (2002) has also looked in detail at the PISA maths test and he concludes that, "it is highly unsuitable as a test of mathematics education or as a guide to improving math education, and that the international comparisons are gratuitously vulnerable to accidental variations and, let us say, subconscious manipulation".

Science

28. The science items have a similar 'feel' to them as the maths items, with again a clearly recognisable difference between TIMSS and PISA. Box 2 provides illustrations of the questions posed.

29. The TIMSS science test was designed to assess five kinds of performance ('understanding simple information', 'understanding complex information', 'theorizing, analysing and solving problems', 'using tools, routine procedures and science processes' and 'investigating the natural world'). The first item in Box 2 is from the 'earth science' category and is intended to test 'understanding complex information', and the second is about 'understanding simple information' in the 'physics' category.

Box 2: Example Science Questions

TIMSS 1999

Which of the following is an important factor in explaining why seasons occur on earth: (A) Earth rotates on its axis; (B) The Sun rotates on its axis; (C) Earth's axis is tilted; (D) The Sun's axis is tilted.

Three drawings show a rocket being launched from Earth then returning, the first is on the launch pad, the second is on the way up, and the third on the way down. Respondents are asked in which of the three positions does gravity act on the rocket: (A) 3 only; (B) 1 and 2 only; (C) 2 and 3 only; (D) 1, 2 and 3.

PISA 2000

A cartoon illustrates the formation of ozone. In the first frame some teletubby-type figures are shown holding hands in pairs under a blazing sun. In the second, some singletons are shown running towards the remaining pairs. In the third, there are two trios. Above the figures are the notations O_2, O and O_3. Respondents are asked to suppose they have an uncle who tries to understand the meaning of the strip. However, he did not get any science education at school and he doesn't understand what the author of the strip is explaining. He knows that there are no little fellows in the atmosphere, but he wonders what those little fellows in the strip stand for, what the notations in each frame stand for, and which processes the strip represents. A maximum of three marks can be awarded for saying (i) oxygen molecules are split into oxygen atoms; (ii) the splitting takes place under the influence of sunlight and (iii) the oxygen atoms combine with other oxygen molecules to form ozone molecules.

An extract from Semmelweis' diary is provided including a graph showing the number of deaths from puerperal fever per 100 deliveries in two hospital wards. Respondents are asked to suppose that they are Semmelweis and to give a reason based on the data Semmelweis collected why puerperal fever is unlikely to be caused by earthquakes. Two marks are awarded for answers that refer to the difference in the number of deaths in the two wards. One point is given to those who do not mention this point, but offer something else that points in the same direction, for example, men experience the earthquakes but do not get puerperal fever.

Sources: International Study Centre, Boston College (2002) *TMISS 1999 Sciences Items*: *Released Set for Eighth Grade.* http://isc.bc.edul; OECD (2002a) *Sample Tasks from the PISA 2000 Assessment*: *Reading, Mathematical and Scientific Literacy*. Paris: OECD.

30. The first item under PISA in Box 2 is an example of a question designed to test the process of 'communicating to others valid conclusions from evidence' in a 'global' context'. It contributes four of the items to booklet 4, which is completed by two items on 'daylight' and three on 'algae'. The second PISA question in Box 2 is intended to test 'critically evaluating scientific evidence/data' in 'an historical context'. It contributes four items to its booklet which is completed by two items on 'tidal power' (one of which was subsequently dropped) and another question of three items. This underlines just how limited the maths and science testing in PISA really was.

Target Population

31. TIMSS and PISA defined their target populations differently. TIMSS specified on a grade-basis, and PISA on an age-basis. The first TIMSS study in 1995 defined its population as "all students enrolled in the two adjacent grades with the largest proportion of 13-year-old pupils at the time of testing" (Keys, Harris and Fernandes, 1996b), in other words, international grades 7 and 8, but the repeat in 1999 concentrated on grade 8. In England this is more commonly referred to as Year 9. PISA (2001), on the other hand, while also targeting those born in 1984 specified the testing of students between the ages of 15 years and 3 (complete) months and 16 years and 2 (complete) months at the beginning of the assessment period, irrespective of where they had got to in the education system. In England, they would have been mainly in Years 10 and 11.

32. In England this definitional difference would have had little effect, because children mainly move up through the schools by age. But in those countries where progression is by performance and it is not unusual for pupils to repeat at least one year, an age-based target group will have led to children being tested at very different stages of their education. In Flemish-speaking Belgium, for example, 72.2 per cent of those tested were in grade 10, 22.8 per cent in grade 9 and 2.5 per cent in grade 8. Not surprisingly, they had very different mean reading scores ranging from 364 for those in grade 8 to 564 for those in grade 10.

33. Advantages are claimed for both approaches. The International Association for the Evaluation of Educational Achievement has favoured a grade-basis since all the students in one system will have experienced the same level of education (albeit more than once in the case of repeaters), and also for the ease of test administration. Brown (2003) has argued strongly for an age-based sample as in PISA on the grounds that it is less liable to bias. She contends that in grade-based sampling a lot of pupils will be missed because of year-repeating, so the pupils tested in some countries will be older than those in others. However, while mean ages may differ more in grade-based approaches, they are not excessively dispersed. In TIMSS 1995, 16 out of the 26 OECD countries taking part lay in the range 14.0-14.3 years (3.6 months) and 19 were in the range 13.9-14.3 years (4.6 months). Only Germany at 14.8 years and Scotland, Iceland and Greece at 13.6-13.7 years were outliers. Neither does age-based sampling ensure similar lengths of schooling; children in England and New Zealand, for example, start school at five, whereas in Scandinavia they start at seven.

Response Rates

34. Whatever the target population, it is important to secure the participation of a high proportion of those drawn in the sample. Otherwise, the results are always open to the criticism that those responding are biased in some way. PISA (OECD, 2001) originally set a requirement for a response rate of 85 per cent of the schools initially selected, ameliorated by allowing replacements to bring participation up to the required level as long as the initial return did not fall below 65 per cent. The Netherlands which had an initial response rate of 27 per cent was dropped from the analysis on these grounds. But the United States and England, both of which fell short, were kept in because of evidence received that the non-participating schools were not likely to deviate significantly from those taking part.

35. Adams (2003), Project Director of the PISA Consortium, in comparing PISA with TIMSS, has noted that England's response rate for TIMSS was not particularly good either. Table 5 summarizes the response rates for England, Germany and Switzerland in the two TIMSS studies and PISA. It is clear that, even assuming that the schools added in from the replacement list are comparable with those in the initial sample who refused to take part, only two-thirds of the targeted pupils in England participated in PISA 2000. Expressed as a percentage of the initial sample of schools it is only 49 per cent. This is in marked contrast to Switzerland where 92 per cent of the initial sample of schools participated rising to 96 per cent when replacements are included, with 95 per cent of students in sampled schools taking the tests. This raises the question of just how comparable are the scores of the two countries.

Table 5: Response Rates in TIMSS and PISA

Country	Study	Initial Sample of Schools	Per Cent Response		Overall
			Schools with Replacement	Students within Schools	
England	TIMSS 1995	56	85	92	78
England	TIMSS 1999	49	85	90	77
England	PISA 2000	59	82	81	66
Germany	TIMSS 1995	72	93	87	81
Germany	PISA 2000	94	94	86	81
Switzerland	TIMSS 1995	93	95	99	94
Switzerland	PISA 2000	92	96	95	91

Source: Adams, R.J. (2003) Response to 'Cautions on OECD's recent educational survey (PISA)', *Oxford Review of Education* 29(3), 378-389.

36. England's Office of National Statistics (Gill, Dunn and Goddard, 2002) has shown that the sample of schools with replacement matched the national distribution in terms of GCSE results and percentage of pupils eligible for free school meals, but there were no pupil-level data available to indicate how participating pupils compared with non-participants. There is, therefore, the risk that the achieved samples are not strictly comparable. There was also considerable variation in the extent to which children with special needs or suffering from disabilities were

included. Adams (2003) appears to take comfort from the relatively low response rates in England in the TIMSS studies, but from our perspective they further point to the limitations which must be borne in mind in interpreting the results of all large-scale international comparative studies.

Impact on Results

37. The differences in countries participating and in the aims, tests, items and target populations could all have had an appreciable impact on the results that were obtained. Prais (2003a) has claimed a swing of about 60 points (when 100 points is one standard deviation) in the maths scores of England/UK and a raft of other countries. "Whereas the IEA surveys of 1995/9 showed the UK at some 40 points in average scores *behind* Switzerland, France, Flemish-speaking Belgium, the Czech Republic and Hungary, in the PISA study of 2000 the UK was some 20 points *ahead* of those countries on average." This calculation involves a certain amount of licence since different countries were involved in the two studies so a point in one is not exactly the same as in the other. But there were similar large shifts in the scores of France, Flemish-speaking Belgium, the Czech Republic and Hungary in comparison with either TIMSS-1995 or TIMSS-1999 (not all the countries participated in both). Prais attributes the turn-around mainly to three factors: the different aims; sampling pupils by age rather than school class (responsible, in his view, for a shift of about 20 points); and the low response rate in England/UK compared with other countries (contributing a swing of about 5 points).

38. Prais' implication that PISA's results are bogus has been hotly contested by Adams (2003). He does not challenge Prais' contention that there are marked differences in the results, but argues that PISA's age-based approach is preferable. He suggests, for example, that: "When comparing PISA and TIMSS results for England, Switzerland and Germany, it should be noted that 60 per cent of the English students were already in Year 11 (international grade 10), the second year of upper secondary education. In Switzerland and Germany over 60 per cent were still in grade 9, the last year of lower secondary education. This fact….provides a possible explanation for the differences in the relative performance of each country in PISA 2000 when compared with TIMSS 1995." Well, yes, but from our point of view it only underlines the difficulty of comparing the results of different countries.

39. PISA 2000 was deliberately framed to be independent of the curriculum. This does not, however, mean that the tests were not more closely related to the curricula of some countries. McGaw (2002b), in a presentation to a symposium on PISA held in Berlin in November 2002, denied there was such an effect. He cited evidence to suggest that countries' ratings of 'the cultural relevance' of the individual items had little bearing on pupils' performance. In only two countries, Norway and Korea, was there a statistically significant effect. This was dismissed as just not credible by Andrew Porter who has edited one of the major texts on cross-national surveys (Porter and Gamoram, 2002). An alternative interpretation of McGaw's findings is that the 'cultural relevance' measure was not robust enough to reveal the effects of curriculum match on the PISA results.

40. Certainly studies in the individual countries do not bear out McGaw. One of the features of PISA has been the success of Finland which Välijärvi *et al* (2002) have attributed, in part, to: "the fact that *the tasks* used in PISA were *well suited to the Finnish curriculum*." (italics original). Rocher (2003) in France used sophisticated regression methods to deduce families of similar countries from their PISA scores and found that "one of the most important sources of 'bias' is the influence of the curricula of the different countries." A similar interpretation emerges in a detailed study of curriculum match conducted in Ireland (Shiel et al, 2002). Ireland did very well in reading and the analysis showed that the PISA questions played to its strengths. On the other hand, it did less well in maths where the curriculum and assessment are traditional, or science where the approach is theoretical.

41. But in the Irish study there is also another very telling observation. Science performance could also have been expected to have been depressed because 11 per cent of 15-year-olds had dropped the subject. But it was found that students not studying science at all performed just as well as those who were taking it at the ordinary level (though not those taking it at the higher level). Shiel *et al* concluded, "it could be argued that some of the PISA science items assess generic reading comprehension and/or problem solving skills rather than purely scientific concepts". A curious aspect of the results for England also point in the same direction. In the analysis by the Office of National Statistics (Gill, Dunn and Goddard, 2002) no difference was found in the reading literacy scores of pupils in Years 10 and 11, suggesting perhaps that the PISA was testing generic skills or else pupils were not learning much in the extra year.

42. Prais (2003a) in his detailed critique of the maths part of the PISA study also questions the validity of the test and suggests that it "may be more a test of 'commonsense', or of 'IQ', than the result of mathematical schooling". Adams (2003) attempts to rebut this by citing results from Germany where a more curriculum-focused test was run in parallel with the PISA test (Baumert, J. *et al,* 2003 a,b). In passing he noted that the national items had a lower correlation with the PISA reading scores than the PISA maths test did, though there was a higher correlation with intelligence. This is consistent with our emerging conclusion that the PISA maths and science tests - both of which develop elaborate verbal contexts - are more tests of reading than of maths or science understanding.

Reading Literacy

43. So far we have been mainly concerned with maths and science since they afforded the best material for comparison, but, in fact, they were minor parts of PISA 2000, occupying only about a quarter of the test session. A more complete comparison for maths will be possible when the 2003 results are published since mathematical literacy formed the major part of that PISA study and there was a TIMSS survey also.

44. The main thrust of PISA 2000 was 'reading literacy' in which most pupils were tested for 90 minutes. But, unlike maths and science, there are few opportunities for comparison. The best available are the IEA Study of Reading Literacy in 1991 (Elley, 1994), the International Literacy Survey (IALS) conducted by Statistics

Canada on a rolling basis between 1994-98, and the Progress in International Reading Literacy Survey (PIRLS) carried out by the IEA in 2001.

45. Although all of these studies were designed to compare reading literacy across countries, there are a number of differences between them. Each construes literacy in its own way, different populations are targeted and different countries participated. The 1991 IEA Study, which may have been the first to use the term 'reading literacy', tested ninth graders in 32 countries, but England did not take part. IALS interviewed representative samples of the populations of 20 countries in their own homes in the period 1994-98, administering scales of prose, document and quantitative literacy. The PIRLS study, under the auspices of the IEA, was a comparative study of the reading achievement in 35 countries, aiming to test 10-year-olds, but with quite a wide range in ages. Scotland's poor showing may be not unconnected with the mean age of its sample being a year younger than Sweden's which came top. Iceland and Greece with young samples also came well down the list.

46. Table 6 shows the ranked results for those countries which also participated in PISA. Perhaps not surprisingly given the differences, there is little correlation between the studies. The only statistically significant association is a weak one ($r=0.51$, $P<0.05$) between PISA and IEA 1991. But even here only Finland and New Zealand do consistently well, and Greece and French Belgium poorly. In contrast, France, Ireland, Hungary, USA and Switzerland emerge very differently on the two occasions.

47. England did not take part in IEA 1991. In PISA 2000 and PIRLS 2001 England/UK scored very well, and ministers and officials have celebrated, but in the IALS study it did very badly. In the results for the age group nearest to that in PISA, the 16-25 year-olds, England was placed 14 out of the 17 countries who took part in both studies. It was claimed that the IALS results showed that 48 per cent of the people tested in the UK had a literacy level so low that they would be unable to handle everyday tasks such as following instructions or understanding a payslip. These dramatic results became the cornerstone of the influential Moser (1999) literacy inquiry.

48. France was even more shocked to find that 65 per cent of its people were classed as functionally illiterate, so much so that it questioned the authenticity of the outcome and withdrew from the reporting stage of the study. This led to a re-evaluation of the findings funded by the European Commission (Carey, 2000). Using results from that investigation, Blum *et al* (2001) were able to show that assigning people to literacy levels on a different - and more plausible criterion - placed only 5 per cent of those tested in France and 3 per cent of those tested in the UK at the lowest level. They also found that a combination of linguistic and cultural differences rendered response patterns unstable, translation led to biases, and the treatment of omissions left the scores vulnerable to the behaviour of interviewers. They concluded "the IALS survey, as it stands, should be treated with caution at a national level and more so at an international level."

49. Fifteen items from the IALS prose literacy scale were incorporated into PISA's 141 item reading literacy measure. When countries are compared on just these items (OECD, 2002c) France and England come out above average, and way ahead of the stars of IALS such as Sweden, Portugal and Norway. The wide discrepancy between the two methods of analysing IALS and the very different findings of IALS and PISA suggests that the scare stories in the 1990s about the low levels of literacy in the UK may not have been justified. But neither is it possible to claim that the PISA and PIRLS results speak of a great improvement. They are cross-sectional studies showing relative standing across countries within the limitations of the methods used. They are best regarded as baselines for future studies.

Table 6: Reading Literacy

Country	PISA[1]		IEA[2]		IALS[3]		PIRLS[4]	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Finland	546	1	560	1	321	1		
Canada	534	2	522	12	295	7=	544	4
Belgium Flemish	532	3			298	5=		
New Zealand	529	4	545	4	288	11=	529	10
Australia	528	5			294	9		
Ireland	527	6	511	15	288	11=		
Scotland	526	7					528	11=
England	523	8			284	14	553	2
Sweden	516	9	546	3	311	2	561	1
Iceland	507	10	536	6=			512	15
France	505	11=	549	2			525	13
Norway	505	11=	516	13	299	4	499	16
USA	504	13	535	8	273	15	542	6
Denmark	497	14	525	9	295	7=		
Switzerland	494	15	536	6=	287	13		
Spain	493	16	490	17				
Czech Republic	492	17			292	10	537	9
Italy	487	18	515	14			541	7
Germany	484	19	524	10	298	5=	539	8
Hungary	480	20	536	5	265	17	543	5
Poland	479	21			270	16		
Belgium French	476	22	481	18				
Greece	474	23	509	16			524	14
Portugal	470	24	523	11	302	3		
Russia	462	25					528	11=
Latvia	458	26					545	3

1. 15y 3m to 16y 2m tested in 2000.

2. Mean ages by country of 13y 9m to 15y 6m tested in 1991.

3. 16-25 year-olds excluding those still in secondary education tested on prose literacy 1994-1998.

4. Mean ages by country of 9y 7m to 10y 9 m tested in 2001.

Sources: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Tables 2.3a, pages 253; Elley, W.B (1994) *The IEA Study of Reading Literacy*, Table 2.3, page 57; OECD and Statistics Canada (2000) *Literacy in the Information Age,* Table 3.5 page 57; NFER (2003) *PIRLS in England*, Figure 2.1, page 9.

Maths, Science and Reading Compared

50. In Table 7 we give the rankings for all three PISA domains separately and in combination. What is striking given the closeness of the mean scores is the similarity of the ranks across the different subjects, with correlation coefficients in each case above +0.9 (0.95 for science and reading, 0.93 for science and maths, and 0.92 for maths and reading). On the face of it, this is unexpected since reading, maths and science call on different abilities and it is not unusual to find students much stronger on one side or the other. In the TIMSS' results, as we saw in Table 4, England tended to come well down the maths rankings, but fare better in science.

Table 7: PISA 2000 Rankings

Country	Maths		Science		Reading		Total	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Japan	557	1	550	2	522	10	1629	1
Korea	547	2	552	1	525	8	1624	2
Finland	536	5	538	3	546	1	1620	3
Canada	533	6=	529	5	534	2	1596	4
New Zealand	537	4	528	6=	529	4	1594	5=
Belgium Flemish	543	3	519	9=	532	3	1594	5=
Australia	533	6=	528	6=	528	5	1589	7
England	529	9=	533	4	523	9	1585	8
Scotland	533	6=	522	8	526	7	1581	9
Ireland	503	17	513	11	527	6	1543	10
Austria	515	12	519	9=	507	12=	1541	11
Sweden	510	16	512	12	516	11	1538	12
France	517	11	500	14=	505	14=	1522	13
Switzerland	529	9=	496	17=	494	18	1519	14
Iceland	514	13=	496	17=	507	12=	1517	15
Norway	499	18	500	14=	505	14=	1504	16
Czech Republic	498	19	511	13	492	20	1501	17
USA	493	20	499	16	504	16	1496	18
Denmark	514	13=	481	23	497	17	1492	19
Liechtenstein	514	13	476	25	483	23	1473	20
Hungary	488	23	496	17=	480	24	1464	21
Germany	490	22	487	21	484	22	1461	22
Spain	476	25	491	20	493	19	1460	23
Belgium French	491	21	467	26	476	26	1434	24
Poland	470	26	483	22	479	25	1432	25
Italy	457	28	478	24	487	21	1422	26
Russia	478	24	460	28=	462	29	1400	27
Portugal	454	29	459	30	470	28	1383	28
Greece	447	30	461	27	474	27	1382	29
Latvia	463	27	460	28=	458	30	1381	30
Luxembourg	446	31	443	31	441	31	1330	31
Mexico	387	32	422	32	422	32	1231	32

Brazil	334	33	375	33	396	33	1105	33

Sources: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Tables 2.3a, 3.1 and 3.3, pages 253, 258 and 261.

Across Countries

51. The similarity in ranks in Table 7 could reflect common features within countries affecting maths, science and reading performance (which could be the quality of the school system, but also many other personal or cultural factors, such as the gene pool, parental commitment to education, or the national value placed on education). But it is also possible that the similarities could reflect common ground between the three 'literacy' tests.

52. The questions do indeed look very similar. Many of the maths and science items depend on establishing detailed verbal contexts. Conversely, a third of the reading literacy test is devoted to 'non-continuous texts', including charts and graphs (11 per cent of total), tables (11 per cent), and diagrams, maps, forms and advertisements (11 per cent). Some of these would not have been out of place on the other papers. For example, one question involved the interpretation of a graph of the changing levels of Lake Chad over the period 11,000 BC to the present in relation to a chronology of Saharan rock art and the changing patterns of wildlife. Respondents are asked to draw certain inferences from the illustrations. Although this question was in the reading literacy test, in essence it could have formed part of either the maths or the science test. Similarly, a series of questions on a tree diagram of the structure of a country's labour force could just have easily appeared on the maths paper.

53. There are nevertheless some differences in the country rankings. Table 7 shows that Japan and Korea which dominate the maths and science lists are down in eighth and tenth places respectively in reading. Whether this is due to a real difference in capability or to some other difference, such as the intrinsic difficulty of oriental languages, is not clear. In Japan, mathematics is used in selection for higher education so it is a high-stakes subject which may explain why although the students do well in international comparisons, they also emerge in TIMSS as disliking it.

Within Countries

54. Comparison of the rank orders of the PISA scores within countries also shows a close correspondence between the three tests.

Table 8: PISA 2000 Rankings for Australian States and Territories

State/Territory	Maths		Science		Reading		Total	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Australian Capital Territory	548	*1*	553	*1*	552	*1*	1653	*1*
Western Australia	547	*2*	544	*2*	538	*3*	1629	*2*
New South Wales	540	*3*	532	*4*	539	*2*	1611	*3*
South Australia	526	*5*	539	*3*	537	*4*	1602	*4*
Queensland	525	*6*	523	*5*	521	*5*	1569	*5*
Victoria	529	*4*	516	*6*	516	*6*	1561	*6*

Tasmania	517	*7*	510	*7*	514	*7*	1541	*7*
Northern Territory	502	*8*	490	*8*	489	*8*	1481	*8*

Source: Lokan, J. et al (2001). *15-Up and Counting. Australian National PISA Report*. Melbourne: ACER.

Table 9: PISA 2000 Rankings for German Lander

Land[1]	Maths		Science		Reading		Total	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Bavaria	516	*1*	508	*1*	510	*1*	1534	*1*
Baden-Württemberg	512	*2*	505	*2*	500	*2*	1517	*2*
Sachsen	501	*3*	499	*3*	491	*3*	1491	*3*
Thüringen	493	*4*	495	*4*	482	*6*	1470	*4*
Rheinland-Pfalz	488	*6*	489	*5*	485	*4*	1462	*5*
Saarland	487	*7*	485	*7*	484	*5*	1456	*6*
Schleswig-Holstein	490	*5*	486	*6*	478	*8*	1454	*7*
Hessen	486	*8*	481	*8*	476	*9*	1443	*8*
Nordrhein-Westfalen	480	*10*	478	*9=*	482	*7*	1440	*9*
Mecklenburg-Vorpommern	484	*9*	478	*9=*	467	*11*	1429	*10*
Niedersachsen	478	*11*	476	*11*	474	*10*	1428	*11*
Sachsen-Anhalt	477	*12*	471	*12*	455	*13*	1403	*12*
Brandenburg	472	*13*	470	*13*	459	*12*	1401	*13*
Bremen	452	*14*	461	*14*	448	*14*	1361	*14*

1. Former East Germany is now the Länder of Brandenburg, Mecklenburg-Vorpommen, Sachsen, Sachsen-Anhalt, Thüringen and Berlin. Of the three city states, results are only available for Bremen.

Source: Baumert, J. *et al* (2003). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublick Deutschland. Zusammenfassung Zentraler Befunde*. Berlin: Max-Planck-Institut für Bildungforschung.

55. Tables 8 and 9 show the results for the Australian states and territories, and the German states. The inter-correlations between literacy in maths, science and reading were all well above 0.9 and highly significant. In Australia, the Capital Territory consistently came top and the Northern Territory bottom, with a similar ordering in between. The former West German states tended to fare better than those originating in East Germany, except in the case of Bremen where 41 per cent of the 15-year-olds are of foreign origin (Baumert *et al*, 2003a).

Gender

56. Gender comparison does, however, suggest that there are some differences between the tests. Table A3 (in the appendix), summarized in Table 10, shows girls to be conspicuously ahead in reading literacy, on a par in science and behind in maths. In reading literacy, they achieved higher scores in all the participating OECD countries, with the gap in their favour ranging from 51 points in Finland to 14 points in Korea. Girls were also ahead in 15 of the 26 countries in science although the overall score difference was negligible. Boys, however, tended to do better in maths with an overall score difference in their favour of 11. Girls were ahead in just two OECD countries (though in Russia also), by 5 points in Iceland and 3 in New Zealand.

57. The results for gender show that the PISA maths, science and reading tests are capable of capturing differences. The overall pattern of results is consistent with psychological measures which show that from the earliest years girls tend to have the advantage, on average, in verbal abilities, and boys in numerical and spatial abilities (Maccoby and Jacklin, 1974). They are also consistent with gender differences in subject choices, with females tending to the humanities and languages, and males to maths and the sciences (Sutherland, 1981).

Table 10: Relative Performance of Girls and Boys[1]

Literacy	Score Difference for Girls	Countries Girls Ahead	Countries Boys Ahead
Reading	29	27	0
Science	0	15	11
Maths	-11	2	25

1. Table summarizes data of Tables A3, which also give the sources.

58. There was a strong correlation across countries with the largest differences in favour of girls in reading associated with the smallest differences in favour of boys in maths ($r=0.74$, $P<0.001$) and science ($r=0.64$, $P<0.001$). However, the outcomes look imbalanced with the girls twice as far ahead in reading literacy as the boys were in mathematical literacy, and overall parity in scientific literacy when the boys might have been expected to be ahead. It is as if we were dealing with a lopsided pendulum, weighted towards reading ability. This could have come from the heavy reliance of all three tests on items that were essentially literacy items. The emphasis on biology and the environment could also have contributed to good performance of the girls in the test of scientific literacy.

Quality and Equity

59. It is sometimes said that a distinguishing feature of England's educational performance is a long tail of underachievement. As well as being something that is often repeated in the press, for example, "the long tail of underachievement that is a recurring feature of British performance in international comparisons" (*The Daily Telegraph*, Chew, 30 August, 2003) and by commentators (for example, National Literacy Trust, 1999; Carol Adams, 2002), it has also become the official view (Ofsted, 2000; DfEE, 2001). The claim is not without empirical support. One of the conclusions of a review of international studies of educational achievement involving England by Reynolds and Farrell (1996) was "English children have a very wide range of achievements, and a greater proportion of low achieving children".

60. But there is no sign of this long tail in the PISA 2000 reading scores. Table A4, in the appendix, shows that the standard deviation for the England/UK score was 100, spot on for OECD countries as a whole. Moreover, when performance is disaggregated into levels of performance, it is apparent that any hint of a wider spread in England is through disproportionately more achieving at the top end. Table A5 shows that among the students achieving at Level 5 (a score above 625), England comes fifth with 15.6 per cent compared with the OECD total of 9.4 per

cent. At the other end of the scale, 12.8 per cent students in England are at Level 1 or below (407 points or fewer) compared with the OECD total of 18.3 per cent. Nineteen countries have more students at Level 1 or below than England, and only seven fewer.

Social Gradient

61. Nevertheless, England/UK finds itself implicitly criticised in OECD analyses of the PISA findings (OECD 2002b, 2003a) for having too steep a social gradient. McGaw (2004) presents, as a fact, that the UK has a 'low equity' education system. Since this is an aspect that has been highlighted it is well worth unpacking. In the main diagram PISA presents (OECD, 2002b, page 42), performance of the countries in the reading test are compared with gradient scores assigned on an Index of Economic, Social and Cultural Status (ESCS), which is interpreted as a measure of inequality. Table 11 adapts the OECD's figure into a word diagram and leaves out those countries with a score close to the mid-point on either dimension.

Table 11: Reading Scores by Measure of Social Inequality[1]

Dimension	High Equity[2,3]	Low Equity
High Attainment	Finland (546, 30) Canada (534,37) Ireland (527,38) Korea (525,21) Japan (522,21) Sweden (516,36) Iceland (507,24)	New Zealand (529,45) Australia (528,46) England (523,49) Belgium (507,48)
Low Attainment	Spain (493,32) Italy (487,32) Poland (479,36) Greece (474,38) Mexico (422,35)	Switzerland (494,49) Czech Rep (492,50) Germany (484,60) Hungary (480,53) Luxembourg (441,46)

1. Adapted OECD (2002b) 'Improving both quality and equity – insights from PISA 2000' *Educational Policy Analysis*, vol 2002, no 9, Edition, pp 46-81. Centre for Educational Research and Innovation. Paris: OECD, page 42.

2. OECD defines inequality in terms of its Index of Economic, Social and Cultural Status which takes into account parental occupation, parental education, family wealth, home educational resources and cultural possessions. The index has been used to calculate social gradient scores ranging in this example, from 21 (low) to 60 (high) which are shown after the countries along with their mean reading scores.

3. Countries close to mid-points not included. USA, France, Denmark, Norway not included because the reading score was in the mid-range 495-505 and Austria, Norway, Denmark, and Portugal not included because the social gradient scores came in the mid-range 39-42.

62. The OECD favourably contrasts the countries in the top-left-hand quadrant (Finland, Canada, Korea, Japan, Sweden, Iceland) with those in the bottom right (Switzerland, Czech Republic, Germany, Hungary, Luxembourg) on the grounds they have achieved above average reading performance with below average impact of family background. It argues, therefore, that "quality and equity need not be considered as competing policy objectives." Bringing together what information it has on the school systems, the OECD document reaches the conclusion that, "the more differentiated and selective an education system is, the larger are the typical

performance differences between students from more and less advantaged backgrounds". Andreas Schleicher (2003), who has emerged as the chief spokesman for PISA, has argued strongly that the poor showing of his own country Germany is due to its selective secondary education system.

63. In PISA's analysis England emerges as one of the countries falling short of the ideal of 'high attainment, high equity', because on the index used it comes out as having a socially unequal education system. However, measuring inequality is by no means straightforward. The OECD itself attempts it in several ways. It has two composite indices of status (the ESCS, already referred to, and the International Socio-Economic Index of Occupational Status), and also a measure of family wealth. Curiously, there is almost no correlation between the status and wealth scores. The wealth score is, however, not a measure of income, but a composite of the ownership of such goods as a dishwasher, television and computer. The ESCS is also an assemblage of responses to different items, so there has to be some uncertainty about what the indices show.

64. Table 11 is based on the ESCS, but in the OECD's presentation of the results equity is extrapolated from status to wealth. It is implied, therefore, that England has a low equity system in which pupils from poor homes are more disadvantaged educationally than in other countries. But this is misleading since the actual figures show England to be one of the more equal countries in terms of performance by family wealth. Smith and Gorard (2002a,b) have also challenged the OECD's interpretation arguing that, "the recent PISA study shows that UK pupils have…one of the smallest gaps between the scores for the richest and poorest".

65. Even accepting the results of Table 11 at face value, it is questionable whether the ideal held up by OECD of a low spread of scores should, in fact, be the ideal. It will depend on how it has been arrived at. Table 12 looks in detail at the distribution of reading scores by parental occupation for England/UK compared with Korea. The means are similar, but Korea has smaller dispersion. It is true that students in Korea in the bottom half on parental occupation score better than their counterparts in England, but in the top half the situation is reversed. In fact, as can be seen in Table A6 in the appendix, in the top quarter England's pupils achieve the highest average scores of all the countries.

Table 12: Reading Scores by Parental Occupation[1]

Index of Parental Occupation[2]	England		Korea		OECD Total
	Score	Rank	Score	Rank	
Top Quarter	579	*1*	542	*16*	543
Second Quarter	543	*4*	531	*8*	515
Third Quarter	513	*7*	524	*3*	492
Bottom Quarter	491	*9*	509	*2*	462
Overall Score	523		525		499

1. Data from Table A.6, ranking out of 26 countries.

2. Derived from students' responses on parental occupation. Father's or mother's taken whichever is the higher.

66. England, rather than having a long tail of underachievement, can claim to walk tall with its highest achievers ahead of the world. In so far as the PISA scores reflect the educational system – and that is debatable – it appears that England has found ways of allowing the talents of its brightest to flourish. It is worth reflecting on how it has managed this before bearing down on 'differentiation' which the OECD sees as increasing social disadvantage.

67. But here we risk falling into the same trap as the OECD in attributing the differences to the educational system. It is clear from the data available in the PISA reports that the high dispersion of reading scores can be associated with a number of other factors. We discuss just two: immigration and differences within countries.

Immigration

68. In Table 13 we examine the right and left halves of Table 11 in relation to the proportion of students of foreign origin in each country (first generation and non-native). Overall, the low inequality group have a much lower proportion of first generation and non-native students than the high inequality group - 4.3 per cent against 14.7 per cent.

Table 13: Immigration and Impact on Reading Scores[1,2]

Country[3]	Immigrants		Native Students Reading Score	Country	Immigrants		Native Students Reading Score
	%	Reading Score			%	Reading Score	
Finland	1.2	468	548	New Zealand	19.6	507	538
Canada	20.6	526	538	Australia	22.6	520	532
Japan	0.1	-	525	England	9.6	495	528
Sweden	10.6	466	523	Belgium	12.0	417	522
Iceland	0.8	-	509	France	12.0	464	512
Ireland	2.3	552	528	USA	13.5	473	511
Spain	2.0	457	494	Switzerland	20.7	428	514
Italy	1.0	445	489	Czech Republic	1.1	-	501
Poland	0.2	-	482	Germany	15.2	423	507
Mexico	3.6	344	427	Hungary	1.7	486	482
Greece	4.8	403	478	Luxembourg	34.2	385	474
Country Average	4.3	458	504	Country Average	14.7	460	511

1. Table displays countries in the same quadrants as Table 10, with France and USA added back in because exclusion on basis of closeness to midpoint of reading scale no longer relevant.

2. Data on students of foreign-origin and reading performance from Table A.7, which gives sources.

3. Question not asked in Korea.

69. The new arrivals in the countries characterised as 'lower equality' generally performed at a much lower level in reading (in what may not have been their first language) than the native students. Germany and Switzerland would have come in the OECD's above average category if only their home students had been counted. The countries in the OECD's favoured quadrant of 'high score, high equity' (top left in Tables 11 and 13) tended to either have very few students of foreign origin

(Finland, Japan and Iceland), or to have high-performing incomers (Canada, Ireland). Even so, the social gradient scores of this second group were noticeably higher, and the same was true of Sweden. Bracey (2002) has discussed the different immigration policies and characterizes Canada's as, "relatively open on the one hand, but energetically recruiting educated people from abroad on the other". Many of its new students will also have spoken English, in contrast to the situation in, say, Germany where the first language of a fifth of the non-native students was Turkish.

70. The associations between reading performance and the extent and type of immigration are one reason why it is premature to attribute country differences to the educational systems. Another is regional patterns.

Region

71. We have already noted the differences between the former West and East Germanies (see Table 8) and the effects on the overall performance of Italy and Canada according to which regions are included (page 6), but the results for Belgium are particularly interesting. Table 14 give the results of the Flemish-speaking and French-speaking communities for all three domains of PISA 2000. As in previous studies (see Tables A1 and A2), Flemish-speaking Belgium did well and, on its own, as we saw in Table 6, would have been placed equal fifth overall, above England. In contrast, French-speaking Belgium comes well down in 24th spot. In Table 14 we can see that the difference on all the tests is more than half a standard deviation. It is not surprising, therefore, that Belgium should come out as one of the high dispersion countries in Table A4 and on OECD's chart encapsulated in Table 11.

Table 14: Belgium

Literacy	Flemish Community	French Community	Total
Reading	532	476	507
Science	519	467	496
Maths	543	491	520

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table B2.1, page 317.

72. The two communities have similar educational systems which separated only a decade ago (Mouton, 2001), so the reasons for the big difference in performance may lie elsewhere. Luc van de Poele, the Flanders' PISA Project Manager, is quoted by Mouton as suggesting the much higher proportion of foreign-origin students in French Belgium – 18 per cent against 7 per cent – played a part. Another possible explanation he offers is that Flanders has stayed "very traditional in its approach to education". An analysis from the perspective of the French community (Deschamps and van den Kerkhove, 2001) found that the difference does not seem to be due to lack of investment since French Belgium is the highest spender on education in the OECD, at 6.8 per cent of GDP. They suggest that French-speaking Belgium needs to develop a culture of evaluation, which implies that students in that part of Belgium may have been less practiced at taking tests.

73. Hazette (2001), the schools minister for French Belgium, is reported as attributing part of the difference to the design of the study. Because of repeating only 53 per cent of the 15-year-olds in the French community had reached at least the fourth grade at the first opportunity compared with 73 per cent in Flanders. Nine per cent of the French-speaking students were double repeaters down in grade 2. But Romainville (2002), in debating the Belgium results with Lafontaine (2002), looks beyond the structure of the education system (grade-repetition, tracking, and school differences) to suggest that the difference between the Flemish north and French south in their performance in international educational studies may be due to more deep-seated features such as the long-term economic decline of the south relative to the north, with higher unemployment rates, greater employment insecurity, and lower regional income.

74. Belgium has the potential to be a fascinating comparative study in its own right. Its importance in our paper is as a further caution against rushing to interpret country differences in terms of the education system, since both communities shared common arrangements until recently.

Schools

75. In so far as differences do emerge with school type in the PISA data, the most striking is the success of independent schools. Table 15 shows that in all the countries studied with independent sectors, except Japan, pupils in the independent schools were considerably ahead in their reading scores.

Table 15: Reading Performance of Government and Independent Schools[1]

Country	Independent Schools[2]		Government Schools	
	Score	% Students	Score	% Students
United Kingdom	614	9.2	515	90.8
New Zealand	599	4.8	528	95.1
Ireland	586	2.9	501	39.5
Canada	568	2.6	532	93.8
Greece	549	4.1	468	95.9
USA	545	4.3	502	94.6
Spain	543	9.2	478	62.0
Austria	532	5.0	504	88.8
Korea	532	33.6	519	50.7
Switzerland	523	4.7	492	94.1
Japan	518	29.6	524	69.6
Italy	513	5.1	486	94.1
Portugal	508	1.5	469	92.6
Poland	500	2.9	478	97.1
Mexico	492	14.0	413	85.1
Brazil	459	10.5	386	89.5

1. Adapted from OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 7.13, page 307. Table only shows countries where independent schools cater for more than 1 per cent of pupils.

2. Does not include independent schools which receive more than 50 per cent of core funding from government. In some countries they can take a large proportion of the pupils, for example, in Ireland 58 per cent, Spain 29 per cent and Denmark 25 per cent.

76. The biggest difference is in England with pupils in independent schools scoring a full standard deviation above their counterparts in the maintained sector (with grammar schools included). But in countries as diverse as Ireland, Greece, Mexico, Brazil, New Zealand and Spain independent-school pupils were more than 50 points ahead.

77. In some countries a high proportion of pupils go to hybrid schools (not shown in Table 15) – private schools which receive over 50 per cent of their core funding from government agencies and do not necessarily charge tuition fees. Interestingly, pupils in these schools also generally performed better than their counterparts in government schools. They were ahead by 83 points in Germany, and by more than 20 points in Canada, Ireland, Switzerland, Spain, and the United States. Only Luxembourg and Denmark had the government schools ahead – by 4 and 2 points respectively.

78. What is the explanation for this striking set of results? Is it something to do with the quality of teaching, the level of investment, freedom from bureaucracy, the history and culture of the schools, or is it just who goes to them? And what is different about independent schools in Japan, and government-funded private schools in Luxembourg and Denmark? Unfortunately, the published analyses are little help. The performance of the independent schools is hardly mentioned in the 322-page publication of first results (OECD, 2001) or the 389-page follow-up (OECD 2003a). The only comment is that pupils are not distributed randomly between schools and tuition fees can be a barrier to attending independent schools. The higher performance of government-funded private schools that charge no tuition fees is dismissed with the comment that "they can cater for a different clientele or apply more restrictive transfer or selection practices." The PISA analysts seem keen to make unjustified inferences about the operation of school systems when the data apparently fit the 'quality-equity' narrative they are wishing to develop. However, when inconvenient results turn up they are dismissed as being nothing to do with the schools, but just an expression of social factors.

79. The mean score of 614 recorded by England's independent schools is the highest of any group – national, institutional or social – in the first PISA report (OECD, 2001). This is not a popular finding even in England. When asked about the contribution of the independent sector to the country's performance in PISA by the House of Commons Education and Skills Select Committee, David Normington (2002), Permanent Secretary at the DfES, replied, "I think the independent sector is in there, but I do not think that is the whole story, because I think you can see improvement in the state sector." Whatever claims are made for the PISA findings we must always be aware that the data are likely to be subordinated to the value systems of those using them.

Participation

80. Discerning newspaper readers will have been puzzled a few months after hearing about England's/UK's apparent success in PISA to be greeted with headlines like that in the *Financial Times*, 'UK tumbles in secondary schools results table' (Timmins, 30 October 2002) and *The Independent*, 'Compared with many other industrialised nations, school standards are far from impressive' (Cassidy, 30 October 2002). The basis for these claims was a briefing given by Andreas Schleicher, in his role as Head of Indicators and Analysis Division, at the OECD Directorate for Education, to the British press marking the publication of OECD's annual statistical compilation, *Education at a Glance*. During the briefing he took the opportunity to press home his criticisms of the German and Swiss systems, and implicitly England's. "The most successful systems are comprehensive and are providing open pathways and highly personalised learning. The least successful are highly institutionally differentiated." (quoted in O'Malley, 2002). But this does not stand up on the participation data presented since Switzerland is placed second and Germany equal fifth out of the 27 OECD countries that we have been considering, though it is true England/UK is down in 16th position.

81. An adaptation of the set of data Schleicher was highlighting is included in the appendix as Table A8 (from the 2003 report based on 2000 data, the year of the PISA study), and an essential point from it is illustrated in Table 16 below. The figures show that only 68 per cent of 25-34-year-olds in the UK had successfully completed upper secondary schooling against 95 per cent in Korea. In the UK the proportion hardly changes with age band so, if it is assumed that these are school leaving qualifications, it does not look as though the UK has improved much over the years. Korea, on the other hand, which was lagging in 22nd place a generation ago with a completion rate of 30 per cent, has now leapt into first place.

Table 16: Per Cent Completing Upper Secondary Education[1,2]

Age	United Kingdom		Korea		OECD Total
	%	Rank	%	Rank	
25-34	68	*20*	95	*1*	74
35-44	65	*16*	77	*14*	69
45-54	61	*14*	49	*19*	60
55-64	55	*12*	30	*22*	49
25-64	63	*16*	68	*14*	64

1. Data from Tables A8 and A9.

2. Countries define upper secondary in their own ways. Some have examinations, others do not. Upper secondary in England is schooling from 14, and further education. Successful completion is 5 or more GCSEs at grades A* -C or NVQ2 or above.

82. The apparent failure of England's performance to progress is hard to square with the August headlines each year announcing (admittedly with some scepticism) by how much GCSE and A-level results have improved. As we have seen in other international comparisons, the figures need careful unpacking. What is upper secondary education and what does successful completion mean? In OECD's (2003) terms it is the final stage of secondary education and it can last from two to five years, with completion ranging typically anywhere from age 16 (Turkey) to 19-20 (Denmark, Switzerland). It does not include vocational training exclusively undertaken in industry.

83. Successful completion is defined by each country as it wishes. In some cases it is defined on the basis of examinations passed, but in others merely as having attended for the requisite number of course hours. In England upper secondary education is taken as schooling from the age of 14 or further education, and successful completion as achieving five or more GCSEs at grades A*-C or an NVQ Level 2 or higher. Those passing at lower levels, for example, four GCSEs A*-C or a foundation GNVQ, will not have been counted, even though they had completed the course hours. The UK's definition is, therefore, a lot tougher than that of many of the other countries in the comparison, and its relatively poor showing may follow from this.

84. There is another puzzle in Table 16. Why have the figures for England remained relatively flat across the age bands when other statistics (for example, Joint Council for General Qualifications, 2003) show the GCSE passes of young people bounding upwards? The explanation may lie in the UK's qualification-based criterion. While it tends to reduce the numbers shown as successfully completing secondary education at school, it makes it easier to record attainment among older groups, thereby tending to smooth out achievement across the age profile. But the country's poor showing has led the Prime Minister (Blair, 2004) to consider raising the school leaving age to age 18. It would not be the first time that an educational policy had been driven by a shaky league table.

85. Participation in upper secondary education is an important issue and the data of Tables 16 and A8 do pose interesting questions. But, as we have seen with PISA, it is wrong to rush to a judgement that one type of educational system is better than another. The way the numbers fall out depends crucially on definition.

Economic Performance

86. PISA is intended to test the application of knowledge, and it is interesting, therefore, to compare its results with economic indicators. Table 17 shows a selection of countries from Table A9 in the appendix (which is adapted from various tables in *Education at a Glance*, OECD, 2003).

Table 17: GDP, Education Spend, Participation and PISA Performance

Country	GDP Per Capita[1]		Spend on Secondary Educ[2]		Completed Upper Sec Education[3]		Overall PISA Score[4]	
	$ (PPP)	Rank	$ (PPP)	Rank	%	Rank	Score	Rank
Luxembourg	48,239	*1*	-		53	*21*	1,330	*26*
Norway	36,202	*2*	8,476	*4*	86	*3=*	1,504	*15*
USA	34,602	*3*	8,855	*2*	88	*1*	1,496	*17*
Finland	25,357	*14*	6,094	*14*	74	*12*	1,620	*3*
Japan	26,011	*15*	6,266	*13*	83	*5=*	1,629	*1*
United Kingdom	24,964	*18*	5,991	*15*	63	*16*	1,584	*7*
Korea	15,186	*23*	4,069	*20*	68	*14*	1,624	*2*

1. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table X2.1, page 418.

2. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table B1.1, page 197.

3. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table A1.2, page 41, age 25-64.

4. Adapted from OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Tables 2.3a, 3.1 and 3.3, pages 253, 258 and 261.

87. Table 17 lists England along with the top three performers on PISA and the top three on wealth creation as indexed by GDP per capita in purchasing power parities as equivalent US$. It also shows for those countries expenditure on secondary education and the proportion of 25-64-year-olds reported as having successfully completed secondary education. Remarkably, the top three countries on wealth creation all come below the mid-point on PISA and the top three on PISA all come below half way on wealth creation.

88. There is thus, as Table 18 shows, no positive correlation across the 27 countries. Norway and the United States - two of the highest earners - also come into the top groups for expenditure on, and participation in, secondary education, but Luxembourg - a small country with the highest GDP per capita - is way down the participation list. It merits a study in itself because it also has the second lowest PISA score. More than a third of its 15-year-olds are of foreign origin, but when they are taken out, as we saw in Table 12, the native students also score poorly.

Table 18: Correlations[1] between Education and Economic Measures[2]

Variable	Overall PISA Score	Completion Upper Second Education	Spend on Secondary Education[3]
GDP Per Capita	0.163	0.384*	0.881**
Secondary Education Spend[3]	0.375	0.469*	

1. Significant beyond 0.001 denoted by two asterisks, beyond 0.05 by one asterisk.

2. Data of Table A9 which also gives the sources.

3. Correlations involving spend on secondary education for 24 countries since Luxembourg, New Zealand and Poland did not provide this information; for the 27 OECD countries.

89. While a country's income does not appear to correlate with its education performance as measured by PISA, it is strongly associated with expenditure on secondary education. This is likely to be because richer countries can afford to spend more on education rather than the other way round, but since it is a correlation it could be either. There are weak relationships between both GDP per capita and expenditure on secondary education on the one hand, and completion of upper secondary education on the other, but a tentative conclusion must be, at this macro-level of analysis, that the money spent on education does not seem to bear directly on educational performance. Looked at from the other direction, neither does good educational performance, as measured by PISA, seem to be associated with high income generation by a country.

Conclusions

90. The main purpose of this study has been to assess whether it is reasonable to draw inferences about the current state of education in England from recent international comparisons. We have traced the development of international testing, identified the two current main players as the International Association for the Evaluation of Educational Achievement (IEA) with its Trends in International Mathematics and

Science Study (TIMSS) and the OECD with its PISA surveys, and looked in detail at the results. Our evaluation is based on six criteria:

- Is PISA technically sound?
- How do the results compare with other studies?
- Can reliable policy conclusions be drawn?
- Have the results been interpreted fairly?
- What is the value of international studies?
- What do international studies tell us about England's educational performance?

Technically Sound?

91. The PISA programme is being carried out under the auspices of a major international organisation, it is well funded and it has drawn together the leading experts in the field from across the world. In questioning whether the study is technically sound, therefore, I feel all the diffidence of the boy in the story of the Emperor's New Clothes. But, nevertheless, there are issues that need to be addressed. In particular:

- Are the measuring instruments valid?
- Is the sampling robust and appropriate?
- Are there perhaps unintended biases?
- Does the mode of presentation lead more to be imputed to the results than is justified?

Validity

92. The validity of an educational test is essentially the extent to which it measures what it purports to measure. The PISA programme has taken great pains to spell out what it is seeking to measure based on a detailed assessment framework (OECD, 2001, 2003c). But the elaboration becomes confusing. In particular, it seems that two separable ideas have been run together. Quite reasonably, the OECD as an economic organisation is interested in labour markets and productivity and has adopted 'knowledge and skills for life' as its focus. Indeed, this is the title given to the first publication of results (OECD, 2001). In measurement terms, PISA's task then becomes to devise valid and reliable methods of quantifying that knowledge and those skills. But, at some point, the tests have become branded as literacy tests, and it is not clear whether they are meant to be mainly about 'literacy' or 'knowledge and skills for life', or whether it is contended that they are synonymous.

93. Reading is measured through a combination of scales of retrieving information, interpreting texts, and reflecting and evaluating, and it is quite possible to defend this approach in terms of its construct validity. But in maths and science the justification is not so obvious. As we have seen (paragraphs 25-30), measuring 'the important knowledge and skills needed in adult life' turns out to involve mainly working with a few wordy texts. The maths and science questions seem to be a

hybrid of elementary maths or science and reading. In devising the tests it looks as if the constructors have taken their own metaphor of literacy too literally.

94. If the PISA maths and science literacy tests are not primarily reflecting knowledge and understanding of these subjects, what are they measuring? We have suggested that, given the similarity of the questions across the three domains, they could all be regarded as reading tests. This receives support from the very high correlations between scores in the three areas both between and within countries. Shiel *et al* (2002) have reached a similar conclusion. They found, in Ireland, that children not taking science at school did as well as those who were following the ordinary level curriculum and suggested that the science test is really assessing "generic reading comprehension and/or problem solving skills rather than science concepts." Prais (2003a) from his detailed scrutiny of performance on the maths test in England and Switzerland suggests that, as the questions were explicitly designed not to test mastery of the school curriculum, "they can, perhaps, be said to be nearer to tests of common sense (or of 'IQ')".

95. Maths and science literacy were minor domains of PISA 2000 and it is possible that as the full scales are rolled out in 2003 and 2006 the constructs will become clearer. But there is an obligation on the PISA programme to provide evidence of the validity of its maths and science tests as measures of knowledge and skills for life rather than asserting that this is the case. In the present study no link was found between a country's performance on PISA and its GDP per capita, though it could be argued that simple correlations would not be expected since wealth creation depends on the skills of the whole workforce and the other factors in production.

Sampling and Response

96. Sampling in PISA has been criticised particularly by Prais (2003a,b) as biasing comparisons between countries. As we saw (paragraphs 31-33), PISA differed from previous international studies in defining its target population in terms of year of birth rather than school class. Prais contends that while this makes little difference in England, in countries like Germany, Switzerland and Belgium, where pupils can enter school later or repeat a year, only about 60-70 per cent of the pupils will have reached the grade for their age. This would have put these countries at a disadvantage compared with England. The point has been conceded by Adams (2003) on behalf of PISA, but he contends that defining the population by age is still preferable because in studies based on school classes countries with repeaters will have the advantage in that some pupils will have been longer in education. This is, however, also true in age-based studies since countries like England and New Zealand start school at a younger age than those in Scandinavia. It is clear that there is no one perfect solution for arriving at a league table. But should league tables be the point? Comparing the educational effectiveness of moving pupils up by readiness or by age would seem to be much more important.

97. Prais (2003a) has also argued that the low response from schools and pupils in England undermines comparisons with other countries. He suggests that only 60 per cent of the randomly-selected sample of schools agreed to participate compared with 95 per cent in Western Europe. Further only 80 per cent of the representative samples in those schools responded, about 10 percentage points lower than in

Western European countries. Adams (2003) has suggested that England's response rate in TIMSS was also low, but as Table 5 (page 12) shows, while lower than Germany and Switzerland, it was higher than in PISA 2000. Sampling in England has been defended by the Office of National Statistics (Gill *et al*, 2002) on the grounds that the replacement schools do not differ significantly from the original sample in term of GCSE results and eligibility for free school meals. Nevertheless, it is likely that the participants and non participants will have differed in other ways and there will have been some impact. Such vagaries are something which studies of people always have to contend with, and it is a further reason to be cautious in interpreting the country rankings.

Unintended Biases

98. As well as these known potential biases which the testing programme will have striven to minimise, it is possible to see in PISA's approach the risk of several unintended biases. PISA explicitly looked beyond the curriculum to everyday life. This is consistent with its stated aims and has the advantage that agreement does not have to be sought across countries on what aspects of the curriculum it is appropriate to test - as in the IEA studies. But ignoring the curriculum does not mean that it is without effect. In reviewing results from PISA 2000 (paragraphs 39-41) we found that the relationship of the tasks to the national curriculum was frequently offered by a country as an explanation for doing well or badly. Other differences are also potentially due to curriculum effects. For example, of the three reading sub-scales, England did best in 'reflection and evaluation' and came equal top, but the 'reading/interpreting' scores were less good. In contrast, 'reflection and evaluation' was France's weakest area. The complete absence of curriculum match analysis in the PISA research severely reduces its explanatory power.

99. Paradoxically, another source of unintended bias could stem from PISA not being able to ignore the curriculum completely. Jan de Lange, Chair of the Mathematics Expert Group reported (de Lange, 2002) that his group had had to leave some intended topics and questions out of the maths test because they had proved too difficult for pupils in some countries. This will have had the effect of making the eventual test less searching and less capable of identifying high-performing education systems.

Presentation

100. The way the results are presented in the official reports is tending to lead commentators to impute more meaning to the numbers that have been obtained than they really carry. Individual scores for countries are listed and these are taken up by the media to produce league tables. Often countries' scores do not differ significantly from those of the countries above or below them. Strictly speaking, in measurement terms their ranks should be the same, and this is more in keeping with the power of the tests which is to roughly order broad groupings rather than make fine distinctions. On this basis, for example, England would have been placed third in PISA maths rather than eighth. PISA (2001) has published such groupings, but in the form of complicated multiple-comparison diagrams which have not attracted much attention.

Comparison with Other Studies

101. We have seen (in paragraphs 10-16) that, in many respects, the PISA findings are in sharp contrast to what has gone before. It was because the PISA results appeared to fly in the face of so much of what Prais (2003a,b) and others believed about the performance of England compared with Switzerland and Germany that the findings have been challenged so strongly. In some ways PISA vs TIMSS is a re-run of the 1960's battles between educational radicals and traditionalists which have left their stamp on curriculum and assessment.

102. The bald story as far as England and Switzerland are concerned is that while in TIMSS 1995 maths Switzerland scored 545 and England 506, in PISA 2000 both scored 529. Prais has been able to show that there were differences between the studies in the types of questions, the target populations and response rates, all of which, he argued, would have tended to do down Switzerland in PISA. He further suggested that the way the scores had been weighted in the item-response procedure tended to obscure Switzerland's true performance. But scrutiny of the detailed results on Zurich's Department of Education's website suggests a more potent factor. There is wide variation between the cantons and it is evident that performance is considerably affected by immigration trends. Native Swiss pupils in St Gallen scored 558 and those in Zurich (where Prais has conducted field studies) scored 554, which are equalled only by the 557 of the top performer Japan, and which are way ahead of native pupils in England who scored 534. Moreover, 21 per cent of Swiss 15-year-olds were non-native compared to the 10 per cent in England.

103. Detailed comparisons between other countries are also likely to show that many of the apparent differences can be more readily interpreted in terms of the methodology of the studies or other factors, rather than the educational policies or schools systems. The differences do not definitively demonstrate that the results of one international study are more credible than those of another, but they do make it difficult to be sure that there have been changes over time.

Relevance to Policy

104. There is a curious contradiction in the design of PISA. It is intended to be a knowledge base for policy analysis, yet it explicitly rejects attempting to assess what pupils have learned in relation to the school curriculum. This puts the onus on PISA to demonstrate that non-curriculum based tests can be used to derive policy conclusions for educational systems.

105. In our comparisons between different studies we have, for example, found Korea and Japan to be consistently at the top in maths and science. Is it reasonable to suppose, therefore, that their education systems are more effective in teaching these subjects and there are lessons to be learned from their approaches? Or could it be due to other differences? The pattern of scores could conceivably reflect inherited abilities, parental support for learning, the seriousness with which the tests had been attempted (they were after all voluntary tests of no consequence to those sitting them), familiarity with the type of question, and a whole host of other factors which are nothing directly to do with the organisation of the schools system.

106. Both TIMSS and PISA have been very disappointing in explaining differences in scores between countries. Both have included what they call contextual measures – questionnaires addressing such features as school and classroom climate, teacher availability, learning outside school, physical infrastructure, school autonomy - but hardly any consistent relationships have been found. This could, of course, actually be the case, but much more likely is that the measures are just not up to capturing the reality. Moreover, the studies have been cross-sectional rather than longitudinal, although PISA is attempting to remedy this.

Interpretation

107. There is a natural wish to be associated with good results. England appears to have done well in PISA and PIRLS (IEA's, 2001, Progress in International Reading Literacy Survey of 10-year-olds), and the results have found their way into a number of ministerial speeches, with the implication that the Labour government's education policies are working. But this is to over-interpret. Even if the PISA results could be shown to represent an improvement on TIMSS they can owe little to recent policies since the 15-year-olds in 2000 will have hardly been touched by them. As regards education, the first Blair government (1997-2001) was mainly concerned with literacy and numeracy in primary schools, so more of a case could have been made for the PIRLS results, but there is no suitable comparative information. Furthermore, in PIRLS it is hard to identify any school factors that might account for the good performance, and it could be that the explanation lies elsewhere. In fact, the PISA and PIRLS results barely correlated, though only half the OECD countries were in PIRLS. Both are best regarded as baselines for future studies.

108. The British government is not alone in tending to see what it wants. Schleicher, as a spokesman for PISA and other OECD compilations, has taken the spread of scores within a country as evidence of the equity of the educational system. He seems to have his own country, Germany, particularly in his sights. Germany comes out with both a relatively low overall score and wide dispersion, and he attributes this to its differentiated and selective secondary education system. Again, as with British politicians, it is making a leap to a causal inference which is not justified. As we have demonstrated (paragraphs 61-64) the inequality measures lack consistency, and many other factors could have contributed to the observed pattern. Germany has, for example, one of the highest proportions of low-scoring immigrants – immigrants who, in many cases, will have been coming to the language afresh, unlike those moving to Anglophone countries who may well have encountered English already.

109. If there is clear evidence that many young people are not achieving what they have the ability to achieve then this is a cause for concern. But underachievement is not the same as a high spread of scores, as we have seen by contrasting England and Korea (paragraphs 65-66). In PISA's terms, the pattern of scores for Korea conforms to the ideal of high average score and low dispersion, while England with a similar average score has a wider spread. But close examination of the distributions suggests the difference is because England's top performers did much better than Korea's, while there was less of a difference at the bottom. A more accurate interpretation than PISA's then is that Korea is not doing as well at the top

end as England. In fact, in the top quartile on parental occupation England's pupils scored highest of all the countries.

110. A factor in England's top ranking is the performance of pupils in independent schools. Across all countries (with the exception of Japan), independent-school pupils did better than those in maintained schools, and those in England did best of all with the gap from the maintained sector the widest. Pupils in government-funded private schools also did better in most countries than those in government schools. These findings are largely ignored in the PISA analyses. In the official reports there is a brief comment about social factors and the results are tucked away in an appendix. It is possible that independent schools have been caught up in the backwash of PISA's implicit dislike of overt selection in state systems. But the failure to explore the apparent success of private schools, whether independent or government-funded, suggests that data attract more attention if they fit in with preconceptions. Such is the mass of numbers generated that there is plenty of scope for projecting patterns on to them.

Value of International Comparisons

111. What then are we to make of large-scale cross-national studies of educational achievement? McGaw (2004) has argued that they provide evidence which leaves us less vulnerable to impression and prejudice. But, as we have seen, ideology can also show through.

112. The OECD is an international quango. The importance it bestows on the PISA findings carries risks. The meaning in a set of numbers can only as good as the methods by which they have been obtained. The PISA programme attempts to operationalize a particular view of the purposes of education: that it can be judged in terms of intended outcomes and chief among those outcomes is the application of knowledge in real-life situations. This is not uncontentious. A good case could be made for education having intrinsic purposes to do with living a rounded life. The danger is that in their desire to move up the league tables countries will be forced by ambitious politicians to adopt the OECD view and become practised in the ways it chooses to frame subjects. Already a number of countries, notably Germany, have been agonising over the PISA findings and wondering what they have to do in order to improve their ranking.

113. Quite an industry has grown up in international comparative studies which has developed a momentum of its own. Both the IEA and PISA were in the field in 2003, and PISA also plans testing in 2006 and 2009. But one wonders how much more cross-sectional studies of this kind will be able to tell us about educational policies and systems. They have already posed some very interesting questions which, in their present form, they themselves do not seem capable of answering. What is it about Finland, Japan and Korea that makes them such high scorers? Why the big difference in the performance of the Flemish-speaking and French-speaking communities in Belgium when the educational systems were originally the same? Belgium would seem to offer the ready-made control groups necessary to make some real discoveries. Why is educational performance in Luxembourg – and indeed in the United States also – apparently so poor when the countries are so good

at wealth creation? Again, why do the pupils of independent schools and government-funded private schools seem to do so well?

114. The answers to these questions, and many others, could have important implications for education policies, but they need to be addressed by in-depth studies with a longitudinal element which offers the prospect of drawing causal inferences. The point of educational research is to have practical application. With better understanding of what lies behind the bare figures of TIMSS or PISA, politicians and policymakers would be in a much better position to decide which elements of the educational systems of successful countries were transferable to their own, and which were inextricably bound up with the people of a particular country and their way of life.

Educational Performance in England

115. International comparisons can be criticised from a number of points of view, but it would be wrong to go to the other extreme and suggest that there is no meaning in the numbers at all. Bearing in mind the caveats, it is possible to come to a broad-brush picture of England's educational performance. Relative to other countries, it is in maths where it looks to be deficient. Setting aside the results from the slimmed down PISA 2000 study England comes consistently near the bottom in the maths comparisons. It is perhaps no coincidence that PISA maths largely ignores the basic arithmetic which gave the English so much difficulty in TIMSS. In the other subjects England emerges not too badly. Its performance in science has been consistently in the top half since the introduction of the National Curriculum. As regards reading, we do not know if there has been any improvement since there are no ready comparators, although it is possible to see hopeful signs in the most recent studies, PISA and PIRLS.

116. But if we also take some of the other studies at face value it is possible to argue that any progress at school is being dissipated in the years afterwards. OECD's figures show that the UK has relatively low participation in formal education and training post 16. The International Adult Literacy Study, whatever its weaknesses, showed that England has nothing to be proud of in respect of adult literacy. Furthermore, the IALS estimates accord with domestic estimates based on Basic Skills Agency surveys (Dearden, McIntosh, Myck and Vignoles, 2000).

117. While policy pointers can be taken from the overall pattern of results, there is little of itself which is conclusive. There is a need now for detailed studies within countries and between groups of countries to test out apparent relationships and discover how much of what is identified as good practice is capable of being transposed. We should be looking for a step-change beyond general description to practical understanding. Maths education in schools would seem to be the prime candidate, with questions to be asked also about learning opportunities beyond school.

References

Adams, C. *Nothing New in Innovation*. www.gtce.org.uk/news/features

Adams, R.J. (2003). Response to 'Cautions on OECD's recent educational survey (PISA)', *Oxford Review of Education,* 29(3), 378-389.

Adams, R.J. and Wu, M. (2002). *PISA 2000 Technical Report*. Paris: OECD.

Baumert, J., Artelt, E., Klieme, E., Neubrand, M., Prenzel, M., Shiefele, U., Schneider, W., Tillman, K.-J. and Weiß, M. (Hrsg.) (2003a). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublick Deutschland.* Oplanden: Leske + Budrich.

Baumert, J., Artelt, E., Klieme, E., Neubrand, M., Prenzel, M., Shiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillman, K.-J. and Weiß, M. (Hrsg.) (2003b). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublick Deutschland. Zusammenfassung Zentraler Befunde*. Berlin: Max-Planck-Institut für Bildungforschung.

Blair, T. (2004). Speech to the Labour Party's Spring Conference, Manchester, 13 March.

Blum, A., Goldstein, H. and Guerin-Pace, F. (2001). 'International Adult Literacy Survey (IALS); an analysis of international comparisons of adult literacy'. *Assessment in Education*, 8, 225-246.

Braams, B.J. (2002). *Mathematics in the OECD PISA Assessment*. www.math.nyu.edu/mfdd/braams/links/pisa0207.html.

Bracey, G.W. (2002). 'Another nation at risk?' *Phi Delta Kappan*, 84(3), 245-247.

Brooks, G., Pugh, A.K. and Schagen, I (1996). *Reading Performance at Nine*. Slough: NFER.

Brown, M. (2003). Presentation to Seminar on Statistical Perspectives on PISA, The Royal Statistical Society, 3 December.

Burghes, D. (1999). 'The Kassel project'. In: *Comparing Standards Internationally* (Eds. B.Jaworski and D. Phillips). Oxford: Symposium Books.

Carey, S. (Ed) (2000). *Measuring Adult Literacy – the International Adult Literacy Survey in the European Context*. London: Office for National Statistics.

Cassidy, S. (2002). 'Compared with many other industrialised nations, school standards are far from impressive, according to new educational research', *The Independent*, 30 October, 2002.

Chew, J (2003). 'Children of few words', *The Daily Telegraph*, 30 August 2003.

Clarke, C. (2003) Speech to NASUWT Conference, Bournemouth, 24 April 2003.

Comber, L.C. and Keeves, J.P. (1973). *Science Education in Nineteen Countries. International Studies in Evaluation I.* Uppsala: Almquist &Wiksell.

Dearden, L., McIntosh, S., Myck, M., and Vignoles, A. (2000). *The Returns to Academic, Vocational and Basic Skills in Britain*. Skills Task Force Research Paper 27. Nottingham: DfES Publications.

De Lange, J. (2002). Presentation to a Symposium on PISA, Berlin 18-20 November, 2002.

Deschamps, R. and van den Kerkhove, M. (2001) 'Education in Europe and French Belgium: what's at stake?' *Reflets et perspectives de la vie économique*, tome XL, no 4, Bruxelles, De Boeck Université

DfEE (2001). *Schools – Building on Success*. Green Paper. Cm 5050. London: The Stationery Office.

DfES (2001). *First Release International Student Assessment.* London: DfES.

DfES (2003). *Participation in Education, Training and Employment by 16-18 Year Olds in England; 2001 and 2002.* London: DfES.

Elley, W.B. (1994). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford (in UK): Published for the International Association for the Evaluation of educational Achievement by Pergamon.

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.

Hazette, P. (2001). Quoted in 'The report of shame', *Le Vif/L'Express*, 7 December 2001.

House of Commons Education and Skills Committee (2003). *Secondary Education: Pupil Achievement*. Seventh Report of the Session 2002-03. HC 513. London: The Stationery Office.

IEA (1999). TMISS 1999 *International Student Achievement in Maths.* http://www.iea.nl.

IEA (1999). TMISS 1999 *International Student Achievement in Science.* http://www.iea.nl.

International Study Centre, Boston College (2002) *TMISS 1999 Maths Items*: *Released Set for Eighth Grade.* http://isc.bc.edul.

International Study Centre, Boston College (2002). *TMISS 1999 Science Items*: *Released Set for Eighth Grade.* http://isc.bc.edul.

Joint Council for General Qualifications (2003), 'GCSE, Entry Level Certificate and GNVQ: Provisional Results', *News Release,* 21 August 2003.

Keys, W., Harris, S. and Fernandes, C. (1996a). *Third International Maths and Science Study. First National Report Part 1.* Slough: NFER.

Keys, W., Harris, S. and Fernandes, C. (1996b). *Third International Maths and Science Study. National Reports Appendices.* Slough: NFER.

Law, N. (2002). 'TIMSS science results for Hong Kong: an ailing dragon with a British legacy'. Chapter 11, pages 157-175, In D. F. Robitaille and A.E.Beaton (Eds.) *Secondary Analysis of the TIMSS Data*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Lafontaine, D. (2002). 'The good (reasoned criticism), the ugly (media excesses) and the villains (Anglo-Saxons).' *La Revue Nouvelle*, Issue 3-4, March – April. (Translated

by Jason Tarsh, Analytical Services, UK Department for Education and Skills, London).

Lokan, J., Cresswell, J. and Greenwood, L. (2001). *15-Up and Counting, Reading, Writing, Reasoning: How Literate are Australian Students? The PISA 2000 Survey of Students' Reading, Mathematical and Scientific Literacy Skills*. Australian National PISA Report. Melbourne: ACER.

Maccoby, E.E. and Jacklin, C.N. (1974). *The Psychology of Sex Differences*. Stanford, California: Stanford University Press.

McGaw, B (2002a). Paragraph 1, Examination of Witnesses, Select Committee on Education and Skills, 20 March 2002, http://www.publications.parliament.uk/pa/cm200102/cmselect

McGaw, B (2002a). Presentation to a Symposium on PISA, Berlin 18-20 November, 2002.

McGaw, B (2004). Learning Power of PISA, *Times Educational Supplement*, 12 March 2004.

Micklewright, J (2003) In discussion at Seminar on Statistical Perspectives on PISA, The Royal Statistical Society, 3 December 2003.

Miliband, D. (2003a). Speech to ATL Conference, Blackpool, 15 April 2003.

Miliband, D. (2003b). Interview in *The Guardian*, 5 May 2003.

Moser, C. (1999). *Improving Literacy and Numeracy: A Fresh Start*. The report of the working group chaired by Sir Claus Moser. London: DfEE, www.lifelonglearning.co.uk/mosergroup.

Mouton, O. (2001). 'The Flemish school of success', *La Libre Belgique*, 19 December 2001.

National Foundation for Educational Research (2003). *The PIRLS England Report.* www.teachernet.gov.uk/pirls.

National Literacy Trust (1999). *Partnerships for Literacy* Annual Conference Report. London: NLT.

Normington, D. (2002a). 'Transforming secondary education', *Education Journal*, issue 62, June 2002.

Normington, D. (2002b). Paragraph 122, Examination of Witnesses, Select Committee on Education and Skills, 26 June 2002, http://www.publications.parliament.uk/pa/cm200102/cmselect .

OECD (2001). *Knowledge and Skills for Life: First Results from PISA 2000*. Paris: OECD.

OECD (2002a). *Sample Tasks from the PISA 2000 Assessment*: *Reading, Mathematical and Scientific Literacy*. Paris: OECD.

OECD (2002b). 'Improving both quality and equity – insights from PISA 2000' *Educational Policy Analysis*, vol 2002, no 9, Edition, pp 46-81. Centre for Educational Research and Innovation. Paris: OECD.

OECD (2002c). *Reading for Change: Performance and Engagement Across Countries*. Paris: OECD.

OECD (2003a). *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris: OECD.

OECD (2003b) *Education at a Glance, OECD Indicators*. Paris: OECD.

OECD (2003c). *The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving, Knowledge and Skills*. Paris: OECD.

OECD and Statistics Canada (2000). *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. Paris: OECD.

Office for Standards in Education (2000). *Improving City Schools*. London: Ofsted.

O'Malley, B. (2002). '…and our results lag behind', *Times Educational Supplement*, 1 November 2002.

Prais, S.J. (1995). *Productivity, Education and Training*. Cambridge: The University Press.

Prais, S.J. (2003a). 'Cautions on OECD's recent educational survey (PISA)', *Oxford Review of Education* 29*(*2), 139-163.

Prais, S.J. (2003b). 'More questions than answers – the OECD Educational Survey'. *National Institute of Social and Economic Research 61st Annual Report.* London: NIESR.

Porter, A. C. and Gamoran, A. (eds.) (2002). *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington, DC: National Academy Press.

Reynolds, D. and Farrell, S. (1996). *Worlds Apart? A Review of International Surveys of Educational Achievement Involving England*. London: HMSO.

Rocher, T. (2003). *On the Methods Used for International Assessments of Educational Competences*. Paris: Office of Statistical Studies, French National Ministry of Education. (Translated by Jason Tarsh, Analytical Services, UK Department for Education and Skills, London, April, 2003).

Romainville, M. (2002). 'On the appropriate use of PISA.' *La Revue Nouvelle*, Issue 3-4, March – April. (Translated by Jason Tarsh, Analytical Services, UK Department for Education and Skills, London).

Schleicher, A. (2003) 'Pisa, engagement and motivation, precursors for performance and destinations'. Presentation at ESRC Seminar Series, London School of Economics, 14 February.

Shiel, G., Cosgrove, J., Sofroniou, N. and Kelly, A. (2002). *Ready for Life? The Literacy Achievements of Irish 15-Year-Olds With comparative International Data*. Dublin: Educational Research Centre.

Smith, E. and Gorard, S. (2002a)'International equity indicators in education; defending comprehensive schools III' Forum, 44, pp 121-122.

Smith, E. and Gorard, S. (2002b) *What does PISA tell us about equity in education systems?* Occasional Paper 54, Cardiff: University School of Social Sciences.

Sutherland, M.B. (1981). *Sex Bias in Education*. Oxford: Basil Blackwell.

Timmins, N. (2002). 'UK tumbles in secondary schools results table', *Financial Times*, 30 October, 2002.

Twist, L., Sainsbury, M., Woodthorpe, A. and Whetton, C. (2003). *Reading All Over The World.* National Report for England on PIRLS, Progress in International Reading Literacy Study. Slough: NFER.

Välijärvi, J., Linnakylä, P., Kupari, P., Reinikainen, P. and Arffman, I. (2002). *The Finnish Success In PISA – And Some Of The Reasons Behind It.* University of Jyväskylä: Institute for Educational Research.

Appendix

In the appendix we include a number of tables adapted from the various data sources which are discussed and sometimes condensed in the text, but which themselves are so large they interrupt the flow of the argument. The tables include:

- A1. Trends in maths rankings.
- A2. Trends in science rankings.
- A3. Gender differences in the PISA 2000.
- A4. Country means and standard deviations in reading scores in PISA 2000.
- A5. Percentage of students at each level of reading proficiency in PISA 2000.
- A6. Ranking of reading scores by occupation of parents.
- A7. Ranking of reading scores by immigration status.
- A8. Percentage successfully completing upper secondary education.
- A9. GDP, education spend, participation and PISA performance.

Table A1: Trends in Maths Rankings[1]

Country[2]	PISA 2000 Age 15	TIMSS 1999 Age 14	TIMSS 1995 Age 14	IAEPM 1990 Age 13	SIMS 1982 Age 13	FIMS 1964 Age 13
Japan	1	5	3		1	2=
Korea	2	2	2	2=		
Belgium (Flemish)	3	6	5		4	3
New Zealand	4	21	15		13	
Finland	5	14=			11	4
Canada	6=	10	12=	10	7	
Australia	6=	13				10
Scotland	6=			11=	8	7
England	9=	20	16	11=	10	6
Switzerland	9=		8	4		
France	11		9	7	5	9
Austria	12					
Iceland	13=		20=			
Denmark	13=					
Liechtenstein	13=					
Sweden	16		14		16	12
Ireland	17		12=	11		
Norway	18		17			
Czech Republic	19	14=	6		14	
USA	20	19	18	15		11
Belgium French	21					
Germany	22					
Hungary	23	9	10	6	3	
Russia	24	12	11	5		
Spain	25		20=	15		
Poland	26					
Latvia	27	18	19			
Italy	28	23		7		
Portugal	29		24	17		
Greece	30					
Luxembourg	31					
Mexico	32					
Brazil	33					
Singapore		1	1			
China				1		
Taipei		3		2=		
Hong Kong		4	4		9	
Netherlands		7			2	8
Slovak Republic		8	7			
Israel		28		9	12	1

1. For explanation of acronyms see Table 1.

2. Countries participating in two or more of the studies, plus notable performances.

Sources: OECD (2001) *Knowledge and Skills for Life: First Results from PISA 2000.* Paris: OECD; *TMISS 1999 International Student Achievement in Maths* http://www.iea.nl; Keys, W., Harris, S. and Fernandes, C. (1996) *Third International Maths and Science Study. First National Report Part 1*. Slough: NFER; Reynolds, D. and Farrell, S. (1996) *Worlds Apart? A Review of International Surveys of Educational Achievement Involving England*. London: HMSO.

Table A2: Trends in Science Rankings[1]

Countries[2]	PISA 2000 Age15	TIMSS 1999 Age 14	TIMSS 1995 Age 14	IAEPS 1990 Age 13	SISS 1982 Age 14	FISS 1964 Age 14
Korea	1	5	4	1	4	
Japan	2	4	3		2	1
Finland	3	10=			5	11
England	4	9	6	9=	16	9
Canada	5	13=	13	9=	6	
Australia	6=	7			10	3
New Zealand	6=	19	15			4
Scotland	8			12		8
Belgium (Flemish)	9=	10=	7	6=		10
Austria	9=					
Ireland	11		9=	16		
Sweden	12		11		12	6
Czech Republic	13	8	2			
France	14=		19	9=		
Norway	14=		14		9	
USA	16	18	12	14		7
Switzerland	17=		16=	3		
Iceland	17=		20			
Hungary	17=	3	5	4	14	2
Spain	20		18	12		
Germany	21					
Poland	22					
Denmark	23					
Italy	24	21		6=	15	12
Liechtenstein	25					
Belgium French	26					
Greece	27					
Russia	28=	16	9	5		
Latvia	28=	20	21			
Portugal	30		22	16		
Luxembourg	31					
Mexico	32					
Brazil	33					
Taipei		1		2		
Singapore		2	1			
Netherlands		6			3	13
Slovak Republic		10=	8			
Hong Kong		15	16		17	
Israel		26		6=	11	
China				14	7	

1. For explanation of acronyms see Table 1.

2. Countries participating in two or more of the studies, plus notable performances.

Sources: OECD (2001) *Knowledge and Skills for Life: First Results from PISA 2000.* Paris: OECD; *TMISS 1999 International Student Achievement in Science* http://www.iea.nl; Keys, W., Harris, S. and Fernandes, C. (1996) *Third International Maths and Science Study. First National Report Part 1*. Slough: NFER; Reynolds, D. and Farrell, S. (1996) *Worlds Apart? A Review of International Surveys of Educational Achievement Involving England*. London: HMSO; Comber, L.C. and Keeves, J.P. (1973) *Science Education in Nineteen Countries*. International Studies in Evaluation I. Uppsala: Almquist & Wiksell.

Table A3: Gender Differences in PISA 2000

Country	Points Score in Favour of Girls[1]		
	Reading	Science	Maths
Finland	51	6	*-1*
New Zealand	46	12	3
Norway	43	7	*-11*
Iceland	40	5	5
Italy	38	9	*-8*
Greece	37	7	*-7*
Sweden	37	0	*-7*
Czech Republic	37	*-1*	*-12*
Poland	36	*-6*	*-5*
Germany	35	*-3*	*-15*
Australia	34	3	*-12*
Belgium	33	2	*-6*
Hungary	32	2	*-7*
Canada	32	2	*-10*
Japan	30	7	*-8*
Switzerland	30	*-7*	*-14*
Ireland	29	6	*-13*
France	29	*-6*	*-14*
USA	28	4	*-10*
Luxembourg	27	7	*-15*
United Kingdom	26	*-4*	*-8*
Austria	26	*-12*	*-27*
Portugal	25	6	*-19*
Denmark	25	*-12*	*-15*
Spain	24	*-1*	*-18*
Mexico	20	*-4*	*-11*
Korea	14	*-19*	*-27*
OECD Total	29	0	*-11*

1. Correlation between reading and science differences 0.619 (P>0.01), reading and maths differences 0.738 (P>0.01), and maths and science differences 0.625 (P>0.01).

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 5.1a, page 276.

Table A4: Country Means and Standard Deviations in Reading Scores in PISA 2000

Country	Standard Deviation		Mean	
	Score	Rank	Score	Rank
Germany	111	*1*	484	*21*
New Zealand	108	*2*	529	*3*
Belgium	107	*3*	507	*10=*
USA	105	*4*	504	*15*
Norway	104	*5*	505	*13=*
Australia	102	*6=*	528	*4*
Switzerland	102	*6=*	494	*17*
United Kingdom	100	*8=*	523	*7*
Poland	100	*8=*	479	*23*
Luxembourg	100	*8=*	441	*26*
Denmark	98	*11*	497	*16*
Greece	97	*12=*	474	*24*
Portugal	97	*12=*	470	*25*
Czech Republic	96	*14*	492	*19*
Canada	95	*15*	534	*2*
Ireland	94	*16=*	527	*5*
Hungary	94	*16=*	480	*22*
Austria	93	*18*	507	*10=*
Sweden	92	*19=*	516	*9*
Iceland	92	*19=*	507	*10=*
France	92	*19=*	505	*13=*
Italy	91	*22*	487	*20*
Finland	89	*23*	546	*1*
Japan	86	*24=*	522	*8*
Mexico	86	*24=*	422	*27*
Spain	85	*26*	493	*18*
Korea	70	*27*	525	*6*
OECD Total	100		499	

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 2.3a, page 253.

Table A5: Percentage of Students at Each Level of Reading Proficiency in PISA 2000

	Level 5 (Above 625)	Level 4 (553- 625)	Level 3 (481-552)	Level 2 (408-480)	Level 1 (335-407)	< Level 1 (< 335)
New Zealand	18.7	25.8	24.6	17.2	8.9	4.8
Finland	18.5	31.6	28.7	14.3	5.2	1.7
Australia	17.6	25.3	25.7	19.0	9.1	3.3
Canada	16.8	27.7	28.0	18.0	7.2	2.4
United Kingdom	15.6	24.4	27.5	19.6	9.2	3.6
Ireland	14.2	27.1	29.7	17.9	7.9	3.1
USA	12.2	21.5	27.4	21.0	11.5	6.4
Belgium	12.0	26.3	25.8	16.8	11.3	7.7
Sweden	11.2	25.6	30.4	20.3	9.3	3.3
Norway	11.2	23.7	28.1	19.5	11.2	6.3
Japan	9.9	28.8	33.3	18.0	7.3	2.7
Switzerland	9.2	21.0	28.0	21.4	13.3	7.0
Iceland	9.1	23.6	30.8	22.0	10.5	4.0
Austria	8.8	24.9	29.9	21.7	10.2	4.4
Germany	8.8	19.4	26.8	22.3	12.7	9.9
France	8.5	23.7	30.6	22.0	11.0	4.2
Denmark	8.1	22.0	29.5	22.5	12.0	5.9
Czech Republic	7.0	19.8	30.9	24.8	11.4	6.1
Poland	5.9	18.6	28.2	24.1	14.6	8.7
Korea	5.7	31.1	38.8	18.6	4.8	0.9
Italy	5.3	19.5	30.6	25.6	13.5	5.4
Liechtenstein	5.1	19.5	30.1	23.2	14.5	7.6
Hungary	5.1	18.5	28.8	25.0	15.8	6.9
Greece	5.0	16.7	28.1	25.9	15.7	8.7
Spain	4.2	21.1	32.8	25.7	12.2	4.1
Portugal	4.2	16.8	27.5	25.3	16.7	9.6
Latvia	4.1	13.8	25.2	26.3	17.9	12.7
Russia	3.2	13.3	26.9	29.2	18.5	9.0
Luxembourg	1.7	11.2	24.6	27.5	20.9	14.2
Mexico	0.9	6.0	18.8	30.3	28.1	16.1
Brazil	0.6	3.1	12.9	27.7	32.5	23.3
OECD Total	9.4	21.8	28.6	21.8	12.1	6.2

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 2.1a, page 246.

Table A6: Ranking of Reading Scores by Occupation of Parents[1]

Country[2]	Socio-Economic Index[1]							
	Top Quarter		Second Quarter		Third Quarter		Bottom Quarter	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
United Kingdom	579	*1*	543	*4*	513	*7*	491	*9*
Finland	576	*2=*	555	*1*	535	*1*	524	*1*
Australia	576	*2=*	538	*5*	523	*4=*	490	*5*
New Zealand	574	*4*	549	*2*	523	*4=*	489	*6*
Canada	570	*5=*	545	*3*	529	*2*	503	*3*
Ireland	570	*5=*	535	*7*	520	*6*	491	*4*
Belgium	560	*7*	537	*6*	497	*12*	457	*16=*
Sweden	558	*8*	522	*10=*	509	*8*	485	*8*
USA	556	*9*	528	*9*	507	*9*	466	*13*
France	552	*10*	520	*12*	496	*12=*	469	*11*
Switzerland	549	*11*	513	*14=*	492	*15*	434	*22*
Austria	547	*12=*	522	*10=*	500	*10*	467	*12*
Norway	547	*12=*	514	*13*	494	*14*	477	*10*
Denmark	543	*14=*	511	*17*	490	*16*	465	*14*
Czech Republic	543	*14=*	499	*20*	487	*17*	445	*18=*
Korea	542	*16*	531	*8*	524	*3*	509	*2*
Germany	541	*17*	513	*14=*	471	*21*	427	*24*
Iceland	540	*18*	513	*14=*	496	*12=*	487	*7*
Poland	534	*19*	493	*22*	472	*20*	445	*18=*
Hungary	531	*20*	504	*19*	461	*22*	435	*21*
Spain	529	*21*	507	*18*	482	*18*	461	*15*
Portugal	527	*22*	485	*24*	452	*24*	431	*23*
Italy	525	*23*	494	*21*	481	*19*	457	*16=*
Greece	519	*24*	486	*23*	460	*23*	440	*20*
Luxembourg	497	*25*	473	*25*	428	*25*	394	*25*
Mexico	471	*26*	435	*26*	408	*26*	385	*26*
OECD Total	543		515		492		462	

1. International Socio-Economic Index of Occupational Status (ISEI) derived from students' responses on parental occupation, father's or mother's taken whichever is the higher.

2. Japan excluded because too many missing cases.

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 6.1a, page 283.

Table A7: Ranking of Reading Scores by Immigration Status[1]

Country	Native Students		First Generation		Non-Native Students	
	Score	% Students	Score	% Students	Score	% Students
Finland	548	98.7	-	0.2	468	1.0
Canada	538	79.5	539	10.8	511	9.8
New Zealand	538	80.4	507	6.4	507	13.2
Australia	532	77.4	528	10.7	513	11.9
United Kingdom	528	90.4	510	7.0	456	2.6
Ireland	528	97.7	519	0.9	573	1.4
Japan	525	99.9	-	-	-	0.1
Sweden	523	89.5	485	4.7	450	5.9
Belgium	522	88.0	411	8.6	431	3.4
Austria	515	90.4	453	3.7	422	5.9
France	512	88.0	471	9.8	434	2.2
Switzerland	514	79.3	460	9.3	402	11.4
USA	511	86.4	478	7.4	466	6.1
Norway	510	95.4	464	1.5	449	3.1
Iceland	509	99.2	-	0.2	-	0.6
Germany	507	84.8	432	5.1	419	10.1
Denmark	504	93.8	409	2.4	433	3.8
Czech Republic	501	98.9	-	0.6	-	0.5
Spain	494	98.0	450	0.6	460	1.4
Italy	489	99.1	-	0.2	445	0.8
Hungary	482	98.3	-	0.1	486	1.6
Poland	482	99.7	-	0.0	-	0.2
Greece	478	95.2	-	0.5	403	4.3
Luxembourg	474	65.8	399	17.8	370	16.4
Portugal	472	96.9	463	1.8	450	1.4
Mexico	427	96.4	378	1.1	329	2.5
OECD Total	503	91.3	479	4.6	452	4.1

1. Based on students' self-reporting.

Source: OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Table 6.10, page 293.

Table A8: Percentage Successfully Completing Upper Secondary Education[1]

Country[2]	Age 25-34		Age 35-44		Age 45-54		Age 55-64	
	%	Rank	%	Rank	%	Rank	%	Rank
Korea	95	*1*	77	*14*	49	*19*	30	*22*
Japan	94	*2=*	94	*1*	81	*6=*	63	*10*
Norway	94	*2=*	91	*1*	82	*5*	71	*6*
Switzerland	92	*4=*	90	*3=*	85	*2*	81	*2*
Czech Republic	92	*4=*	90	*3=*	84	*3*	76	*3=*
Sweden	91	*6*	86	*6=*	78	*9*	65	*8=*
Canada	89	*7*	85	*8*	81	*6=*	67	*7*
USA	88	*8*	89	*5*	89	*1*	83	*1*
Finland	87	*9*	84	*9*	70	*13*	51	*13*
Denmark	86	*10*	80	*11=*	80	*8*	72	*5*
Germany	85	*11*	86	*6=*	83	*4*	76	*3=*
Austria	84	*12*	81	*10*	73	*11*	65	*8=*
New Zealand	82	*13*	80	*11=*	75	*10*	60	*11*
Hungary	81	*14*	79	*13*	72	*12*	44	*16=*
France	78	*15*	67	*15*	58	*15*	46	*14=*
Belgium	76	*16*	64	*17*	53	*18*	39	*19*
Ireland	73	*17=*	62	*18*	48	*20*	35	*21*
Greece	73	*17=*	60	*19=*	43	*23*	28	*23*
Australia	71	*19*	60	*19=*	55	*17*	44	*16=*
United Kingdom	68	*20*	65	*16*	61	*14*	55	*12*
Iceland	61	*21*	60	*19=*	56	*16*	46	*14=*
Luxembourg	59	*22*	57	*22*	47	*21*	42	*18*
Italy	57	*23=*	49	*23*	39	*24*	22	*24*
Spain	57	*23=*	45	*25*	29	*25*	17	*25*
Poland	52	*25*	48	*24*	44	*22*	36	*20*
Portugal	32	*26*	20	*27*	14	*27*	9	*27*
Mexico	25	*27*	25	*26*	17	*26*	11	*26*
Mean	74		69		60		49	

1. Countries define upper secondary in their own ways. Some have examinations others do not. Upper secondary in England is schooling from 14 and further education. Successful completion is 5 or more GCSEs at grades A* -C or NVQ2 or above.

2. Not including the Netherlands, Slovak Republic, Turkey which have not been covered in the other tables.

Source: OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table A1.2, page 41.

Table A9: GDP, Education Spend, Participation and PISA Performance

Country	GDP Per Capita[1]		Spend on Secondary Educ[2]		Completed Upper Sec Education[3]		Overall PISA Score[4]	
	$ (PPP)	Rank	$ (PPP)	Rank	%	Rank	Score	Rank
Luxembourg	48,239	*1*	-		53	*21*	1,330	*26*
Norway	36,202	*2*	8,476	*4*	86	*3=*	1,504	*15*
USA	34,602	*3*	8,855	*2*	88	*1*	1,496	*17*
Switzerland	29,617	*4*	9,780	*1*	87	*2*	1,519	*13*
Denmark	28,755	*5*	7,726	*5*	80	*9*	1,492	*18*
Ireland	28,285	*6*	4,618	*19*	58	*19*	1,543	*8*
Iceland	28,143	*7*	6,518	*11*	57	*20*	1,517	*14*
Canada	28,130	*8*	5,947	*16*	82	*7*	1,596	*4*
Austria	28,070	*9*	8,578	*3*	77	*10*	1,541	*9*
Belgium	26,392	*10*	6,889	*9*	59	*17=*	1,523	*11*
Australia	26,325	*11*	6,894	*8*	59	*17=*	1,589	*6*
Sweden	26,161	*12*	6,339	*12*	81	*8*	1,538	*10*
Germany	26,139	*13*	6,826	*10*	83	*5=*	1,461	*20*
Finland	25,357	*14*	6,094	*14*	74	*12*	1,620	*3*
Japan	26,011	*15*	6,266	*13*	83	*5=*	1,629	*1*
Italy	25,095	*16*	7,218	*7*	43	*24*	1,422	*23*
France	25,090	*17*	7,636	*6*	64	*15*	1,522	*12*
United Kingdom	24,964	*18*	5,991	*15*	63	*16*	1,584	*7*
New Zealand	20,372	*19*	-		76	*11*	1,594	*5*
Spain	20,195	*20*	5,185	*18*	40	*25*	1,460	*21*
Portugal	16,780	*21*	5,349	*17*	20	*27*	1,383	*24*
Greece	15,885	*22*	3,859	*21*	51	*22*	1,382	*25*
Korea	15,186	*23*	4,069	*20*	68	*14*	1,624	*2*
Czech Republic	13,806	*24*	3,239	*22*	86	*3=*	1,501	*16*
Hungary	12,204	*25*	2,446	*23*	70	*13*	1,464	*19*
Poland	9,547	*26*	-		46	*23*	1,432	*22*
Mexico	9,117	*27*	1,615	*24*	22	*26*	1,231	*27*

1. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table X2.1, page 418.

2. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table B1.1, page 197.

3. Adapted from OECD (2003b) *Education at a Glance,* OECD Indicators. Paris: OECD, Table A1.2, page 41.

4. Adapted from OECD (2001) *Knowledge and Skills for Life,* First Results from PISA. Paris: OECD, Tables 2.3a, 3.1 and 3.3, pages 253, 258 and 261.