

---

## Can We Believe the International League Tables?

PETER WILBY

**ABSTRACT** This article, updated and expanded from one written for *The Times Educational Supplement*, 10 December 2010, asks whether politicians are right to quote the country's performance in international tests in support of such policies as re-introducing O levels. It finds reasons to doubt that the tests give an adequate picture of children's learning, in comparison with either older cohorts or overseas peers. Nor do they provide a fair measure of schools' success or the validity of examination qualifications.

The Coalition Government's belief that English schools require transformational change is based almost entirely on England's performance in tests carried out on 15-year-olds by the Programme for International Student Assessment (PISA). We know this because the Prime Minister and his deputy, as well as Michael Gove, the education secretary, have told us so. In their introduction to the 2010 schools White Paper, David Cameron and Nick Clegg wrote: 'We are standing still while others race past.' In January 2011, Gove told the World Education Forum that PISA 'shows that we are falling further and further behind other nations'. He made the same point, more theatrically, in the House of Commons a month later: 'Literacy, down; numeracy, down; science, down; fail, fail, fail.' In June 2012, after Gove's plans to scrap the GCSE and bring back O levels were leaked to the *Daily Mail*, the PISA results were again quoted in the subsequent Commons debate.

On a first and superficial glance, the results appear to support ministers' concerns. In 2000, England was seventh in the world in reading, eighth in maths and fourth in science. By 2009 (the tests are at three-yearly intervals), the nation had fallen to 25th, 27th and 16th respectively. Uncoincidentally – Cameron and Gove, if not Clegg, would like us to believe – the 2000 cohort had received all but three years of their education under Conservative governments, while the 2009 cohort had received all but three years of theirs under Labour.

But there are several reasons for questioning whether the decline in England's PISA results bears any relation to reality. First, there were sampling errors in the 2000 results which, most likely, led to the English performance being overestimated. The OECD, which runs the PISA tests, at first took an indulgent view of the errors but, when further errors occurred in 2003, decided to scrub both years from the record. A note to its report on the 2009 results states that 'no trend comparisons are possible with these years'. That still leaves England with a decline between 2006 and 2009 but, in all three subjects, one that is well within the bounds of measurement error.

Second, English pupils took the 2000 and 2003 tests in the spring. Schools complained that the timing interfered with preparation for GCSE exams – not a problem for most other countries which, if they have such exams at all, do not place nearly as much weight on them as Britain does – and this, indeed, partly accounted for the inadequate response rate. From 2006, therefore, the tests were taken in the autumn, five months earlier. In an Institute of Education study, Dr John Jerrim has estimated that the extra schooling time would have given the earlier cohorts an advantage of 15 points on the PISA scales. If we 'award' these points to the 2009 pupils, they move up to 10th in reading, 18th in maths and joint eighth in science, still down from the (unreliable) 2000 results that ministers quote but hardly 'fail, fail, fail'.

Third, the PISA tests cannot offer a guide to the relative success of different school systems. It tests 15-year-olds on the 'cumulative impact of learning ... experiences both in school and at home'. The tests are not about knowledge and understanding of school subjects as conventionally defined; they are about what PISA rather confusingly calls 'literacy' – in maths and science, as well as in the native language – meaning the ability to use these subjects in everyday life. Pupils who haven't studied science often do better on the PISA science tests than do pupils who have. PISA, you may say, tests exactly what should be tested: a 15-year-old's fitness to cope with the adult world. But since it is impossible to tell whether children have acquired the relevant skills and knowledge inside or outside school, it hardly seems fair to use the results to judge the effectiveness of schools, teachers, examination systems and governments.

That takes us to a fourth issue: England's apparently poor performance on the PISA tests does not accord with their results in another international survey: the Trends in International Mathematics and Science Study (TIMSS). This study gears its tests to subject knowledge and may therefore be regarded as more academic than PISA. In 2007, English year 9 pupils were outscored on maths by only five countries, all from the Pacific Rim, and on science by only four, again from the Pacific Rim. In science, the English scores were almost identical to those of the previous survey in 2003 and, in maths, they were significantly higher. The results were similarly impressive for year 5 children. I can find no record of Gove ever quoting TIMSS. The survey contradicts his stated opinions not only on the overall failure of English children but also on the source of that

alleged failure. He thinks learning in English schools is insufficiently academic. England's success in TIMSS, but failure in PISA, suggests the precise opposite.

If Gove is guilty of cherry-picking the findings that suit him, we should not fall into the same trap. We should note that a third survey – the Progress in International Reading Literacy Study (PIRLS) – showed a significant decline in the scores of English 10-year-olds between 2001 and 2006, and a slump in their league table position from fourth to 15th. But there were problems with that study, too, not least that, in nine of the 14 countries that scored higher than England, the children taking the test were older, sometimes by more than a year. Moreover, of 10 blocks of items in the tests, only four were common to both 2001 and 2006. When the National Foundation for Educational Research (NFER) examined these, it concluded that 'none of the differences between 2001 and 2006 is statistically significant'. Establishing whether the trend in test results is up or down depends enormously on the methods used to equate different items in tests given in different years. Referring to both PISA and PIRLS, the NFER wrote: 'There is no one single equating methodology which both links countries and produces robust estimates of change over time.' You have to delve very deep into the appendices and footnotes of international test reports to discover these wrinkles. The testing industry has no more interest than the politics industry in confessing that these expensively obtained results cannot give a definitive verdict on different education systems.

The truth is that international comparisons in education are fraught with difficulties. The validity and reliability of the league tables for English schools are endlessly debated. What allowances should be made for home background or for children whose mother tongue isn't English? Should children with special needs be included? How do we account for absentees? How can we stop teaching to the test or outright cheating? All these and other issues – for example, national differences in the seriousness with which children approach tests – affect the results on PISA and other international studies. Problems with sampling are not confined to England. Austria, for example, was alarmed to discover its league table position had slumped in 2003 compared with 2000. Only then did it find that the 2000 sample underweighted vocational schools and the decline was therefore illusory. The inconsistencies between tests sometimes border on absurdity. Hungary comes near the top in TIMSS, but well down the PISA league; for New Zealand, it's other way round. Italy comes fifth in PIRLS (10-year-olds), and 29th in the PISA reading test (15-year-olds). Is Italy to conclude that its primary schools are brilliant and its secondary schools sensationally inept? Or that the tests are measuring different things?

Everybody loves league tables. They are harmless enough (and probably accurate enough) in a sporting context. But what is the point of these international comparisons? Politicians argue that, in the era of globalisation, they have a duty to ensure their citizens acquire the necessary levels of skills and knowledge to compete in international markets. PISA, TIMSS and PIRLS help to assess how the next generation of workers in each country compares with its overseas rivals. Yet the world's richest country, the United States, performs

*Peter Wilby*

badly in all international educational comparisons and has done for years; Norway and Luxembourg, two of the top eight in GDP per capita, struggle in the lower echelons of the PISA league; the Russian Federation stands proudly at the top of the PIRLS league (though one wonders whether its test results are any more reliable than its election results) but 70th in GDP per capita. As the Americans say, go figure.

---

**PETER WILBY**, a former education correspondent and a former editor of *The Independent on Sunday* and *New Statesman*, is now a columnist for *The Guardian* and *New Statesman*. Correspondence: [peter.wilby@gmail.com](mailto:peter.wilby@gmail.com)