

REPORT

BENCHMARKING THE ENGLISH SCHOOL SYSTEM

AGAINST
THE BEST
IN THE WORLD

Jonathan Clifton

July 2011
© IPPR 2011



ABOUT THE AUTHOR

Jonathan Clifton is a research fellow at IPPR.

ACKNOWLEDGMENTS

With thanks to the NASUWT for their generous support of this research paper.

ABOUT IPPR

IPPR, the Institute for Public Policy Research, is the UK's leading progressive thinktank. We produce rigorous research and innovative policy ideas for a fair, democratic and sustainable world.

We are open and independent in how we work, and with offices in London and the North of England, IPPR spans a full range of local and national policy debates. Our international partnerships extend IPPR's influence and reputation across the world.

IPPR
4th Floor
14 Buckingham Street
London WC2N 6DF
T: +44 (0)20 7470 6100
E: info@ippr.org
www.ippr.org
Registered charity no. 800065

This paper was first published in July 2011. © 2011
The contents and opinions expressed in this paper are those of the author(s) only.

IDEAS to
CHANGE MINDS

Contents

| | |
|---|----|
| Introduction | 2 |
| Methods | 2 |
| Why should we benchmark the English school system against other countries? | 3 |
| Which benchmarks could be used? | 3 |
| How have other countries used benchmarks? | 5 |
| The issue of relevance: five questions to ask before introducing a benchmark..... | 7 |
| What should be benchmarked? | 8 |
| Some pitfalls of international benchmarks | 10 |
| Conclusion: developing international benchmarks in England..... | 11 |
| References | 13 |

Introduction

‘A generation ago, a British Prime Minister had to worry about a global arms race. Today a British Prime Minister has to worry about a global skills race.’

Gordon Brown, 2008

‘Instead of comparing ourselves with the past, we should compare ourselves with the best.’

Michael Gove, 2010

While the threat of foreign competition has long been used to justify education reforms, the use of international comparisons has reached something of a crescendo in recent years. In an attempt to improve their school systems, policymakers are increasingly turning to other countries for inspiration and ideas.

If global competition is the rhetoric driving international comparisons, then the OECD’s Programme of International Student Assessment (PISA) is the tool that enables them to happen. First developed in 1998 with 28 countries taking part, PISA now covers 74 education systems that make up 86 per cent of the world economy. Almost half of OECD countries now have some form of international benchmark or target in place in their school systems.

In England, the reaction to international assessments has generally been dominated by media hysteria and short-lived political panic. Only in the last couple of years have attempts been made to systematically benchmark our school performance against other countries. For the previous government, this culminated in the Prime Minister setting a target to be ‘in the top three in science and top five in maths out of all OECD countries by 2015’ (Brown 2010). The current government has subsequently legislated for Ofqual to benchmark English exams against those in other countries (Ofqual 2011), and the ongoing curriculum review has promised to draw heavily on lessons from overseas (Oates 2010).

This interest in using international benchmarks to drive improvement has the potential to move debates beyond ‘shock reactions’ to our rank position in league tables. But it also raises fundamental questions about the purpose and design of international assessments, and how they can be used in a progressive way that reflects their limitations.

Given these are recent developments in the English context, a better understanding of international benchmarking is needed. This paper explores the principles behind international benchmarking, how it is being carried out in different countries, what factors policy makers need to take into account when introducing benchmarks, and some potential pitfalls that they need to be aware of.

Methods

This briefing paper is based on an expert seminar on international benchmarking held by IPPR and NASUWT in London in March 2011. The seminar included presentations from Andreas Schleicher, Head of the OECD’s Programme of International Student Assessment (PISA), and Tim Oates, Group Director of Assessment, Research and Development at Cambridge Assessment. It was followed by a roundtable discussion with 20 leading experts in this field. Research for this paper was supplemented with an additional literature review and four expert interviews. The views expressed in this paper are those of the author and do not necessarily reflect the views of everyone participating in the research.

While the paper draws on a number of different international comparisons, the analysis largely focuses on the OECDs PISA assessment. This reflects the dominance of PISA in current policy debates, as well as the focus of the expert seminar on which the research was based.

Why should we benchmark the English school system against other countries?

The recent turn to compare our school performance against other countries has been driven by five factors:

- It enables England to get a sense of the potential for its system to improve. This is not something that can be achieved through more traditional domestic assessments. As Andreas Schleicher argues, without international comparisons ‘you never know how good is good enough’.
- It provides insights that cannot be garnered solely from looking within the English system. It can help us to understand the relative strengths and weaknesses of our own system, identify best practice, and inject energy into making improvements.
- It can provide a source of ‘external challenge’ to hidden assumptions within our own system that may be holding it back. For example, in Germany there was an overwhelming assumption that children should be streamed at an early age – something that wasn’t challenged until they looked at the international evidence against it.
- It can provide a ‘peg’ to measure progress against. Just as countries peg their currencies, so having an external mark to judge educational improvements against can prevent grade inflation or the easing of standards.
- It enables us to compare our skills base against other economies. While the role of skills in making an economy competitive has been overemphasized in recent debates (Keep 2008), it is true that the emergence of many countries into the knowledge economy since the 1970s has increased competition for skilled workers.

The benefits of international benchmarking therefore lie in its ability to improve England’s own school system. This is not something that can be achieved through a simple ranking of how we compare against other countries. It also requires identifying how other countries achieved their outcomes and whether this can inform practice at home.

Which benchmarks could be used?

While PISA is the most widely used international assessment, there are a number of different studies that can be used for benchmarking performance. Each study has a slightly different design and focus. Box 1 (over) summarises the main assessments that could be used.

Box 1: International assessments of school systems***Programme for International Student Assessment (PISA)***

PISA is run by the OECD and takes place every three years. It is a sample survey that assesses 15–16 year olds in three areas: literacy, maths and science. Each wave of the study includes a particular focus on one of those areas to give more detailed results. Covering 74 school systems it is the largest international assessment.

PISA puts less emphasis on whether a student can reproduce content, and focuses more on their ability to apply knowledge to solve tasks. This is important in the modern labour market which privileges people's ability to transfer their knowledge and skills to new situations. Pupils and head teachers also complete a survey so scores can be linked with contextual data about a pupil's social background, attitudes towards learning, and the nature of the institutions they learn in.

Trends in International Mathematics and Science Study (TIMMS)

Run by the International Association for the Evaluation of Educational Achievement, TIMMS assesses 9–10 year olds and 13–14 year olds on their skills in both maths and science. TIMMS takes place every three years and more than 50 countries participate. It focuses on curriculum and as a result tends to test pupil's content knowledge rather than their ability to apply it.

Progress in International Reading and Literacy Study (PIRLS)

PIRLS assesses 9–10 year old pupils on their reading literacy. Using a similar design to TIMMS, it focuses on assessing their knowledge and content of the curriculum. It takes place every five years and there are currently 35 countries participating. PIRLS is also run by the International Association for the Evaluation of Educational Achievement.

UNICEF Child Wellbeing indicators

Unicef have developed a broader set of wellbeing indicators that include health and safety, education, family and peer relationships, attitudes, behaviours and risks. Many of the education indicators are drawn from PISA and TIMMS and therefore do not represent new assessments. However they are brought together with the other indicators of wellbeing to give a more holistic assessment.

The indicators described in Box 1 all get students from different countries to take a common test and then compare the results. An alternative model of benchmarking can also be used, where a country conducts its own research into the key features of leading systems and asks how that can inform domestic practice. This allows a country to target the use of international comparisons to particular issues. Hong Kong, for example, constantly compares its curriculum, qualifications, teacher assessments and content of its textbooks to those in other countries. This is the model of international benchmarking being adopted by Ofqual as they seek to ensure England's qualifications keep pace with the best in the world. They have selected qualifications in four disciplines (English, mathematics, chemistry and history) and will compare them across 22 school systems (Ofqual 2011). The ongoing National Curriculum Review has taken a similar approach, asking what specific systems do differently to England and what we can learn from them (Oates 2010).

The advantage of the latter approach is that it allows comparisons to be targeted. So a particular problem with our curriculum can be identified and lessons can be sought from countries that have a similar context to our own. The disadvantage is that it can lead to comparisons focusing on a handful of factors and ignoring other things that might explain performance. So discussion of a top performing country might focus on changes it made to the curriculum, but miss the fact that this only worked because the country also trained its teachers in new pedagogical techniques to work alongside the new curriculum. Comparisons that focus on parts of a school system in isolation can miss how they work together. As Gray (2008, 8) has argued, in many studies 'too little attention was paid to how the parts made up the whole'.

Bearing in mind that every assessment or approach to benchmarking has some weaknesses, policy makers should not be tempted to rely too heavily on one particular measure. It can be beneficial to triangulate observations from different tests and countries in order to validate them.

How have other countries used benchmarks?

Drawing inspiration from industrial benchmarking that took place in the 1970s and 1980s, many countries have started to systematically compare their school systems with those of their competitors. They have put systems and institutions in place to ensure that best practice from other countries is reflected in their own schools. The case studies below demonstrate how three systems have used international comparisons to benchmark their performance.

Box 2: How other countries have used international benchmarks

Brazil

Brazil's stable economic growth over the last decade has been accompanied by a concerted effort to improve its education system. Many of the reforms they introduced were based on specific problems that Brazil faced, such as low high school completion rates, low school funding and poor quality teachers.

An important part of Brazil's reforms involved aligning their domestic targets for schools with those of PISA. Their main dataset for monitoring schools, the Basic Education Development Index (IDEB), was adjusted to ensure that each school's progress in relation to PISA could be measured. This included putting items from PISA into domestic assessments. A number of safeguards have been put in place to ensure this high stakes system does not create perverse incentives, for example progression rates are included in the IDEB to ensure that schools are not incentivised to hold weaker students back from the year groups that are tested. This method of benchmarking has allowed them to use the aggregate picture of national performance generated by PISA to inform practice at the level of individual schools (OECD 2011).

Germany

With an education system stretching back to Humboldt and the German Enlightenment, Germany had long prided itself on the effectiveness and fairness of its school system. However the PISA 2000 results told quite a different story, highlighting substantial problems in terms of both attainment and equity. Germany's misplaced belief in the excellence of its education system was in part the result of a system that did not put much focus on measuring and comparing standards.

The reaction to PISA 2000 was one of national shock, with blanket news coverage and national debate. This prompted Germany to look at what it could learn from other countries. As the OECD have argued, 'after the PISA 2000 results, Germany became an avid, determined, international benchmarker' (OECD 2011: 9). It did this in three ways.

First, it conducted research into what leading school systems were doing and applied it at home. This led to changes to its rigidly streamed school structure, increasing school hours, expansion of preschool and the introduction of common standards. Second, it built several of the international assessments (including PISA, TIMMS and PIRLS) into its own national testing regime. National tests now emphasize the skills and competencies measured in international assessments such as PISA. Third, Germany created a new institution – the Institute for Educational Progress – in order to provide the necessary infrastructure to carry out regular benchmarking. This institute develops standards and assessments, as well as analyzing information and monitoring progress. It helps to 'fix' the use of international comparisons into the education system – ensuring their insights have a more lasting impact than a few newspaper headlines (OECD 2011).

United States

The United States has a federal system where individual states pride themselves on their independence from the national government. This means that, historically, it has not tried to put in place a set of common education standards across the whole country, preferring to leave each state to be responsible for their own school system.

In 2001, the No Child Left Behind Act required states to put in place measurable standards and goals for all children in certain grades. However it stopped short of setting a national benchmark, since each state is able to set its own standards.

Following No Child Left Behind, a number of states have worked together to try and implement their own set of Common Core State Standards. The standards were developed with teachers and experts, and set out a core framework for what K-12 students are expected to learn. As a state-led initiative, this was a 'bottom-up' attempt to ensure there was an appropriate benchmark for all students, regardless of which state they lived in. More recently, the attempt to implement the standards has been taken up by Barack Obama. His flagship education programme, 'Race to the Top', includes a financial incentive for states to adopt the standards (US Department of Education 2009).

While drawing up the detail of the core standards, educators were aware that they needed to be based on an assessment of what the USA could learn from other countries. They therefore established an International Benchmarking Advisory Group to ensure the standards were informed by other top performing countries. Drawing on international research by Schmidt and Prawat (2006), the group called for the standards to reflect the importance of a focused and coherent curriculum with improved materials. They also aligned their assessments with PISA, TIMMS and PIRLS (NGA et al 2008).

By exploring how the countries in Box 2 have responded to the growth of international comparisons, we can see benchmarking takes a number of forms, including:

- Linking national assessments to international tests such as PISA
- Setting national targets to raise a country's score or rank on international assessments
- Establishing institutions that can systematically apply learning from overseas into the national context
- Using international comparisons on a more ad hoc basis to inform specific education reforms in a country.

It is important to remember that the use of international evidence to inform education policy is not new. In England, for example, the Qualifications and Curriculum Development Authority used the 'INCA' dataset to examine international policy and practice around curriculum and assessment. The difference with the developments outlined above is that they involve systematic benchmarks being put in place.

The way in which international indicators are used will vary depending on a school system's approach to data and accountability (Husbands et al 2008). Broadly speaking there are two approaches to using indicators in an education system.

In the first approach, indicators are used to support a system of 'high stakes accountability', where schools are judged on their ability to meet the benchmark and are held to account for their performance. This is the approach adopted in Brazil, where school performance data was explicitly linked to PISA and school leaders were set targets to raise their outcomes on specific measures. This approach would be difficult to put in place in England, because it is unlikely to get the support of exam boards (who are resistant to using common questions in national assessments) and teaching unions (who are resistant to data being used for 'high stakes' accountability).

In the second approach, international indicators are used to 'hold a mirror' to a country's education system as a whole. This model uses indicators to identify where there are shortcomings in the domestic system, and where effort, reforms, and resources should therefore be focused. This is more akin to the German model, where teaching unions supported the use of international benchmarks on the condition they would not be used to hold teachers to account (OECD 2011). This approach would be easier to put in place in England, but would require some form of institutional arrangement that enabled the findings of international comparisons to effect change in the rest of the system. At present, the government is arguing that schools should be 'self-improving', and it is not clear how the findings of international comparisons could be used to drive change in such a devolved system. One option is to leave individual schools (or chains of schools) to identify and learn from international best practice themselves. Another is to use the performance frameworks that still remain (such as Ofsted and league tables) to ensure schools reflect lessons from overseas. A third option would be to create new tools or institutions that encourage schools to reflect international best practice in a systematic way.

The issue of relevance: five questions to ask before introducing a benchmark.

Before putting any benchmarks in place, it is important to make sure they are measuring the right things. This section highlights five important questions a school system should ask to ensure it establishes benchmarks that are meaningful.

1. What are the aims of our education system? Until the system's aims have been established, it is impossible to identify which lessons from overseas will be useful for achieving these aims.
2. Which benchmarks are relevant for our aims? Benchmarks should only be chosen if they help a country to reach its aims. For example, it would be meaningless for a country to align its domestic exams with PISA if its main objective was to improve pupil's group-working skills, since PISA does not assess team work. Or more broadly, PISA links its assessment strongly to a pupil's ability to succeed in the labour market. If this is not the aim of the education system, then PISA is unlikely to be a useful measure.
3. What new information will the benchmark tell us? Comparing outcomes against those in another country is of limited value unless it also helps to identify which factors have driven improvements in that system. This was seen following the early waves of PISA, when countries complained that PISA assessed their overall performance, but did not provide any explanation for what was driving improvement in competitor nations. PISA has subsequently invested heavily in being able to account for changes in performance, as explained in the section below.
4. How does the benchmark compare with what other assessments are saying? Any comparison or assessment can only focus on a certain number of things. Even with large studies such as PISA, trade-offs have to be made when deciding what the assessment will focus on. It is therefore important that a benchmark can be 'triangulated' with other international and domestic assessments, to ensure the outcomes are both useful and valid.
5. How do these international comparisons relate to our specific context? This question is often overlooked by policymakers. It is tempting to identify what helped another system improve and assume that the same reforms can simply be 'transplanted' across borders. In fact, many of the policies and techniques in leading systems were developed to address a specific set of circumstances. These are not necessarily the same circumstances faced by the country that is benchmarking their performance. Lessons from overseas need to be adapted to fit the domestic institutional context. A benchmark is only useful if it allows a country to 'bring the lessons back' to its own system.

What should be benchmarked?

Historically, international comparisons focused on the outcomes of a school system. It was relatively easy, for example, to collect data on the proportion of children who completed secondary school, or how many children could pass a standardized test, and compare these data across borders.

While this form of benchmarking allowed comparisons of outcomes, it did not shed light on what was driving changes in each school system. More recent comparisons have tried to address this by identifying the 'control factors' that account for improved outcomes in different countries. These are the factors that policymakers can change to improve their performance.

In order to establish which factors in the school system are driving improvements, studies also have to collect background information on context (for example the socio-economic status of pupils). This allows them to distinguish the impact the school system has on outcomes, as opposed to the impact of these other contextual factors.

Extending the focus of benchmarking to explain what drives performance, rather than simply describing it, brings a whole range of other factors into view. Table 1 (over) explains the different dimensions that benchmarking needs to consider in order to explain improvements:

Table 1
Dimensions of an education system that need to be considered for benchmarking

| | | Domain 1 Antecedents (Contextualise education) | Domain 2 Policy levers (Shape outcomes) | Domain 3 Outputs and outcomes of school system |
|---------|-------------------------------|---|--|--|
| Level A | Individual learner | Socio-economic background of learners | Individual attitudes, engagement and behaviour | Quality and distribution of knowledge and skills |
| Level B | Instructional setting | Student learning and teacher working conditions | Teaching, learning practices and classroom climate | Quality of instructional delivery |
| Level C | Institutions (schools) | Community and school characteristics | The learning environment at school | Output and performance of institutions |
| Level D | Country or system | National educational, social and economic context | Structures, resource allocation and policies | Social and economic outcomes of education |

Source: Adapted from Schleicher 2011

As can be seen in Domain 2, this approach to benchmarking captures the processes that might explain what drives change in a system. PISA can now statistically account for 85 per cent of performance variance across the OECD. But while it can account for which factors are driving the changes, it can't account for the causal nature of these relationships. It should also be noted that PISA is better at measuring some of these factors than others. For example it is good at measuring the resource allocation in a school system, but finds it harder to measure the individual attitudes and engagement of a pupil.

By incorporating these different dimensions to benchmarking, studies can start identifying in detail the factors that policymakers can change in order to improve their school system. Table 2 provides a more detailed list of these factors.

Table 2
The factors that control a school system's performance

| |
|--|
| Curriculum content and materials |
| Assessment and qualifications |
| National framework and system shape (e.g. classes of qualifications) |
| Inspection |
| Pedagogy |
| Professional development of teachers |
| Institutional development |
| Institutional forms and structures (eg size of schools and education phases) |
| Allied social measures (eg linking health care, welfare and education) |
| Funding |
| Governance structures (eg level of autonomy) |
| Accountability arrangements (eg league tables) |
| Selection and gatekeeping (eg university admissions requirements) |
| Labour market pull (eg professional licensing, regulation) |

Source: Oates 2011

It is essential that all the different dimensions listed in Tables 1 and 2 are captured in a benchmarking exercise. Otherwise there is a danger that people use the comparisons selectively, picking one or two dimensions while ignoring other equally important factors.

Some pitfalls of international benchmarks

While it is important for countries to learn from international best practice, there are also some potential pitfalls with this process. Broadly speaking, three sets of criticisms can be identified:

The first set of problems relate to the **design of international assessments**:

- The sampling methods of international assessments have been criticised for being too small to reliably judge a whole system's performance, and for being open to countries 'gaming' the sample by excluding pupils who are likely to perform poorly (Hormann 2009, Mortimore 2009). A large amount of effort has been taken to address these concerns by assessments such as PISA, which asks for a high response rate compared to other transnational studies and scrutinizes the samples in each country. In 2003, for example, the United Kingdom was rejected from PISA for having an unrepresentative sample.
- International assessments only provide system-level data, which makes it hard to apply the lessons at a more local level. Some elements of PISA cannot even be disaggregated below UK level, which means it is hard to separate lessons for England and the devolved nations.
- Country-specific factors – including the nature of curriculum, testing and teaching – can mean some pupils are better prepared for the format of international assessments than others. Results can therefore reflect whether students were familiar with the format of the test, rather than a substantive assessment of the quality of education they receive (Bradshaw et al 2006).
- International assessments are cross-sectional studies and therefore can't measure a student's outcomes over time. This makes it hard to ensure that they are assessing the right things, because they can't match a student's test score with what they went on to achieve later in life. Canada has conducted a longitudinal study linked to PISA, which showed the assessment was a very good predictor of labour market success (OECD 2010). Until a similar study is conducted in England, it will not be possible to prove that scoring well on PISA relates directly to outcomes in later life for English students. Nevertheless, the data from the Canadian study, coupled with correlations between national scores and outcomes, suggests that PISA is a better measure of whether a child is likely to succeed in the jobs market than many national assessments.

The second set of criticisms relate to **how international assessments are appropriated or used**:

- A common problem is that commentators and policy makers treat international studies as definitive assessments of an education system. In fact, each assessment is designed in a different way and can produce quite different results. The NASUWT (working document) point out that in PISA 2006 and 2009, New Zealand scored significantly better than England in reading, maths and science. Conversely in TIMMS 2007 and PIRLS 2005, England's results were significantly better than those of New Zealand in the same subjects. The difference can be explained by the design of the tests. For example, PISA has a heavier reading load, less focus on content knowledge, fewer multiple choice questions, and is taken at a later age than TIMMS.

This goes to show that whether an education system is described as ‘world class’ will inevitably depend on the design and content of the test being used. The choice of international benchmark to some extent relies on a value judgment about which assessment fits national aims. It should also be triangulated with other forms of assessing performance.

- International assessments are mistakenly used as a measure of a country’s skill level. In fact, they are focused on certain subject areas (literacy, maths and science) and particular age groups (aged 10–15). They therefore exclude children’s performance in other subjects, and the vast array of informal learning, further education and work-based learning that take place later in life, that are also essential for success in the labour market (Keep 2008).
- A desire to improve in international assessments can drive a narrowing of the curriculum towards the areas that they assess. For example, if PISA becomes the definitive measure of quality in an education system, the back-wash effect could be that schools focus on science, maths and literacy and marginalise other subjects in the process (Mortimore 2009, Mansell 2007).
- Commentators and policymakers focus on their country’s rank in an assessment, rather than their score. The closeness of scores in international assessments can mean statistically insignificant differences result in large falls or rises in rankings. Any benchmark must reflect this, for example by being based on statistically significant difference in relation to the international average.
- Commentators and policy makers focus on their headline attainment score, and ignore the vast array of other measures included in international assessments. For example, PISA results are presented in six volumes and can be broken down by gender, social background, equity measures, variance between high and low performers, immigrant status, school resources and a host of other factors. Benchmarks could include a country’s performance against many of these specific criteria, as well as overall score.

The third set of criticisms relate to **the aims and assumptions that lie behind international assessments:**

- International studies prove their relevance by looking at the links between learning outcomes and the labour market. Some question that labour market success should be the primary aim of an education system (Figazzolo 2009).
- Applying a uniform test to all students ignores the fact that the skills and knowledge a student needs to succeed in life are not necessarily the same everywhere on the globe (ibid).

Conclusion: developing international benchmarks in England

Every three years, the publication of the OECDs ‘world league tables’ is greeted with hysteria in England. Headlines bemoan the performance of English school students and politicians rush to explain or deplore the results (see for example Shepherd 2010, Paton 2010). The amount of attention afforded to the rankings shows the growing importance of international comparisons in education.

Policymakers should move beyond this media hysteria and develop a more considered and systematic approach to using international comparisons in the English school system. While international comparisons have been used in England in the past, they have not involved systematic benchmarking. The first step in this direction is Ofqual’s work to benchmark English qualifications to ensure they are ‘world class’ (Ofqual 2011).

Other countries have developed more comprehensive approaches to benchmarking. Broadly speaking, there are three ways in which other countries have incorporated systematic international comparisons into their school systems:

- Linking national assessments to international tests such as PISA
- Setting national targets to raise a country's score or rank on international assessments
- Establishing institutions that can systematically apply learning from overseas into the national context

While trying to learn from the world's top-performing school systems is a welcome move, it raises thorny questions over how the government can ensure schools reflect these lessons in their day-to-day practice. It is not clear how the use of international benchmarking will fit with the government's desire for schools to be 'self-improving', with parents and teachers driving changes in the system. Unlike other countries, England does not yet have an institutional framework for assessing and implementing international lessons on school reform.

If benchmarking is to be developed in a more systematic way in England, two steps need to be taken:

- Benchmarks must be relevant to the aims of the English school system, but the cross-sectional design of many international assessments means it is hard to prove whether they accurately predict a student's future success in life. Introducing a longitudinal survey in England, that sits alongside international assessments, would provide much greater evidence on whether they assess the things that matter for the English school system and whether they are a useful benchmarking tool.
- Lessons from overseas are only useful if they can inform the English system. An institutional framework needs to be developed that allows these lessons to inform the day-to-day practice of schools in England.

References

- Bradshaw J, Sturman L, Vappula H, Ager R and Wheeler R (2007) Achievement of 15-year-olds in England: PISA 2006 National Report (OECD Programme for International Student Assessment) Slough: NFER. <http://www.nfer.ac.uk/publications/NPC02/>
- Brown G (2008) speech, quoted in the *Guardian*, 28 January 2008. <http://www.guardian.co.uk/politics/2008/jan/28/labour.uk>
- Brown G (2010) 'Excellence in education', speech to Woodberry Down Primary School, London, 23 February 2010. <https://community.oecd.org/docs/DOC-6026.pdf>
- Commission of the European Communities (CEC) (2009) *Progress towards the Lisbon objectives in education and training: indicators and benchmarks 2009*. http://ec.europa.eu/education/lifelong-learning-policy/doc/report09/report_en.pdf
- Figazzolo L (2009) *Impact of PISA 2006 on the education policy debate* Brussels: Education International. <http://download.ei-ie.org/docs/IRISDocuments/Research%20Website%20Documents/2009-00036-01-E.pdf>
- Gove M (2010) speech to Westminster Academy, 6 September 2010. <http://www.education.gov.uk/inthenews/speeches/a0064281/michael-gove-to-westminster-academy>
- Gray J (2008) 'Forward' in G Ruddock and M Sainsbury (eds) *Comparison of the core primary curriculum in England to those of other high-performing countries*, DCSF research report RW048, London: DCSF. <http://www.education.gov.uk/publications/eOrderingDownload/DCSF-RW048v2.pdf>
- Hormann B (2009) 'Disappearing students: PISA and students with disability' in S Hopmann and G Brinkek (eds) *PISA according to PISA*, 157–174, Vienna: University of Vienna. <http://www.univie.ac.at/pisaaccordingtopisa/pisazufolgepisa.pdf>
- Husbands C, Shreeve A and Jones N (2008) 'Accountability and children's outcomes in high-performing education systems' in *Research Evidence in Education Library* London: EPPI-Centre, Institute of Education, University of London. <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=eFHWat1OOPM%3D&tabid=2427&mid=4505&language=en-US>
- Keep E (2008) *From Competence and Competition to the Leitch Review: the utility of comparative analyses of skills and performance* London: Institute for Employment Studies. <http://www.employment-studies.co.uk/pdflibrary/wp17.pdf>
- Mansell W (2007) *Education by numbers: the tyranny of testing* London: Politico's
- Mortimore P (2009) *Alternative models for analyzing and representing countries' performance in PISA* Brussels: Education International. <http://download.ei-ie.org/Docs/WebDepot/Alternative%20Models%20in%20PISA.pdf>
- NASUWT (working document) *A framework for judging the education system: supporting information* Birmingham: NASUWT
- National Governors Association [NGA], Council of Chief State School Officers and Achieve Inc (2008) *Benchmarking for success: ensuring US students receive a world-class education* Washington DC: NGA. <http://www.achieve.org/files/BenchmarkingforSuccess.pdf>

Oates T (2010) *Could do better: using international comparisons to refine the National Curriculum in England* Cambridge: Cambridge Assessment. http://www.cambridgeassessment.org.uk/ca/digitalAssets/188853_Could_do_better_FINAL_inc_foreword.pdf

Oates T (2011) 'Reflecting on PISA', presentation at IPPR, March 2011

OECD (2010) *Pathways to success: how knowledge and skills at age 15 shape future lives in Canada*. <http://www.pisa.oecd.org/dataoecd/59/35/44574748.pdf>

OECD (2011) *Lessons from PISA for the United States: strong performers and successful reformers in education* OECD Publishing. <http://www.oecd.org/dataoecd/32/50/46623978.pdf>

Ofqual (2011) *International comparisons in senior secondary assessment: progress report* Coventry: Ofqual. <http://www.ofqual.gov.uk/files/2011-02-24-international-comparisons-in-senior-secondary-assessment-progress-report.pdf>

Paton G (2010) 'OECD school league tables: UK ranked 28th for maths', *Telegraph*, 7 December 2010. <http://www.telegraph.co.uk/education/educationnews/8185935/OECD-school-league-tables-UK-ranked-28th-for-maths.html>

Schleicher A (2011) 'International benchmarking: Holy Grail or Alchemists' Stone?', presentation at IPPR, March 2011

Schmidt W and Prawat R (2006) 'Curriculum coherence and national control of education: issue or non-issue?' *Journal of Curriculum Studies* 38(6): 641–658

Shepherd J (2010) 'UK schools slip down world rankings', *Guardian*, 7 December 2010. <http://www.guardian.co.uk/education/2010/dec/07/uk-schools-slip-world-rankings>

US Department of Education (2009) *Race to the Top Program: Executive Summary* Washington DC. <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>